9.10 ASSESSING THE LONG-TERM REPRESENTATIVENESS OF SHORT WIND RECORDS

Karsten A. Shein *
NOAA National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

Wind speed data are utilized for many purposes, including dispersion modeling, wind energy, and structure loading. However, due to restrictive time frames for some applications, these approaches must utilize relatively short wind speed data records to model a longer-scale wind climatology.

While a number of methods exist for constructing an extended time series from a shorter record of wind speeds, such as the Measure-Correlate-Predict approach of Derrick (1992), or the joint probabilistic technique presented by Garcia-Rojo (2004), such methods require establishing a stationary empirical relationship between the short wind record and a corresponding record from a nearby so-called reference station. These approaches assume a stationary relationship between the wind speeds at the two locations and do not account for the possibility that the wind speeds of the short record may not be statistically representative of the long-term wind speed climatology at that location. Shein and Robeson (1995) have illustrated the non-stationarity that exists in long-term wind speed records and have shown that substantial inter-annual variability may exist in annual density function parameters.

In other cases, a researcher may wish to determine whether a regional wind regime for a given period of time might be representative of the longer-term wind field behavior over that region. Because of the continuous spatial nature of wind, such analysis may permit a short-term record from a dense network of stations to be evaluated from a much sparser network of long-term stations. For example, Shein (2005) utilized 30-year records of 20 First-order National Weather Service stations to establish the representativeness of two years of wind speed data over the Great Lakes region from 113 ASOS stations.

If the period of record for the short series is in fact anomalous, it may be erroneous to assume that any statistics or empirical relationships obtained from those data would apply to speeds at other times. Additionally, a consideration that often is overlooked when comparing a short record of climatological data to a longer one is that the two series may have been collected using different methods and systems. This may lead to unaccounted bias in the data, which may invalidate assumptions about representativeness, empirical relationships, or subsequent analysis.

Thus, it is important to establish how representative a short wind record is relative to the long-term wind series. Assessing long-term representativeness can be accomplished by first ensuring the homogeneity of the short- and long-term data, and applying a goodness-of-fit test to the density functions from the series.

2. DATA AND METHODS

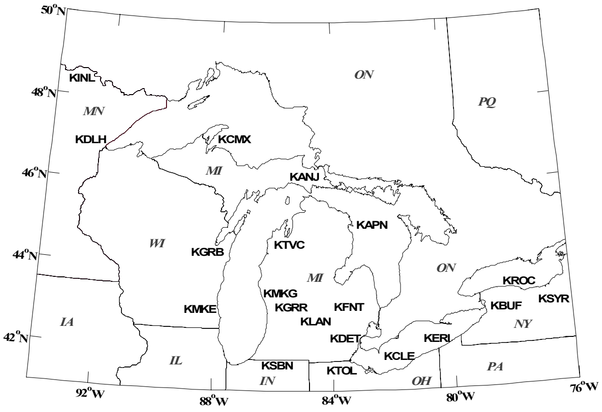

Fig. 1. Locations of the 20 anemometer sites. Each is indicated by its 4-letter ICAO identifier.

Hourly ASOS wind speed data at 20 stations in the Great Lakes region of the United States (Fig. 1) for the period 1 Jan. to 31 Dec. 2003 was compared to hourly First-Order National Weather Service observations at those same locations for the period 1 Jan. 1961 to 31 Dec. 1990. The ASOS data was obtained in raw format from NOAA's National Climatic Data Center (NCDC) and processed by the author to remove obvious errors (see Shein, 2005). Pre-ASOS data were

---

* *Corresponding author address:* Karsten A. Shein, NOAA/NESDIS/NCDC, 151 Patton Ave., Asheville, NC 28801; e-mail: karsten.shein@noaa.gov.

extracted from the Solar and Meteorological Surface Observation Network dataset, or SAMSON (NREL, 1993). Both the ASOS data and the SAMSON dataset are available from NCDC.

