APPENDIX 2

Technical Study of Weather and Normals

I. Introduction

A number of tests were performed on the degree day record available to the Companies. The Companies calculate weighted average degree day values for different areas in their three Illinois service territories. None of the data goes back beyond 1971. For studies of weather normalization, it is highly desirable to have a long period of record. The primary reason being that to compute 30-year normals, data from 1971-2000 would be the first 30-year normal and only a limited number could be calculated.

As a result, based on discussions with Company personnel we determined that Champaign-Urbana ("CU") could be used as a proxy for studying 10-year versus 30-year normals. The station has a long period of record (back to 1889), it is a first order station and it is fairly centrally located in reference to the service territories of the three Ameren Illinois subsidiaries. As indicated in the Direct testimony, I ran correlation analysis between the Company data used in the three filings versus CU. The correlations were high and Company witness Willen and I agreed that for my long term study, CU data would be a good proxy. Data is available for a period of record beginning in 1889 through 2006 or 118 data points. Due to this very long period of record and its geographical location, the degree days analysis provides reasonable conclusions. The analysis was based on the degree days in a 12 calendar month period. The primary goal of the research was to determine if 10-year normals were more predictive and accurate than 30-year normals, for purposes of normalizing billing determinants in the Companies' gas rate cases. The normals are determined by calculating the average annual heating degree days in each of the yearly groupings over the entire data set.

This document also includes definitions for the various statistical terms and tests that were used and summarized on Ameren Exhibit 14.4G. See Section IV. below. Other statistical terms, mostly basic terms, are defined and/or described in Section VIII.

II. Data and Basic Statistics

Ameren Exhibit 14.4G shows the summary results of the major tests performed, contrasting 10- and 30-year normals. One of the most noteworthy findings, particularly for the past 30 years, is the continuing decline in the mean degree days as the groupings include fewer years. For example, Page 1 of Ameren Exhibit 14.3G clearly shows that 30-year or other longer-term normals (rolling or moving averages) are not representative of recent weather. Looking at Page 3 of Ameren Exhibit 14.3G, one can see that there has been a substantial declination of heating degree-days over the past 15 years or so. It is also apparent from Page 4 of this exhibit that the 30-year average has been in decline since about 1990.

I performed a number of regression analyses, the results of which are shown at Ameren Exhibit 14.4G Page 5 and in the Workpapers. This graph illustrates three periods: warming,

cooling and warming. This indicates that data is not necessarily flat over the entire record and thus not necessarily stationary. Regardless, in weather analysis for weather normalization, one is more concerned with recent data, not what happened many decades ago.

III. Summary and Conclusions

Our primary goal was to determine if 10-year normals are more appropriate than 30-year normals. We used a bank of statistical tests, of which the primary tests are more important. Secondary tests are of lesser importance and mostly useful for tie-breaking. Ameren Exhibit 14.4G shows that in comparing 10- and 30-year normals, 10 years are best - they win 5 of the 6 primary tests. They win 12 of 14 secondary tests and 17 of 20 total tests.

IV. Accuracy and Stability Tests

A. Introduction

Ameren Exhibit 14.4G shows a summary of the results of the main tests used in this study. Determining weather normals as applied in a utility rate case is simply forecasting what next year's degree-days will be. In forecasting, one is concerned with **accuracy** and **stability**. Accuracy pertains to how accurate the forecasted variable is compared to the actual measure. Stability pertains to variation. In general, one is most concerned with accuracy first and then stability. Thus, given two methods that are equally accurate, one would opt to use the method that has the least variability. I used a battery of tests composed of those considered primary and secondary.

In this work, 10-and 30-year normals were compared. The normals are determined on the basis of the average of the prior x years where x is 10 or 30. Thus, the 10 year normal to be used in 2006 reflects the average of the degree days in the 10 calendar years 1996-2005. The primary determinant of accuracy is the difference between the forecasted value and the actual degree days for the year. This is classically referred to as the **error** or **deviation from actual** ("DFA").

Mathematically, $DFA = F_i - A_i$ where F is the forecasted value, A is the actual value and i = the year. In this case, i = 1 to 88 corresponding to forecast years from 1919-2006. In the following formulas unless otherwise shown, summations (denoted by Σ) are taken from 1 to 88. The number of years, N, equals 88.

While the full data record is 1889-2006, the tests are based on data for 1918-2005, since 1918 is the first year for which test normals may be calculated for 30-year normals. It takes the data from 1889-1918 to generate the first 30-year normals. Normals through a given year (e.g. 1990) are compared to the actual degree day data for the next year (e.g. 1991).

