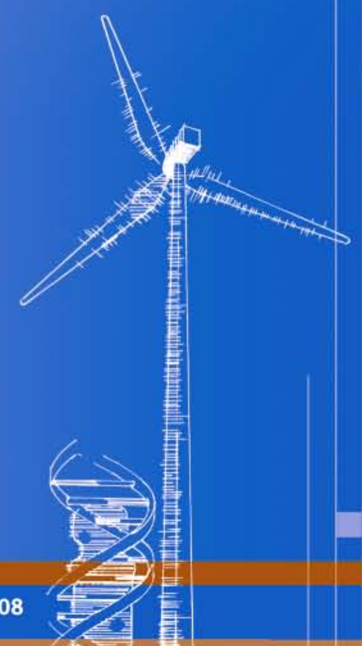
NREL National Renewable Energy Laboratory
Office of Energy Efficiency & Renewable Energy
Innovation for Our Energy Future

# Development of Regional Wind Resource and Wind Plant Output Datasets

## Final Subcontract Report
## 15 October 2007 – 15 March 2009

*3TIER*
*Seattle, Washington*

***Subcontract Report***
**NREL/SR-550-47676**
**March 2010**



NREL is operated for DOE by the Alliance for Sustainable Energy, LLC    Contract No. DE-AC36-08-GO28308

# Development of Regional Wind Resource and Wind Plant Output Datasets

## Final Subcontract Report
## 15 October 2007 – 15 March 2009

*3TIER*
*Seattle, Washington*

NREL Technical Monitor: Debra Lew

Prepared under Subcontract No. AAM-8-77556-01

*Subcontract Report*
**NREL/SR-550-47676**
**March 2010**

**National Renewable Energy Laboratory**
**1617 Cole Boulevard, Golden, Colorado 80401-3393**
**303-275-3000 • www.nrel.gov**

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

Contract No. DE-AC36-08-GO28308

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/ordering.htm

**This publication received minimal editorial review at NREL**

Printed on paper containing at least 50% wastepaper, including 20% postconsumer waste

# Table of Contents

This page left intentionally blank

## Executive Summary

This is the Final Report for the project "*Development of Regional Wind Resource and Wind Plant Output Datasets*"(NREL subcontract number: AAM-8-77556-01 under prime contract number: DE-AC36-99GO10337). The report covers the period of the contract from October 15, 2007 through March 15, 2009. The final delivered outcomes of this project include:

- This report detailing the work produced.
- 30 validation reports for the purpose of tuning the mesoscale models.
- The numerical weather prediction (NWP) simulations for 2004-2006 in NetCDF format with a spatial resolution of one arc-minute and a temporal resolution of ten minutes.
- 30,544 original sites and 1499 additional that were extracted into time series data files for 2004-2006. These sites had the following information provided:
    - Wind speed at 100 meters (m)
    - Rated power output at 100 m
    - SCORE-lite[1] power output at 100 m
    - Mesoscale forecasts at 100 m
    - "Perfect" forecasts
    - "Two-hourly" persistence forecasts
    - By-hour monthly climatology forecasts
- Solar forecasts from mesoscale models for a regular grid of 8736 points
- 28 validation reports for the final data set on publicly available data
- 2 validation reports on confidentially sourced data
- SCORE-lite validation report
- Web-based graphical file server
- Extension to the contract where each of the 32,043 sites had wind speed and wind direction pulled from the model runs at 10 m, 20 m, 50 m, 100 m, and 200 m.

A paper was written based on this collaboration between the National Renewable Energy Laboratory (NREL) and 3TIER. The paper was presented at the *7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms*, Madrid, May 2008. It was subsequently invited for publication in the *Wind Engineering Journal*, Volume 32, Number 4, 2008.

---

[1] SCORE (Statistical Correction to Output from a Record Extension); SCORE-lite models each grid point, instead of each turbine, by aggregating ten individual samples from the original SCORE probability density functions (as though ten turbines were being modeled) to develop new probability density functions that represent ten turbines instead of one.

# 1. Status of the Project

The project contract was executed on October 25, 2007. All variances from the scope of work were discussed and agreed upon well in advance, and consequently, some of the reports were delivered behind the original schedule. Due to the complexity of this study and the large quantity of data, a number of problems arose during the creation and transfer of the data. A majority of the errors were caused by storage failures and subsequent "quick fixes."

The storage failures occurred too often with this particular project to be mere bad luck – and towards the end of the project, 3TIER surmised that the problems were likely due to the number of concurrent input/output demands on the storage servers. Since 3TIER changed the way that these scripts were accessing the data, there have been no further data storage problems. Thus, it was concluded that the problems were coming from the sheer size of the dataset.

The other sources of error were the quick fixes that were applied to try to correct the data without affecting the timelines. These fixes all worked for a vast majority of the cases – but due to the size of the dataset, it was difficult to quickly and conclusively test the dataset. Unfortunately, the drive to meet timelines ended up causing large delays as the fixes were applied, and then when the data was examined, some errors were apparent. The final iteration of the data creation was much slower, and also much more exhaustively tested. Due to the size of the dataset, it was still impossible to visually check every possible data point, yet a number of bulk statistic checks were performed on each variable (range checks, difference checks). Some more advanced checks such as histograms, difference histograms, and diurnal cycle plots were performed on a randomly selected (yet significant portion) of the data points. The final checks were as close to exhaustive as could be feasibly performed, and there were no data points that were outside our expected parameters for acceptable data.

This project was led at 3TIER by:


**Cameron Potter, Power Prediction Engineer**

Project manager and main contact for NREL

e-mail: cpotter@3tier.com

**Bart Nijssen, Chief Technology Officer**

Overall project responsibility and oversight

e-mail: bnijssen@3tier.com


During the life of the project, 26 staff members of 3TIER were directly involved with producing deliverables.

## 2. Variances and Proposed Variances from the Contract

Through discussions with NREL counterparts, the following decisions were made regarding the project. None of these decisions materially changed the nature of the contract or the deliverables, but instead served to clarify parts of the contract and deliverables:

### Wind site selection:

The site selection algorithm was extensively discussed, undergoing several iterations between 3TIER and NREL (especially regarding transmission zones and renewable energy zones). The final algorithm was agreed on and implemented, as described in *Section 6, Wind Site Selection*.

After site selection was complete, an additional 1499 points were added to the scope (at no additional cost) to allow for sites that were not captured by the original site selection process.

### Solar site selection:

The solar site selection was originally going to be made by NREL, but, following discussions between NREL and 3TIER, it was decided to produce a gridded dataset covering the entire region. This grid spacing was at 12 arc-minutes. This was deemed acceptable since solar variation over the area is generally far less than wind variation (within the limits of the modeling process, i.e., shading due to small features cannot be resolved).

### SCORE-lite validation:

The Statistical Correction to Output from a Record Extension (SCORE) validation was not completed on the schedule of six (6) weeks after subcontracting – but this had been discussed and understood by both parties.

The delay was caused by a more concerted effort on the site selection algorithm (which took longer than expected) and limited access to useful SCORE-lite validation data.

### Final validation reports:

The scope of work stated that the final validation reports were due at the end of the 18$^{th}$ week of the project – at that stage, the data was not complete, so it was agreed to delay the validation reports for delivery with the final data.

