WindLogics®

1021 Bandana Blvd. East
St. Paul, MN, 55108

Tel: 651.556.4200 Fax: 651.556.4210

The Long-Term Wind Resource

Part 1 - Data Sources for Predicting the Performance of Wind Plants

By Dennis A. Moon, Ph.D.

Table of Contents

The Long-Term Wind Resource
Part 1 - Data Sources for Predicting the Performance of Wind Plants

By Dennis A. Moon & Scott E. Haynes

Introduction

Understanding the long-term wind resource is critical to developing and operating wind power plants. A growing body of work has shown the need for long-term reference wind data covering a span of 30 years or more. Due to natural climatic variability on the decadal scale, shorter time spans may provide misleading results when characterizing long-term mean wind speeds and understanding wind variability at a site.

To understand the value of wind resource assessment over long time scales, WindLogics has been doing extensive research to compare different sources of long-term data, studying their correlation with multi-year tall tower data and analyzing the errors associated with wind resource estimates based on different sources and correlation techniques. A theme that has emerged from this work is that regardless of the type of reference dataset or the methods used to extend the data to reflect long-term trends at the site, it is the *predictive ability* of the data and methods used that matters most. In other words, statistics that characterize the correlation between the site and reference datasets are not sufficient for evaluating the relationship between the site and reference datasets – the underlying distributions (time-series, histograms, wind roses, etc.) *must* be examined as well.

A summary of these results is discussed below, and a more detailed technical paper is available from the authors.

Long-term Data Sources

Obviously, some long-term data source is required as a starting point. Sources of long-term reference data for wind plant assessment include airports and other weather service surface sites, weather surface radiosonde (weather balloon) measurements, the NCAR/NCEP Global Reanalysis dataset (RNL), the NCAR/NCEP North American Regional Reanalysis dataset (NARR), and the European center for Medium-Range weather Forecasting (ECMWF) Global Reanalysis dataset (ERA40).

Surface METAR Observations

Observed wind speeds from surface observing (METAR) stations are frequently used for long-term reference data. If the proposed wind plant site is sufficiently close to the surface site, and the location is sufficiently representative of the on-site conditions, this can yield useful information regarding the wind variability.

WindLogics Inc.
1021 Bandana Blvd. E., St. Paul, MN 55108 Tel: 651.556.4200 Fax: 651.556.4210
www.WindLogics.com

