Uncertainties in Results of Measure-Correlate-Predict Analyses

Anthony L. Rogers, Ph. D.*
John W. Rogers, M.S.**
James F. Manwell, Ph. D.*

*Renewable Energy Research Laboratory
Dept. of Mechanical and Industrial Engineering
University of Massachusetts, Amherst, MA 01003
Telephone: 413-545-4866, fax. 413-545-1027, Email: arogers@ecs.umass.edu

**Westat, 1650 Research Blvd., Rockville, MD 20850
Telephone: 301-294-2804, Email: rogersj1@westat.com

Abstract

Measure-correlate-predict (MCP) algorithms are used to predict the wind resource at target sites for wind power development. MCP methods model the relationship between wind data (speed and direction) measured at the target site, usually over a period of up to a year, and concurrent data at a nearby reference site. The model is then used with long-term data from the reference site to estimate the long-term wind speed and direction distributions at the target site. An MCP analysis results in estimations of long-term mean wind speed and wind speed distributions at the target site. In order for MCP estimations to be most useful for wind power development, the uncertainties in the process need to be understood by land owners, lenders, developers and policy makers. This paper explores the uncertainties in MCP predictions. First, prediction uncertainties as a function of concurrent data length are explored using reference and target sites for which long term concurrent data sets exist. Statistical models are then investigated that estimate the uncertainties in MCP estimations, when only short-term concurrent data exist. These include the use of linear regression statistics and the jackknife estimate of variance. The success of these approaches for estimating the uncertainty of the MCP predictions is evaluated. The results of using the various statistical models on wind data are then used to illuminate important issues and to suggest approaches to the solution of this problem.

Introduction

Measure-correlate-predict (MCP) algorithms are used to predict the wind resource at target sites for wind power development. MCP methods model the relationship between wind data (speed and direction) measured at the target site, usually over a period of up to a year, and concurrent data at a nearby reference site. The model is then used with long-term data from the reference site to predict the long-term wind speed and direction distributions at the target site. An MCP analysis results in predictions of long-term mean wind speed and wind speed distributions at the target site. In order to be most useful for wind power development, the uncertainties in the predictions need to be understood.

This paper explores possible approaches to estimating the uncertainties in MCP predictions. First prediction uncertainties as a function of concurrent data length are explored using reference and target sites for which long term concurrent data sets exist.

Statistical models are then investigated that estimate the uncertainties in an MCP prediction when only short-term concurrent data exist. The first of these uses the variances and covariance of the slope and offset of a linear regression MCP model to estimate the uncertainty of the MCP estimates of mean wind speed. An alternate approach, using the jackknife estimate of variance is then introduced in this paper. A method for evaluating the success of these approaches is also described. The results of using the statistical models on long-term data sets are then used to illuminate important issues and suggest approaches to the solution of this problem.

MCP Method

The MCP approach uses data from a reference site, such as shown in Figure 1, a period of concurrent data from a nearby target site, as in Figure 2, and a model to determine a relationship between the wind speeds at each site (Figure 3). That relationship is then applied to the longer-term reference site data to predict what the wind speeds were at the target site during the period of data collection at the reference site. Typically the analysis is performed with the data binned into direction sectors.

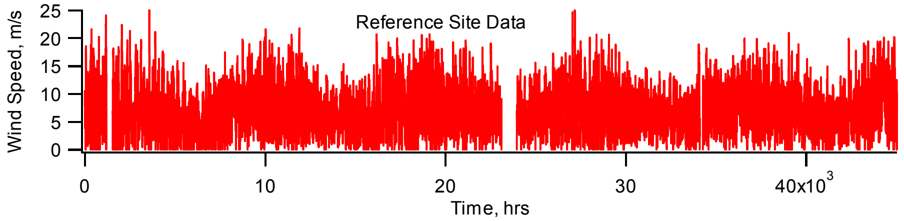

Figure 1. Sample reference site data

Figure 2. Sample target site data

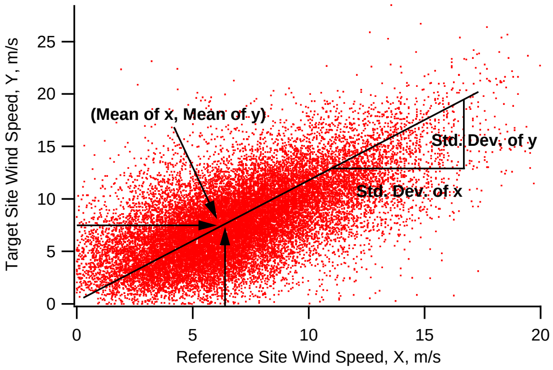

Figure 3. Sample MCP relationship

MCP Algorithms

A variety of MCP algorithms have been proposed in the literature [see references in Rogers, et al., 2005] for characterizing the relationship between the concurrent wind data at the two sites. Four were investigated by the authors and the methods of Mortimer and one proposed by the authors, the "Variance" method were shown to reliably predict a variety of site characteristics important to wind power applications [Rogers, et al., 2005]. The two methods that used linear regression did not predict the wind speed distributions as well as the Variance or Mortimer's method. Linear regression is often used in the wind industry to predict mean wind speed characteristics. In this study, the Variance method will be used almost exclusively to characterize the relationship between the data from the two sites. Linear regression method will be used when the applicability of using the uncertainties of the slope and offset for estimating the standard deviation of the results is being explored.

Data Used For Analysis

Twelve sets of data have been used in this analysis. The details of the data are included in Table 1. Each data set consists of numerous years of long term concurrent data from two sites. The sites range from 9 to 566 km apart and the periods of good concurrent data at each site range from 3.3 years to 16.9 years. These sites are often more distant from each other than desirable for an MCP analysis, but they are just as useful as nearby sites for testing the acceptability of statistical techniques.

Table 1. Details of sites used for analysis

Data Set	File	State / Region	Latitude, N	Longitude, W	Distance	Heading	Mean Speed m/s	Height	Years of Good Data
1	Kennewick - 86 ft	OR	46-06-15	119-07-43	112 km	262	7.80	86 ft	11.92
	Goodnoe - 195 feet*	OR	45-57-18	120-33-46			6.32	195 ft	
2	Red Oak	IA	41	95.17	219 km	84	7.07	50 m	4.66
	Cedar*	IA	41.19	92.57			6.19	50 m	
3	Estherville	IA	43.27	94.86	100 km	89	7.69	50 m	4.45
	Forest City*	IA	43.28	93.63			6.84	50 m	
4	Inwood	IA	43.25	96.47	66 km	75	6.85	50 m	4.52
	Sibley*	IA	43.4	95.68			7.31	50 m	
5	Radcliffe	IA	42.3	93.44	186 km	295	7.29	50 m	3.31
	Sutherland*	IA	43	95.52			7.10	50 m	
6	Buoy 44005	NE	43.18	69.18	87 km	270	6.46	5 m	16.70
	Buoy 44007*	NE	43.53	70.14			5.50	5 m	
7	Buoy 44013	NE	42.35	70.69	231 km	152	5.98	5 m	14.01
	Buoy 44008*	NE	40.5	69.43			6.59	5 m	
8	Buoy BUZM3	NE	41.4	71.03	178 km	11	7.74	25 m	14.74
	Buoy IOSN3*	NE	42.97	70.62			7.08	19 m	
9	Buoy MDRM1	NE	43.97	68.13	62 km	250	8.01	23 m	16.94
	Buoy MISM1*	NE	43.78	68.86			7.91	17 m	
10	Buoy 51001	HI	23.43	162.21	497 km	161	6.70	5 m	13.76
	Buoy 51003*	HI	19.16	160.74			6.03	5 m	
11	Buoy 51002	HI	17.14	157.79	566 km	85	7.72	5 m	13.83
	Buoy 51004*	HI	17.52	152.48			7.37	5 m	
12	US176x14 Data	ID	43.0638°	116.7581°	9 km	14	3.73	2 m	10.60
	US127x07 Data*	ID	43.1381°	116.7327°			3.04	2 m	

Measuring Uncertainties in MCP predictions

The uncertainties of the MCP prediction can be explored by using data from sites for which concurrent long-term data sets exist. When long-term target site data exist, multiple lengths of non-overlapping, shorter concurrent data sets from the longer set can be used, with the long-term reference site data to predict the characteristics of the long-term target site data set. The standard deviation of the estimates for different concurrent data sets can then be used to estimate the precision of the predictions. This can be done for different lengths of concurrent data. For example, multiple separate years of concurrent data, as is indicated in Figure 4, could be used with the long-term reference site data (Figure 1) to predict the precision of the long-term target site wind speed prediction.

