Validation of Long-Term Predicted Wind Speeds

Matthew Breakey P. Eng, Elizabeth Traiger Ph.D, A. Stat – CanWEA 2012

GENIVAR

Summary

At many wind power projects, multiple reference sites are candidates for use in generating long-term wind speed predictions. A systematic method to rank the quality of a reference dataset is a critical part of the wind resource analysis procedure.

A case study at a site with more than four years of on-site data was completed. Three different methods to rank the quality of available Environment Canada sites as long-term references were completed. The results of the three methods were critically compared and analyzed.

Site Description

Seven Environment Canada reference sites were evaluated as potential references to extend on-site data. Table 1 summarizes the period of record, validity and distance of the reference sites to the measured data. Figure 1 shows mean annual wind speeds and the general wind speed trends in the region.

Table 1: Summary of Reference Stations

Reference Station	Site Description		
	Period [Months]	Valid	Distance [km]
Reference A	59	96.4%	31
Reference B	61	98.6%	66
Reference C	103	93.0%	66
Reference D	115	94.1%	98
Reference E	116	97.5%	64
Reference F	115	92.1%	87
Reference G	115	91.4%	41

Figure 1: Annual Wind Speeds of Reference Stations

Method 1: Ranking of References by R^2

The coefficient of determination (R^2) is a commonly used metric to determine the quality of correlation between on-site data and a reference site. The influence of the following parameters on the r^2 were compared:

1) Measurement period of on-site data
2) Averaging period
3) Measurement height (diurnal corrected)

1) The averaging period used when calculating R^2 effects the resulting ranking of reference sites as seen in Table 2. Such variation is likely due to diurnal differences.

Table 2: Rank of Coefficient of Determination by Averaging Period

Reference Station	10 m Hourly	10 m Daily	10 m Weekly
Reference A	2	1	1
Reference B	4	5	5
Reference C	3	3	3
Reference D	5	4	4
Reference E	6	7	7
Reference F	7	6	6
Reference G	1	2	2

Table 3: Coefficient of Determination and Corresponding Rank by the Measurement Period.

Reference Station	Daily R^2 by Synchronised Period of On-Site Data						Daily R^2 Ranks					
	3 Months	6 Month	1 Year	2 Year	3 Year	+4 Year	3 Months	6 Month	1 Year	2 Year	3 Year	+4 Year
Reference A	0.64	0.64	0.65	0.68	0.69	0.69	2	2	2	2	2	2
Reference B	0.52	0.52	0.61	0.66	0.66	0.64	5	5	5	3	3	3
Reference C	0.61	0.61	0.62	0.63	0.64	0.63	3	3	3	4	4	4
Reference D	0.42	0.42	0.62	0.62	0.62	0.60	7	7	4	5	5	5
Reference E	0.44	0.44	0.53	0.55	0.56	0.54	6	6	6	6	6	6
Reference F	0.53	0.53	0.45	0.50	0.51	0.50	4	4	7	7	7	7
Reference G	0.75	0.75	0.74	0.76	0.75	0.74	1	1	1	1	1	1

2) The significance of the on-site measurement period is shown in Table 3. The ranking changes across 6 months to 1 year to 2 years before stabilizing from 2 years onward. Notably, the R^2 values change significantly with time for Reference D. The R^2 values for Reference D jumps from 0.42 at six months to 0.62 after a whole year of data is gathered.

3) Environment Canada wind speed and direction data is typically collected at 10 meters (m) where on-site wind speed measurements are made at 50 m or higher. The two locations are not directly comparable as wind speeds should change throughout the boundary layer due to the diurnal shear pattern. Both the reference and on-site datasets were sheared to the same height to generate the ranks in Table 4. To correct for the difference in diurnal behavior, a shear profile binned by time of day and season was applied to the reference data set. Note the ranking changes for all the reference sites.