### Final report:

Due to a number of data re-processing delays, the final report was rewritten to more accurately reflect the overall nature of the project. Consequently, the final report was delivered significantly after the original due date.

### Delivery of the mesoscale model simulations:

Delays in delivery to the NREL server slowed this process from the original timeline, although this was fortuitous in the end because it meant that the data was not sent until everything had been carefully checked (performing range checks on EVERY value in the entire dataset).

### Solar forecasting:

Because the "persistence" forecasts would be directly derived from data that was already available to NREL (and was simple to perform), it was agreed that 3TIER would direct their efforts to improvement of the mesoscale model solar forecasts.

## 3. Meetings and Other Important Contacts during the Contract

During this project, there was extensive contact between staff at NREL and 3TIER to ensure that the project scope of work was met to the satisfaction of both parties and, where possible, to call on the experience of both parties to produce better results.

The chief contacts at NREL were:

### Debra Lew, Senior Project Leader

Project manager and main contact
e-mail: debra_lew@nrel.gov/ voice: 303-384-7037

### David Corbus, Senior Engineer

Overall project responsibility and oversight
e-mail: david_corbus@nrel.gov / voice: 303-384-6966

### Neil Wikstrom, Senior Contract Administrator

Contract management
e-mail: neil_wikstrom@nrel.gov / voice: 303-384-6960

Every week, internal meetings were held at 3TIER from October 31, 2007 onwards to ensure that all project staff were in constant contact.

Every week, calls were scheduled between NREL and 3TIER staff from October 25, 2007 onwards to ensure that the project remained on track and also to ensure that NREL was kept abreast of project developments.

The following is a list of the additional scheduled calls/meetings:

- November 7, 2007 = Commencement call with NREL
- November 9, 2007 = Western Wind and Solar Integration Study Stakeholders meeting
- November 13, 2007 = Model configuration selection call
- November 15, 2007: Brief in-person visit to NREL wind group by Bart Nijssen, Pascal Storck, and Bernard Walter (as part of a 3TIER visit to the NREL solar group).
- December 3, 2007 = Site Selection Discussion
- December 3, 2007 = Photovoltaic solar discussion
- December 3, 2007 = Concentrating solar discussion
- March 4, 2008 = Call with NREL and GE
- March 12, 2008 = Western Wind and Solar Integration Study Technical Review Committee (TRC) call
- March 19, 2008 = Stakeholders meeting
- April 1, 2008 = NREL and GE call

In addition, a large number of unscheduled calls occurred for clarification and to improve collaboration during the project.

During this contract, staff from 3TIER and NREL also made extensive use of e-mail contact to keep the communication channels open:

- 1200+ e-mails regarding this contract were exchanged.

During the contract, there was also extensive use of e-mail for exchanging working documents:

- 70+ e-mails with attachments from NREL
- 75+ e-mails with attachments from 3TIER

Finally, use of conventional mail for over-night delivery was also used between 3TIER and NREL (as well as for interaction with GE Consulting). Conventional mail was used for sending large amounts of data on LaCie 1-TB external hard drives, 32-GB USB flash drives, and DVD media.

The entirety of the dataset was much too large to send on normal drives, so a server was used to transport the datasets. This server was provided by NREL, sent to 3TIER for data transfer, and tested to ensure that the data transfer was successfully and accurately completed before being returned to NREL.

In addition, 3TIER staff will continue communication with NREL staff through the power system modeling portion of this project.

## 4. Wind Modeling

*"Wind energy is the fastest growing source of energy in the United States. As this important energy source continues to grow, evaluating its impacts on the operation of electrical systems becomes increasingly important."*

– Quoted from the Statement of Work, 07/17/07, Development of Regional Wind Resource and Wind Plant Output Datasets

The entire Western Wind Integration dataset was created in two separate stages with a consistent modeling technique to allow for a smooth combination of the datasets. The first stage modeled the Pacific Northwest and was performed for the Northwest Wind Integration Action Plan (NWIAP), jointly sponsored by the Bonneville Power Administration (BPA) and NREL. It covered the states of Washington, Oregon, and Idaho, as well as most of Montana and Wyoming. Fig. 1 shows the area covered by the NWIAP modeling effort bounded by a red box. The second stage expanded the area covered by the modeling runs south to the southern border of California, Arizona, and New Mexico, and out to the eastern border of Colorado.

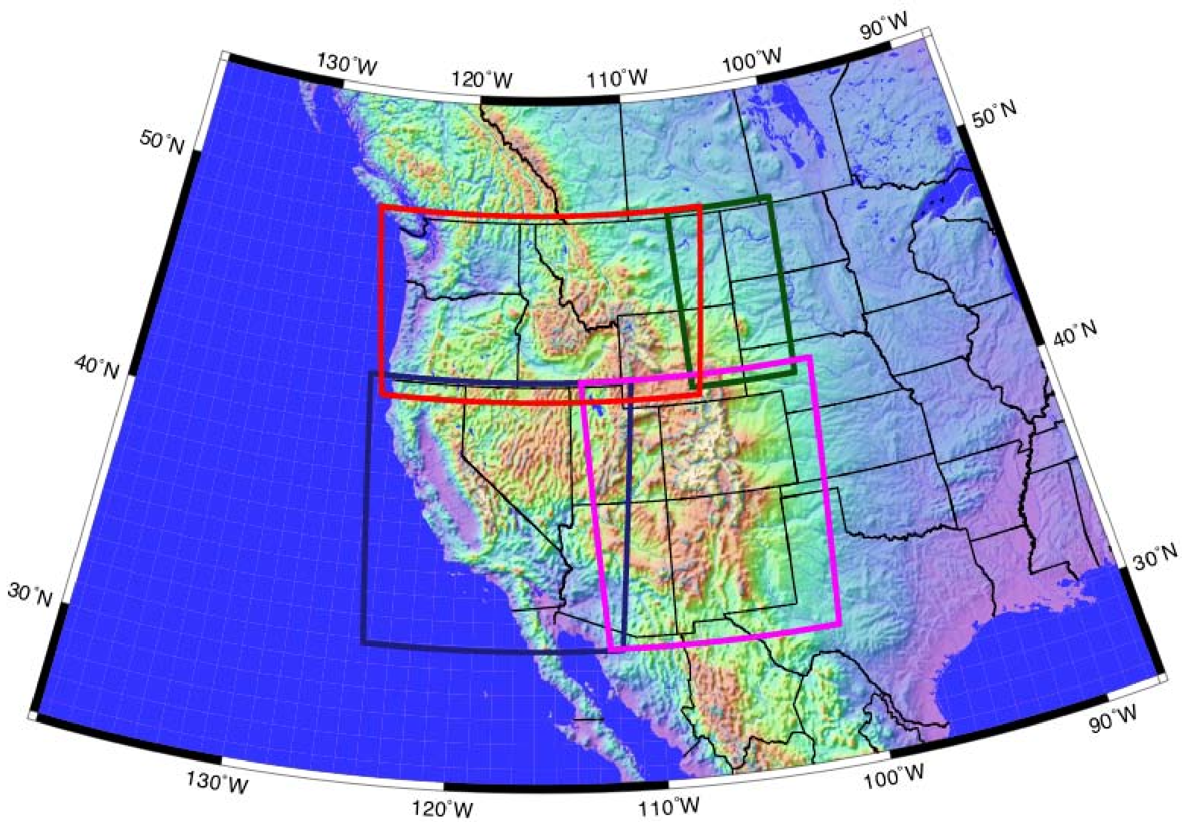



**Fig. 1 –** The map above shows the modeling domains in the Western Wind and Solar Integration Study. The red bounding box shows the NWIAP region and the other domains, green, blue and magenta, are called Domains 1, 2 and 3 respectively.