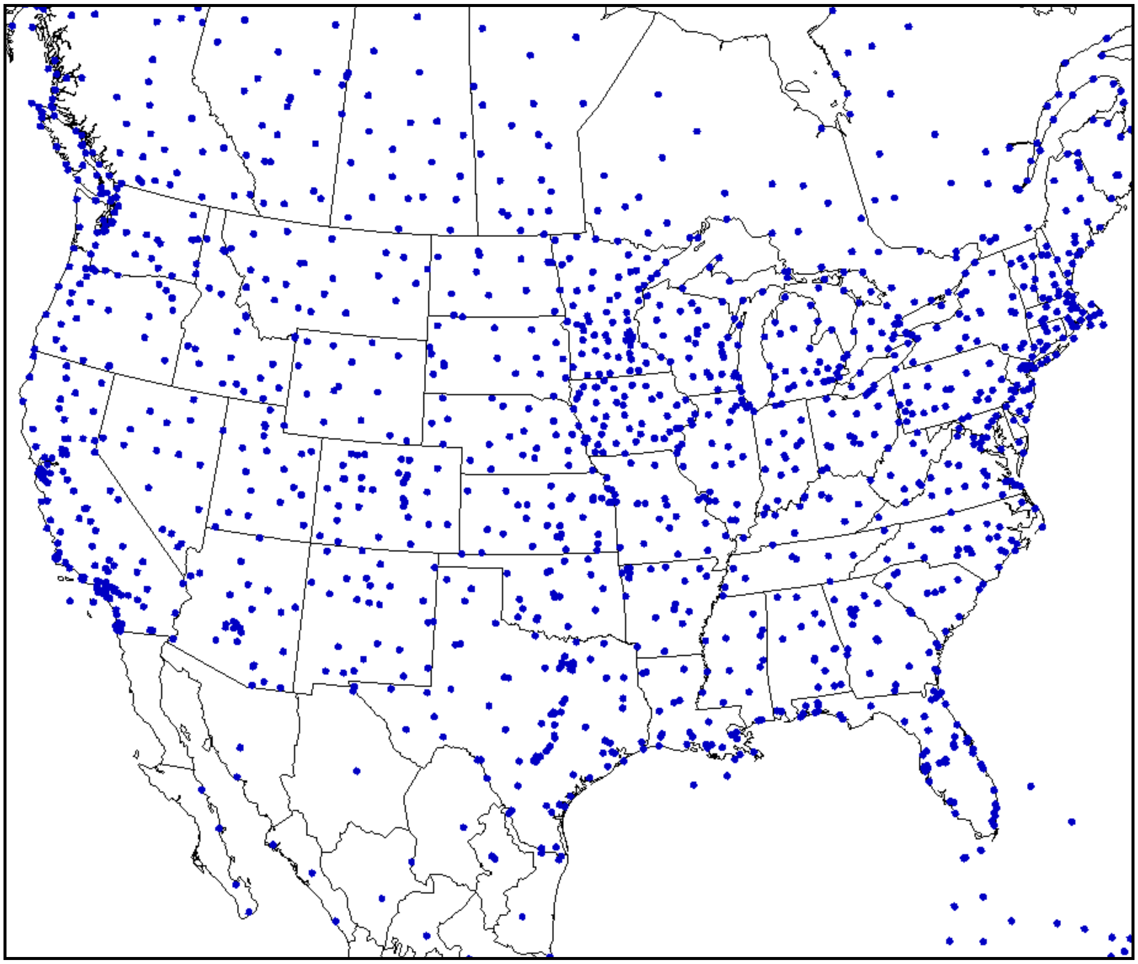

Figure 1. Location of surface measurement (METAR) stations.

METAR stations in North America are typically located at airports and observations are usually collected at hourly or 20-minute increments at a height of 10 meters (m) Above Ground Level (AGL). The length of observations varies between stations, and it is typical to see changes in equipment or location over a 10 to 12 year range. Many of the most recent changes were part of the Automated Surface Observing System (ASOS) upgrade in the 1990s.

The brevity of the METAR observation period makes it impossible to infer information about variability over decadal timescales. It also makes it difficult to accurately estimate the standard deviation in the annual mean speed; a quantity frequently used in estimating the predictive intervals (P75, P90, etc.) for wind energy financing.

Another major challenge stems from differences in the diurnal variation of the speed at different heights. At the majority of sites, the maximum wind speed at the 10 m AGL occurs during the daytime. The pattern at 80 m AGL (a typical hub height) is very different, with maximum wind speeds at many sites occurring during the night and minimum wind speeds during the day.

Figure 2 illustrates this phenomenon with data from a tower near Finley, ND. The times are in local standard time, and the green line indicates the wind speed at 10 m AGL while the purple line shows the speed at 80 m AGL. This dramatic difference based on height can make it difficult to use 10 m data (or even data at 40 m) as a long-term reference, especially when it is important to understand the daily energy production pattern.

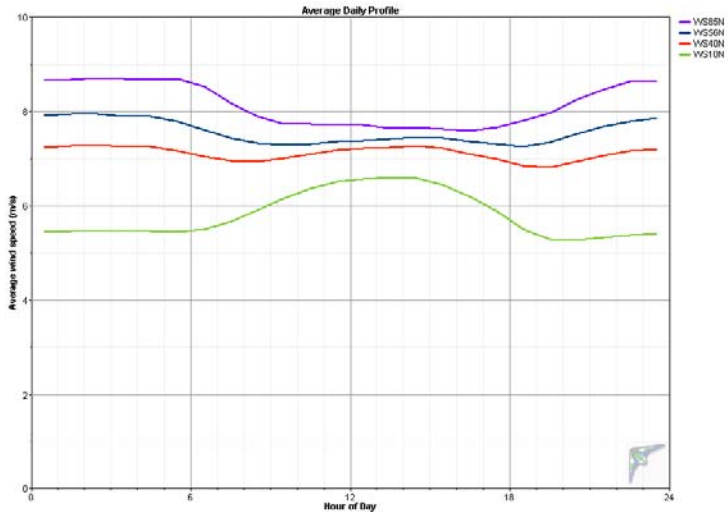

Figure 2. Tower-based diurnal pattern near Finley, ND at 10, 40, 56, and 85m AGL

Radiosonde Observations

Radiosonde instruments on weather balloons are used to collect information about atmospheric thermodynamic parameters and winds in the troposphere and lower stratosphere. The spatial density (Figure 3) and the sampling period (typically 12 hours) are coarser than the surface station METAR network.