In reviewing the results, I have grouped certain tests in boxes. For example, the best predictor based on Mean Prediction error would also be the winner of Absolute Prediction error or Mean

Absolute Percent error.

B. Primary Tests

Prediction Error (Absolute)

Absolute prediction error is determined by taking the normal value for a year, subtracting the actual value and determining the deviation from actual or error. The absolute value (denoted as | |) is then taken for each of the DFAs. Finally, the sum or cumulative value is determined for the period 88 years ending 2006. 10 -year normals performed best under this test.

$$PE= \Sigma|(DFA_i)|$$

Mean Prediction Error

This method is also referred to as Mean Error or Mean Absolute Deviation or Average Absolute Deviation. It is determined as the average of the absolute prediction errors. Both prediction errors would be divided by the same number of years to determine the averages.

$$MPE= \Sigma\ |(DFA_i)|/N$$

The standard deviation of the absolute prediction errors were determined. The 10-year normals performed best under this test.

Mean Absolute Percent Error

Here the DFA is calculated as a percentage difference from the actual value. The absolute value is determined and finally the average over the 88 years is determined.

$$MAPE= \Sigma\ [(DFA_i)/A_i]*100|/N$$

The standard deviation of the MAPE shows that 10-year normals performed best.

Sum (Total) of the Squared Error

Each of the DFA values (not the absolute values) is squared and then summed over the 88 years.

$$SSE= \Sigma\ (DFA_i)^2$$

The standard deviation of the squared errors for 10-year normals performed best.

Mean Squared Error

This is also referred to as the Extrapolation Variance. It is simply an average of the squared errors. 10-year normals performed best under this test.

$$MSE = \Sigma (DFA_i)^2/N$$

Root Mean Squared Error

RMSE is also referred to as the Standard Error of Extrapolation. It is the square root of the MSE. 10-year normals performed best under this test.

$$RMSE = \sqrt{\Sigma (DFA_i)^2/N}$$

Closest Predictor

This test sets up a frequency of the normal that is closest to actual for each year. 10-year normals won.

Mean Percent Error of Closest Predictor

This test shows the average percent anomaly between normal and actual for each of the years that the respective normal (10 vs.30) was best.his was the only primary test under which 30-year normals performed better.

C. Secondary Tests

Most of the secondary tests are based on the DFA, not the absolute DFA and are therefore less important. In some cases, they can be misleading.

Net Total Bias

Net total bias is determined by taking the normal value for a year, subtracting the actual value and determining the deviation from actual or error. The sum or cumulative value is determined for the study. It is the same as absolute prediction error except DFAs are used rather than absolute DFAs. 10-year normals performed best under this test.

$$NTB = \Sigma (DFA_i)$$

Average Bias

AB is the average of the net total bias. 10-year normals performed best under this test.

$$AB = \Sigma (DFA_i)/N$$

Other Bias Analyses

Maximum and minimum values were determined. The maximum is the highest positive DFA and the minimum is the largest negative DFA. The range is the maximum value minus the minimum value.

Positive bias is the summation of the positive DFAs and negative bias is the sum of the negative DFAs. The number of positive and negative DFAs was calculated and is used in the turning point test.

Turning Point Test

The turning point test is the ratio of the number of negative DFAs to the number of positive DFAs. The closet value to unity is best. 10-year normals performed best under this test.

V. Coefficient of Determination - R^2

In performing a regression analysis one must be careful to not read too much into the relationship. Since x is independent and y is the dependent variable, there may be little or no cause and effect. For example, in performing a linear regression of degree-days against time, time does not cause the degree days. In analyzing a time series, regression can in fact be used to fit an equation. One should not assume or contend that a regression line or equation can be used to "predict" degree-days in the future. Similarly, with low R^2 values, one must be careful in judging the fit of the data. Degree-days may be "predicted" by using various methods; e.g. 10-year normals, 20-year normals, etc. The only purpose of computed values based on a regression should be to illustrate the possibility of a trend in degree days.

In viewing a time series of data, one always questions whether the data portrays a trend or not. If it does not, one might conclude that it is random. If the data does in fact show a trend, it is not random.

The R^2 falls out of a regression performed to determine the coefficients for a "best fit" line of degree day data. In other words, using a simple linear relationship one could not draw a better line (or equivalently fit a better first-degree equation). One should not assume that degree-days can, in fact, be predicted from this line. Even if the R^2 were high, say 90% or better, due to the type of data, degree-days versus years, one should not infer that an increase of a year will thus categorically lead to a decrease or increase of degree days.