## A. Modeling Domains

Fig. 1 shows four domains: the NWIAP domain, and three other domains numbered in the following order: 1) north-easterly, 2) south-westerly, and 3) south-easterly. The use of multiple domains (especially the splitting of the southern region into two domains) was dictated by the magnitude of the area being modeled at a high resolution.

The mesoscale model was operated by allocating sub-sections of the model domain (i.e., sub-domains) to individual computer processors on a supercomputing cluster. When operating a NWP model, the model runs are often too large (especially in this case) to run in the memory of a single processor. Parallelisation is used to overcome the memory constraints as well as to provide more powerful computing power. However, the processors acting on the sub-domains cannot do the calculations entirely independently. Each processor must communicate with the other processors that are calculating adjacent sub-domains to allow "advection" and "diffusion" operators to transfer information about weather events from neighbouring sub-domains. Sub-domains allow these models to be run accurately and relatively quickly, but there is still a limit to the number of sub-domains that can practically be accommodated. The size of the sub-domain is memory-limited, and the number of sub-domains is limited by the bandwidth of the inter-node links. If too many sub-domains are used, the communication channels in the computing cluster become clogged, resulting in latency issues. For this project, the latency restrictions required that the southern region identified in Fig. 1 had to be split into two separate modeling domains, each with their own sub-domains.

## B. Configuring the Models

The Weather Research and Forecasting (WRF) model was used as the mesoscale model in this study. WRF is generally considered to be the most advanced mesoscale model in North America and has superseded the previous industry standard, the MM5 model. The WRF model has a number of configurations that can be chosen to model the atmosphere. Four different model configurations were tested. These different models, selected based on 3TIER's expertise and experience, are detailed in Table 1.

**Table 1 –** NWP Configurations Using the Advanced Research WRF Core

| Model Configuration | Vertical Levels | Planetary Boundary Layer Parameterisation | Elevation Data Set | Land Surface |
|---|---|---|---|---|
| A | 31 | Yonsei University | 30 arc-second USGS | 5-layer soil diffusivity |
| B | 31 | Mellor-Yamada-Janjic | 30 arc-second USGS | 5-layer soil diffusivity |
| C | 31 | Yonsei University | 30 arc-second USGS | Oregon State University |
| D | 37 | Yonsei University | 30 arc-second USGS | 5-layer soil diffusivity |

Configuration A was used as the baseline model configuration with configurations B, C, and D all having a single parameter of deviation. Configuration B used the Mellor-Yamada-Janjic boundary layer parameterisation, which features explicit prognostic equations for boundary layer turbulence. Configuration C used the Oregon State University land surface model, a more sophisticated physical process model for estimating surface fluxes. Both Configurations B and C should theoretically be better than Configuration A. However, the extra sophistication in the models introduces additional assumptions and unconstrained parameters that can adversely affect the accuracy of the model. Configuration D adds extra vertical levels in the boundary layer in an attempt to better simulate the vertical profiles of wind and temperature near the surface.

As running these models is computationally expensive, the trial runs to evaluate the various configurations had to be simplified. The trials were run at a coarser spatial resolution of 6-km by 6-km grid spacing instead of 2-km by 2-km. The temporal resolution of the output was also reduced by only saving the hourly data in the trials. Finally, the model was only run for three out of every nine days for the year 2006. Nonetheless, the trial model runs of the different configurations showed different skill and were used to determine the best configuration in each domain.

Trial runs were executed for each of the four domains. The NWIAP domain was modeled first and validated against data from six tall towers. The validation showed that the default configuration, A, was optimal. Runs for the other three domains that were modeled as part of this study were validated against data from a total of 30 tall towers. Each of the different configurations was judged qualitatively "best" (over a number of parameters) for at least one tower. The validation reports (previously supplied to NREL) were discussed with the engineers and meteorologists at NREL and the following consensus was reached:

- NWIAP Domain – Configuration A was previously selected.
- Domain 1 – Configurations A and D performed at a similar level of accuracy, but it was decided that Configuration D would be used because there was greater consistency between Domains 1, 2, and 3.
- Domain 2 – Configuration D outperformed the other configurations most consistently.
- Domain 3 – Configuration D outperformed the other configurations most consistently.

### C. Producing the Dataset

With the model configurations selected, the models were run on the supercomputing cluster. Each grid point's location is defined by latitude, longitude, and elevation. The model simulations produced a time series at each grid point, including:

- Wind speed and direction at 10 m, 20 m, 50 m, 100 m, 200 m and at 500 mb (higher in the atmosphere).
- Temperature at 0 m, 2 m, 20 m and 50 m.
- Specific humidity at 2 m.
- Pressure at 0 m.
- Precipitation at 0 m.
- Down-welling radiation (longwave and shortwave) at 0 m.

### D. Re-gridding the Dataset to a Consistent Spacing

The original model run was performed at 2-km by 2-km grid spacing across each domain; however, the edges of the domains were not perfectly aligned because each domain was defined individually. The original datasets were re-gridded to a consistent grid spacing across the entire area covered by the Western Wind and Solar Integration Study. This was re-gridded to one arc-minute spacing so that the grid points were easily identified using regular latitudes and longitudes.

## E. Blending the Dataset into a Single Dataset

The desired outcome from this project was to produce a single, consistent dataset that presented three years of ten-minute resolution data at a grid spacing of one arc-minute for the years of 2004 to 2006. However, the datasets were modeled as four separate domains, with some differences between the simulations at the boundaries. To produce a seamless dataset, data from the individual model domains were blended at the overlapping boundaries (see Fig. 1). The result was a single large dataset with over 1.2 million individual grid points, each grid point having a time series of 157,680 points for each of the variables listed in *Section 4.3, Producing the Dataset*. This dataset, even when stored in efficient netCDF format, used more than 24 TB of storage space in its final form.

The sheer size of this dataset caused significant problems. To maintain the integrity of the dataset, each time a process was implemented that altered the core dataset (e.g., re-gridding, blending, etc.), the dataset was first copied and the alteration was performed on the duplicate. This way the original dataset was maintained until the altered duplicate could be thoroughly verified. This procedure was very reliable, but required many TBs of duplicate data to be stored as a safety back up. The process was difficult to manage and time consuming, because even the process of copying 24 TB of data is non-trivial. However, the production of the dataset was a major cost of the project (both in time and money), and losing the original dataset was not an acceptable risk.

Later in the project, some flaws in the data transfer and subsequent processing came to light and the maintenance of the original dataset meant the project could continue without needing to re-run the entire project.

## 5. Wind Validation

Validation was carried out at 28 towers (with publicly available data) over the model domains. An extensive validation report (in excess of 10 pages of analysis) was produced for every tower. The validation reports show a comparison of the observed data from the observation anemometer and the model data scaled to match the height of the anemometer. The data that was used in these validations was the corrected model data that is available in the comma-delimited files discussed in *Section 6, Wind Site Selection* and *Section 7, Wind Energy Modeling*.