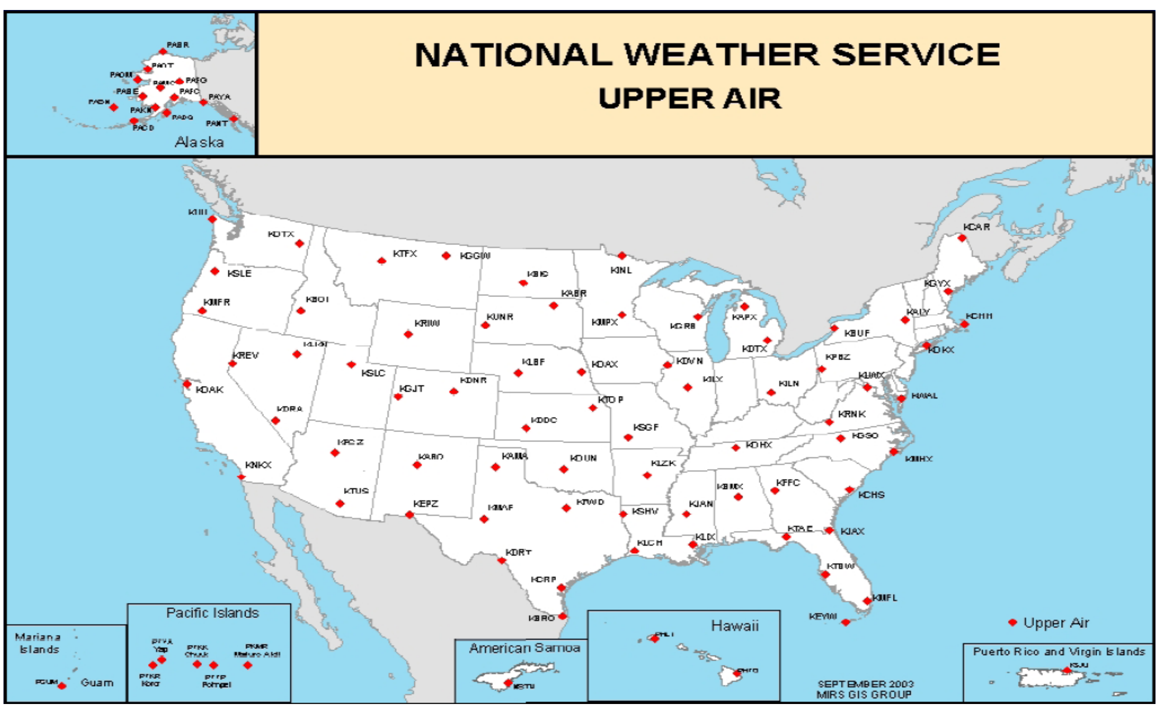

Figure 3. Location of radiosonde observation stations.

The most complete and recently compiled source for radiosonde data is the Integrated Global Radiosonde Archive (IGRA). The average IGRA inter-station distance is on the order of 300 kilometers (km) with the length of record being quite variable, with some stations having data back as far as the 1950s.

Unfortunately, great care must be taken in ensuring consistency of the data over the observation period due to re-location and changes in equipment. Since 1990, virtually all of the radiosonde sites in the United States have experienced changes to equipment and nineteen of the sites have also changed location.

According to the NOAA 2005 Radiosonde Replacement System (RRS) Workstation User Guide, "*The main obstacle to using radiosonde data in long-term climate change research is the effect of changing instrumentation and processing… homogenizing radiosonde data is evidently more difficult than other sources of data as the results from various radiosonde groups around the world do not converge the way SSTs [sea-surface temperatures] and land surface temperature adjustments do.*"

Regarding the use of long-term radiosonde wind data, the same NOAA report states, *"Wind data can be important in climate analyses that look at fluxes of energy. Most analyses of fluxes don't use radiosonde (wind) data directly. Instead they use reanalysis (wind) data which has assimilated the data into model dynamics."*

Obtaining consistent data at wind turbine hub height, or something close to it, is another challenge when attempting to use radiosonde data as a long-term wind speed reference. There is no single level in the dataset that can consistently be used, so the data are typically interpolated to a level that will usually have valid data both above and below. In practice, a height of 700-800 meters is frequently used. In some cases, wind data from a given pressure level (i.e. 850 millibars - mb) has been used as a reference, but because the actual height of this pressure surface above the ground varies greatly with the passage of weather systems (up to several hundred meters), this is a questionable practice.

