Evaluation of Alternative Methods to Demonstrate Long-Term Representativeness of Baseline-Period Meteorological Monitoring at In-Situ Uranium Project Sites

May 2014

Prepared by:

IML Air Science

A division of Inter-Mountain Labs, Inc.

555 Absaraka Street

Sheridan, WY 82801

(307) 674-7506

INTER–MOUNTAIN LABS

TABLE OF CONTENTS

Background

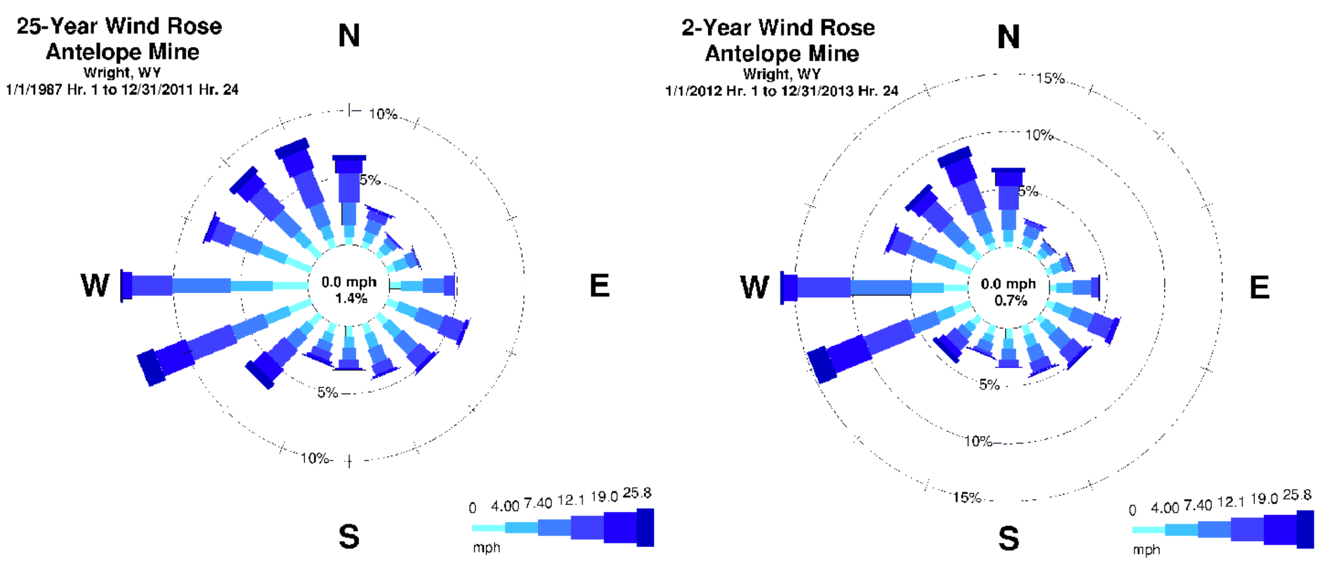
NRC has recently questioned the validity of using regression analysis to demonstrate that baseline-period wind patterns are representative of the long term at any given monitoring site. In a meeting with AUC LLC (NRC 2014), NRC staff commented, “Based on the information submitted, the NRC staff cannot determine that the regression analysis is the appropriate analysis to demonstrate representativeness. The NRC staff finds that other statistical approaches should be considered or discuss why they are not applicable, such as: 1) testing summary statistics, such as the mean from the short and long term data; and 2) testing the statistics for similarity or validity of the data by using a statistical method such as the Student’s T-test, Chi Square test for distribution, Kolmogorov-Smirnov test for distribution, etc., as appropriate.”

This report explores the merits of alternative tests for comparing short and long-term, hourly wind and atmospheric stability data. It provides examples of 2-year and long-term hourly data from Antelope Coal Company (ACC) and the NWS stations at Gillette and Casper. These three sites have been used as long-term references for several ISR license applications in eastern Wyoming.

Objectives

The fundamental measures of atmospheric dispersion potential are wind speed and wind direction, with a third measure being provided by atmospheric stability. The stability class can be derived from wind speed and variability in wind direction, using the σ_θ method. The goal of demonstrating similarities between baseline-period and long-term wind conditions is to enable the use of on-site, baseline meteorological data for modeling and monitoring of pollutant dispersion. In particular, the MILDOS model accepts a joint frequency distribution that combines 6 wind speed classes, 16 wind directions, and 6 atmospheric stability classes. The wind speed and direction category frequencies are illustrated graphically in a typical wind rose (Figure 1). To validate modeling with baseline-period data, we must show that baseline period and long-term frequency distributions are statistically similar. NRC allows this demonstration to be made at sites in the general vicinity of an ISR project, where long-term data are available. Figure 1 compares long and short-term wind roses for the ACC site.

Figure 1 – ACC Long-Term and Short-Term Wind Roses



Beyond this visual demonstration of temporal uniformity, a useful statistical test should accomplish the following:

1. Quantify the goodness of fit between short and long-term sets of categorical wind speed, wind direction or stability class data
2. Avoid false positives (concluding that closely matched frequency distributions are statistically different)
3. Avoid false negatives (concluding that dissimilar frequency distributions are not statistically different)
4. Exhibit low sensitivity to sample size (total number of observations), in this case driven by the period of record
5. Exhibit low sensitivity to data classification (number of categories), in this case driven by the frequency distributions relevant to modeling

To accomplish these objectives, several prospective methods are considered:

1. Summary statistics
2. Chi-square (χ^2) test
3. Student's t-test
4. Kolmogorov-Smirnov (K-S) test
5. Linear correlation/regression analysis

Period of Record

The choice of the long-term period of record has often been restricted by the availability of hourly wind data in electronic form. There are 27 years of continuous data for ACC (IML 2014), 17 years for Casper (NCDC 2014) and 15 years for Gillette (NCDC 2014). This generic analysis assumes a two-year baseline period (2012-2013) and honors the requirement for sample independence by using non-overlapping short and long-term data sets. This leaves 25 years for the long-term period of record for ACC, 15 years for Casper and 13 years for Gillette.

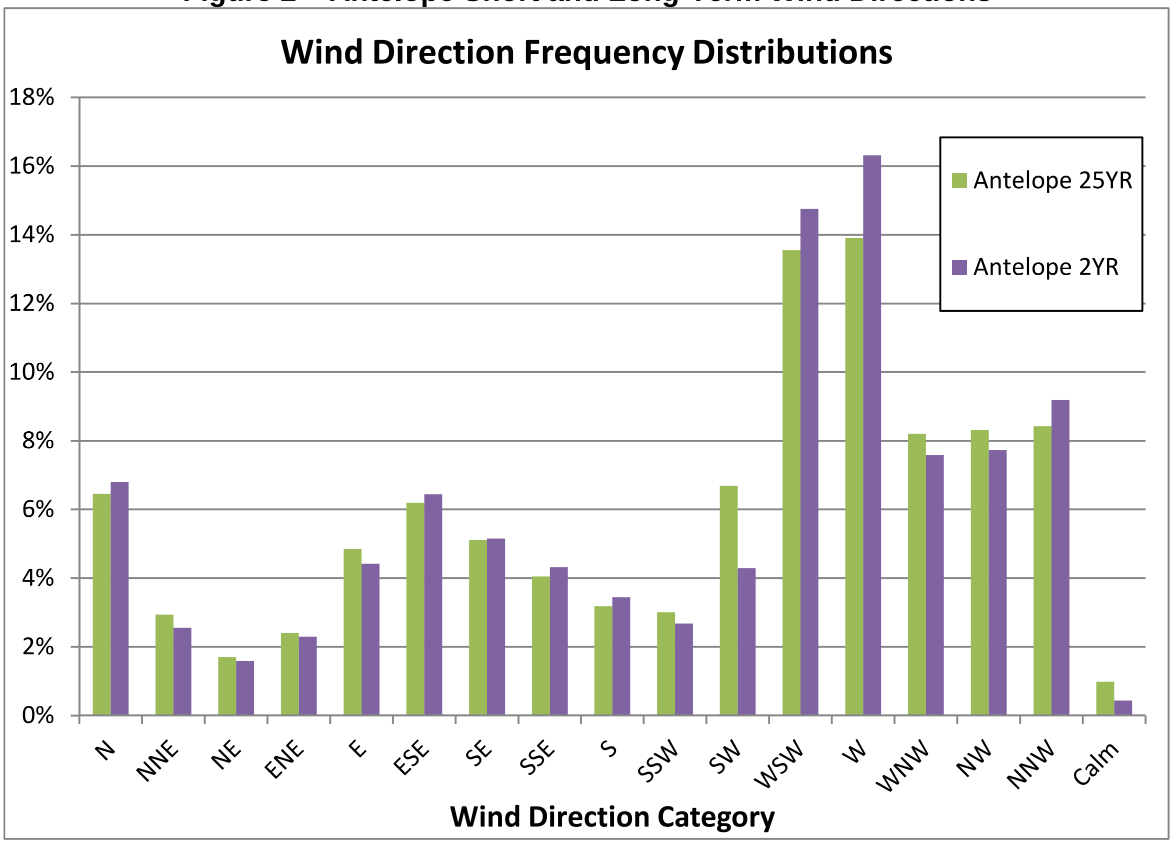
A publication by the U.S. Air Force Climatology Center suggests a tradeoff between too few and too many years of data to establish long-term climate characteristics. "As the POR expands, maintaining homogeneity of the data becomes more difficult. Climatological statistics obtained from too long a period may not be representative of contemporary conditions" (Coffin 1996). The authors cite Panofsky and Brier, who claim that a mean based on 15 years of data gives the best estimate for next year's mean and is therefore preferable to climatic normals based on more than 15 years. They also cite the findings of Rubinstein, Kuznetsova, and Shvec, who claim that estimates of the mean wind speed can be calculated with a sufficient degree of accuracy on the basis of 20 to 25 years' worth of data. These recommendations appear to be generally compatible with the long-term data sets described above for ACC, Casper and Gillette.