The correction was performed statistically by comparing with the Rapid Update Cycle (RUC) dataset produced by the National Oceanic and Atmospheric Administration (NOAA) and the National Center for Environmental Prediction (NCEP). The RUC dataset makes extensive use of the low-level (10-m) towers, and this can mean that in some areas of high terrain variability, the use of the RUC correction is less valid. 3TIER weighted the relative value of the RUC dataset according to the terrain before applying the corrections.

Each validation report includes the following information:

- Tower location (latitude, longitude and state),
- Height of anemometry on the observation tower,
- The period of useful observational data,
- The average windspeed and standard deviation of the observations and the corrected model data,
- A brief explanation of the modeling used in the NREL work and a map showing the validation tower locations,
- The correlation value and the root mean square (RMS) error of the corrected model data compared with the monthly-mean and daily-mean observations,
- A plot of the monthly-means of the observed and corrected model data,
- The wind speed histograms of the observed and corrected model data with fitted Weibull distributions,
- Comparison of the observed and simulated wind rose plots, both over the entire period of observed data and broken out by month,
- Plot of the diurnal cycle of the observed and corrected model data, both over the entire period of observed data and broken out by month,
- In the appendices, the data used to develop many of the plots is tabulated to a more quantitative comparison.

Table 2 shows a brief summary of the full reports. After comparison with all validation towers, the corrected wind speeds have a slightly lower bias and the corrected standard deviations are also closer to the observations. The overall wind speed bias was low, but some of the validation reports showed large errors. These errors may be due to sub-grid terrain variability, unreliable observations, or simple model inaccuracy. Furthermore, the improvement from using the correction is not universal. However, it is important to note that the correction was not developed using any of these tower's data as inputs. This means that the validation integrity is maintained and the validation results are broadly applicable.

**Table 2 –** Summary of Validation Data

| Validation Tower | State | Latitude | Longitude | Start Date | End Date | Anemometer Height [m] | Observed Mean Wind Speed [m/s] | Observed Wind Speed Standard Deviation | Simulated Mean Wind Speed [m/s] | Simulated Wind Speed Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|---|
| 3001 | AZ | 34.957 | -111.160 | Nov-04 | Jul-06 | 30 | 5.47 | 3.58 | 5.95 | 3.73 |
| 3002 | AZ | 35.668 | -111.750 | May-05 | Oct-06 | 30 | 6.40 | 3.58 | 6.23 | 3.20 |
| 3003 | AZ | 35.040 | -114.534 | Jun-05 | Jun-06 | 30 | 5.43 | 3.46 | 5.47 | 3.73 |
| 3006 | AZ | 34.493 | -109.444 | Jul-05 | Dec-06 | 30 | 4.80 | 3.38 | 4.47 | 3.40 |
| 4402 | CA | 38.247 | -121.502 | Aug-04 | May-05 | 47 | 4.24 | 2.60 | 4.40 | 2.60 |
| 4403 | CA | 38.762 | -122.841 | Jun-04 | Jan-05 | 61 | 5.89 | 4.12 | 4.55 | 3.72 |
| 6001 | CO | 40.920 | -102.256 | Jan-04 | Aug-04 | 44 | 7.37 | 3.30 | 6.47 | 2.95 |
| 6008 | CO | 40.666 | -103.345 | Jul-05 | Nov-06 | 20 | 5.29 | 3.39 | 4.84 | 2.71 |
| 6009 | CO | 40.525 | -107.596 | Aug-05 | Nov-06 | 20 | 4.30 | 2.91 | 4.36 | 2.87 |
| 6013 | CO | 39.911 | -105.235 | Jan-04 | Dec-05 | 80 | 4.80 | 3.53 | 6.02 | 5.07 |
| 6029 | CO | 38.575 | -102.669 | Aug-04 | Nov-06 | 20 | 5.87 | 3.09 | 5.27 | 2.70 |
| 6039 | CO | 37.711 | -104.928 | Aug-05 | Oct-06 | 50 | 7.71 | 4.64 | 7.56 | 4.57 |
| 12111 | ID | 43.407 | -111.748 | Jan-04 | Mar-05 | 60 | 6.49 | 3.28 | 6.03 | 3.05 |
| 12131 | ID | 42.062 | -116.078 | Jan-04 | Nov-05 | 20 | 5.90 | 3.08 | 5.01 | 2.09 |
| 12439 | ID | 43.022 | -115.275 | Jan-04 | Oct-05 | 20 | 5.51 | 3.76 | 4.25 | 3.22 |
| 12500 | ID | 47.367 | -116.933 | Jul-05 | Sep-06 | 20 | 6.29 | 3.47 | 4.39 | 2.36 |
| 12505 | ID | 42.905 | -112.349 | May-05 | Jul-06 | 20 | 5.45 | 3.17 | 3.79 | 1.88 |
| 26007 | MT | 46.457 | -110.114 | Jan-04 | May-05 | 30 | 7.39 | 4.53 | 6.24 | 3.87 |
| 26010 | MT | 47.262 | -110.625 | Jul-04 | Feb-06 | 40 | 7.62 | 4.44 | 6.59 | 3.47 |
| 28001 | NV | 39.046 | -117.001 | Jan-04 | Dec-06 | 50 | 4.62 | 3.37 | 4.82 | 3.49 |
| 28002 | NV | 38.541 | -118.294 | Jan-04 | Dec-06 | 50 | 4.46 | 2.92 | 5.50 | 3.88 |
| 28003 | NV | 38.372 | -117.472 | Jan-04 | Dec-06 | 50 | 5.46 | 3.38 | 4.97 | 3.44 |
| 31010 | NM | 34.793 | -104.025 | Jun-05 | Dec-06 | 70 | 8.54 | 3.72 | 8.07 | 3.49 |
| 31011 | NM | 34.827 | -104.999 | Jan-04 | Aug-04 | 39 | 7.02 | 3.11 | 6.96 | 3.27 |
| 34018 | ND | 48.705 | -102.571 | Jan-04 | Jul-06 | 40 | 7.07 | 3.46 | 6.98 | 2.92 |
| 44003 | UT | 39.277 | -111.683 | Jan-04 | Sep-04 | 20 | 3.98 | 2.76 | 3.76 | 2.16 |
| 44022 | UT | 37.734 | -109.288 | Jan-04 | Mar-05 | 20 | 4.90 | 3.07 | 4.48 | 2.81 |
| 44999 | UT | 38.940 | -112.881 | Jun-06 | Dec-06 | 50 | 6.42 | 5.23 | 6.31 | 3.76 |

**NOTE:** The BPA validation reports were completed earlier than the other validation reports. Consequently, they did not use the correction technique – although the actual data derived for the power data files in this region still used the correction, so there is no inconsistency in the final dataset.

# 6. Wind Site Selection

The creation of the modeled dataset was the first phase of the project. To make the data accessible for power systems modeling, the dataset had to be converted into synthetic wind energy project data.

The initial request for proposals required 300 GW of synthetic wind energy with a variety of project sizes spread across the Western Electricity Coordinating Council (WECC) area. 3TIER decided to produce a superset of 900 GW worth of sites so that the desired 300 GW could be chosen interactively by NREL, GE Consulting, and project stakeholders. In fact, the 300 GW was itself a superset from which 70 GW would be chosen for power systems modeling.