Furthermore, the entire NOAA upper air measurement system is in the process of being replaced with the Radiosonde Replacement System (RRS), and of all the weather variables measured, the wind data will be most strongly affected by this change. Currently, the radiosondes are tracked by measuring an azimuth and elevation angle and inferring the position from the range and direction (RDF location). In the new system, the winds will come directly from GPS-based instruments. The RRS Workstation Users guide states that *"Accuracy of the wind data will improve as much as five times that of wind calculations using RDF equipment".* While this is a welcome improvement, homogenizing current observations with previous years will become even more problematic.

Reanalysis Datasets

There are several reanalysis datasets currently available for long-term meteorological study. Reanalysis datasets are produced by national weather modeling centers to provide an accurate and consistent gridded representation of the past weather. The quality of these datasets is critical to research in climate change and variability, and significant resources are devoted to this research work in government laboratories, academic institutions and private companies worldwide.

Background

Reanalysis datasets are created using a combination of the best available modeling technology and historical weather data, with the data carefully rechecked for quality using the best available methods. The entire past period is modeled using the same state-of-the-art weather model and the observational data are used to adjust the model fields on an hourly basis. The model imposes fundamental physics-based constraints on the variables, ensuring that the weather fields are dynamically consistent with conservation of energy, mass, momentum, and the equations of state. The modeling process acts as something of a "flywheel", rejecting the influence of spurious observations. The use of a fixed modeling system over the entire time-span also helps to ameliorate the effects of changing observing systems, such as changes to specific equipment or the increasing availability of satellite data.

As with any reference data source, the process is not perfect and changes in observing systems can create artifacts, so the data should be used and interpreted with care. But because of the three-dimensional "gridded" nature of these datasets, they are powerful tools for understanding the *patterns* and *regional relationships* that drive the wind flow in the atmosphere. As we will show below, this has dramatic benefits over "one-to-one" linear methods for extending onsite observations from wind energy sites. By relating to multi-gridpoint patterns with appropriate techniques, we can both achieve better correlation to onsite data and reduced sensitivity to potentially spurious speed trends at a given grid cell.

Available Datasets

The reanalysis datasets vary in time span, domain, spatial resolution and time resolution. The most studied and used reanalysis dataset is the NCEP/NCAR Global Reanalysis (RNL). This dataset has global coverage from 1948 to present, but with relatively coarse spatial resolution of 2.5 degrees (roughly 175 miles or 280 km). This resolution is insufficient to directly resolve mesoscale flow patterns, such as sea breezes, slope-related flows, mountain influences or thermally forced flows, however, as will be seen, it works quite well for long-term normalization.

The ECMWF ERA40 dataset shares the same resolution and grid as the RNL dataset, but the ERA40 dataset only extends from 1957 to 2002. Since the wind energy industry normally deals with onsite observations taken in the last few years, it is typically not possible to establish a relationship to the ERA40 data.

The new North American Regional Reanalysis dataset uses a 32km grid (about 20 miles), but only for the region shown in Figure 4 and only for the period from 1979 to present. This dataset has yet to be widely applied to climatological normalization of wind data, although our new results indicate it can be used successfully.