Also, a trend line computed from a linear regression does not explain variation in degree-days. Annual degree-days are not explained by the progression of time.

VI. Data Variability

Some might suggest that a weather normalization should depict a range of variability in annual degree-day measures. No analyst should set out to intentionally increase variability in these

analyses. The primary goal, which might be tempered by other goals, should be to attempt to predict with greatest accuracy the degree-days in an ensuing year. Given two otherwise equal choices, ceteris paribus, one should choose the method with the least variation, not set out intentionally to maximize variability.

I determined a number of measures of standard deviation. In ,general they tend to support use of 10-year normals versus 30-year normals in that the 10- year normals reflect a lower standard deviation than the 30-year normals.

VII. Graphs

In addition to the graphs included in my Exhibits, a variety of additional graphs are included in my Workpapers. Most are self-explanatory. Virtually all of the graphs for various measures show the superiority of 10-year normals.

VIII. Definitions of Other Terms

Mean (arithmetic)

$$\mu = \Sigma X/N$$

Median

M_d is the middle item in a group. For groups with an even number of items it is determined by averaging the middle two items.

Mode

M_o is the value which occurs most frequently in a series of numbers.

Empirical Mode

$$M_{OE} = \mu - 3(\mu - M_d)$$

Frequency Mode

$$M_{OF} = L + [(d_1/(d_1 + d_2))*i] \quad \text{where}$$

L = lower limit of modal class
d_1 = # of items in modal class minus # of items in prior class
d_2 = # of items in modal class minus # of items in next class
i = class interval of modal class

Range

A simple measure of dispersion of data.

R = difference between the largest and smallest values of a variable

Standard Deviation

The square root of the variance.

$$\sigma = \sqrt{\Sigma (X - \mu)^2/N}$$

Variance

The arithmetic average of the squared deviations of individual items about their mean. Also called the second moment.

$$\sigma^2 = \Sigma (X - \mu)^2/N$$

Third Moment

$$M_3 = \Sigma \frac{(X - \mu)^3}{N}$$

Coefficient of Determination - R^2

$R^2 = \Sigma (y_c - y_{bar})^2 / \Sigma (y - y_{bar})^2$ or equivalently

$R^2 = b^2 \Sigma (x - x_{bar})^2 / \Sigma (y - y_{bar})^2$ where

$y_c = a + bx$ where

$$b = \frac{n \Sigma xy - (\Sigma x)(\Sigma y)}{n \Sigma x^2 - (\Sigma x)^2}$$

The extended form is:

$$R^2 = \frac{[\Sigma (x - x_{bar})(y - y_{bar})]^2}{\Sigma (x - x_{bar})^2 \Sigma (y - y_{bar})^2}$$

Coefficient of Variation

A measure of relative dispersion or variation as a percent of the average about which they are computed.

$C_v = (\sigma / \mu)*100$

Pearson's Coefficient of Skewness

Sk_1 - Based on Median

$Sk_1 = 3(\mu - M_d)/ \sigma$

Sk_2 - Based on Mode

$Sk_2 = (\mu - M_o)/ \sigma$

Third Moment - Absolute Skewness

For frequency data where m = midpoint of each class interval

$$M_3 = \Sigma \frac{f(m - \mu)^3}{N}$$

Relative Skewness - Beta One

$ß_1 = M_3^2/M_2^3$

Where M_2 = variance or σ^2

Relative Skewness - Alpha Three

$\alpha_3 = \sqrt{ß_1}$

Z Stat

Testing means:

$$z = \frac{(\mu_1 - \mu_2)}{\sqrt{[\sigma_1^2/N_1 + \sigma_2^2/N_2}}$$

As used in the runs test (and standard form):

$z = (R - \mu_R)/ \sigma_R$ where

$\mu_R = [(2N_1N_2)/(N_1+N_2)] +1$

$\sigma_R = \sqrt{[(2 N_1 N_2)(2 N_1 N_2 - N_1 - N_2)/(N_1 + N_2)2(N_1 + N_2 - 1)]}$

R = number of runs N_1 = number of belows N_2 = number of aboves

Turning Point Test

In a bias analysis, the number of positive and negative forecasting errors can be determined; i.e. how many were above or below the actual. The ratio of the number of negatives to positives is the test statistic. The closer this value is to 1, the better the model.