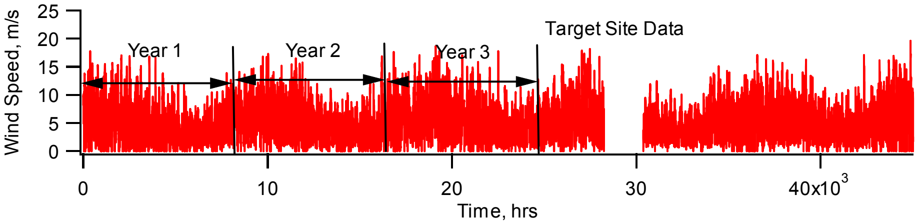

Figure 4. Example of the use of multiple short-term concurrent data sets from the long-term data

As an example, the results of such an analysis are illustrated in Figure 5, using the data from Inwood and Sibly in Iowa. These two sites are 66 km apart. For this analysis, the number of separate MCP estimates ranged from 39 for a data length of 1000 hours to 4 for data lengths of 9000 hours. In general, the longer the length of the concurrent data, the smaller the standard deviation of the results and, thus, the less uncertainty in the results. It would be expected that the standard deviation should decrease as the square root of the concurrent data length, if the MCP model that is used correctly models the relationship between the two data sets. However, the effect of un-modeled seasonal and other characteristics affects the results.

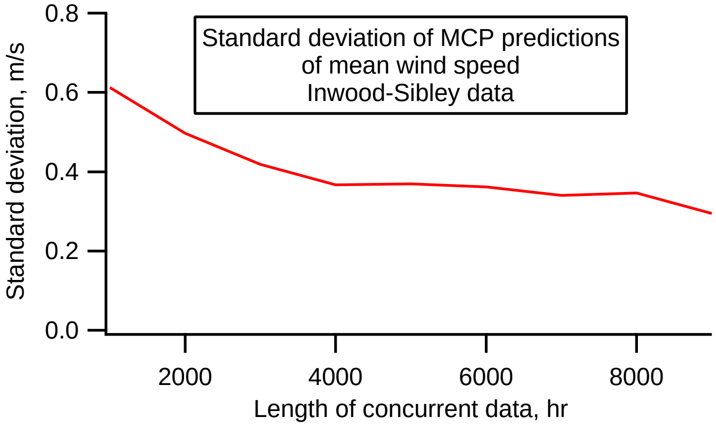

Figure 5. Sample standard deviation of the MCP estimates as a function of concurrent data length

The nature of the statistics for all of the 12 data sets is illustrated in Figure 6, which illustrates the range of uncertainty that may be encountered. Here the mean across all of the 12 data sets of the standard deviation of the mean wind speed is indicated, including error bars indicating the range of standard deviations found (one standard deviation of the standard deviation values). The results using eight 45 degree direction sectors are indicated by a thick line. Those when only one 360 degree direction sector is used are indicated by the dotted line and thinner error bars.

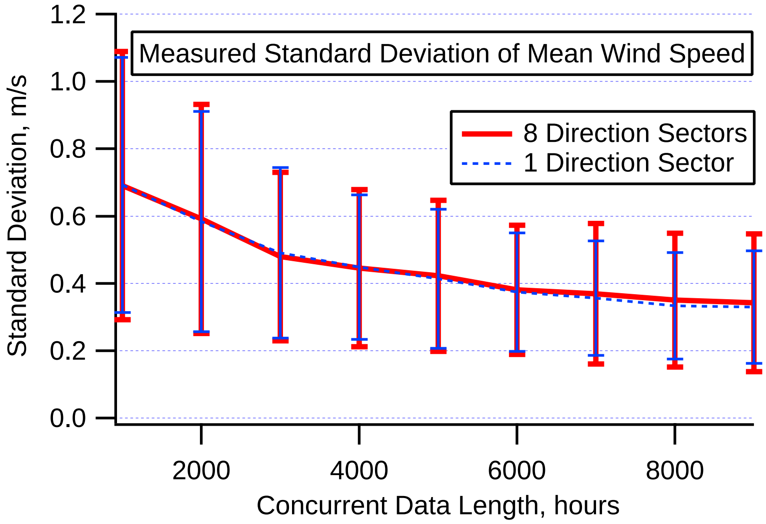

Figure 6. Measured standard deviation of mean wind speed. Mean and variability (one standard deviation) over 12 data sets

The magnitude of the variability of the MCP estimates is, in part, a function of the magnitude of the wind speeds at the target site. The standard deviation for each site and for each concurrent data length can be normalized by the site mean wind speed to illustrate the kinds of variability that might be found at any site. Figure 7 shows the mean and +/1 one standard deviation of these normalized MCP estimates of the mean wind speed across the 12 data sets.

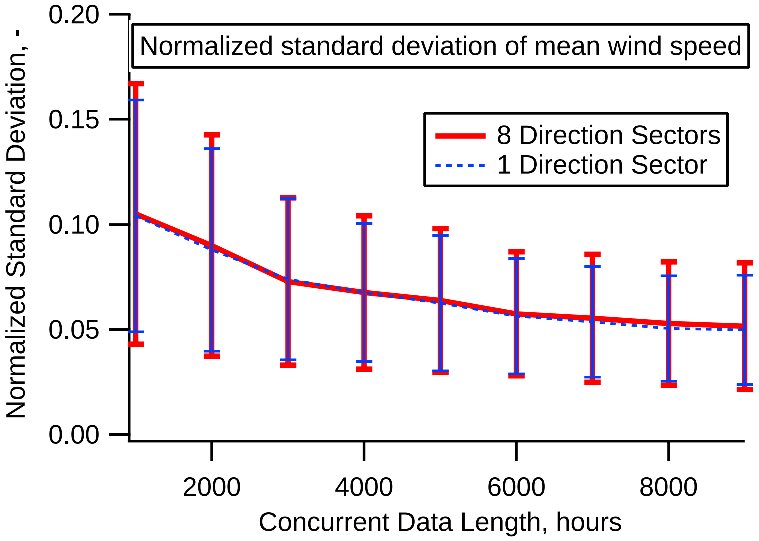

Figure 7. Measured standard deviation of mean wind speed normalized by the mean wind speed and variability (one standard deviation) over 12 data sets

Figures 6 and 7 indicate that

- The uncertainty of the MCP estimates of long-term wind speed typically decrease as the length of the concurrent data increases, over the range of concurrent data lengths tested.
- Standard deviation of MCP estimates, using 1000 hours of concurrent data are typically about 0.7 m/s but may be below 0.3 m/s or over 1.1 m/s
- The standard deviation of the MCP estimates is typically about 11% of the mean wind speed when only 1000 hours of concurrent data are available and 5% of the mean wind speed when 9000 hours of concurrent data are used in the analysis.
- The standard deviation at a given site may be as high as 20% of the mean wind speed, or more, when only 1000 hours of concurrent data are available and may be up to 15% of the mean wind speed when 9000 hours of concurrent data are used in the analysis.
- On the other hand, the standard deviation at any given site may be as low as only a few percent of the mean wind speed.
- Standard deviations of mean wind speed predictions using eight 45 degree direction sectors or only one 360 degree direction sector are similar

Figures 6 and 7 indicate that the uncertainty found at any one site might vary over a fairly large range with respect to the uncertainty found at other sites. On the other hand, the uncertainty at any given site using shorter lengths of concurrent data is correlated with the uncertainty using longer concurrent data sets. Typically, the higher the uncertainty using shorter concurrent data lengths, the higher it will be with longer concurrent data lengths. Figure 8 illustrates the average and the range of standard deviations when the standard deviations are normalized by the value at 9000 hours. It can be seen that whatever the level of uncertainty, the uncertainty using 9000 hours of concurrent data is typically about half that using 1000 hours of concurrent data.