Evaluation of Summary Statistics

The best way to make a preliminary study of a series of meteorological observations is to form a frequency distribution (Brooks 1978). The frequency distributions discussed above constitute summary statistics, categorized so as to correspond directly to the MILDOS model. The distributions of categorical data are more useful than population means or standard deviations. The mean annual wind speed can be calculated from hourly data, but it is already implied by the wind speed frequency distribution (time-weighted average of central speeds in each class). The latter offers insight into both the central value and the temporal allocation of wind speeds. Similar wind speed frequency distributions will always yield similar average wind speeds (for example, the 25-year average wind speed at ACC is 11.0 mph, while the 2-year average wind speed at ACC is 11.4 mph). Average wind directions or stability classes have no statistical meaning, since stability classes are categorized non-numerically and wind directions are vector quantities. Dominant wind directions or stability classes, made apparent by frequency distributions, are more useful.

While wind speeds often approximate a Weibull distribution, the distributions of wind directions and stability classes are inherently asymmetric and generally multi-modal – characteristics best demonstrated by frequency distributions. Figure 2 illustrates this for wind directions at the ACC site, making apparent the similarity between 2-year and 25-year monitoring data.

Figure 2 – Antelope Short and Long-Term Wind Directions



Evaluation of the Chi-Square (χ^2) Test

The χ^2 test is used to evaluate the null hypothesis that two frequency distributions are similar. Brooks and Carruthers (Brooks 1978) state that any two observed distributions may be compared with the chi-square test without reference to the theoretical forms of the distributions. They further state that a useful application of the test in meteorology is testing the similarity of frequency distributions of two or more sets of observations.

The χ^2 test comes with some constraints, however. Relative frequency distributions do not lend themselves directly to the χ^2 test, which generally requires a minimum of 5 expected occurrences in each data category. NRC guidance reduces this requirement to 2 (NRC 2011), but relative frequencies are by definition less than or equal to 1. Even if the category frequencies are converted to percentages, some categories still fall below the 2% occurrence rate. A possible remedy is to use the number of hours observed in each wind category rather than the relative frequencies. But short and long-term data sets contain vastly different total hours (i.e., sample sizes), violating another condition of the χ^2 test. This difficulty can be surmounted by converting relative frequencies to equivalent annual hours. Long-term values can be regarded as the expected counts, and short-term (baseline period) values as the observed counts. Table 1 shows the resulting analysis of wind speeds at the Antelope mine. The calculated χ^2 value is much larger than the 95% confidence statistic for 6 degrees of freedom, leading to a strong rejection of the null hypothesis (H_0). The tentative conclusion: the short and long-term wind speed distributions are significantly different (with a confidence close to 100%). It is worth noting that the greatest contribution to the χ^2 value comes from the calm category, even though it comprises less than 1% of the annual observations.

Table 1 – χ^2 Test for Annual Wind Speed

Wind Speeds - ACC Annual Hours				
mph	25-Yr WS	2-Yr WS	$(ST-LT)^2/LT$	Chi-Square
0 - 3	1063	931	17	71
4 - 7	2127	1993	8	$\chi^2_{0.95}(6) = 12.59$
8 - 12	2374	2462	3	**Reject H_o**
13 - 18	1778	1902	9	p-value = 0.000
19 - 24	807	871	5	Min Count = 86
> 24	524	563	3	
Calm	86	38	27	

Table 2 shows a similar test for 25-year vs. 2-year wind directions at Antelope, with the same result.

Table 2 – χ^2 Test for Annual Wind Direction

Wind Directions - ACC Annual Hours				
Sector	ACC 25Yr WD	ACC 2Yr WD	$(ST-LT)^2/LT$	Chi-Square
N	566	596	1.63	180
NNE	258	224	4.37	$\chi^2_{0.95}(16) = 26.30$
NE	149	139	0.65	**Reject H_o**
ENE	211	201	0.48	p-value = 0.000
E	426	387	3.54	Min Count = 86
ESE	543	564	0.83	
SE	448	452	0.02	
SSE	354	378	1.59	
S	279	302	1.94	
SSW	263	234	3.24	
SW	586	376	75.11	
WSW	1,187	1,292	9.27	
W	1,218	1,430	36.62	
WNW	719	664	4.20	
NW	729	677	3.68	
NNW	738	805	6.25	
Calm	86	38	26.56	

The outcome of the χ^2 test for the ACC site seems to contradict the visual evidence in Figure 1 and Figure 2. Such counterintuitive results are an artifact of sample size; the choice of annual hours was arbitrary. Multiplying the wind direction frequencies by 250 (the minimum multiplier that yields long-term counts of 2 or more) instead of 8,760, produces a different test outcome, shown in Table 3 for wind direction and speed distributions at all three reference sites. The Antelope, Gillette and Casper data all show low χ^2 values relative to the test statistic at 95% confidence (insufficient evidence to reject H_0). For this embodiment of the test, we cannot justify a conclusion that short and long-term wind speed and direction distributions are statistically different.

That a sample size of 250 leads to a different test outcome than a sample size of 8,760 reveals a weakness of the χ^2 test. When differences are practically insignificant but statistically significant, it is due to a very large sample size. In a smaller sample the differences would not be enough to be statistically significant. Dealing with a large number of observations, Hessen, Dolan, and Wicherts noted that χ^2 values are inflated by large total sample sizes rendering the test results "of little use" under these circumstances (Hessen 2006). A statistical text (Sharpe 2012) warns: "Beware large samples! With a sufficiently large sample size, a chi-square test can always reject the null hypothesis."

With the right scaling factor to reduce sample size, it may appear that the χ^2 test still has merit in demonstrating long-term representativeness of wind frequency distributions. Table 4 shows the test of direction frequencies scaled by 250 does differentiate between wind direction distributions from the same site and those from different sites (a p-value greater than 0.05 does not warrant rejection of H_0 at 95% confidence). But scaling by 250 fails to differentiate wind speed distributions when comparing Casper and Gillette. Moreover, this technique places the outcome of the χ^2 test at the mercy of an arbitrary choice of scaling factor. Depending on this choice, the same data sets can lead either to rejection or non-rejection of the null hypothesis.

A less arbitrary adaptation of the χ^2 test involves the use of the phi coefficient of association to neutralize the test's sensitivity to sample size. To evaluate the degree to which two distributions differ, the χ^2 statistic can be converted to the phi coefficient:

$$\emptyset = \sqrt{\left(\frac{\chi^2}{N}\right)}$$

An analysis of categorized cloud cover by the U.S. Air Force employed this technique, and established a critical phi coefficient of 0.20. "Values [of phi] close to zero indicate the distributions are nearly identical, while those approaching either 1 or -1 are significantly distinct. In most of the cases, the phi coefficient was less than or equal to 0.20, implying a large degree of similarity in the two distributions" (Lowther 1991). Table 5 shows low phi coefficients for the ACC long-term/short-term χ^2 tests (0.09 for wind speeds and 0.14 for wind directions). It also illustrates the typically high phi coefficients obtained from applying the χ^2 tests to inter-site comparisons (0.97 for ACC/Gil wind speed distributions). While the phi coefficients in Table 5 cannot prove conclusively that the short-term and long-term wind measurements are statistically equivalent, neither can the χ^2 test itself, as applied in Table 3. Brooks and Carruthers state that "χ^2 is a guide to significance not an exact measure of it" (Brooks 1978, p. 160).

Table 3 – χ^2 Test for Wind Direction and Speed Frequencies Multiplied by 250

Wind Directions - Antelope LT/ST Freq Scaled Up

ACC 25Yr WD	ACC 2Yr WD	$(ST-LT)^2/LT$	Chi-Square
16.15	17.01	0.044	7.14
7.36	6.40	0.143	$\chi^2_{0.95}(16) = 26.30$
4.25	3.97	0.020	Can't reject H_o
6.03	5.74	0.014	p-value = 0.995
12.15	11.04	0.111	Min Count = 2
15.50	16.10	0.023	
12.80	12.89	0.001	
10.11	10.79	0.043	
7.95	8.61	0.051	
7.52	6.69	0.104	
16.71	10.73	3.340	
33.88	36.87	0.243	
34.77	40.80	0.891	
20.53	18.96	0.130	
20.81	19.33	0.113	
21.05	22.99	0.163	
2.45	1.09	1.709	

Wind Directions - Gillette LT/ST Freq Scaled Up

Gil 13Yr WD	Gil 2Yr WD	$(ST-LT)^2/LT$	Chi-Square
16.13	16.28	0.001	2.87
7.76	8.66	0.095	$\chi^2_{0.95}(16) = 26.30$
4.92	5.17	0.013	Can't reject H_o
3.42	4.36	0.204	p-value = 1.000
4.76	5.09	0.020	Min Count = 3
4.63	3.58	0.309	
9.02	6.82	0.708	
18.52	16.77	0.184	
30.86	28.76	0.154	
11.53	12.29	0.047	
20.46	18.82	0.143	
12.86	11.61	0.136	
19.70	19.48	0.002	
14.98	14.19	0.045	
23.18	25.77	0.260	
22.46	25.82	0.437	
24.80	26.55	0.115	