2.1 *Homogenization*

ASOS wind speed observations are made at 10 meters above the ground and may be reported every 20 minutes or more depending on weather conditions. At the least, observations are reported every hour, normally between 15 and 5 minutes of the top of the hour. Each observation is actually a report of a two-minute moving average of 5-second averages of 1-second instantaneous wind speed observations. Pre-ASOS First-Order observations, on the other hand are simply 2-minute running means taken around 5 minutes before the top of the hour (Koren, 1973). Unlike the standardization of the ASOS anemometers, the heights of pre-ASOS anemometers varied widely. At the 20 stations used in this study, the First-Order anemometers ranged in height from 5.79 meters to 25.60 meters. At 14 stations, anemometer heights changed during the period of record.

To homogenize the two data sets, the SAMSON data were adjusted to a 10 meter height using the wind profile power law (Eq. 1) with an exponent of 1/7th, where u_Z is the wind speed at height Z, and u_{10} is the estimated wind speed at 10 m height (Touma, 1977).

$$u_{10} = u_Z (10/Z)^{1/7} \tag{1}$$

Thereafter, the ASOS data was reduced to a single observation per hour. To correspond with the SAMSON observations, only ASOS observations occurring within 10 minutes of the top of the hour were included. Because there is no overlap of data between the two series, it was not possible to examine the two series for systematic bias as a function of the changeover from old to new instrumentation around 1996. However, EPA (1997) demonstrates minimal difference between ASOS and pre-ASOS wind speeds.

2.2 *Establishing Representativeness*

Two approaches to establishing the long-term representativeness of the short-term wind speed series are examined. The first is a graphical examination of the short record relative to the averages of the station's long record. This analysis is accomplished by plotting the time series of each record (short and average-long) against each other. Additionally, the cumulative probability distributions of the short and long records are plotted against each other. Large deviations in the two plots would tend to indicate a lack of agreement between the two.

Unfortunately, graphical methods, while useful in an exploratory sense for quickly dismissing obviously unrepresentative data, do not provide a statistical measure of the confidence that might be ascribed to a statement of representativeness. Thus, statistical approaches are used. The first is to calculate a summary statistic, such as the mean, from both the short and long series. The means are then tested for similarity using the Student's T-test (Rogerson, 2001). However, to ensure the validity of the results, the assumptions of the T-test must be satisfied. Namely, the data must be independent and come from a normal distribution. Wind speeds generally do not adhere to either assumption, and thus the data must be transformed to a normal distribution, and a random sample be taken from the series to minimize the influence of serial autocorrelation.

A more appropriate statistical alternative to assessing the representativeness of the short record relative to the long record is to subject the empirical distribution of the short record to a goodness-of-fit test against the probability density function of the long series.

Although wind speeds distributions are widely approximated by the highly versatile Weibull distribution (a variation on the Gamma distribution), other density functions have been used with success. Thus it is necessary to utilize a goodness-of-fit test that is not dependent upon the distribution being modeled. Such a test is the Kolmogorov-Smirnov (K-S) test, which identifies the maximum difference between the cumulative probability of a wind observation (*S(u)*) and the theoretical or reference cumulative probability of that speed (*F(u)*) (Snedecor and Cochran, 1989).

$$KS = \max\left|S(u) - F(u)\right|, \text{ for all } u \tag{2}$$

There are two important assumptions of the K-S test. The first is that the test distribution must be fully specified. This means the distribution parameters may not have been estimated from the data being evaluated. Because the data being evaluated are from 2003 and the data used to specify the distribution end in 1990, this assumption is satisfied. The second assumption is that the data are independent. To ensure independence, repeated random samples of n = 720 ASOS values (approx. 1 month) are drawn until the lag-1 autocorrelation coefficient is not significantly different from zero (T-test for Pearson's r at α = 0.05). The empirical cumulative distribution of the ASOS sample is then compared against a best fit Weibull distribution (by the method of Justus et al., 1978) from the entire SAMSON series, with one exception.