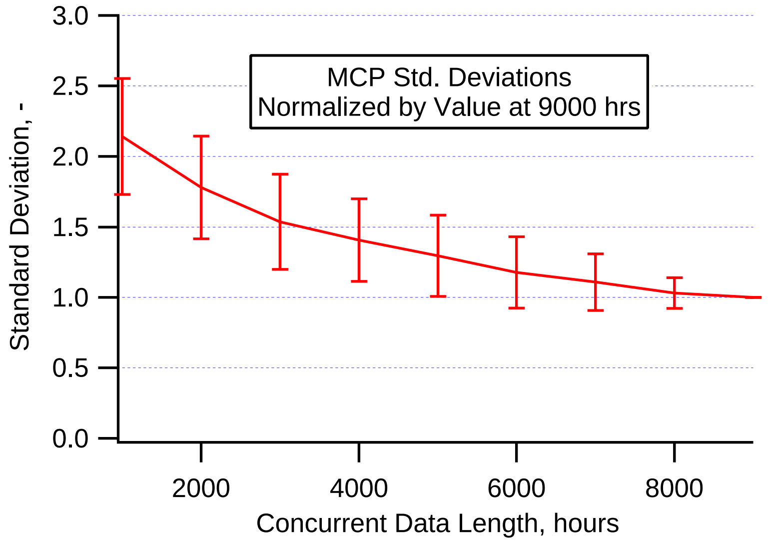

Figure 8. Measured standard deviation of mean wind speed normalized by standard deviation at 9000 hours and variability (one standard deviation) over 12 data sets of

MCP Uncertainties for Weibull K Estimates

A similar analysis was done with the uncertainty of the Weibull parameters derived from the predicted wind speeds. The following graphs illustrate the results for the Weibull shape parameter, k. The results of the Weibull scale parameter, c, are very similar to those for the mean wind speed. Figure 9 shows the results for the standard deviation of the Weibull scale parameter, k. The results using 8 direction sectors are in bold and those for one direction sector are dotted and narrow. The results indicate that the number of direction sectors affects the uncertainty of the Weibull k results, but not substantially.

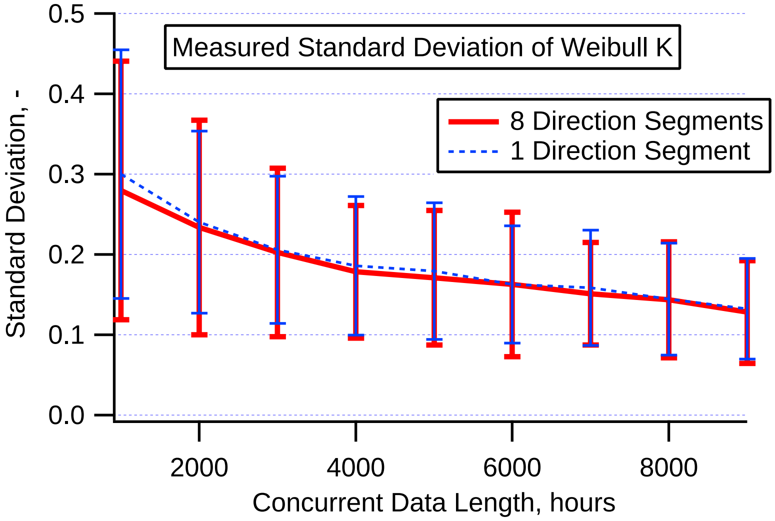

Figure 9. Mean of measured MCP uncertainties of Weibull K estimates across data sets, with error bars showing +/- 1 standard deviation

Figure 10 shows the results for the standard deviation of the Weibull scale parameter, k, normalized by the Weibull k value. The results using 8 direction sectors are in bold and those for one direction sector are dotted and narrow. The results indicate that the number of direction sectors does not substantially affect the uncertainty of the Weibull k results. The data indicate that typically the uncertainty in the Weibull k estimate is on the order of 12% of the Weibull k value using concurrent data lengths of 1000 hours. This drops to about 6% when 9000 hours of concurrent data are used.

Estimating Uncertainties in MCP Predictions

Users of an MCP algorithm do not have long-term data at each site with which to estimate the uncertainty of the results. Thus, the question arises how the uncertainties in the MCP predictions can be estimated, with only the long-term reference site data and the shorter term concurrent data at the target site.

When linear regression is used to model the relationship between the reference and target sites, the uncertainty of the slope and offset are often used to determine the uncertainty of the results. For example, Derrick [Derrick, 1992 and 1993] has detailed the approach for estimating the standard deviation of the estimate of the predicted wind speed, using the variances and covariance of the slope and offset, when the analysis is performed for different direction bins.

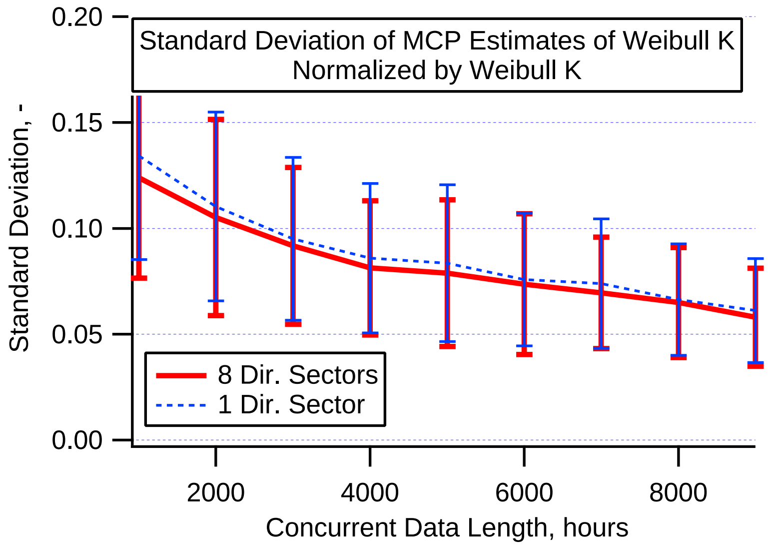

Figure 10. Mean of measured MCP uncertainties of Weibull K estimates divided by site Weibull k across data sets, with error bars showing +/- 1 standard deviation

Applying this method to a sample data set, it can be seen, in Figure 11, that the estimates of the standard deviation do not match the actual measured results. Indeed, this approach significantly underestimates the actual standard deviation in all cases studied, often by about a factor of 10.

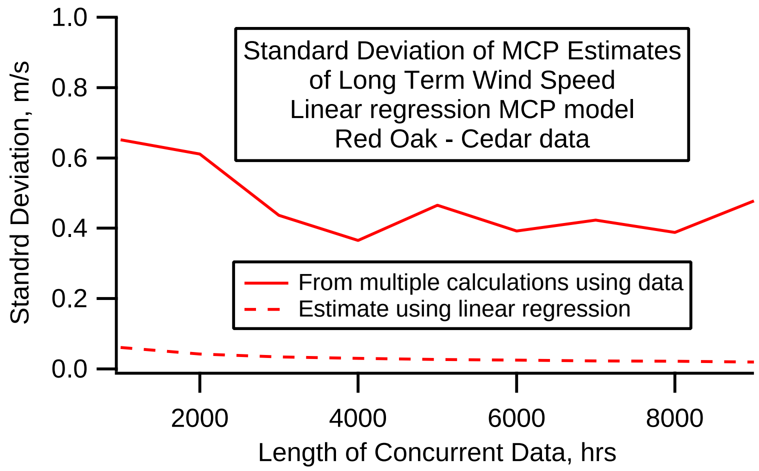

Figure 11. Comparison of standard deviation of MCP predictions of long-term mean wind speed as a function of concurrent data length with estimates using linear regression statistics

Using the uncertainties derived from linear regression is a standard statistical approach that assumes that the data are not serially correlated, i.e. that their cross-correlation is zero at and beyond a lag of one data point. Wind data and the data at two nearby sites are indeed serially correlated. This can affect the results significantly. Figure 12 illustrates the result of using a linear regression fit for the MCP analysis, as above, and for calculating the estimate of the standard deviation of the predicted mean wind speed. In this case, the results of the analysis using two different sets of data are compared. One data set is the original Red Oak – Cedar data set. One is the same data but with the paired data points randomly ordered to remove the serial correlation. When the randomized data is used, the linear regression approach does correctly estimate the uncertainty of the randomized data set. When the original data is used, the linear regression estimate approximates that of the random data which underestimates the uncertainties of the original data set.