Wind Directions - Casper LT/ST Freq Scaled Up

Csp 15Yr WD	Csp 2Yr WD	$(ST-LT)^2/LT$	Chi-Square
11.19	10.31	0.074	5.81
16.46	12.48	1.267	$\chi^2_{0.95}(16) = 26.30$
12.58	9.23	1.213	Can't reject H_o
10.30	8.13	0.577	p-value = 0.995
14.71	12.20	0.514	Min Count = 2
5.19	4.25	0.209	
2.43	2.02	0.081	
1.95	1.82	0.010	
5.02	5.70	0.080	
24.89	25.23	0.005	
49.84	56.10	0.699	
32.20	35.16	0.250	
20.70	21.66	0.042	
10.30	9.68	0.039	
8.39	8.34	0.000	
8.16	8.15	0.000	
15.70	19.53	0.752	

Wind Speeds - Antelope LT/ST Freq Scaled Up

ACC 25Yr WS	ACC 2Yr WS	$(ST-LT)^2/LT$	Chi-Square
30.35	26.56	0.541	3.04
60.70	56.89	0.256	$\chi^2_{0.95}(6) = 12.59$
67.76	70.26	0.089	Can't reject H_o
50.74	54.27	0.229	p-value = 0.916
23.03	24.87	0.136	Min Count = 2
14.97	16.07	0.076	
2.45	1.09	1.709	

Wind Speeds - Gillette LT/ST Freq Scaled Up

Gil 13Yr WS	Gil 2Yr WS	$(ST-LT)^2/LT$	Chi-Square
13.07	13.77	0.036	6.00
53.62	46.06	1.242	$\chi^2_{0.95}(6) = 12.59$
73.83	64.74	1.276	Can't reject H_o
59.04	63.95	0.377	p-value = 0.280
17.00	20.38	0.562	Min Count = 9
8.64	14.55	2.397	
24.80	26.55	0.115	

Wind Speeds - Casper LT/ST Freq Scaled Up

Csp 15Yr WS	Csp 2Yr WS	$(ST-LT)^2/LT$	Chi-Square
12.83	14.30	0.152	7.10
60.49	50.37	2.033	$\chi^2_{0.95}(6) = 12.59$
70.18	61.73	1.157	Can't reject H_o
55.29	58.30	0.155	p-value = 0.225
21.77	24.83	0.378	Min Count = 13
13.74	20.94	2.472	
15.70	19.53	0.752	

Table 4 – χ^2 Temporal and Spatial Correlation Summary

Chi-Square Test Discriminating Power: Scaled Wind Speed Distributions (freq x 250)			
Site	p-value	Paired Sites	p-value
ACC LT/ST	0.916	ACC-Csp	0.000
Csp LT/ST	0.225	ACC-Gil	0.000
Gil LT/ST	0.280	Csp-Gil	0.150
Chi-Square Test Discriminating Power: Scaled Wind Direction Distributions (freq x 250)			
Site	p-value	Paired Sites	p-value
ACC LT/ST	0.995	ACC-Csp	0.000
Csp LT/ST	0.995	ACC-Gil	0.000
Gil LT/ST	1.000	Csp-Gil	0.000

Table 5 – χ^2 Tests with Phi Coefficients

Wind Speeds - ACC Annual Hours				
mph	25-Yr WS	2-Yr WS	$(ST-LT)^2/LT$	Chi-Square
0 - 3	1063	931	17	71
4 - 7	2127	1993	8	$\chi^2_{0.95}(6) = 12.59$
8 - 12	2374	2462	3	**Reject H_o**
13 - 18	1778	1902	9	p-value = 0.000
19 - 24	807	871	5	Min Count = 86
> 24	524	563	3	**Phi-value = 0.09**
Calm	86	38	27	

Wind Speeds - ACC vs Gil Annual Hours				
mph	ACC 25-Yr WS	Gil 13-Yr WS	$(ACC-Gil)^2/ACC$	Chi-Square
0 - 3	1053	461	332	8284
4 - 7	2116	1841	36	$\chi^2_{0.95}(6) = 12.59$
8 - 12	2381	2542	11	**Reject H_o**
13 - 18	1788	2093	52	p-value = 0.000
19 - 24	812	612	49	Min Count = 82
> 24	528	332	72	**Phi-value = 0.97**
Calm	82	878	7732	

Wind Directions - ACC Annual Hours				
Sector	ACC 25Yr WD	ACC 2Yr WD	$(ST-LT)^2/LT$	Chi-Square
N	566	596	1.63	180
NNE	258	224	4.37	$\chi^2_{0.95}(16) = 26.30$
NE	149	139	0.65	**Reject H_o**
ENE	211	201	0.48	p-value = 0.000
E	426	387	3.54	Min Count = 86
ESE	543	564	0.83	**Phi-value = 0.14**
SE	448	452	0.02	
SSE	354	378	1.59	
S	279	302	1.94	
SSW	263	234	3.24	
SW	586	376	75.11	
WSW	1,187	1,292	9.27	
W	1,218	1,430	36.62	
WNW	719	664	4.20	
NW	729	677	3.68	
NNW	738	805	6.25	
Calm	86	38	26.56	

Short and long-term atmospheric stability class distributions, comprising six categories, can be analyzed in the same fashion. The histogram in Figure 3 shows the Antelope stability class distributions to be quite similar between the 2-year and 25-year periods. The χ^2 test can be applied to these distributions by first converting them to annual hours. The test results in Table 6 initially reject the hypothesis that the short and long-term stability class distributions are not statistically different. The low phi coefficient of 0.11, however, adjusts for the large sample size of 8,760 and suggests a large degree of similarity between short and long-term stability class distributions.

Figure 3 – Antelope Short and Long-Term Stability Class Distributions

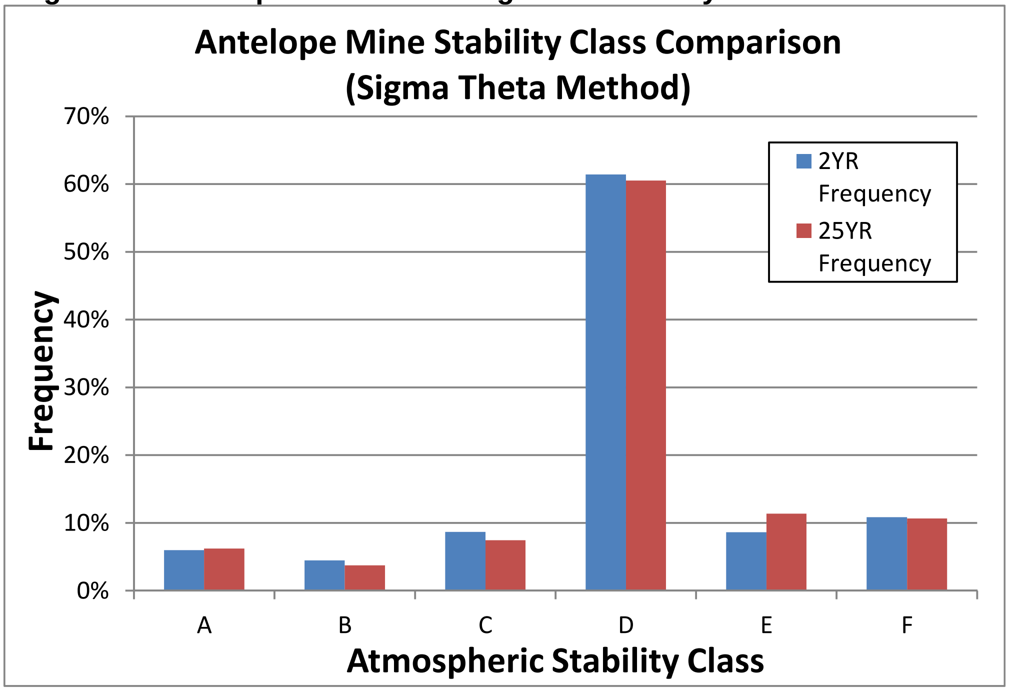


Table 6 – χ^2 Test for Antelope Stability Class Distributions

Class	2YR	25YR	$(ST-LT)^2/LT$	Chi-Square
A	527	547	0.78	107.90
B	391	327	10.55	$\chi^2_{0.95}(5) = 11.07$
C	759	654	14.65	**Reject H_o**
D	5377	5299	1.12	p-value = 0.000
E	755	998	77.76	Min Count = 391
F	950	935	0.25	**Phi-value = 0.11**

Evaluation of the Student's T-Test

The Student's t-test can be used to assess similarity between two sets of data, with the assumption that they both approximate normal distributions. The assumption of normality does not apply to wind and atmospheric stability parameters, but the t-test is fairly robust to this assumption. A more serious limitation is that all frequency distributions have a mean of 1/n, where n is the number of categories. A two-sample Student's t-test will therefore show no statistical difference between any two wind speed frequency distributions, since they have the same mean of 1/7. Likewise, the t-test will fail to differentiate between any two wind direction distributions (same mean of 1/17), or any two stability class distributions (same mean of 1/6). The paired t-test offers no improvement, since the mean difference between paired frequencies will always be zero.

This limitation can be lifted by defining a distribution of ratios between short and long-term frequencies. Two similar distributions should yield frequency ratios with a mean near 1, which can be tested using a one-sample t-test. As expected, the results in Table 7 indicate that a mean of 1 falls within the 95% confidence interval for wind direction frequency ratios at ACC. However, the same conclusion is wrongly inferred for the ratio of ACC-to-Casper frequencies (Table 8), and for all other inter-site comparisons. Very dissimilar frequency distributions will yield ratios with sufficiently large standard deviations to widen the confidence interval and thereby accommodate a mean substantially different than 1.