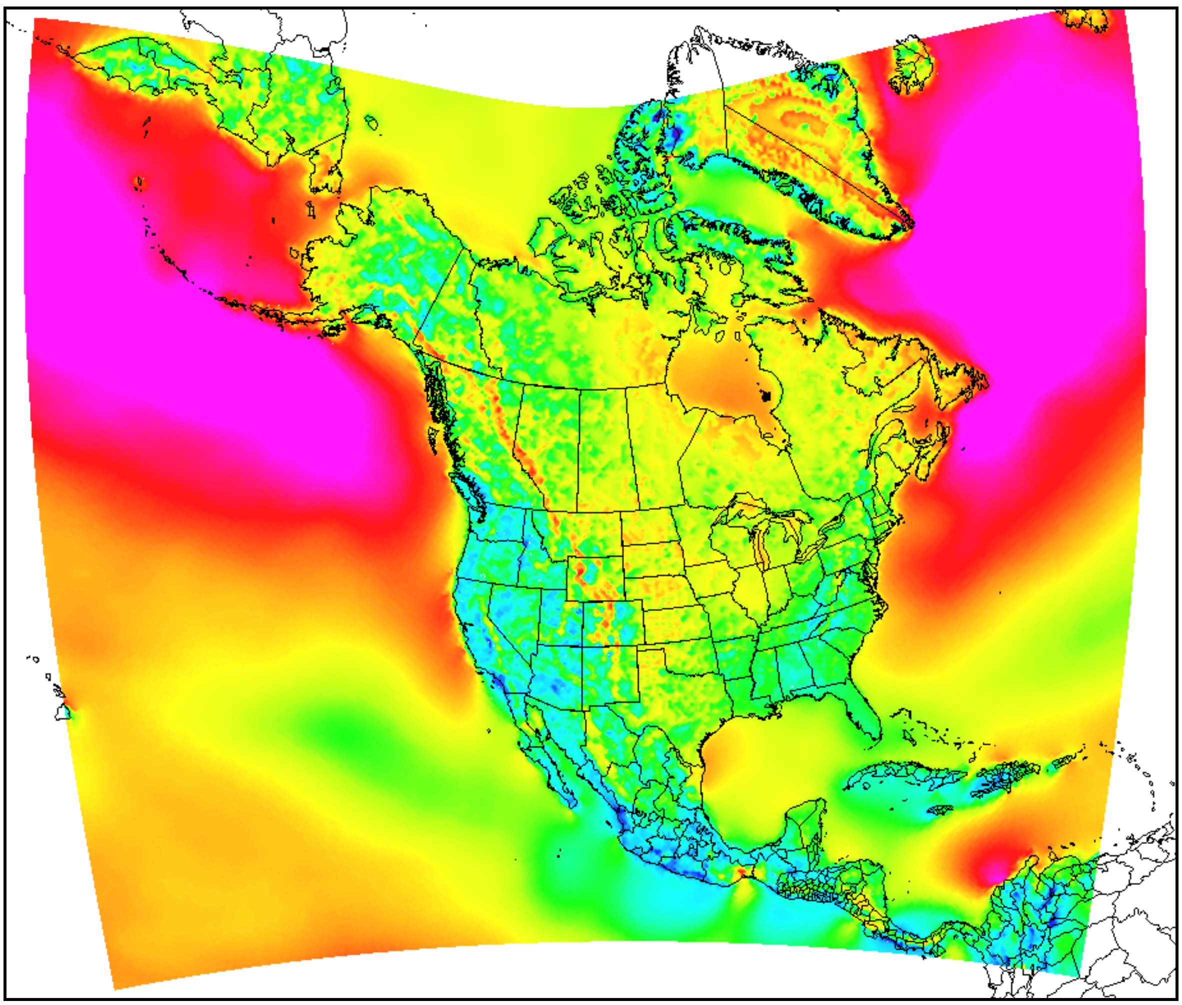

Figure 4. NARR grid coverage.

Consistency of Reanalysis Datasets

The ECMWF ERA40 dataset shares the same resolution and grid as the NCAR/NCEP Reanalysis (RNL) dataset, but the ERA40 dataset only extends from 1957 to 2002. Since the wind energy industry normally deals with onsite observations taken in the last few years, it is typically not possible to establish a relationship to the ERA40 data.

But because the ERA40 was independently developed with a different model and data quality control, it provides an excellent opportunity for comparison between the two datasets. A consistency check can highlight areas of agreement and disagreement between them.