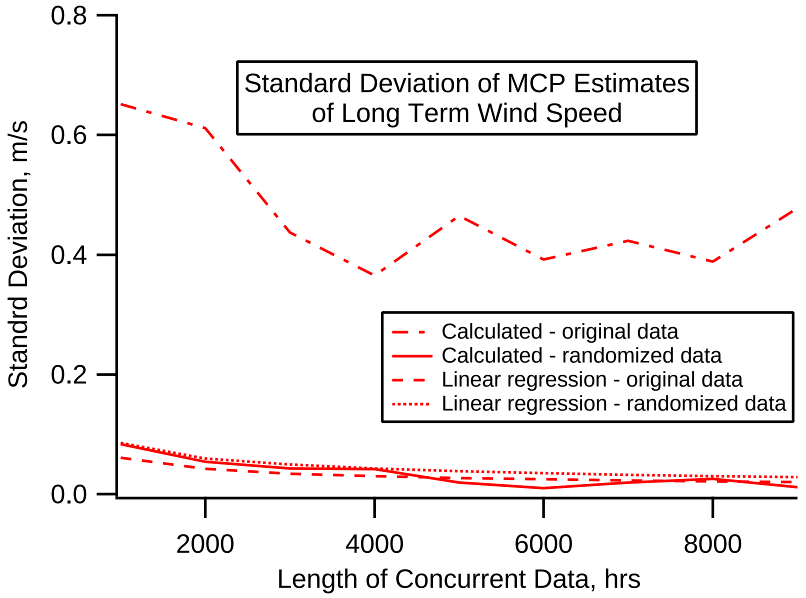

Figure 12. Comparison of linear regression estimates of uncertainty with measured uncertainty using original and randomly sorted Red Oak – Cedar data

Issues to be Addressed

The approach of using the uncertainty of the slope and offset from the linear regression fit to estimate the uncertainty of the mean wind speed works for the data that are independent, but not for serially correlated data. The approach has additional drawbacks. It can only be used with a linear regression MCP model. It can also not be used to predict the uncertainty of other values of interest, for example Weibull parameters or capacity factors that one might derive from an MCP analysis. Finally, when used on serially correlated data, underestimation of the MCP prediction uncertainties leads to unjustified confidence in the results. These concerns raise the question if there are more appropriate alternate uncertainty estimation methods.

An alternate statistical approach, the jackknife estimate of variance [Bradley, 1982], is investigated in the rest of this paper to evaluate MCP predictions for mean wind speed and Weibull parameters of the wind speed distribution. The jackknife estimate of variance assesses the prediction uncertainty using the variability of the MCP predictions made by dropping out subsets of the concurrent data as described below The investigation of this approach will be accomplished by:

- Applying the jackknife method to the 12 available data sets
- Using the root-mean-square prediction error to determine the most appropriate number of subsets to use in the jackknife method

- Using the F-distribution to evaluate whether an estimate of the standard deviation is acceptable
- Exploring and describing issues yet to be solved

Jackknife Estimate of Variance

All statistical approaches to estimating the MCP prediction uncertainty must necessarily use the variability in the relationship between the data from the two sites to predict the overall variability of the predictions. For example, the linear regression approach uses the variances of the slope and offset of the linear regression fit to the concurrent data to estimate the standard deviation of the predictions. The jackknife estimate of variance also determines the standard deviation of the MCP predictions by estimating the variation in the relationship between the two sites, using the concurrent data. As mentioned above, the jackknife approach assesses the prediction uncertainty using the variability of the MCP predictions made by dropping out segments of the concurrent data as described below.

If Y^* is an estimate based on one set of concurrent data and if Y_i^* are *n* estimates of the long term mean using the same concurrent data sets that have each a different, non-overlapping 1/*n*th of the data file missing, then an asymptotically unbiased estimate of the variance of the estimate of Y^*, based on the complete concurrent data set, can be expressed as:

$$\sigma^2\left(Y^*\right)=\frac{(n-1)}{n}\sum_{i=1}^{n}\left(Y^* + Y_i^*\right)^2$$

The approach assumes that the *n* subsets that are removed are independent. This may not be true if the *n* subsets are of such short duration that they are correlated. In that case, the serial correlation of the data will affect this method. In any case, annual cycles and patterns in the data of longer period than the concurrent data length, that are not modeled, will also affect the results. It can be shown that, if the data are independent (not correlated), then the jackknife variance estimate converges to the regression estimate as sample sizes increase. The following steps are used to determine the measured standard deviation of the MCP estimates and the jackknife estimate:

1. Using sets of long term data at each site, the measured standard deviation of the predicted MCP results is determined by repeating the MCP analysis with multiple, *m*, separate lengths of concurrent data. This estimate has (*m*-1) degrees of freedom.

2. Within each separate length of concurrent data, multiple jackknife sub-sets, *n*, are used to estimate the uncertainty in the predicted results. Multiple (*n*) estimates of the predicted results are obtained by dropping out each jackknife subset and determining the predicted results. The estimated prediction variance is then determined, using the jackknife equation, above. This results in *m* estimates of variance of the prediction, one for each *m* lengths of concurrent data. From the *m*

variance estimates, a pooled estimate of the predicted standard deviation, $\sigma^*(Y^*)$, is obtained, using

$$\sigma^*\left(Y^*\right)=\sqrt{\frac{\sum_{k=1}^{m}\sigma_k{}^2\left(Y^*\right)}{m}}$$

This estimate has *m*(*n*-1) degrees of freedom.

Number of Jackknife subsets

As mentioned above, the length of the jackknife subsets may have some effect on the results. The best choice of number of jackknife subsets may depend on the length of concurrent data and the serial correlation characteristics of the data. To determine the most useful number of jackknife subsets, jackknife estimates of the standard deviation of the MCP estimates of mean wind speed and Weibull parameters were made using a variety of numbers of jackknife subsets and concurrent data lengths and using each of the 12 data sets. The number of jackknife subsets ranged from 2 to 32 in steps of approximately the square root of 2. For the investigation of the best number of jackknife subsets, only one direction sector was used to minimize computation time. This exercise results in multiple separate MCP estimates of mean wind speed and of Weibull parameters and in an equal number of jackknife estimates of variance, one for each concurrent data length. The best number of jackknife subsets for each data length and each data set was determined based on the number of jackknife subsets resulting in the least root-mean-square prediction error (RMSE) for the standard deviation of the mean wind speed. If $\sigma_M\left(Y^*\right)$ is the measured standard deviation of the predicted MCP estimates and $\sigma_k^*\left(Y^*\right)$ is the *k*th pooled estimated of the predicted standard deviation, then the RMSE is defined as:

$$RMSE=\sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(\sigma_M\left(Y^*\right)-\sigma_k^*\left(Y^*\right)\right)^2}$$

For example, Figures 13 and 14 shows the RMSE for two data sets. This data is typical of the results from all of the data sets. At different lengths of concurrent data, one or another choice of number of jackknife subsets would provide the least RMSE, but often numerous choices would provide very similar performance. None of the RMSE values were very different from the rest of the RMSE values and the RMSE values usually decreased as the concurrent data length increased.