Table 7 – One-Sample T-Test for Antelope Short/Long-Term Wind Direction Ratio

```
Test of mu = 1 vs not = 1


Variable      N     Mean    StDev  SE Mean        95% CI            T      P
AntRatio_WD  17  1.10918  0.32712  0.07934  (0.94099, 1.27737)  1.38  0.188
```

Table 8 – One-Sample T-Test for Antelope/Casper Wind Direction Ratio

```
Test of mu = 1 vs not = 1


Variable     N     Mean    StDev  SE Mean        95% CI            T      P
ACRatio_WD  17  1.72114  1.56156  0.37873  (0.91826, 2.52402)  1.90  0.075
```

Class-wise T-test Applied to Wind Frequency Data

The two-sample T-test can yield meaningful results when applied separately to each wind speed and wind direction category. Here we are interested in the variation among annual frequencies observed over time for a given category, rather than how the frequencies are distributed over all categories for an aggregated time period. This scenario requires 7 t-tests for wind speeds (6 speed classes plus a “calm” class) and 17 t-tests for wind direction (16 directions plus “calm”). A demonstration of representativeness between frequency distributions will be made if each of the 24 tests fails to reject the null hypothesis that the short and long-term data populations are different. The first sample in each test consists of annual frequencies for a given category over a long period (e.g., 25 years would yield 25 frequencies). The second sample consists of annual frequencies for the same category over a shorter period (e.g., 2 years).

Assume that for a given wind speed or direction category (for example, southerly wind direction):

N_1 = the number of years in the long-term data set

N_2 = the number of years in the short-term or baseline data set

$\bar{X}_1$= the mean annual frequency in the long-term data set

$\bar{X}_2$= the mean annual frequency in the short-term data set

$\bar{S}_1$= the standard deviation of the frequencies in the long-term data set

$\bar{S}_2$= the standard deviation of the frequencies in the short-term data set

$\bar{S}_P$= the pooled standard deviation from short and long-term frequencies

Then the T-statistic for the category of interest is be given by:

$$T = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p\sqrt{\frac{1}{N_1} - \frac{1}{N_2}}} \qquad \text{where} \quad S_p = \sqrt{\frac{(N_1 - 1){S_1}^2 + (N_2 - 1){S_2}^2}{N_1 + N_2 - 2}}$$

The quantity $N_1 + N_2 - 2$ represents the degrees of freedom between the two samples. To ensure sample independence, we might compare a two-year period with the previous 25-year period. The degrees of freedom would then be 25. The critical value $T_{(0.95,25)} = 2.06$, which represents the 95% confidence level in a two-tailed t-test, that the two populations are different. Any t-statistic between -2.06 and +2.06 (or p-value > 0.05) signifies that insufficient evidence exists, at the 95% confidence level, to justify a conclusion that the two populations are different.

Table 9 presents the results of individual t-tests performed on wind speed categories at ACC, using a pooled standard deviation. The long-term wind data span 25 years, from 1987 through 2011. The baseline period spans 2 years, 2012 and 2013. The p-values for each wind speed class are much greater than 0.05, indicating insufficient evidence to conclude a difference between the short-term and long-term wind speed data.

Table 9 – 25-Yr vs. 2-Yr Relative Wind Speed Frequencies and t-test Results

Speed (mph)	25YR Mean	25YR Stdev	2YR Mean	2YR Stdev	Stdev-Pooled	T-Statistic	P-Value
0 - 3	0.1210	0.0173	0.1062	0.0005	0.0170	1.18	0.248
4 - 7	0.2418	0.0331	0.2275	0.0088	0.0325	0.60	0.555
8 - 12	0.2706	0.0183	0.2810	0.0119	0.0181	-0.79	0.439
13 - 18	0.2038	0.0205	0.2171	0.0101	0.0202	-0.90	0.379
19 - 24	0.0929	0.0161	0.0995	0.0016	0.0158	-0.57	0.577
> 24	0.0601	0.0154	0.0643	0.0094	0.0152	-0.37	0.713
Calm	0.0099	0.0118	0.0043	0.0003	0.0116	0.65	0.523

Table 10 presents the results of individual t-tests performed on wind direction categories for the same periods at ACC, using a pooled standard deviation. Again, the p-values for each wind direction class are much greater than 0.05, indicating insufficient evidence to conclude a difference between the short-term and long-term wind direction data.

The use of a pooled standard deviation assumes the unknown variances of the two populations are equal. This would appear reasonable given that wind data at any given location are influenced by fundamental climatological and topographical conditions that do not change appreciably from year to year. Histograms of the yearly frequencies in each wind speed and direction category confirm that these data are approximately normally distributed. Nonetheless, the assumption of equal variances can be supported or refuted by an F-Test or Levene's test performed for each two-sample t-test. This step is particularly relevant given that the short-term data set in this case contains only two wind frequencies. The F-test assumes normally distributed data, while Levene's test relaxes this assumption. For conservatism, Levene's test is employed to validate the assumption of equal variances in both sets of t-tests. If the p-value for Levene's test is less than 0.05, we conclude that the variances are unequal and repeat the t-test based on unequal variance. If the p-value is greater than 0.05, we preserve the t-test results based on the pooled estimate of the standard deviation, S_p.

Table 10 – 25-Yr vs. 2-Yr Relative Wind Direction Frequencies and t-test Results

Wind Direction	25YR Mean	25YR Stdev	2YR Mean	2YR Stdev	Stdev-Pooled	T-Statistic	P-Value
N	0.0643	0.0187	0.0681	0.0086	0.0184	-0.28	0.781
NNE	0.0292	0.0114	0.0256	0.0056	0.0113	0.43	0.670
NE	0.0169	0.0047	0.0159	0.0034	0.0046	0.29	0.777
ENE	0.0239	0.0097	0.0230	0.0032	0.0095	0.13	0.899
E	0.0484	0.0189	0.0442	0.0034	0.0186	0.31	0.758
ESE	0.0620	0.0117	0.0644	0.0029	0.0115	-0.28	0.782
SE	0.0507	0.0120	0.0515	0.0071	0.0118	-0.10	0.923
SSE	0.0399	0.0116	0.0431	0.0036	0.0114	-0.38	0.705
S	0.0314	0.0108	0.0345	0.0019	0.0106	-0.39	0.701
SSW	0.0299	0.0092	0.0267	0.0031	0.0091	0.48	0.637
SW	0.0677	0.0313	0.0429	0.0177	0.0309	1.09	0.286
WSW	0.1364	0.0235	0.1475	0.0301	0.0238	-0.64	0.530
W	0.1392	0.0246	0.1632	0.0078	0.0242	-1.35	0.190
WNW	0.0823	0.0115	0.0758	0.0099	0.0114	0.77	0.446
NW	0.0836	0.0164	0.0773	0.01	0.0162	0.53	0.604
NNW	0.0844	0.0116	0.0919	0.0042	0.0114	-0.90	0.378
Calm	0.0099	0.0118	0.0043	0.0003	0.0116	0.65	0.523

Table 11 lists the p-values from Levene's test applied to the short and long-term wind speed frequencies for each class. Since all of the p-values are greater than 0.05, the test confirms the assumption of equal variances among relative frequencies for each wind speed, and thereby validates the t-test results in Table 9.

Table 11 – Levine'sTest for Equal Variance Among Wind Speed Frequencies

Wind Speed (mph)	Levene's p-value
0 - 3	0.135
4 - 7	0.800
8 - 12	0.570
13 - 18	0.324
19 - 24	0.246
> 24	0.433
Calm	0.455

Table 12 lists the p-values from Levene's test applied to wind direction frequencies. Since all p-values are greater than 0.05, the test confirms the assumption of equal variances among relative frequencies for each wind direction, and thereby validates the t-test results in Table 10.

Table 12 – Levine’sTest for Equal Variance Among Wind Direction Frequencies

Wind Direction	Levene’s p-value
N	0.326
NNE	0.634
NE	0.819
ENE	0.546
E	0.204
ESE	0.340
SE	0.442
SSE	0.071
S	0.085
SSW	0.233
SW	0.547
WSW	0.708
W	0.380
WNW	0.875
NW	0.499
NNW	0.268
Calm	0.455

Notably, the south (S) and south-southeast (SSE) directions exhibit marginally low p-values in Table 12. The t-test can be repeated for these directions, assuming unequal variances. The resulting p-values are 0.420 and 0.262. As expected, these are lower than the p-values of 0.705 and 0.701 shown in Table 10 (which assumed equal variances). However, they are still substantially greater than the critical p-value of 0.05. Thus, in those two cases where the assumption of equal variances may be in doubt, we find the assumption is not needed. The t-test has demonstrated, for all wind directions, a lack of significant difference between short and long-term frequencies.

Since the t-test failed to show a statistically significant difference between short and long-term wind speed and direction frequencies, we conclude that the two-year baseline period is representative of the previous 25-years at ACC.

Support for the class-wise t-test performed on wind frequency distributions can be found in the literature of meteorological statistics. Brooks and Carruthers (Brooks 1978, p. 66) offer an example that seeks to determine whether the frequency of occurrence of gale-force winds over a 3-year period is the same as the frequency of gale-force winds over a previous 9-year period. A two-sample t-test is used to demonstrate a significant difference between the two frequencies. This approach is equivalent to the above analysis, except that Brooks and Carruthers applied it to only one wind speed category.