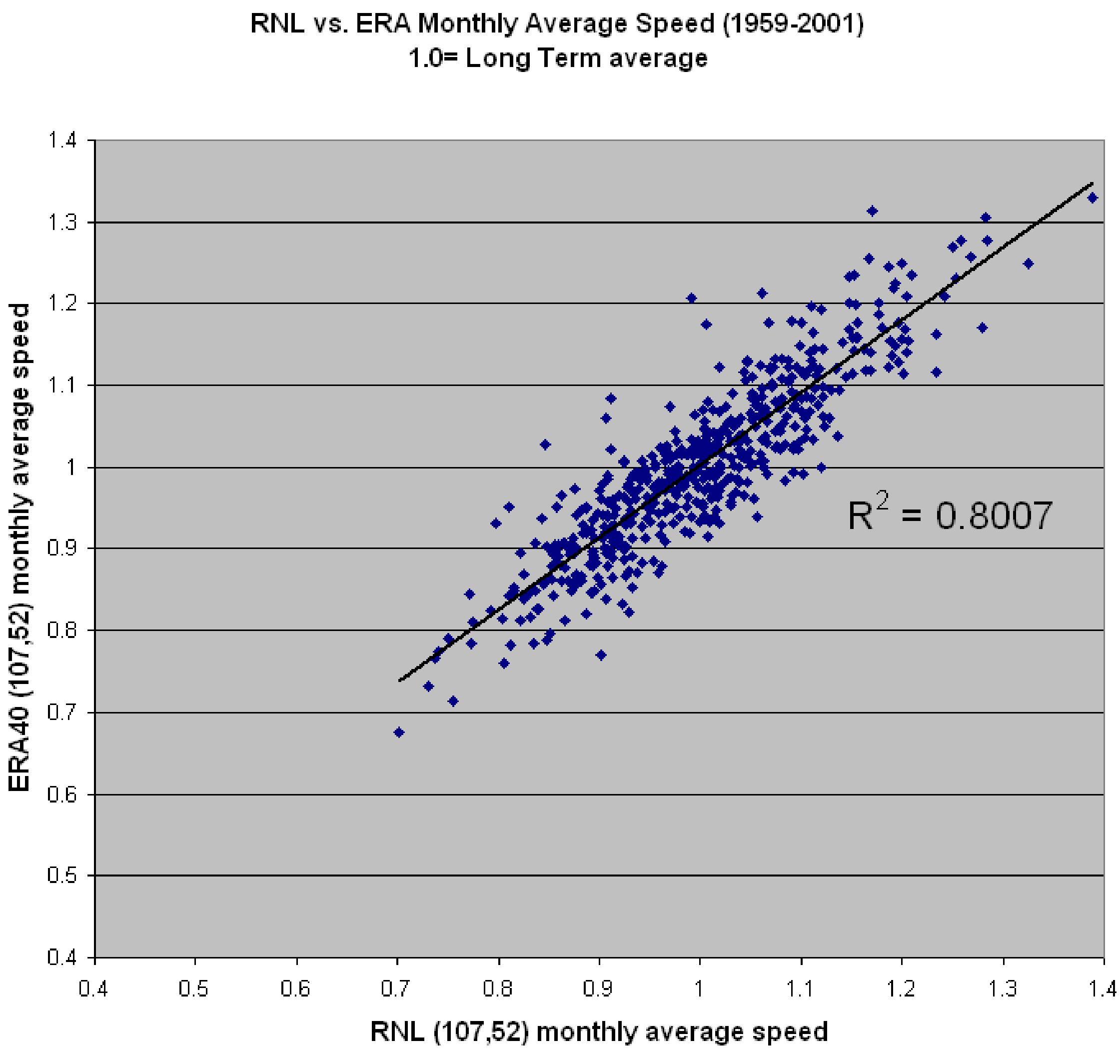

Figure 5. Sample scatter plot showing monthly mean wind speeds from RNL and ERA40 at cell (107,52).

Figure 5 shows a sample scatter plot of monthly mean wind speed for the two data sources for a grid cell in Kansas for all months from 1959 to 2001. The coefficient of determination (R^2) is a widely used measure of the degree of correlation between two datasets and describes the degree to which variations in one dataset account for the variability in the other. Perfectly correlated data, (all the points of the scatter plot falling exactly on a line) would be indicated by a value of 1.0, while a value of zero implies that dataset “A” does no better job of accounting for the variability in dataset “B” than would be obtained by simply using the mean value of dataset B.