It should be noted that although the Weibull distribution is zero-bounded, it is not tractable at zero. Additionally, ASOS winds below 2 kts are reported as calm and given a value of 0 kts. Therefore, all adjusted SAMSON speeds below 2 kts were also set to zero, and only non-zero values were used to graphically and statistically evaluate the goodness of fit of the distributions. However, the percentage of calm (zero) values was also evaluated by simple comparison.

3. RESULTS AND DISCUSSION

The graphical examination of cumulative ASOS distributions relative to the corresponding SAMSON Weibull distributions show that the 2003 ASOS data do appear to adhere to the 30-year SAMSON wind speeds (Fig. 2). It appears that some stations do experience small, but notable differences between the two distributions, particularly in the tails of the distribution. However, given the general agreement in the graphical analysis, it would not appear to be inappropriate to use the year of ASOS data as representative of the long-term wind regime at these stations or over the region at large.

The results of the application of the K-S test to empirical distributions of random samples of the ASOS data corroborated the graphical fit analysis. Each station was subjected to 1000 trials and the results are presented in Table 1. The K-S test was evaluated for statistical significance at α = 0.05. The null hypothesis is that the data come from the specified distribution. Thus, a good fit is one that fails to reject the null hypothesis at the stated significance level.

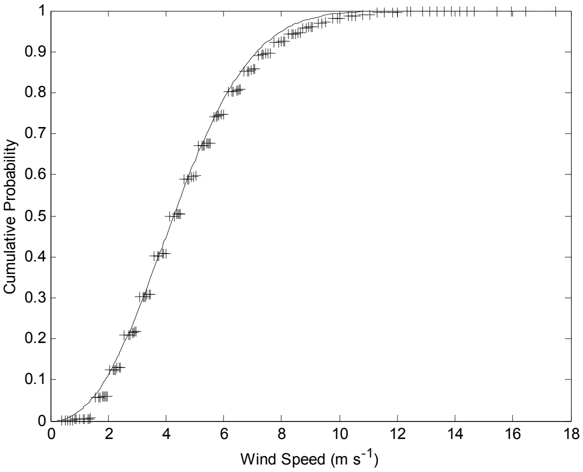

Figure 2. Graphical comparison between the 2003 (+) and 1961-1990 Weibull (-) cumulative probability distributions for all non-zero data at Lansing, MI (KLAN).

Table 1. Kolmogorov-Smirnov test results for 20 Great Lakes (US) stations, comparing 2003 and 1961 – 1990 hourly wind speed distributions. A probability level < 0.05 is necessary to reject the null hypothesis.

Station (ICAO)	Median Test Val.	Median Prob. Level	% Ho Accept
KANJ	*0.638*	*0.23*	*99.6*
KAPN	*0.054*	*0.35*	*99.9*
KBUF	*0.057*	*0.31*	*98.0*
KCLE	*0.069*	*0.18*	*81.6*
KCMX	*0.059*	*0.29*	*99.8*
KDET	*0.053*	*0.36*	*100.0*
KDLH	*0.055*	*0.34*	*99.8*
KERI	*0.054*	*0.35*	*98.3*
KFNT	*0.052*	*0.37*	*100.0*
KGRB	*0.053*	*0.36*	*100.0*
KGRR	*0.058*	*0.30*	*96.9*
KINL	*0.053*	*0.36*	*99.9*
KLAN	*0.064*	*0.23*	*75.5*
KMKE	*0.062*	*0.25*	*99.4*
KMKG	*0.060*	*0.27*	*99.6*
KROC	*0.063*	*0.24*	*99.8*
KSBN	*0.056*	*0.32*	*99.9*
KSYR	*0.055*	*0.33*	*99.9*
KTOL	*0.055*	*0.33*	*100.0*
KTVC	*0.054*	*0.35*	*99.6*

Several stations exhibited a perfect agreement, with all 1000 trials failing to reject the null hypothesis. The exceedingly high null acceptance rate confirms that it was unlikely the (non-zero) distributions of the 2003 ASOS data differed significantly from the long-term (1961-1990) wind speed climatology. As this high level of agreement was noted at all available stations, it is plausible to assume that not only was 2003 a representative wind year at these stations, but barring any substantial local alterations of the surface characteristics, wind speed data measured at any location within the network domain might also be considered representative of the long-term wind regime.