The class-wise t-test will reject the null hypothesis of similarity between wind data from different sites. For example, the t-test leads to the conclusion that wind speeds and wind directions are not similarly distributed between the ACC and Gillette sites. Table 13 shows that only 3 of the 7 wind speed classes do not exhibit a statistical difference between the two sites (p-values > 0.05).

Table 13 – Inter-site t-test Results for Wind Speed

Antelope 25YR Comparison to Gillette 2YR							
Speed (mph)	ACC 25YR Mean	ACC 25YR Stdev	Gil 2YR Mean	Gil 2YR Stdev	Stdev-Pooled	T-Statistic	P-Value
0 - 3	0.1210	0.0173	0.0551	0.0015	0.0170	5.29	0.000
4 - 7	0.2418	0.0331	0.1842	0.0039	0.0325	2.41	0.023
8 - 12	0.2706	0.0183	0.2590	0.0004	0.0179	0.88	0.387
13 - 18	0.2038	0.0205	0.2559	0.0126	0.0203	-3.50	0.002
19 - 24	0.0929	0.0161	0.0816	0.0054	0.0159	0.97	0.340
> 24	0.0601	0.0154	0.0581	0.0047	0.0151	0.18	0.860
Calm	0.0099	0.0118	0.1061	0.0074	0.0117	-11.22	0.000

Likewise, Table 14 shows that only 6 of the 17 wind direction classes do not exhibit a statistical difference between the two sites (p-values > 0.05).

Table 14 – Inter-site t-test Results for Wind Direction

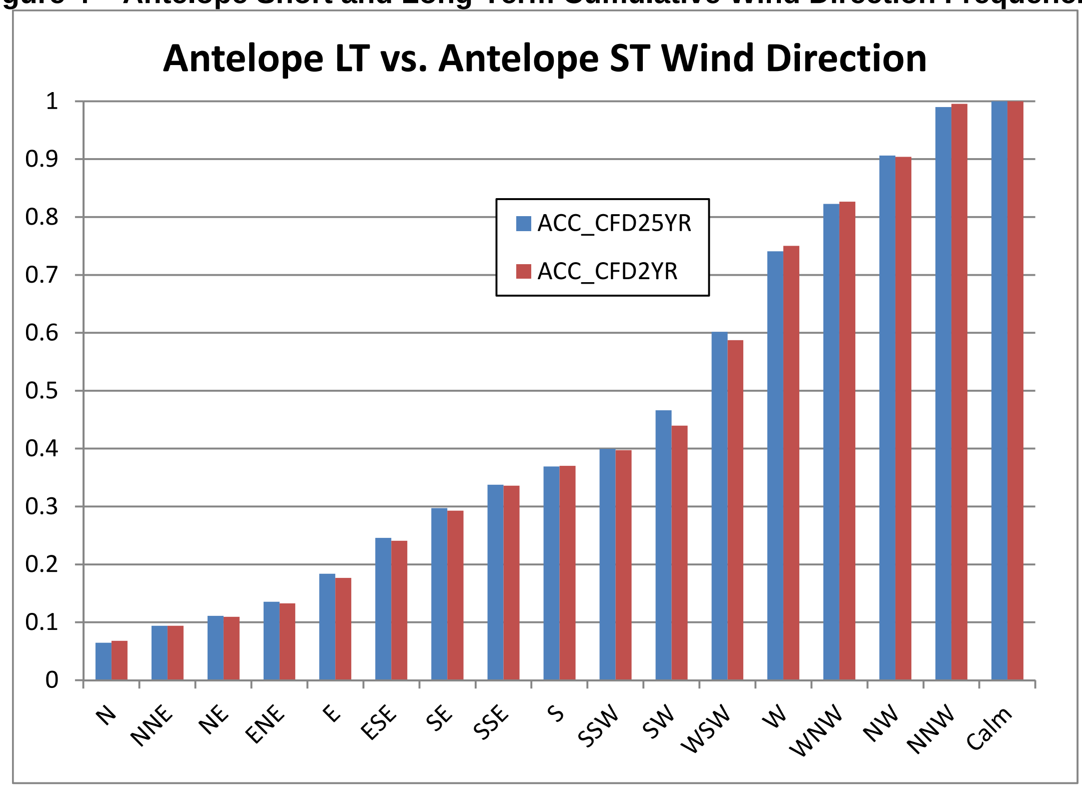
Antelope 25YR Comparison to Gillette 2YR							
Wind Direction	ACC 25YR Mean	ACC 25YR Stdev	Gil 2YR Mean	Gil 2YR Stdev	Stdev-Pooled	T-Statistic	P-Value
N	0.0643	0.0187	0.0651	0.00171	0.0183	-0.07	0.948
NNE	0.0292	0.0114	0.0347	0.00224	0.0112	-0.67	0.509
NE	0.0169	0.0047	0.0207	0.00216	0.0046	-1.13	0.270
ENE	0.0239	0.0097	0.0174	0.00385	0.0095	0.92	0.365
E	0.0484	0.0189	0.0203	0.00322	0.0186	2.06	0.050
ESE	0.0620	0.0117	0.0143	0.00103	0.0115	5.65	0.000
SE	0.0507	0.0120	0.0273	0.00046	0.0117	2.72	0.012
SSE	0.0399	0.0116	0.0671	0.00591	0.0114	-3.24	0.003
S	0.0314	0.0108	0.1152	0.01599	0.0111	-10.29	0.000
SSW	0.0299	0.0092	0.0493	0.00954	0.0092	-2.85	0.009
SW	0.0677	0.0313	0.0755	0.02027	0.0309	-0.35	0.733
WSW	0.1364	0.0235	0.0464	0.00286	0.0230	5.32	0.000
W	0.1392	0.0246	0.0779	0.00627	0.0242	3.46	0.002
WNW	0.0823	0.0115	0.0567	0.00454	0.0113	3.09	0.005
NW	0.0836	0.0164	0.1029	0.01573	0.0164	-1.60	0.121
NNW	0.0844	0.0116	0.1032	0.00811	0.0115	-2.22	0.036
Calm	0.0099	0.0118	0.1061	0.00741	0.0117	-11.22	0.000

Evaluation of the Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) test is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare two samples. The Kolmogorov–Smirnov (K-S) statistic quantifies a distance between the cumulative distribution functions (CDF's) of two samples. It is most effectively applied to continuous data, but will generally accommodate arbitrarily classified data (such as wind directions and stability classes) and ordinal data (such as wind speed frequencies). In exchange for this broad applicability, it sacrifices statistical efficiency. For the meteorological data considered in this report, the K-S test errs on the side of inferring similarity even where little exists.

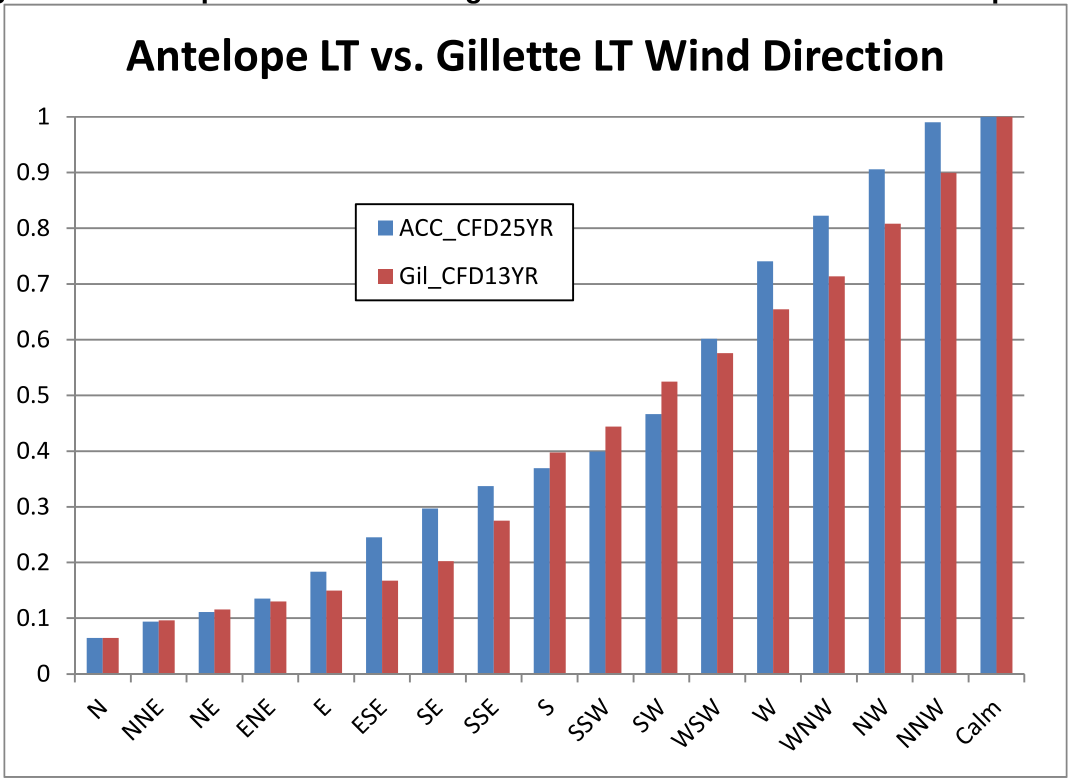
The CDF's for Antelope wind direction are shown in Figure 4. The K-S statistic is expressed here as the maximum $|LT_k – ST_k|$ where LT_k and ST_k are the cumulative long-term and short-term frequencies through class "k." The K-S statistic is seen in Figure 4 as the greatest vertical difference between adjacent blue and red bars, in this case 0.026 (the SW direction). This test statistic is compared to the critical value for a sample size of 17 and a confidence level of 95%. The critical K-S value in this instance is 0.318, as found in a standard statistical table (NRC 2011). Because the K-S statistic is less than the critical value, we conclude that the short and long-term wind direction frequencies come from a common distribution. In other words, one is representative of the other.