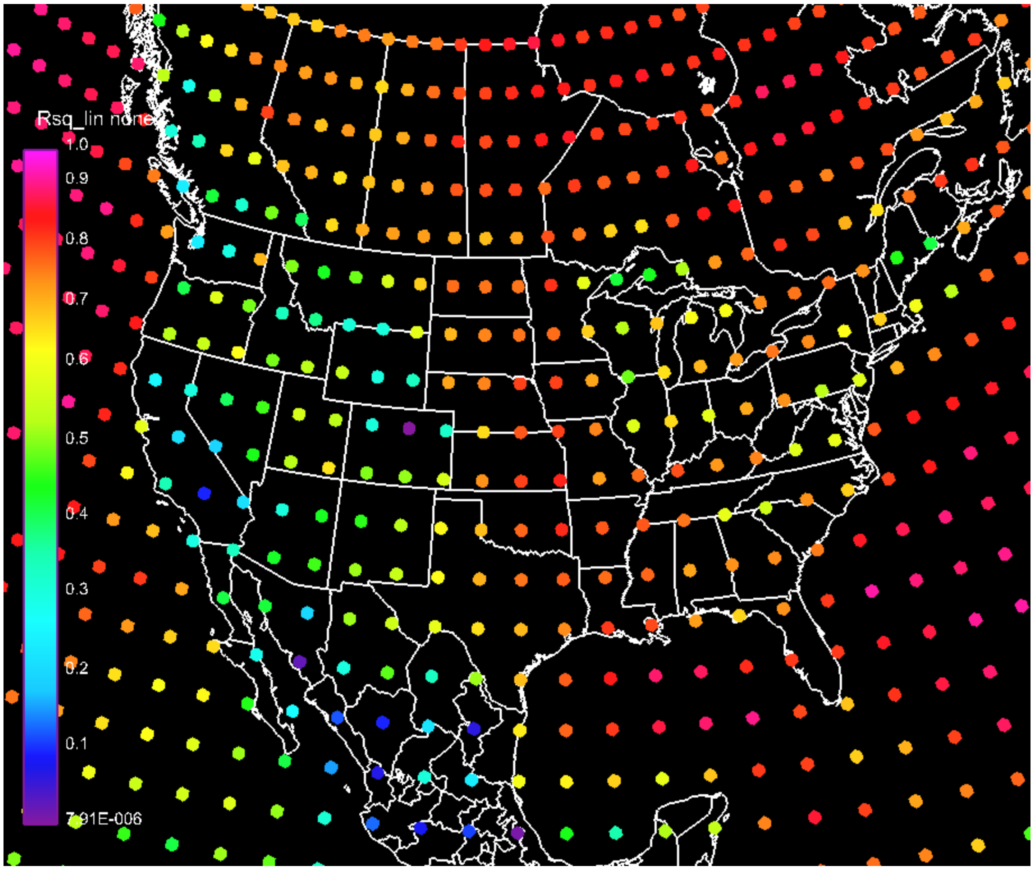

Figure 6. Coefficient of determination (R^2) between RNL and ERA40 monthly mean wind speeds for the period 1958-2001.

Figure 6 shows the R^2 value between the two datasets across North America. The agreement is quite good in general, with an average R^2 value of 0.65 and the majority of cells exceeding an R^2 value of 0.7. Not surprisingly, the correlation is somewhat weaker in mountainous areas. The cell in Northern Colorado stands out as a singular example where the coefficient of determination is essentially zero, indicating significant disagreement between the two data sources.

Correlation of Reference Datasets versus Public Tower Data

Thirty publicly available “tall-tower” sites, each with between three and seventeen years of wind speed data, were used to test various reference sources and methods. The process was designed to objectively compare the predictive ability of various reference data sources including reanalysis, METAR and radiosonde data. Many additional details about these towers and tests are available in the complete technical report.

The test first simply measured the degree of correlation, based on monthly mean speeds, between the tower sites and various reference datasets. Radiosonde data were interpolated to 750 m AGL and the site or grid point showing the best correlation to the tower data was selected for each of the data sources.