Rather than modeling each synthetic project as a unit, 3TIER assumed that each grid point could be a potential wind project. The points could then be aggregated to become whatever size project was desirable. Ideally, each grid point in the modeled dataset would be converted into a synthetic wind energy project. However, this was impractical given the number of individual grid points. Instead, a subset of the potential sites was selected for modeling as synthetic wind projects.

To determine how many sites needed to be selected, it was first important to determine the number of MWs that each site could represent. It was decided that each synthetic wind project would be modeled using the same turbine for consistency across the dataset, and that this turbine should be large because the dataset was designed to represent build-outs of wind energy production up to 2017 (ten years in the future from the commencement of the project), and there is a general trend in the United States towards larger turbines. The Vestas V-90 3-MW turbine was chosen as a good middle ground between today's mean turbine size and those likely to be used in the future. With the turbine selected, it had to be determined how many turbines could fit in a model grid cell. To achieve this some simple heuristics were used:

- Spacing of ten rotor diameters between strings of wind turbines and,
- Spacing of four rotor diameters between turbines on the same string,
- An appropriate buffer zone at the edge of each grid cell was also required because the turbine layouts could be tiled next to adjacent areas without violating the turbine spacing guideline.

These heuristics indicate that ten turbines can fit into a 2-km by 2-km area as indicated in Fig. 2. Ten turbines at 3 MW per turbine meant that each grid point could represent a 30-MW project. Thus, 30,000 points were required to model the desired 900 GW of wind energy. Finally, as planned, multiple sites could then be aggregated to obtain wind energy projects of a larger size that were still modeled in such a way to allow for varying wind speed across the project.

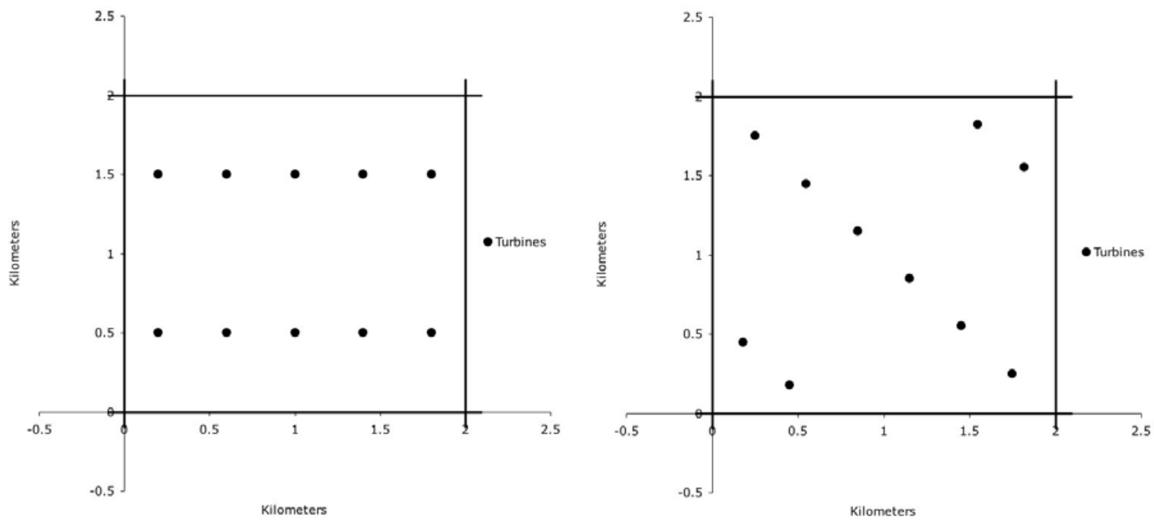



**Fig. 2 –** The graphic above shows example layouts of wind turbines with ten rotor diameters between strings and four rotor diameters between turbines on the same string.

To select which of the 1.2 million sites to use, a multi-phase selection algorithm was developed in conjunction with NREL staff. Each phase selected only from previously unselected points in order to have the final number meet the desired goal. While this study would include modeling of the entire WECC area, the “study footprint” was defined as the WestConnect group of utilities outside of California. This represented Nevada, Arizona, New Mexico, Colorado, and Wyoming, and the number of selected sites was intentionally biased to select more points from these states.

1. The first phase of site selection was to pre-select a set of points that would represent existing wind energy projects and those under development. This information was obtained and compiled by NREL and resulted in 404 sites (or approximately 12 GW). Some of the information about these plants was incorrect, which led to a similar process once the site selection was complete in which an additional 1499 sites were added to the total selected sites.

2. The next phase was to identify the sites with the highest wind energy potential based on wind energy density at 100 m within 80 km (50 miles) of existing or planned major transmission networks or within pre-identified high-potential renewable energy zones (REZ) in the study footprint. 200 GW of sites (6667 sites) were selected in the transmission corridors or REZ areas. This phase was done without regard to geographic dispersion.

3. The third selection phase aimed to find the sites that had the best correlation with the load profile of WestConnect (limited to sites with a wind energy density of greater than or equal to 300W/m$^2$). The load correlation measure was evaluated by calculating the difference between the average normalised load profile and average normalised wind energy density on an hourly basis; the smaller the difference, the better the site. This phase, and the next phase, attempted to promote geographical diversity. This was achieved by NREL assigning each state (and two offshore regions) an approximate number of MWs that should be selected. Table 3 shows the approximate number of MWs modeled in each state.

4. The fourth selection phase was a simple selection by highest wind energy density, again selected according to the allocations in Table 3.

5. Finally, after the planned site selection algorithm was complete, it became apparent that some sites were missed from the pre-selected set of sites in the first selection phase. To rectify this oversight, a further set of "post-selected" sites was identified with input from stakeholders in this project and an additional 1499 points were selected.

**Table 3 –** Sites for Selection Determined By Load Correlation and Power Density

| State/Offshore Region | Selected by load correlation [MWs] | Selected by power density [MWs] |
|---|---:|---:|
| Arizona* | 18,000 | 18,000 |
| California | 8,000 | 74,000 |
| Colorado* | 28,000 | 28,500 |
| Idaho | 8,000 | 13,500 |
| Montana | 13,000 | 35,000 |
| North Dakota | 4,000 | 5,000 |
| Nebraska | 8,000 | 5,000 |
| New Mexico* | 32,000 | 40,500 |
| Nevada* | 33,000 | 48,000 |
| Oklahoma | 7,000 | 7,000 |
| Oregon | 4,000 | 36,000 |
| South Dakota | 7,000 | 10,000 |
| Texas | 8,000 | 10,000 |
| Utah | 8,000 | 11,000 |
| Washington | 4,000 | 44,000 |
| Wyoming* | 54,000 | 69,000 |
| Offshore CA | 1,000 | 4,000 |
| Offshore WA/OR | 500 | 1,000 |
| **TOTAL MWs** | **245,500** | **459,500** |

*In the WestConnect study footprint

In the initial phases of the site selection, 30,544 sites were chosen and an additional 1,499 sites were selected at the end of the process. Therefore, the final number of points that were selected for further study was 32,043. The selected sites are shown in Fig. 3. With the sites selected, each synthetic wind energy site needed modeling as an individual project, producing comma-delimited files with a three-year time series at a ten-minute resolution.