Figure 4 – Antelope Short and Long-Term Cumulative Wind Direction Frequencies



The question arises whether the K-S test will discriminate between similar and dissimilar wind distributions. Repeating the K-S test for Antelope and Gillette wind direction distributions, we obtain the same outcome. Figure 5 shows substantial differences between the CFD's, yet the K-S statistic of 0.109 (WNW direction) is still well under the critical value of 0.318.

Figure 5 – Antelope and Gillette Long-Term Cumlative Wind Direction Frequencies



The bottom half of Table 15 confirms the poor efficiency of the K-S test for wind speed and direction comparisons. In all cases, the comparisons show no statistically significant difference. The inability of the K-S test to distinguish between clearly dissimilar wind patterns eliminates this method as a viable alternative.

Table 15 – Comparative K-S Test Results

Site (s)	Parameter	K-S Statistic	Critical K-S Value	Inference
Antelope LT/ST	Wind Speed	0.030	0.318	Similar Distributions
Antelope LT/ST	Wind Direction	0.026	0.483	Similar Distributions
Casper LT/ST	Wind Speed	0.068	0.318	Similar Distributions
Casper LT/ST	Wind Direction	0.047	0.483	Similar Distributions
Gillette LT/ST	Wind Speed	0.064	0.318	Similar Distributions
Gillette LT/ST	Wind Direction	0.026	0.483	Similar Distributions
Ant LT/Csp LT	Wind Speed	0.071	0.318	Similar Distributions
Ant LT/Csp LT	Wind Direction	0.149	0.483	Similar Distributions
Ant LT/Gil LT	Wind Speed	0.097	0.318	Similar Distributions
Ant LT/Gil LT	Wind Direction	0.109	0.483	Similar Distributions
Csp LT/Gil LT	Wind Speed	0.036	0.318	Similar Distributions
Csp LT/Gil LT	Wind Direction	0.175	0.483	Similar Distributions

Evaluation of Linear Correlation and Linear Regression

The following discussion combines linear correlation and regression since they yield closely related statistics. Under the assumptions applied to wind frequency distributions the Pearson's correlation coefficient R is equal, or very nearly equal to the square root of the linear regression coefficient of determination R^2. While linear regression has not been commonly employed to demonstrate the degree of similarity between two meteorological frequency distributions, linear correlation coefficients have (Coffin 1996). Using either approach to assess a linear association between two relative frequency distributions is indicated, since their component frequencies each sum to 1. If two such distributions are similar, they will necessarily be linearly associated and the scatterplot of their frequencies will cluster around the identity line.

We do not seek a causal relationship between long-term and short-term frequencies, as linear regression is often misconstrued to do. Brooks and Carruthers point out that a correlation coefficient is merely a mathematical expression of the "correspondence" between two series of numbers, whose relationship may be indirect and tied to a third variable (Brooks 1978, p. 226). This applies particularly to wind distributions, the short and long-term measurements of which only approximate the true distributions. The long-term variable (X) might be thought of as "independent" and the short-term variable (Y) as "dependent." This is consistent with assigning the dependent variable to the unknown or most uncertain of the two (Brooks 1978). But because both variables are measurements with nonparametric distributions of roughly equal variances, reversing this assignment will also work. Chapter 19 of NU-REG 1475 (US NRC 2011) states that with linear correlation neither variable is necessarily dependent or independent. For the cases analyzed in the context of this report, the distinction becomes mostly arbitrary. The choice affects the R^2 values very slightly due to the nature of a least-squares fit (switching the short and long-term variables altered the R^2 values by less than 0.01 in all cases).

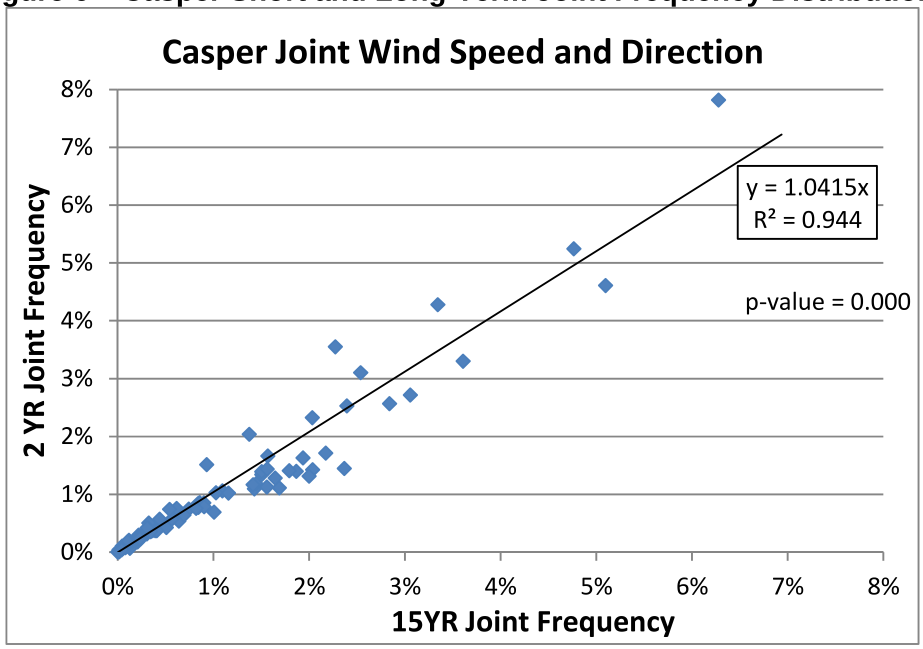
If a zero intercept is enforced, the slope of the regression line approaches unity as the linear relationship approaches equality (i.e., Y = X). The coefficient of determination R^2 (or the correlation coefficient R) reflects the strength of the linear relationship. The p-value in the regression analysis reflects the statistical significance of this conclusion.

Accordingly, several refinements to the regression analyses in previous submittals to NRC should be noted:

1. Adopting the convention of assigning long-term frequencies to the independent variable
2. Using non-overlapping short-term and long-term periods to enforce strict sample independence
3. Forcing the regression line to pass through the origin (zero intercept) in recognition of the fact that two relative frequency data sets that each sum to 1 cannot exhibit a systematic bias

Figure 6 illustrates the linear association between joint wind speed and direction frequencies at Casper, representing a 2-year baseline period (1/1/2012 - 12/31/2013) and a 15-year long-term period (1997-2011). The hourly data for each distribution fall into one of 97 categories. In this two-way classification, the product of the 6 speed classes and 16 direction sectors, plus a calm category equals 97. The graph illustrates the degree to which the 2-year joint frequencies match the 15-year frequencies. In this case, the right-most point on the graph happens to correspond to the calm category, which occurred 6.3% of the time during the 15-year period (X), and 7.8% of the time during the past 2 years (Y). The other points represent the remaining 96 categories. The R^2 value of 0.94 confirms a strong linear relationship, and the slope of 1.04 indicates substantial equivalence between short and long-term frequencies. A p-value of zero leaves little doubt that this relationship is significant. Linear correlation and regression analysis weight each pair of frequencies in proportion to their magnitude. Unlike the χ^2 test, which can weight very small frequencies disproportionately, with linear regression large relative differences between frequencies in seldom-occurring wind categories do not distort the overall strength of association.

Figure 6 – Casper Short and Long-Term Joint Frequency Distributions



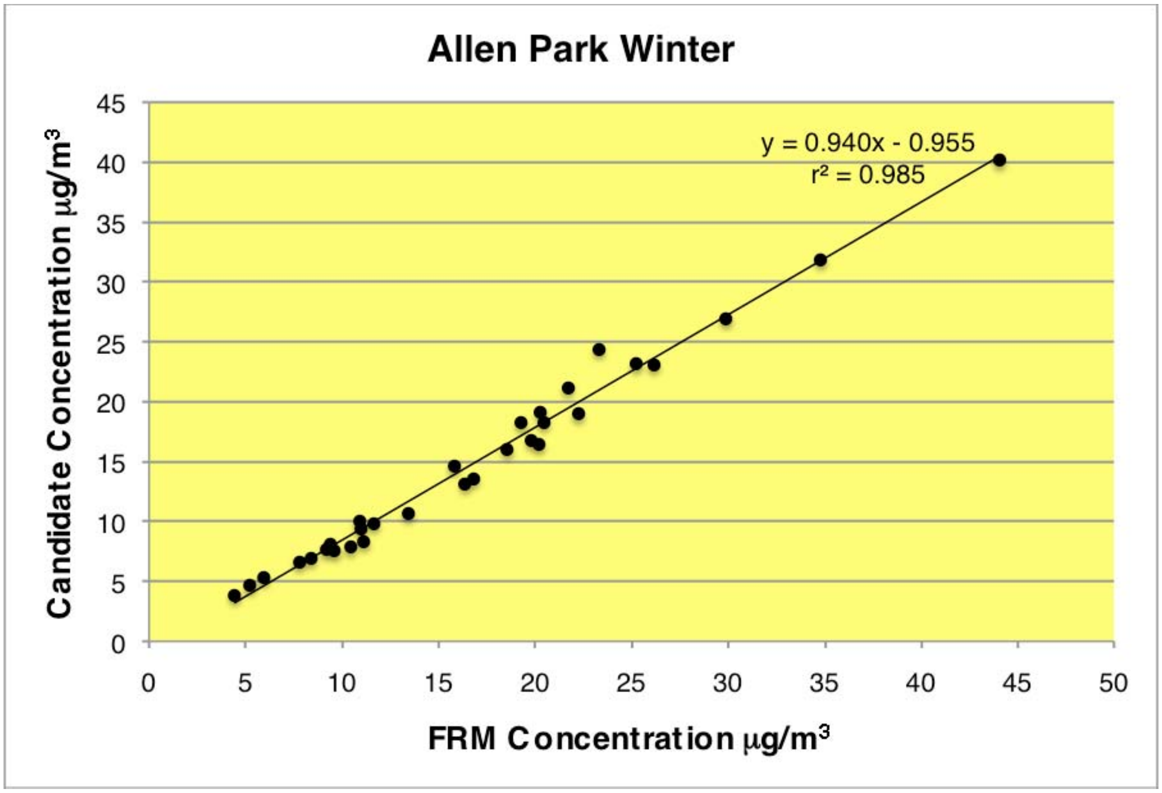
The statistical significance of the correlation coefficient R can be found using a t-test.