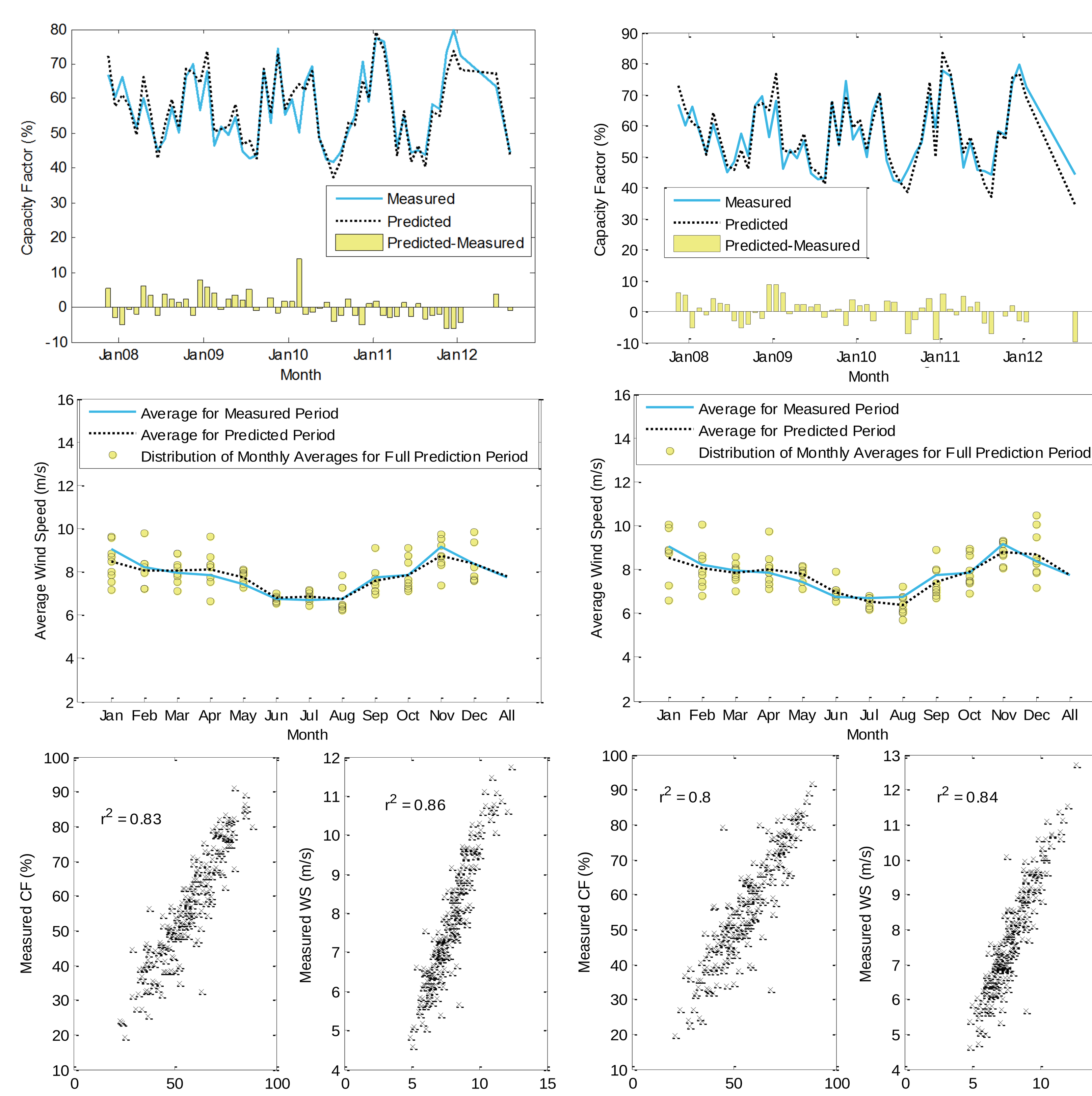
Table 4: Rank of R^2 with Height

Reference Station	10 m Daily	80 m Daily
Reference A	1	2
Reference B	5	3
Reference C	3	4
Reference D	4	5
Reference E	7	6
Reference F	6	7
Reference G	2	1

Method 2: Ranking of Visual Comparison Between Measured and Predicted Datasets

The R^2 value is not conclusive for ranking the quality of reference sites. Rather, R^2 is a useful to determine of which reference site warrants further investigation. Once the prediction process has been completed, a number of different tests can be employed to validate the usefulness of the reference dataset, three of which are shown in figure 2. Two sites were selected as an example, Reference G which consistently had one of the highest R^2 values and Reference C which had a lower R^2 value. Figure 2 shows Reference C has a better correlation to on-site data once the long-term transformation had been applied, contrary to initial indications of reference suitability based purely on R^2 values.

- **Row 1:** The time series of monthly average capacity factors with the residuals plotted below show the predicted data set follows measured data on a temporal basis with no obvious bias, no step changes and few outliers
- **Row 2 :** The monthly average wind speeds for both sites follow the same seasonal trend, something not typically verified with an R^2 analysis.
- **Row 3:** Capacity factor and wind speed correlations for the predicted data set with measured data can show how well the process of generating a long-term prediction has performed. Here, the prediction process has improved the correlation.