Lastly, the frequencies of calm periods as identified in the preceding section were compared at each station. Calms represent a small, but not inconsequential percentage of overall data. The median calm frequency in the ASOS data was 10.0%, while for the SAMSON data it was 4.8%. The median absolute difference between the two frequencies was 4.9%. Although calm behavior was well correlated (r = 0.89), it does appear that differences in ASOS vs. pre-ASOS data must be taken into account.

Considerations that were not explored in this research include the division and analysis of the ASOS record by season. It is possible that deviations from the distribution may have been a function of an anomalous season within the year, but that the influence of that season was damped by the more “normal” conditions of the other seasons. Additionally, although the 1961 – 1990 data did not demonstrate any notable systematic trends or cycles that would indicate a long-term non-stationarity in the region’s wind climatology, more than a decade exists between the end of the long-term record and the short record. In establishing the representativeness of a particular year of data, a comparison to more recent historical records should be included such that recent climatic fluctuations can be taken into account. Lastly, no attempt was made to assess the temporal behavior of wind speeds within either the ASOS or long-term SAMSON data. Although the ASOS wind climatology appears to be representative of the long-term wind, this analysis does not reveal whether the intra-annual composition of the data is similar to that of the long-term SAMSON data.

Overall, it appears that the non-zero, 2003 wind speed distribution for the Great Lakes region of the US does not deviate significantly from the long-term distribution of wind speeds. Thus, the wind speed distributions of 2003 are indeed representative of the long-term wind resource distributions, and might reasonably be utilized to produce an accurate wind resource analysis of the region.

4. REFERENCES

Derrick, A., 1992: Development of the measure-correlate-predict strategy for site assessment. In *Wind Energy Conversion 1992*. B.R. Clayton, ed. Mechanical Engineering Publications. London, pp. 259 - 265.

EPA (U.S. Environmental Protection Agency), 1997: *Analysis of the affect of ASOS-derived meteorological data on refined modeling.* EPA-454/R-97-014. 95 p.

Garcia-Rojo, R., 2004: Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach. *Wind Eng.*, **28**, 213 – 224.

Justus, C.G., W.R. Hargraves, A. Mikhail and D. Graber, 1978: Methods for estimating wind speed frequency distributions. *J. Appl. Meteor.*, **17**, 350 – 353.

Koren, O., 1973: Digital output wind system for airport use. *J. Appl. Meteor.*, **12**, 529 – 536.

NREL (National Renewable Energy Laboratory), 1993: *Solar and Meteorological Surface Observation Network 1961-1990*. Version 1.0. NSRDB-Volume 1. Distributed by the National Climatic Data Center, Asheville, NC.

Rogerson, P.A., 2001: *Statistical Methods for Geography*, Sage Publications, London, 236 p.

Shein, K.A., 2005: *On the Use of a Regional-Scale Numerical Climate Model in Wind Energy Applications*. Ph.D. Dissertation, Michigan State University, 289 p.

Shein, K.A. and S.M. Robeson, 1995: Interannual, seasonal, and diurnal variability of wind in the Midwestern United States. *Proceedings of Windpower ’95*, American Wind Energy Association, Washington, DC.

Snedecor, G.W. and W.G. Cochran, 1989: *Statistical Methods, 8th Ed.* Iowa State University Press, Ames, IA. 503 p.

Touma, J.S., 1977: Dependence of the wind profile power law on stability for various locations. *J. Air Poll. Ctrl. Assn.*, **27**, 863 – 866.