$$t = \frac{R}{\sqrt{\frac{(1 - R^2)}{N - 2}}}$$

The t-test performed on joint wind frequency data in Figure 6 yields a p-value of 0.000, indicating extremely high confidence in the correlation coefficient R.

There is ample precedent for using linear regression to demonstrate statistical equivalence between two separate measurements of the same fundamental variable. For example, EPA specifies linear regression to demonstrate that non-certified monitors adequately measure ambient air quality. According to Part 58 of 40 CFR, Appendix G, particulate matter measurements from non-Federal Reference Method (FRM) monitors may be used for the purpose of reporting the Air Quality Index if a linear relationship between these measurements and reference method measurements can be established by statistical linear regression. In its data quality objectives (EPA 2002) EPA determined that the statistical parameter of interest is the R^2 parameter which measures the square of the correlation coefficient between measured and modeled FRM $PM_{2.5}$ data. Figure 7 presents such a demonstration by Met One in its application for federal designation of its BAM-1020 air particulate monitor. EPA subsequently granted the designation.

Figure 7 – BAM-1020 Regression Fit to FEM Monitor (Gobeli 2008)



The nature of the two variables in the EPA example is similar to the NRC long-term representativeness requirement. There is no causal relationship between independent and dependent variables. Both $PM_{2.5}$ monitors provide approximate measurements of a third variable, which is the true (and unknown) ambient concentration. There is a higher confidence in the FRM measurements; therefore they are assigned to the independent variable and the candidate monitor measurements are compared to them to demonstrate equivalency. For the meteorological application, baseline and long-term meteorological measurements approximate the true (and unknown) distributions of wind speed, wind direction, joint frequency, and atmospheric stability class. The shorter baseline period invites more uncertainty than the longer period, so we validate baseline frequencies by demonstrating a strong linear relationship with long-term frequencies.

The MILDOS model accepts meteorological inputs in the form of joint wind speed, wind direction and stability class frequency distributions, also known as STAR distributions. An important subset of the STAR distribution is the two-way wind classification, which categorizes hourly wind data by both speed and direction. Hypothesis testing is generally unworkable in comparing joint wind speed and direction frequencies because the wind data are partitioned into too many categories. Brooks and Carruthers offer a general rule, that the number of categories in hypothesis testing should not exceed $5*\log_{10}(N)$, where N is the sample size (Brooks 1978). For a one-year sample of hourly averages (N = 8,760) the maximum number of categories would be 20. This limit is consistent with 7 wind speed classes or 17 wind directions, but not with 97 joint frequency categories. Thus, hypothesis testing methods such as the χ^2 test and the t-test are limited to one-way classified wind data.

Among the statistical methods considered, joint wind speed and direction distributions are only amenable to linear regression or correlation (see Figure 6). Analyzing these distributions can strengthen the case for long-term representativeness of baseline wind data. The joint analysis offers the most rigorous comparison between short and long-term wind frequency distributions, because a conclusion of representativeness demands not only that wind direction frequencies be similarly distributed, but that they also be similarly distributed within each wind speed class. In this sense, linear regression analysis of short and long-term joint wind distributions is the best quantitative measure of the similarity between the associated wind roses.

To summarize, linear regression analysis is deemed appropriate to demonstrate long-term representativeness of wind and stability class frequency data, for the following reasons:

1. The scatterplot and fitted line provide visual confirmation of a linear correlation between short and long-term wind category frequencies. R^2 gives the strength of a linear relationship, and the p-value gives the degree of confidence in the result.
2. Inherent to every relative frequency distribution is the requirement that all of the individual frequencies must sum to 1; therefore, demonstrating a strong linear association also demonstrates equivalence.
3. Linear regression can be applied to joint wind speed and direction frequencies, which correspond more closely than wind speed or wind direction frequencies by themselves, to the wind data format used by MILDOS.
4. Linear regression isolates the sources of variation among category frequencies. When multiplied by 100, R^2 signifies the percent of variation from a mean frequency that is common to both short and long-term distributions. In Figure 6, for example, 94% of the variation among 2-year joint frequencies can be predicted based on measured long-term frequencies, while only 6% is attributed to random, year-to-year fluctuations and/or measurement error.
5. Linear regression distinguishes well between generally weak spatial correlation and generally strong temporal correlation.

The last of these reasons can be demonstrated by summarizing the results of regression analyses performed on wind direction and joint wind speed and direction frequencies at all three reference sites (Table 16). The "LT/ST" designation signifies a comparison of long-term/short-term data at a given site. All comparisons between two sites were made using long-term data.

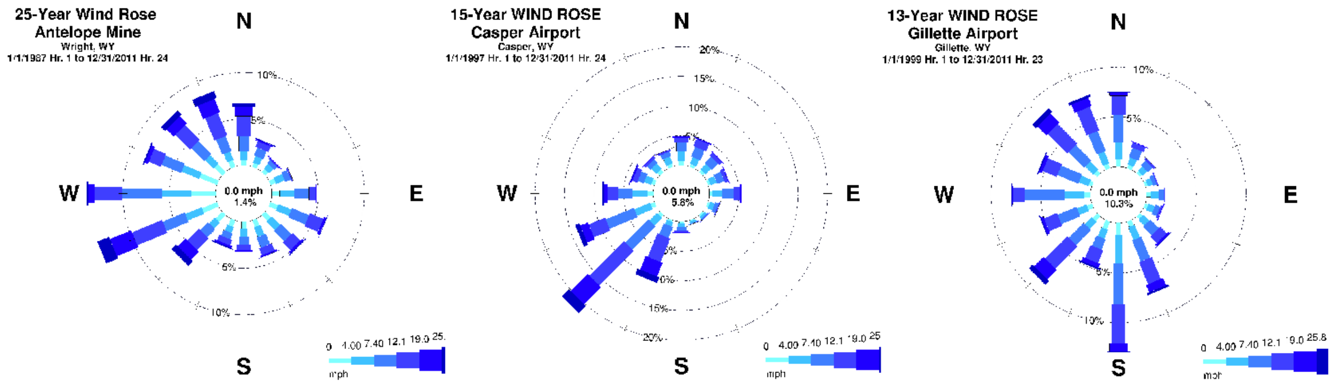
Table 16 – Linear Regression Temporal and Spatial Correlation Summary

Discriminating Power of Linear Regression: Wind Direction Distributions			
Site	R^2	Paired Sites	R^2
ACC LT/ST	0.960	ACC-Csp	0.078
Csp LT/ST	0.980	ACC-Gil	0.034
Gil LT/ST	0.974	Csp-Gil	0.005

Discriminating Power of Linear Regression: Joint Wind Speed and Direction Distributions			
Site	R^2	Paired Sites	R^2
ACC LT/ST	0.933	ACC-Csp	0.142
Csp LT/ST	0.962	ACC-Gil	0.165
Gil LT/ST	0.973	Csp-Gil	0.278

An R^2 greater than 0.9 indicates strong correlation; an R^2 less than 0.5 indicates weak correlation, if any. Table 16 shows that short and long-term joint frequencies correlate very strongly at each of the three locations, but very weakly between locations. Thus, regression analysis discriminates between similar and dissimilar wind roses (see Figure 8).

Figure 8 – Long Term Wind Roses for Antelope, Casper and Gillette



Linear correlation produces Pearson's correlation coefficient R, based on the assumption of normally distributed data. This assumption can be relaxed by ranking the data and computing Spearman's correlation coefficient, a method commonly applied to nonparametric data. Analyzing 25-year and 2-year wind frequency distributions at ACC, we obtain Spearman's R = 0.986 for wind speed and 0.983 for wind direction. Hence, the assumption of normality does not introduce appreciable error in the linear correlation analysis.

<u>Conclusion</u>
In fulfillment of NRC guidelines, the combination of linear correlation and hypothesis testing provides a comprehensive demonstration of long-term representativeness. For the ACC, Casper and Gillette sites, the most recent two years of hourly wind data are statistically no different than the previously recorded long-term data at each site. This conclusion is corroborated by three tests, which have been jointly applied by others to analyze seasonally classified cloud cover (Lowther 1991, pp. 32-33):

1. χ^2 test, with the phi coefficient to adjust for large sample size
2. The Student's t-test
3. Linear correlation coefficient

Not only do these methods demonstrate temporal uniformity in wind patterns at any of the three sites, but they confirm statistically significant differences in wind patterns between these sites.