Based on the result from all of the data sets, a final choice for number of jackknife subsets was determined. For each concurrent data length, the number of jackknife subsets was chosen to be the median of the best performing number of jackknife subsets among the 12 data sets. The results are illustrated in Figure 15. This choice of number of jackknife subsets was used in the rest of the analysis in the paper. It represents the choice of number of jackknife subsets that provides the least overall RMSE for these data sets.

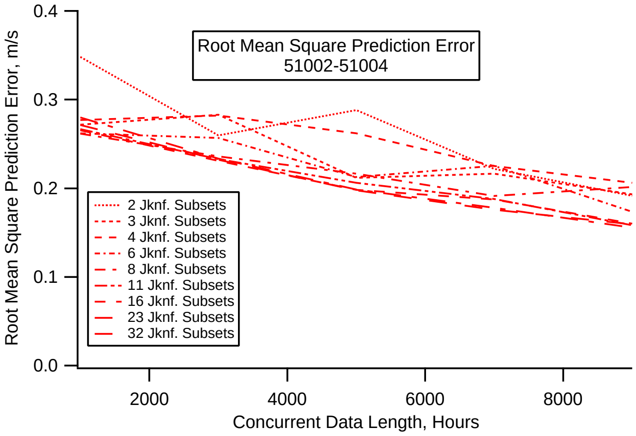

Figure 13. Sample root-mean-square prediction error for buoys 51002 and 51004

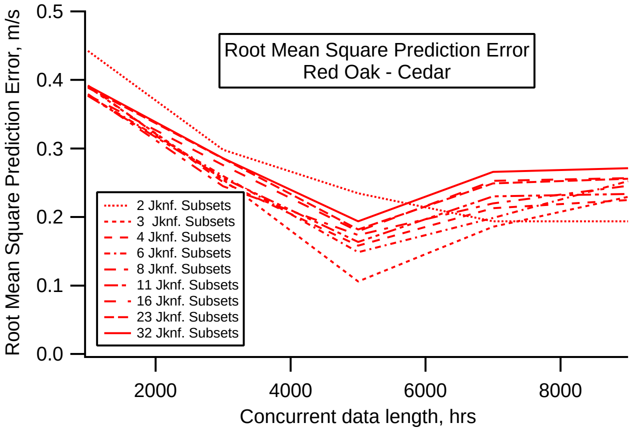

Figure 14. Sample root-mean-square prediction error for Red Oak - Cedar

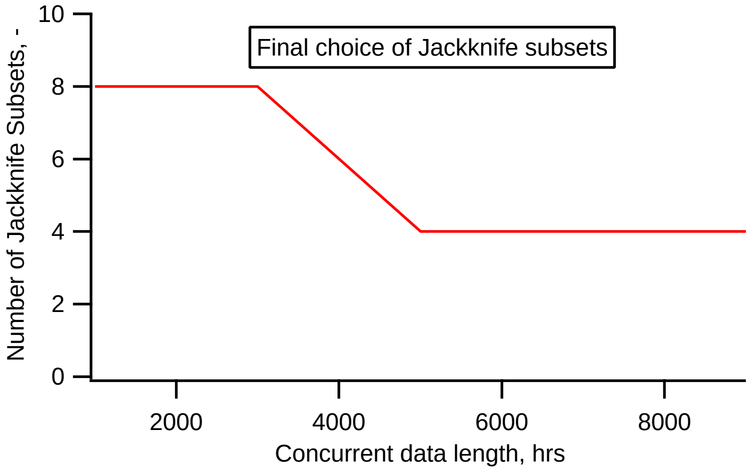

Figure 15. Optimum number of jackknife subsets as a function of concurrent data length

Results of Jackknife Estimates of Variance

The jackknife estimate of variance, using the numbers of jackknife subsets in Figure 15, was then used to estimate the uncertainty in estimates of long-term mean wind speed and Weibull parameters. The MCP model was the Variance method using 8 direction sectors. In the following sections the results for mean wind speed and the Weibull shape parameter, k, are explained and illustrated. The results for the Weibull scale parameter, c, are very similar to those of the mean wind speed and are not shown. The results were obtained for all 12 data sets.

Jackknife Estimates of Uncertainties for Mean Wind Speed Predictions

The analysis of the results of the use of the jackknife estimate of variance on all of the 12 data sets is presented below. In general, the jackknife method is a vast improvement over using the results of linear regression, but not perfect.

Figure 16 shows the mean and variability of the measured standard deviation across the 12 data sets and also of the jackknife estimates. The results are, in general, much better than with the use of the linear regression statistics, but the jackknife, in general, still under estimates the correct uncertainty. The jackknife method often does better with longer lengths of concurrent data.

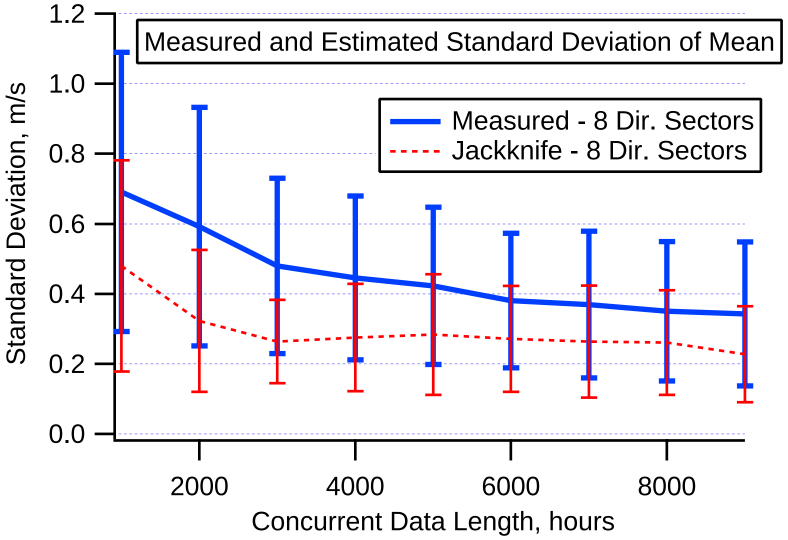

Figure 16. Measured and estimated standard deviation of the mean wind speed estimates, including error bars

Figure 17 shows the RMS prediction error of the jackknife estimate as a function of length of concurrent data, with the number of jackknife subsets fixed as described above, for 8 and for only1 direction sectors. The results show that the typical RMS error of the jackknife estimates is on the order of 0.45 m/s using concurrent data lengths of 1000 hours, but it ranges from higher than 0.75 m/s to lower than 0.15 m/s at different sites. Typically, the RMS error of the jackknife estimates decreases by a factor of 2.5 as the length of the concurrent data set increases from 1000 to 9000 data points.

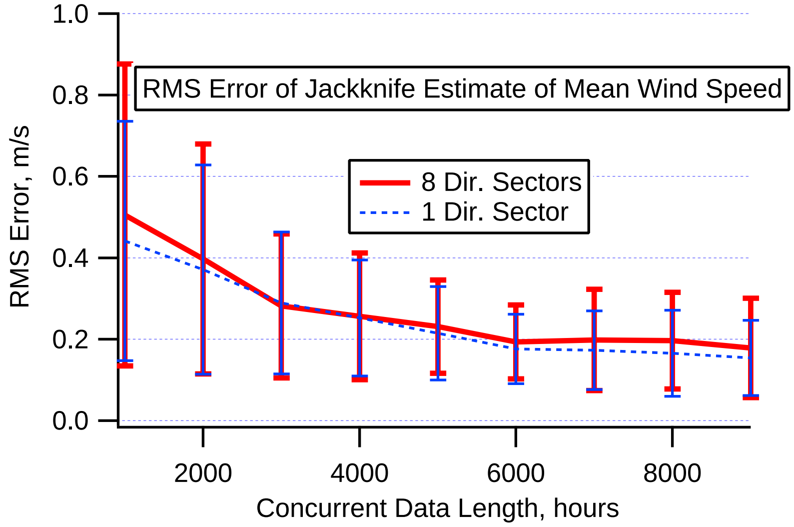

Figure 17. RMS Error of jackknife estimate of mean wind speed

Finally, Figure 18 illustrates the trends of the ratios of the jackknife estimate at each site to the actual measured standard deviation of the mean wind speed prediction at each site. Figure 18 shows the means of these ratios and +/- one standard deviation across the 12 data sets. The ratio of the measured uncertainty of the MCP prediction to the estimated uncertainty typically decreases from about 2, using data lengths of 1000 hours, to about 1.5 using data lengths of 4000 hours or greater, although these ratios vary at each site about these values by about +/-0.5. This ratio is fairly consistent across data sets and among different concurrent data lengths. The overall mean of the ratios across the 12 sites and lengths of concurrent data is 1.61 for these data sets.