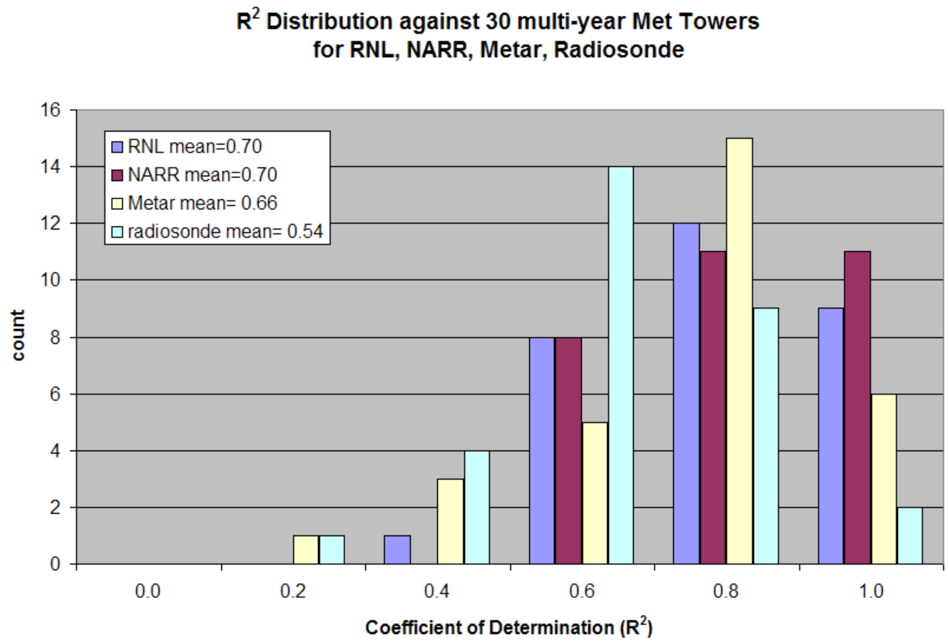

Figure 7. Distribution of Coefficient of Determination between 30 public tower sites and four sources of reference data

Figure 7 shows that the RNL and NARR datasets have the strongest correlation with tower data while the IRGA radiosonde has the weakest correlation with tower data. The average R^2 for both RNL and NARR datasets was found to be 0.70 while the average R^2 for METAR and IGRA datasets was found to be 0.66 and 0.55 respectively.

Downscaling

To use such reanalysis data in valid ways, it is critical to use appropriate *downscaling* methods. To extract a specific point from a reanalysis dataset and use it as if it were an airport anemometer can be inappropriate and misleading. Those skilled in meteorological and climatological modeling know this (or at least, they should).

More sophisticated downscaling methods relate the large-scale patterns in the reanalysis dataset to smaller-scale features at a particular site. This is an active area of research in the climate community, and while several statistical methods have been shown to exhibit considerable skill, the common thread with the successful methods is that they consider the relationship between the gridded *pattern* and the onsite data, rather than simply extracting a time series from the data and correlating with onsite measurements.

As we will show next month in Part 2 of this article, non-linear methods and pattern-based techniques from the data-mining field, such as Support Vector Regression, provide significant improvements over standard linear regression approaches. Methods for detection of patterns in regional and three-dimensional spatial data, as in advanced patent-pending methods developed by WindLogics, are demonstrated to work very well even in areas where a specific data point is suspect or in situations where linear methods show poor correlation.

Summary for Part 1

Several sources of long-term reference data were described and compared above, and next month we will show results from investigation of various correlation techniques. While all of the data sources must be applied with care and can be susceptible to false trends due to changes in the instrument system, the results will clearly demonstrate the predictive abilities of various combinations of data sources and methods as validated against long-term measured tall-tower values.

We will show how the proper use of reanalysis datasets – using appropriate downscaling methods and techniques that exploit the powerful regional, 3-dimensional and weather pattern attributes for which such datasets were designed – can provide results with very high predictive abilities even in areas where single-point analysis may appear to show inconsistent trends or low statistical correlations..

Dr. Dennis Moon is the Chief Scientist at WindLogics Inc. He is located in the WindLogics Sciences Center in Grand Rapids, Minnesota. Dennis can be reached at dmoon@windlogics.com. A research report with additional detail on these topics was published in the proceedings of the AWEA WINDPOWER 2006 conference and is available from the authors upon request.

References

[1] Albers, A., 2006: Long Term Variation of Wind Potential: How Long is Long Enough?, *European Wind Energy Association 2006 Annual meeting*

[2] Albers, A., 2004: Long Term Variation of Wind Potential: Are we Moving into a Low NAO Cycle, *European Wind Energy Association 2004 Annual meeting*

[3] Pryor S.C., Barthelomie R.J., Schoof J.T., 2004 : How coherent is the inter-annual variability of wind indices across Europe and what are the implications for large scale penetration by wind energy of electricity markets., *European Wind Energy Association 2004 Annual meeting*

[4] Johansson C, Bergstrom H, 2004: Variability in Energy Content over Scandinavia, a 101-year perspective, *European Wind Energy Association 2004 Annual meeting*

[5] Peterson T.C., Durre I., 2004: A Climate Continuity Strategy for the Radiosonde Replacement System Transition, *AMS Eighth Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surfaces*

[6] NOAA, U.S. Dept of Commerce, 2005: Radiosonde Replacement System (RRS) Workstation User Guide