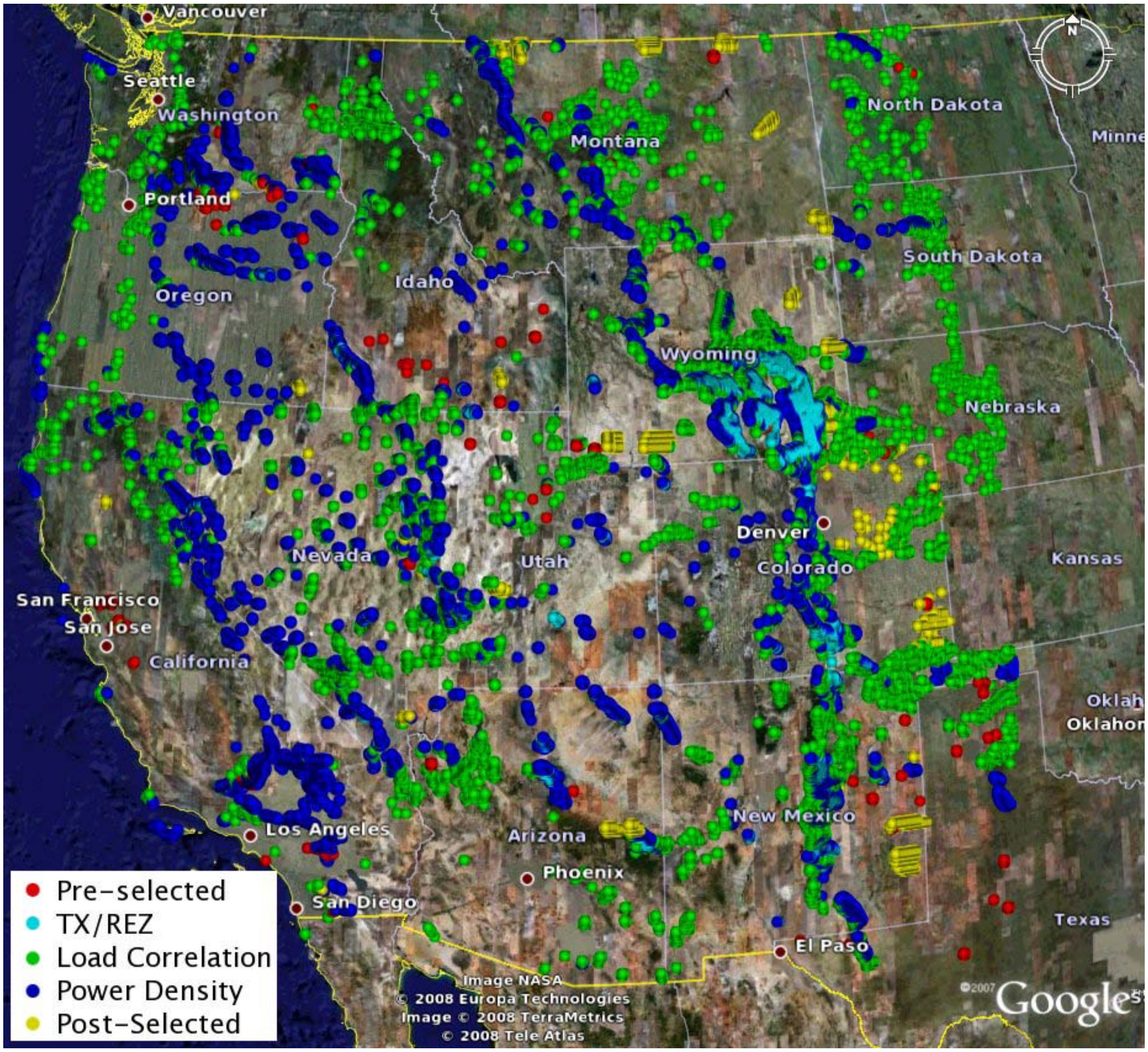



**Fig. 3 –** Shown above is a map of the selected sites, with each point coloured differently depending on selection technique.

## 7. Wind Energy Modeling

A key component of this project was to develop realistic power output at over 900 GW of wind energy sites. This meant that the wind speed data had to be converted to power output data. The industry standard is to produce a "rated" power output using a simple power curve. This power curve can be the manufacturer's power curve or some kind of "smoothed" power curve based on the behaviour of other wind plants.

Having decided on the power curve to use, the wind speed is converted to an "effective" wind speed based on a reference air density. This results in each wind speed being converted to a single power output value. However, NWP models have a tendency to produce wind speed time series that are excessively smooth, that is, they do not produce sufficient wind speed variation at short timescales. As a result, the overall behavior of wind plant output directly derived from wind speeds from a mesoscale model and put through a rating curve results in excessively smooth plant output. For this project, we produced two forms of power output: the rated power output and a statistically corrected power output that better models the variation.

The statistical correction used in this work was called SCORE-lite and is based on SCORE (Statistical Correction to Output from a Record Extension). SCORE was developed by 3TIER and originally proposed in a paper presented at the IEEE Power Engineering Society General Meeting in 2007[2]. Prior to the completion of this project SCORE has been used to model several GWs of potential wind energy installations; the technique is gaining industry acceptance as people become more familiar with the process.

The SCORE process uses observed statistical deviations from a mean value to create probability density functions of deviation from some central point. It is run on each turbine and produces a time series of power output data for each turbine, which can then be aggregated to sub-project or entire project output. However, trying to run a probabilistic process on 32,043 x 10 turbines would be an extremely time consuming process and the turbine locations would need to be approximated anyway, meaning that the individual turbine locations would provide no extra information. SCORE-lite was developed to solve this problem.

SCORE-lite models each grid point instead of each turbine. This is achieved by multiple sampling from the original SCORE probability density functions (PDFs), once for each turbine per grid point. The re-sampling process is carried out ten million times to create new PDFs. SCORE-lite takes the "rated" power as an input and modifies it such that the overall change characteristics more closely resemble those observed in reality. As part of this project, SCORE-lite was validated[3]. It was found that SCORE-lite produced a more realistic change histogram than the use of a rating curve alone – without any appreciable loss of accuracy in modeling the diurnal cycle.

---

[2] C. W. Potter, H. A. Gil and J. McCaa, "Wind Power Data for Grid Integration Studies", Proc. 2007 IEEE Power Engineering Society General Meeting, Tampa, FL, USA, paper No. 07GM0808, Jun. 2007.

[3] The SCORE-lite validation has already been submitted as a deliverable to this project and further describes the SCORE-lite power modeling technique.

# 8. Wind Forecast Data

A wind energy forecast was required at each synthetic wind energy site to adequately model operation of the power system with the hypothetical wind plants. Wind energy production is not random, and many studies have shown that accurate wind energy forecasts can reduce the costs of integrating wind energy into a power system. To adequately assess the integration cost/impact in a wind integration study, forecasting plays a major role. Four different forecasts were provided as part of the final dataset for this project. These four forecast methodologies represent the scope of forecasting possibilities.

## A. Persistence Forecast

A persistence forecast is made by assuming that the variable does not change with time. This is a simple forecast, yet it does provide a good benchmark for a look-ahead period of the next few hours' worth of wind energy production. The persistence forecast for this project provides a one-hour forecast with a two-hour look-ahead period. This time delay was chosen as a generally representative delay, allowing time for the forecast to be created and inspected and still leaving time for a human operator to react before the power had to be scheduled on the hourly timescale. However, for forecasts of more than a few hours ahead, other techniques must be used.