Brooks and Carruthers (Brooks 1978) stated, "All statistical tests must be interpreted in the light of common sense." We should not discount the role of graphical comparisons such as wind roses, histograms, and scatter plots with fitted regression lines, in validating statistical test outcomes. In the final analysis, graphical illustrations may be the most useful of these (Gardiner 1979). Figures 9 and 10 below leave little doubt that wind speed and direction patterns in eastern Wyoming do not vary nearly as much from year to year as they do from place to place.

Figure 9 –Long-Term and Short-Term (Baseline) Wind Roses

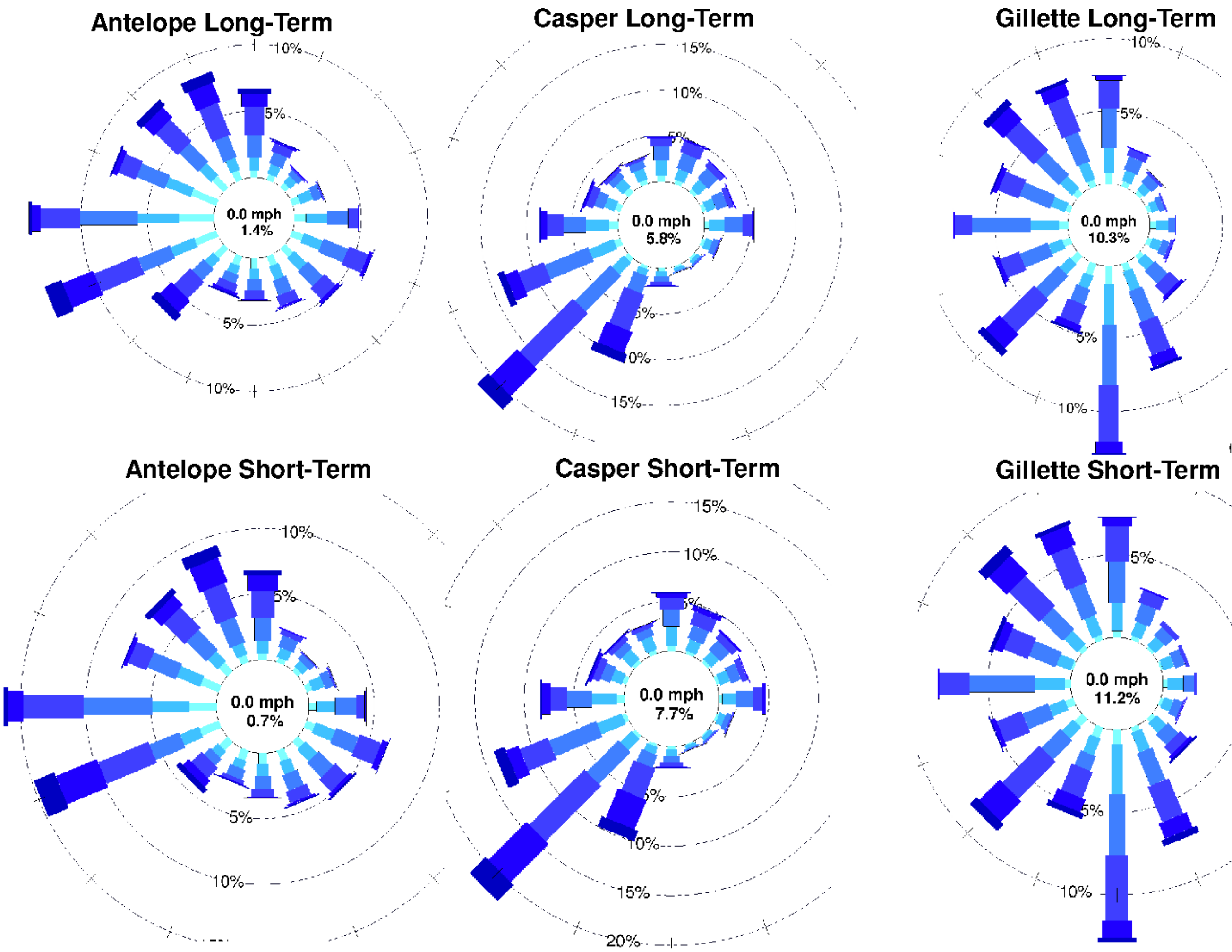
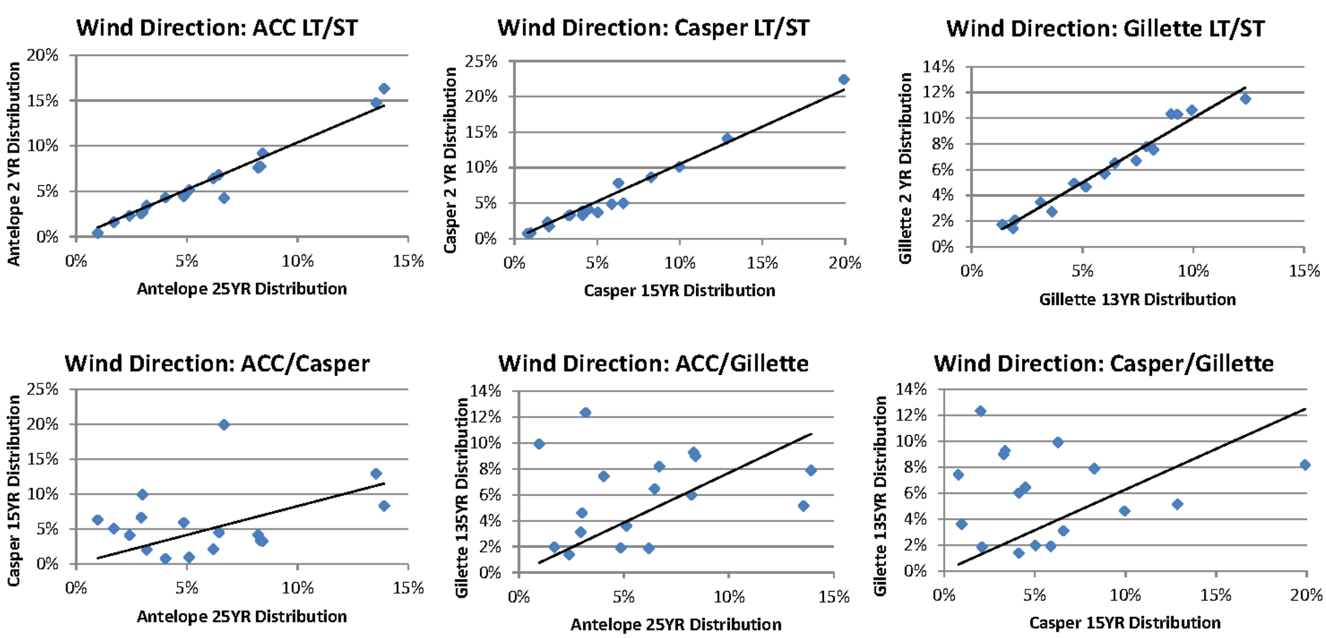


Figure 10 – Temporal and Spatial Wind Direction Frequency Comparisons



References

Brooks 1978, Handbook of Statistical Methods in Meteorology, C. E. P. Brooks and N. Carruthers, Reprint of 1953 Edition, 1978.

Coffin 1996, Consolidated Statistical Background Papers, U.S. Air Force Climatology Center, Charles R. Coffin, November 1996

EPA 2002, Data Quality Objectives (DQOs) For Relating Federal Reference Method (FRM) and Continuous PM2.5 Measurements to Report an Air Quality Index (AQI), Shelly Eberly,Terence Fitz-Simons, Tim Hanley, Lewis Weinstock, November 2002.

Gardiner 1979, Analysis of Frequency Distributions, Concepts and Techniques in Modern Geography, V. Gardiner & G. Gardiner, 1979.

Gobeli 2008, Met One Instruments BAM-1020 Beta Attenuation Mass Monitor US-EPA PM2.5 Federal Equivalent Method Field Test Results, Paper # 2008-A-485-AWMA, David Gobeli, Ph.D., Herbert Schloesser, Ph.D. and Thomas Pottberg, 2008.

Hessen 2006, The Multigroup Common Factor Model With Minimal Uniqueness Constraints and the Power to Detect Uniform Bias. Applied Psychological Measurement, Vol. 30 No. 3, Hessen, D. J., Dolan, C. V, & Wicherts, J. M., May 2006.

IML 2014, Meteorological Monitoring Database for Antelope Coal Company, IML Air Science, POR 1987-2014.

Lowther 1991, RTNEPH Total Cloud Cover Validation Study, Capt Ronald P. Lowther, Mr. Mark T. Surmeier, Capt Richard W. Hartman, Mr Charles R. Coffin, Capt Anthony J. Warren, November 1991.

NCDC 2014, National Climate Data Center, National Oceanic and Atmospheric Administration, Quality Controlled Local Climatological Data (after 1/1/2005), Unedited Local Climatological Data (pre 2005), accessed April 2014.

NRC 2011, NU-REG 1475, Revision 1, Applying Statistics, March 2011.

NRC 2013, Thomas Lancaster (NRC) Letter to Josh Leftwich (Cameco Resources), ML13177A344, July 3, 2013.

NRC 2014, Public Meeting Summary (NRC), New Request for Additional Information (RAI) - Meteorology, ML14099A027, April 10, 2014.

Sharpe 2012, Business Statistics Second Edition, Chapter 15, Inference for Counts: Chi-Square Tests, Sharpe, DeVeaux and Velleman, 2012.