Reference C

Reference G

Figure 2: Comparison of Predicted Data to On-Site Measured Data.

Method 3: Ranking of References by Uncertainty in the MCP Process

The first two methods, R^2 and visual verification, can be used to rank reference sites for synchronized periods, but neither method provides insight into other important qualities of a reference site including length of the period of record or data validity. The uncertainty in the generation of long-term predicted data sets are presented and ranked in Table 5.

It is often hard to intuitively weight the merits of a slightly better correlation versus a slightly better validity or slightly longer period of record. For example, Reference C has the lowest MCP uncertainty as reflected in the visual verification method, but has a shorter period of record that adds to the overall uncertainty dropping its overall rank to 4. Note that reference A has one of the best R^2 values, but is last in the uncertainty ranking due to its short period of record, a quality not reflected in the previous two methods.

Table 5: MCP Uncertainty Assessment by Reference

Reference Station	MCP[1]	Period[2]	Validity[3]	Total	Rank
Reference A	0.6%	2.0%	0.1%	2.1%	7
Reference B	0.7%	1.8%	0.0%	1.9%	6
Reference C	0.6%	1.4%	0.1%	1.5%	4
Reference D	0.7%	1.3%	0.0%	1.5%	3
Reference E	0.7%	1.3%	0.0%	1.5%	2
Reference F	0.8%	1.3%	0.1%	1.5%	5
Reference G	0.6%	1.3%	0.1%	1.5%	1

1) The MCP uncertainty was calculated by using a Jackknifing technique via comparison between measured and predicted datasets.
2) The period uncertainty was calculated based on the period of record and the site inter-annual variation.
3) The validity uncertainty was calculated based on the sensitivity if all of the missing data was filled using data from similar time of day and season bins.

Conclusions

Comments on R^2 Method

- Is a good indicator of what reference site might be suitable, but is not definitive.
- R^2 values may change significantly depending on the amount of measured data and pre-processing
- Does not incorporate reference data set data validity or length of period.
- Requires further validation of the correlation including verification of temporal and seasonal trends as well as final predicted data set quality.

Comments on Visual Verification Method

- Can be used to verify temporal variation, residuals, seasonal trends, and long-term biases.
- Can be used to verify correlation of the predicted production or wind speed to measured production which adds confirmation the prediction process worked.
- Is restricted to synchronized records.
- Does not incorporate data validity or length of the reference data set.

Conclusion of Uncertainty Method

- Identifies the reference site that minimizes uncertainty .
- Incorporates a number of important parameters such as correlation quality, period of record and data availability.
- Can be used to quantify how changes in data collection and MCP process will affect confidence intervals. Figure 3 is an example of how the period of record changes uncertainty.

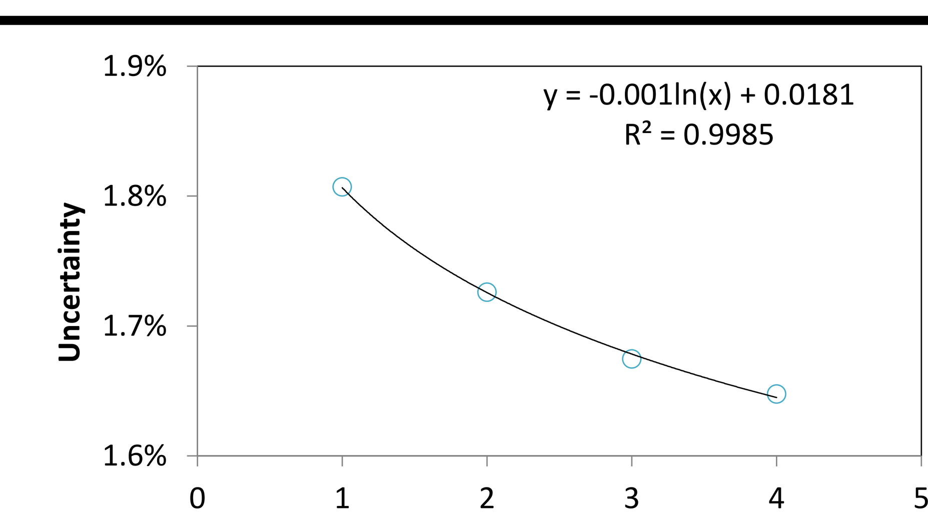


Figure 3: MCP Uncertainty with Period of Record