## B. Hourly Climatological Forecast

An hourly climatological forecast is produced by averaging the wind energy production for each hour of the day over some representative period of time; it is designed to capture the average hourly diurnal cycle for the present weather regime. Climatology is used as the basic benchmark for day-ahead prediction. For this project, each month of each year had its own climatological trace of 24 one-hour values. It is important to note that this approach actually includes "future" information in the forecast and cannot be produced operationally. However, for this project, the climatological forecast is only a baseline forecast. The mesoscale model forecast should provide more accurate forecasts than the climatological forecast, unless the mesoscale model has a large bias (i.e., needs correcting with onsite data to perform a Model Output Statistics − MOS correction).

## C. Mesoscale Model Forecast

The mesoscale model forecast represents the state of the art in day-ahead forecasting. However, it is important to note that the mesoscale model forecasts produced for this project only represent baseline accuracy for state-of-the-art forecasting. To produce the optimal forecasts, each forecast must be tuned to the data from each specific project, and such extensive data manipulation was beyond the extent of the forecasting portion of this scope of work.

The mesoscale model forecast is run in a very similar method to *Section 4, Wind Modeling*. This may raise questions about the validity of the forecast. However, since different data were used to drive the models, the results remain quite independent. The synthetic data modeling was driven using the National Center for Environmental Prediction/National Center for Atmospheric Research reanalysis dataset – an archive representing the overall state of the atmosphere over the entire planet and derived from sophisticated computer analysis of available surface and upper air observations. The forecast data modeling was driven using the Global Forecast System (GFS) information, the actual information used to perform state-of-the-art forecasting.

The mesoscale model forecasting was meant to be a smaller scope of work than the simulation of synthetic wind energy data, so the same granularity of the models could not be afforded. Instead, the models were run with a 6-km by 6-km resolution and the model output was stored at the hourly timescale, which enabled 3TIER to run the wind forecast model as a single large domain.

As mentioned before, true state-of-the-art forecasting is specifically tuned to operate optimally at the desired forecast location through the use of a MOS correction. Due to modeling over 32,000 sites, such a detailed procedure was impractical, but the mesoscale forecast is a good indicator of the kinds of forecasts obtained from a state-of-the-art model and also highlights characteristic errors – but it is not as good as a true state-of-the-art forecast.

## D. "Perfect" Forecast

The perfect forecast is an hourly resolution forecast that perfectly represents the hourly average of the six ten-minute values that comprise an hour of modeled data. It is used as an upper bound on forecast accuracy. The "perfect" forecast is an artificial forecast that cannot be produced in reality, but can be used to find the minimum wind integration cost. Wind is a variable resource, and so even if it is forecast perfectly, the resulting variation will still require some of the generators on the system to operate away from maximum efficiency (or change the generation mix). This change away from optimal operation has a cost, even if project performance is perfectly predicted. The true state-of-the-art forecast will lie between the simplified mesoscale model forecast produced for this project and the perfect forecast produced for this project.

## 9. Solar Forecast Data

This study is not just a wind integration study; the final power system analysis will also include the effect of solar energy on the power system. To this end, solar forecasts were also required as part of the integration study. However, solar energy forecasting is less mature than wind energy forecasting and consequently, new techniques had to be developed. The original scope of work asked for a day-ahead forecast and a persistence forecast. Once the requirements were better understood, NREL and 3TIER agreed that the persistence forecast would be developed at NREL, leaving 3TIER to focus its allocated time on improving the mesoscale model solar forecasts.

The approach that was used (and vetted by NREL) was a combination of methodologies principally derived from three publications:

- R. Perez et al, "Forecasting Solar Radiation – Preliminary Evaluation of an Approach Based on the National Forecast Data Base," Solar Energy, Vol. 81, Iss. 6, June 2007, pages 809-812
- R. Perez et al, "A new operational model for satellite-derived irradiances: description and validation," Solar Energy, Vol. 73, Iss. 5, November 2002, pages 307-317
- P. Banacos, "BTV_SkyTool Documentation," National Oceanic and Atmospheric Administration and National Weather Service Smart Tool Repository, www.mdl.nws.noaa.gov/~applications/STR/generalappinfoout.php3?appnum=1104

The radiation reaching the earth's surface can be represented in a number of different ways. Three values have been used to represent the solar irradiation for this project: global horizontal irradiation (GHI), direct normal irradiation (DNI), and diffuse irradiation.

The GHI is the total amount of shortwave radiation received from above by a horizontal surface. This value is of particular interest to photovoltaic installations and includes both direct radiation and diffuse radiation.

The DNI is the amount of direct radiation received per unit area by a surface that is always held perpendicular (normal) to the rays that come directly from the direction of the solar disk in the sky. By keeping the surface normal to the incoming radiation, you maximize the amount of energy received. This quantity is of particular interest to concentrating solar installations and installations that track the position of the sun.

Diffuse Irradiance is the amount of diffuse radiation received per unit area by any surface that is not subject to any shade or shadow. Since the diffuse component of radiation is more or less equal from all directions, there is no distinction between a normal and horizontal component.

The 2007 paper by Perez et al shows how to calculate the actual global horizontal irradiance (GHI) given the clear sky GHI and the sky cover. The clear sky GHI was calculated using a function from the 2002 paper by Perez et al. The procedure is a function of point elevation, solar zenith angle, Linke turbidity index, and elevation-adjusted air mass. The sky cover is derived using the technique described by Banacos to convert the simulated relative humidity to sky cover.

With the GHI calculated, the direct normal irradiance (DNI) and diffuse irradiance were calculated. The DNI was calculated as described in the 2002 paper by Perez et al, and is a function of GHI, solar zenith angle, elevation, and the day of the year. Diffuse irradiance was calculated by subtracting the DNI divided by the cosine of the solar zenith angle from the GHI.

During validation, it was found that the (empirically derived) diffuse irradiance calculation sometimes produced unrealistically low values. A second pass was used in such cases that calculated the minimum diffuse irradiance using a physically based equation, also described in the 2002 paper by Perez et al. This minimum diffuse irradiance value was used to recalculate the DNI. GHI was assumed to be correct and remained unchanged.

## 10. Graphical Dataset Interface

The final stage of the project was to produce a web-based time-series database interface to host the modeled wind data and allow stakeholders and the general public to have access to the modeling output for the 32,043 synthetic wind sites. The software distribution was written such that the visual tiles were pre-rendered as flat images with the icons merged onto the background. Therefore, rather than having to render thousands of images every time the map is moved, the navigation around the map is comparably seamless.

Even though the images are pre-rendered, the map is still interactive. The top left corner shows four arrows, which move the map (although the map can also be moved by clicking the left mouse button and dragging). Just under this is the zoom feature that can be used to zoom in and out (the scroll-wheel on the mouse or double-clicking the left mouse button also change the zoom). Furthermore, each icon still responds to the mouse cursor or can be located using a simple site ID number search. Fig. 4 shows an interface screenshot. When a site is selected, a larger turbine image is displayed at that location and the metadata from the site is shown, including:

- Site ID number,
- Location in latitude and longitude,
- Average annual capacity factor,
- Power density,
- Wind speed,
- Elevation,
- State.