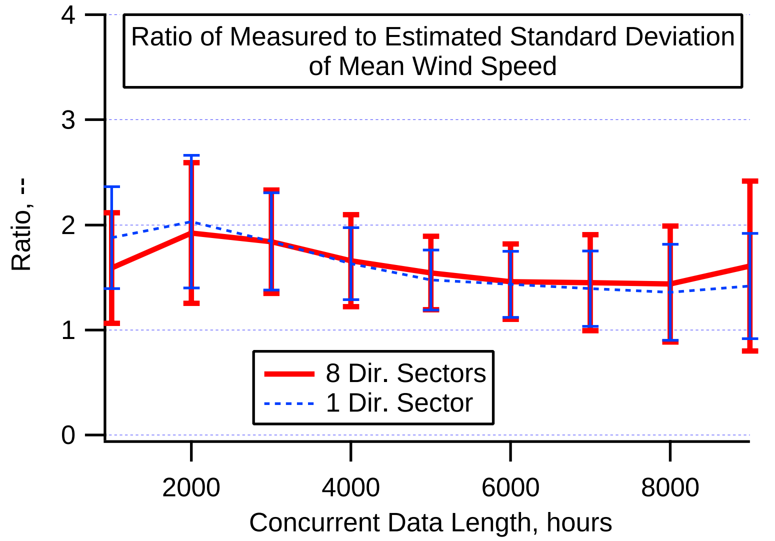

Figure 18. Ratio of measured to estimated standard deviation of mean wind speed

F- Distribution

The standard deviations and the jackknife estimates have been compared to determine if, with 95% confidence, the jackknife estimate can be shown to incorrectly estimate the true standard deviation. The F-distribution is used for this test. Specifically, the F-distribution is used to test the ratio of the two variances. The ratio of two variance estimates, both of which estimate the same variance, has an F distribution, assuming that 1) the variance estimates are independent, and 2) the data have a normal distribution. For the wind data means, for example, the normal distribution is reasonable (by the central limit theorem). If the data are normally distributed and independent (uncorrelated) then the between period and within period (jackknife) estimates are independent. So as long as the jackknife subsets are long enough that the subset estimates are uncorrelated, the formulas should work. Using the F-Distribution, the confidence interval on the ratio of two independent variance estimates of the same variance is only a function of the precision of the estimates, as measured by the degrees of freedom. Here an estimation method is assumed to be incorrect if the standard deviation is outside of the 95% probability limits that the estimate is too low, as determined by the F-distribution. The F-distribution is a function of the incomplete beta function and is described in many statistics textbooks.

Figure 18 shows the mean of the F-Test results across the 12 data sets. A probability of 95%, indicated on the graph, is the point at which the measured standard deviation estimates (from different concurrent data lengths) can be said to be significantly different from the jackknife estimates (based on the variability within concurrent data periods). It can be seen that the F-test significance for the jackknife estimates of the standard deviation of the mean predicted wind speed is typically less than 0.95 when concurrent data lengths are greater than about 5000 hours. This indicates that the jackknife approach does a better job with longer concurrent data lengths and that the underestimation at concurrent data lengths less than a half of a year is most likely not a random occurrence and needs to be understood better.

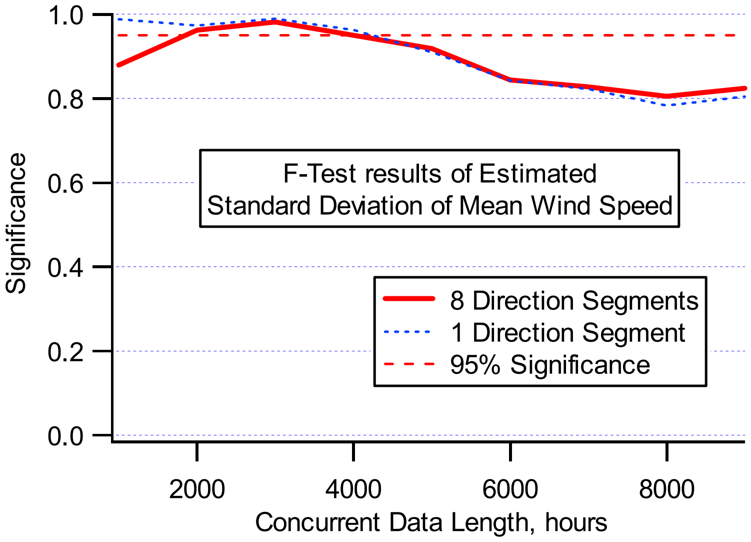

Figure 19. F-test results for jackknife estimates of standard deviation of mean wind speed estimates

Jackknife Estimates of Uncertainties for Weibull K Parameters

A similar analysis was performed for the estimates of Weibull parameter uncertainty. The Weibull shape parameter results are very similar to those of the mean wind speed and are not shown here.

Figure 20 shows the mean and variability of the measured standard deviation of Weibull k parameter of the wind speed distribution across the 12 data sets. It also shows the mean and variability of the jackknife estimates. The jackknife method does a pretty good job of estimating the standard deviation of the Weibull k estimates, even with short lengths of concurrent data.

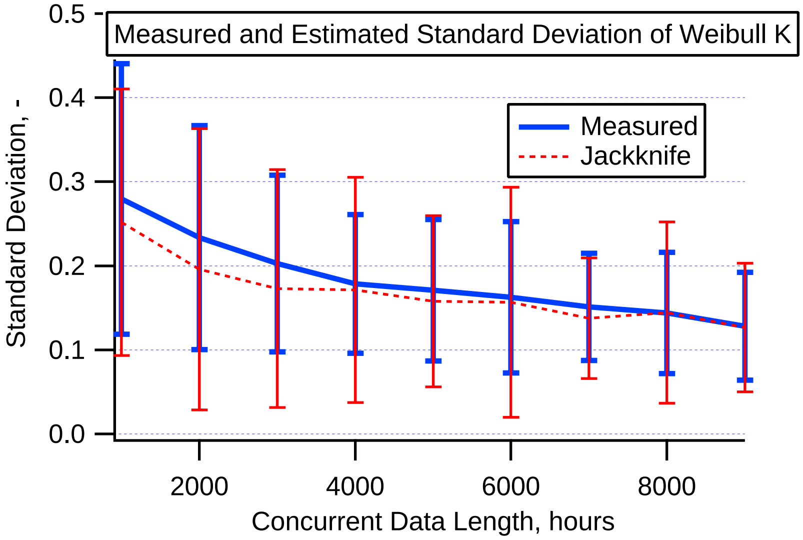

Figure 20. Measured and estimated standard deviation of the Weibull K estimates, including error bars

Figure 21 shows the RMS prediction error of the jackknife estimate as a function of length of concurrent data, with the number of jackknife subsets fixed as described above, for 8 and for only1 direction sector. The results show that the typical RMS error of the jackknife estimates of Weibull k is on the order of 0.16 using concurrent data lengths of 1000 hours, but it ranges from higher than 0.26 to lower than 0.06 at different sites. Typically, the RMS error of the jackknife estimates decreases by a factor of about 2 as the length of the concurrent data set increases from 1000 to 9000 data points. The RMS prediction error of the for the Weibull k parameter is often much less variable and slightly lower when using one direction sector than when using 8 direction sectors with this data.

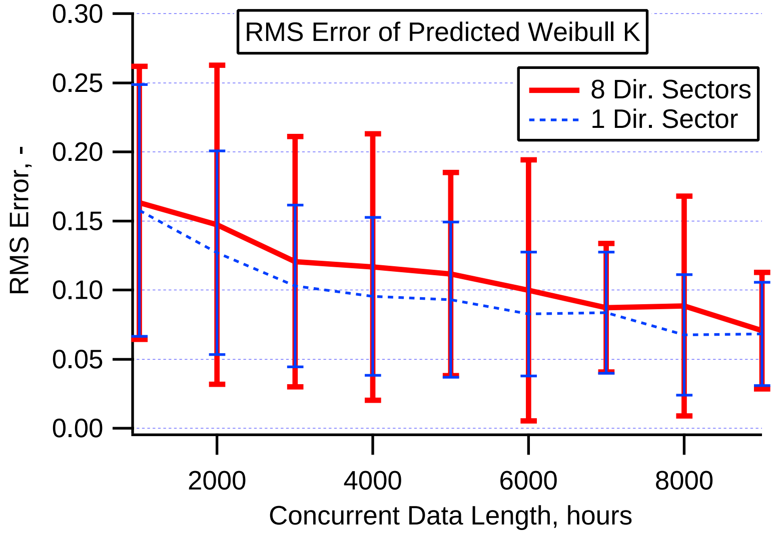

Figure 21. RMS Error of jackknife estimate of Weibull K

Finally, Figure 22 illustrates the ratio of the measured standard deviation of Weibull k to the jackknife estimate of the standard deviation. The ratio of the measured uncertainty of the MCP prediction to the estimated uncertainty varies between about 1.0 and 1.5 and is fairly constant after 3000 concurrent data points. The ratio is a little lower when 8 direction sectors are used.

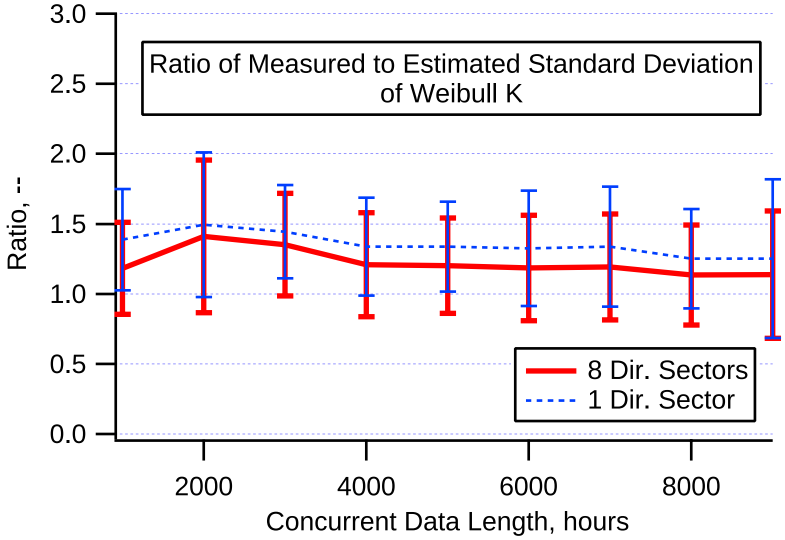

Figure 22. Ratio of measured to estimated standard deviation of Weibull K

Figure 23 shows the F-Test results. A probability of 95% is indicated on the graph. It can be seen that the F-test significance for the jackknife estimates of the standard deviation of the Weibull k parameter is typically less than 0.95 for all concurrent data lengths. This indicates that the jackknife approach does a fairly good job estimating the Weibull k uncertainty and that 8 direction sectors provide better results than just 1 sector.

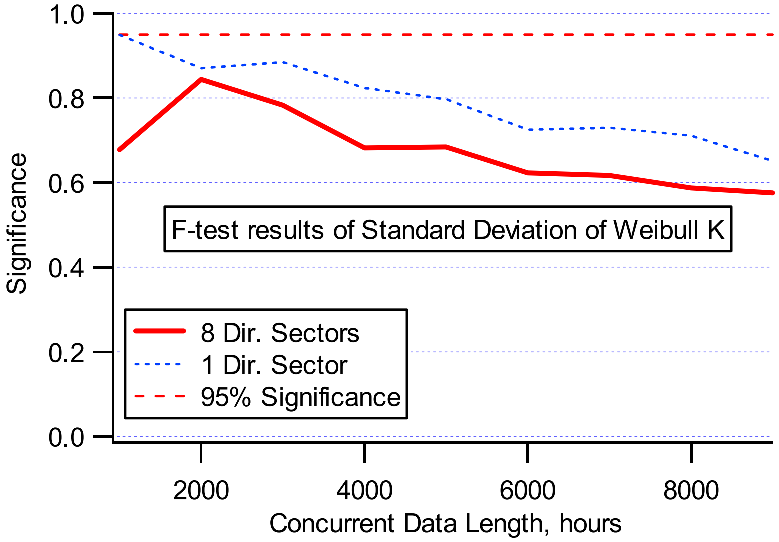

Figure 23. F-Test results for jackknife estimates of standard deviation of Weibull K estimates

Discussion

The analysis shows that the jackknife method does a much better job than using linear regression. Nevertheless, it tends to underestimate the uncertainty of the mean wind speed prediction by about 38% and of the Weibull k prediction by about 18%. A closer look at the nature of the jackknife estimate and the behavior of the data is useful for considering how to improve the estimates.

Figure 24 shows the MCP estimates of mean long-term wind speed, for successive concurrent data segments for the two USDA sites, US176x14 and US127x07. These sites are inland and 9 km distant from each other. The jackknife estimate of uncertainty for the concurrent data segments is indicated by the error bars. The graph shows a thin solid line for successive lengths of 1000 data points and a thicker dotted line for lengths of 9000 points. The data for 1000 points shows a clear annual variability. The jackknife estimates of uncertainty for each 1000 point segment usually significantly underestimate the standard deviation of all of the 1000 point segments. In this case the variability within the 1000 point segments is significantly less than the variability among 1000 point segments and the jackknife estimate is lower than the true standard deviation. The jackknife estimate of the variability of the 9000 point segments, based on the variability within each 9000 point segment is within the range of the actual variability of the 9000 point predictions. Thus, the jackknife prediction for 9000 points is reasonably good. Figure 25 shows the measured standard deviation and jackknife estimate this data. As discussed above, it can be seen that the jackknife estimate greatly underestimates the variability using 1000 concurrent data points, but does a good job at 9000 points.

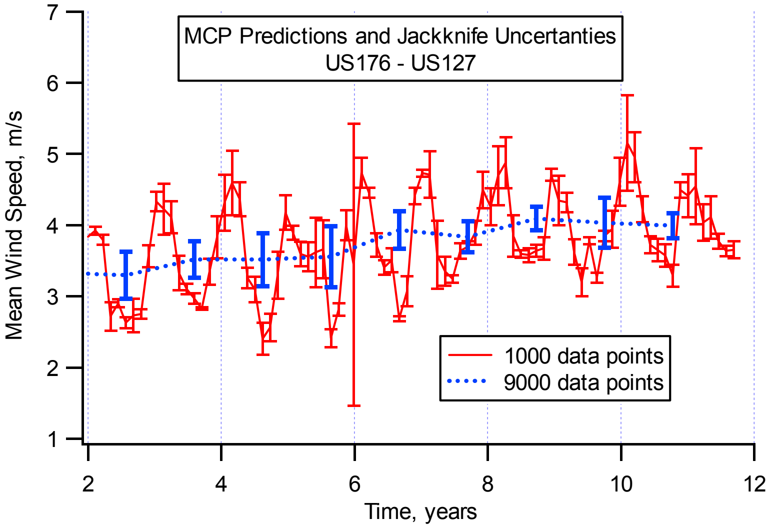

Figure 24 MCP predictions of mean wind speed and jackknife uncertainties for data lengths of 1000 and 9000 hours