The option is also provided to download the dataset for individual locations for 2004, 2005, or 2006. The dataset will be downloaded as a simple comma-delimited file, identified by the site ID number. The entire metadata file containing the summary information for each site can also be downloaded from the gray bar on the left of the page.

At the top of the display there is a blue bar titled *Capacity:* that allows the user to toggle between displaying all locations or some subset of locations. The blue bar also acts as a legend; every turbine icon is color-coded according to the site capacity factor. Fig. 4 shows an example region where at least one turbine of every color is visible.

The entire process behind the interface was carefully designed for ease of installation. In fact, the distribution is entirely "plug-and-play;" the files need to be in the proper directories, but once in the correct directory structure, the fully interactive website could be served from a static storage device. This means that NREL need only host the *index.html* file in a location that can be accessed externally (found in */nrel-distrib/Web_nrel/index.html*) and the graphical dataset interface will operate as intended.

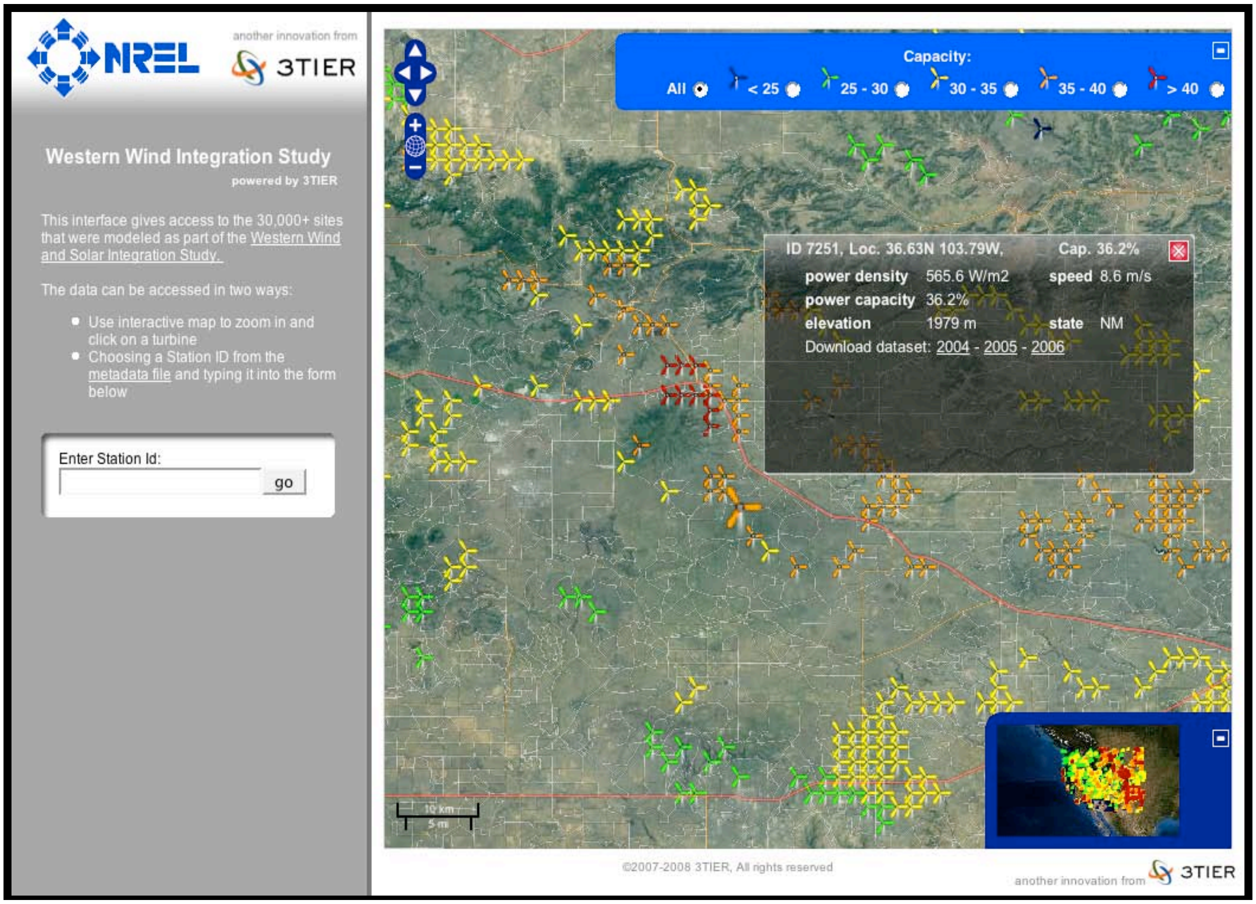



**Fig. 4 –** A screenshot of the graphical dataset interface showing a central orange turbine selected from within New Mexico.

## Appendix – Implementing the Graphical Dataset Interface

The Graphical Dataset Interface is designed to run on any static web server on any computing platform. The Graphical Dataset Interface was delivered with the directory structure already in the proper configuration, as shown below:

| | |
|---|---|
| /nrel-distrib/Web_nrel/ | Document Root |
| /nrel-distrib/Web_nrel/scripts/ | Javascript helper files |
| /nrel-distrib/Web_nrel/json/ | Static index files |
| /nrel-distrib/Web_nrel/images/ | Static imagery |
| /nrel-distrib/Web_nrel/css/ | Style information |
| /nrel-distrib/Web_nrel/cache/tiles/merged/ | Static tile information |
| /nrel-distrib/Web_nrel/data/ | Wind site data (under sub-directories by year) |

To host the Graphical Dataset Interface, simply configure the *document root directory* of your web server to point to "/nrel-distrib/Web_nrel/".

E.g. Apache Web Server configuration:
*DocumentRoot {path to directory}/nrel-distrib/Web_nrel/*

**Note:** Different web servers may require different commands to set the document root directory. Please see the documentation for your particular web server for more information.

| **REPORT DOCUMENTATION PAGE** | *Form Approved* *OMB No. 0704-0188* |
|---|---|



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| March 2010 | Subcontract Report | 10/2007 to 3/2009 |

**4. TITLE AND SUBTITLE**
Development of Regional Wind Resource and Wind Plant Output Datasets: Final Subcontract Report, 15 October 2007 – 15 March 2009

**5a. CONTRACT NUMBER**
DE-AC36-08-GO28308

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
3TIER

**5d. PROJECT NUMBER**
NREL/SR-550-47676

**5e. TASK NUMBER**
WW10.4311

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
3TIER
2001 Sixth Avenue, Suite 2100
Seattle, WA 98121

**8. PERFORMING ORGANIZATION REPORT NUMBER**
AAM-8-77556-01

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401-3393

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

**11. SPONSORING/MONITORING AGENCY REPORT NUMBER**
NREL/SR-550-47676

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

**13. SUPPLEMENTARY NOTES**
NREL Technical Monitor: Debra Lew

**14. ABSTRACT** *(Maximum 200 Words)*
This report describes the development of the necessary and needed wind and solar datasets used in the Western Wind and Solar Integration Study (WWSIS).

**15. SUBJECT TERMS**
WWSIS; datasets; WestConnect; wind; integration study; utilities; power industry; renewables

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

**Standard Form 298** (Rev. 8/98)
Prescribed by ANSI Std. Z39.18