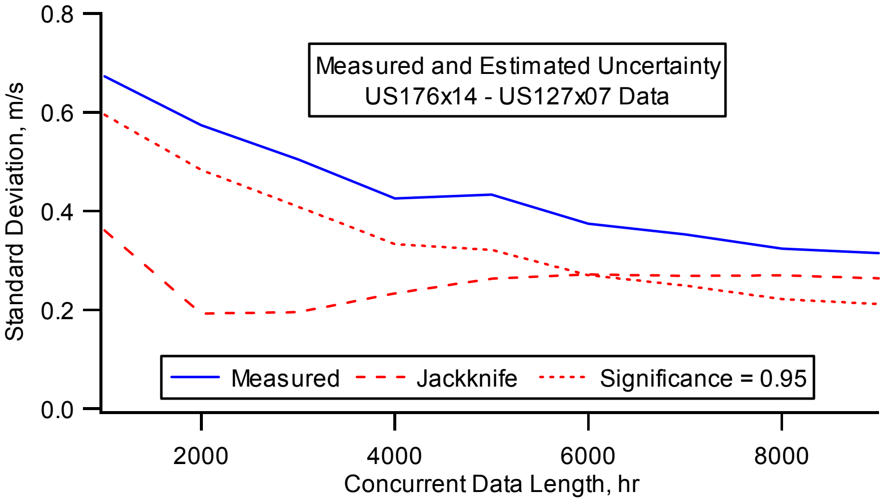

Figure 25. Measured and estimated standard deviation of the mean wind speed estimates, including 95% significance level

Figure 26 shows similar data for two buoys off of the east coast of New England, BUZM3 and IOSN3. These two sites are 178 km apart. In this case, there is significant variability among the 1000 point segments and even within them, as the jackknife estimate of variability is only a little less than that found among all of the 1000 point data sets. In this case, the jackknife estimate of the variability of the 9000 point segments is slightly lower than the actual variability among the segments. As a result (see Figure 27) the jackknife method does a much better job of estimating the uncertainty for all concurrent data lengths with this data.

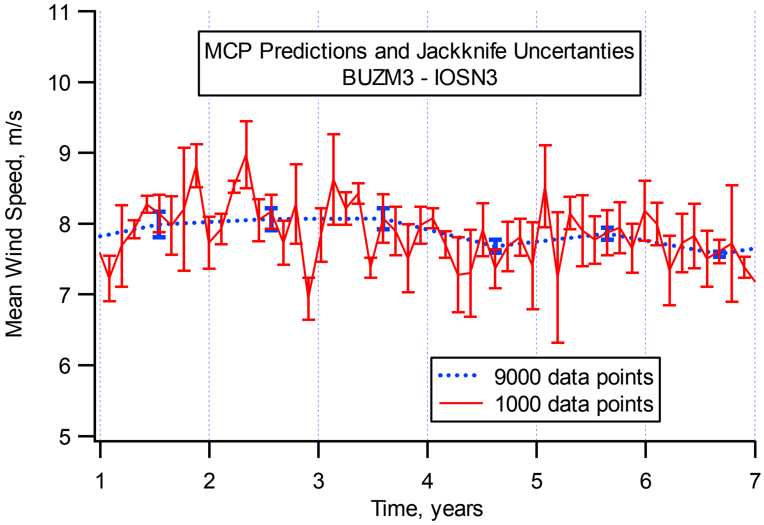

Figure 26. MCP predictions of mean wind speed and jackknife uncertainties for data lengths of 1000 and 9000 hours

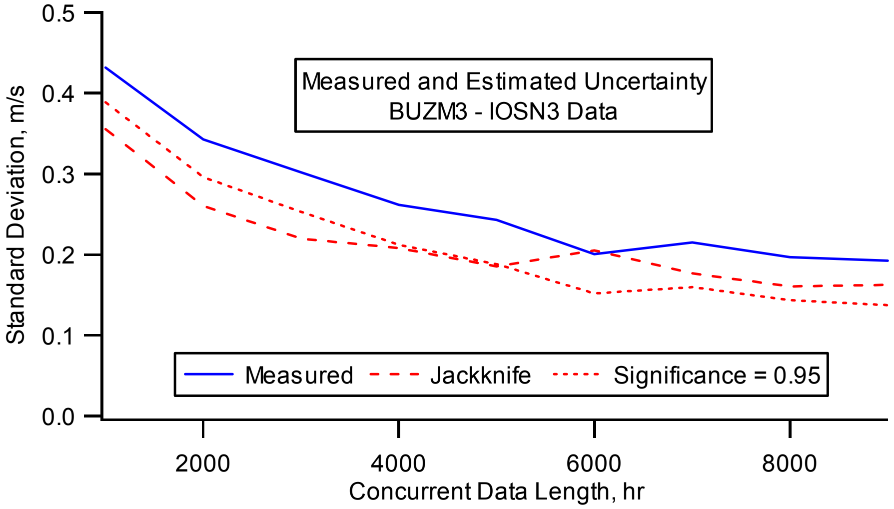

Figure 27. Measured and estimated standard deviation of the mean wind speed estimates, including 95% significance level

A number of the data sets show strong seasonal trends in the relationship between the two sites and the variability within shorter concurrent data segments is usually much less than the seasonal variability. The magnitude of the seasonal trends at some sites was unexpected and complicates the use of the jackknife estimate of variance.

The variability of the relationship, that is measured, can be divided into two parts, 1) the variability about the correct MCP model, if one could find it, that could be used to characterize the relationship between the two sites (say an MCP algorithm with seasonal terms characterized all of the variability in the relationship, then the measured variability about the results would only be random variability), and 2) the variability due to the difference between that correct MCP model and the model that was fit (i.e. we use an MCP algorithm without seasonal terms and we measure more variability in the results, because we are not modeling all of the factors in the relationship).

One approach to resolving the discrepancies between the measured and jackknife estimates would be the use of a better MCP model to represent the relationship between the data from the two sites. If the MCP model does not accurately model the relationship between the two sites, then the variability among the concurrent data segments includes not only random variability, which one might assume would be roughly constant over a variety of time scales, but also variability about an incorrect model. The buoy data above has significant variability that is about the same within the concurrent data segments as among data segments. The USDA data has significant annual variability that is not reflected within the data segments. Thus, the inclusion of temporal terms to the MCP model, whether seasonal, annual or semiannual might improve the jackknife estimates by ensuring the use of an appropriate model for the relationship between the data at the two sites.

Conclusions

In conclusion:

- The use of linear regression statistics seriously underestimates uncertainty estimates due to serial correlation of data
- The jackknife estimate of variance is a great improvement for estimating uncertainty of MCP predictions of mean wind speed and Weibull parameters of the wind speed distribution
- The performance of jackknife estimate of variance is limited by an inadequate MCP model (needs addition of seasonal terms)
- One model variation that could be explored is including seasonal variation in the MCP prediction. That might entail identifying a seasonal pattern in the data, extrapolating the pattern to a year, if necessary, and assuming that it applied generally over time to the site.
- The causes of the strong seasonal variability between the some sites needs to be better understood.

Acknowledgements

This work has been conducted with the support from the Massachusetts Technology Collaborative.

References

Derrick A., Development of the measure-correlate-predict strategy for site assessment, *Proc. BWEA*, 1992.
Derrick A., Development of the Measure-Correlate-Predict Strategy for Site Assessment, *Proc. EWEC*, 1993.
Mortimer A. A., A new correlation/prediction method for potential wind farm sites, Mortimer, *Proc. BWEA*, 1994
Bradley, E., The Jackknife, the Bootstrap, and Other Resampling Plans, SIAM 1982, ISBN 0-89871-179-7
Rogers, A. L., Rogers, J. W., Manwell, J. F., Comparison of the Performance of Four Measure-Correlate-Predict Algorithms, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 93/3, pp. 243-264, 2005