Portland General Electric Company
Legal Department
121 SW Salmon Street • Portland, Oregon 97204
(503) 464-8926 • Facsimile (503) 464-2200

Douglas C. Tingey
Assistant General Counsel

July 16, 2010

Via Electronic Filing and U.S. Mail

Oregon Public Utility Commission
Attention: Filing Center
550 Capitol Street NE, #215
PO Box 2148
Salem OR 97308-2148

Re: UM 1355 – Investigation into Forecasting Forced Outage Rates

Attention Filing Center:

Enclosed for filing in the captioned docket are an original and five copies of:

- **Opening Testimony of Jay Tinker and David Weitzel (PGE/Tinker-Weitzel/300)**

Also enclosed are an original and three copies of:

- **Work Papers on CD (both confidential and non-confidential portions)**

These documents are being served upon the UM 1355 service list. Confidential portions are subject to Protective Order No. 08-549 and therefore not to be posted on the OPUC website. The Word file, Overview-Guide, in the non-confidential work papers provides a guide to all work papers.

This document is being filed by electronic mail with the Filing Center. An extra copy of the cover letter is enclosed. Please date stamp the extra copy and return to me in the envelope provided.

Thank you in advance for your assistance.

Sincerely,

DOUGLAS C. TINGEY
Assistant General Counsel

DCT:cbm
Enclosures
cc: UM 1355 Service List

CERTIFICATE OF SERVICE

I hereby certify that I have this day caused **Opening Testimony of Jay Tinker and David Weitzel (PGE/Tinker-Weitzel/300) with Work Papers** to be served by electronic mail to those parties whose email addresses appear on the attached service list, and by First Class US Mail, postage prepaid and properly addressed, to those parties on the attached service list who have not waived paper service from OPUC Docket No. 1355.

Dated at Portland, Oregon, this 16th day of July, 2010.

DOUGLAS C. TINGEY, OSB # 044366
Assistant General Counsel
Portland General Electric Company
121 SW Salmon St., 1WTC1301
Portland, OR 97204
(503) 464-8926 (telephone)
(503) 464-2200 (fax)
doug.tingey@pgn.com

SERVICE LIST
OPUC DOCKET # UM 1355

Citizen's Utility Board of Oregon OPUC Dockets dockets@oregoncub.org (*Waived Paper Service)	Citizen's Utility Board of Oregon Robert Jenks bob@oregoncub.org (*Waived Paper Service)
Citizen's Utility Board of Oregon Gordon Feighner gordon@oregoncub.org (*Waived Paper Service)	Citizen's Utility Board of Oregon G. Catriona McCracken catriona@oregoncub.org (*Waived Paper Service)
Davison Van Cleve Irion Sanger 333 SW Taylor- Ste 400 Portland, OR 97204 ias@dvclaw.com	Davison Van Cleve Melinda J. Davison 333 SW Taylor- Ste 400 Portland, OR 97204 mail@dvclaw.com
Department of Justice Michael T. Weirich 1162 Court Street NE Salem, OR 97301-4096 michael.weirich@state.or.us	Idaho Power Company Scott Wright swright@idahopower.com (*waived paper service)
Idaho Power Company Tim Tatum ttatum@idahopower.com (*waived paper service)	Idaho Power Company Christa Bearry cbearry@idahopower.com (*waived paper service)
Idaho Power Company Rex Blackburn rblackburn@idahopower.com (*waived paper service)	Idaho Power Company Gregory W. Said gsaid@idahopower.com (*waived paper service)
Idaho Power Company Lisa D. Nordstrom lnordstrom@idahopower.com (*waived paper service)	McDowell, Rackner & Gibson PC Wendy McIndoo 520 SW Sixth Ave, Suite 830 Portland, OR 97204 wendy@mcd-law.com
McDowell, Rackner & Gibson PC Katherine McDowell 520 SW Sixth Ave, Suite 830 Portland, OR 97204 katherine@mcd-law.com	McDowell, Rackner & Gibson, PC Lisa F. Rackner 520 SW Sixth Ave, Suite 830 Portland, OR 97204 lisa@mcd-law.com
McDowell, Rackner & Gibson, PC Adam Lowney 520 SW Sixth Ave, Suite 830 Portland, OR 97204 adam@mcd-law.com	Pacific Power and Light Michelle R. Mishoe Michelle.mishoe@pacificorp.com (*waived paper service)
Pacificorp	Public Utility Commission of Oregon

Oregon Dockets oregondockets@pacificorp.com (*waived paper service)	Kelcey Brown PO Box 2148 Salem, OR 97301 Kelcey.brown@state.or.us
RFI Consulting Randall J. Falkenberg PMB 362 8343 Roswell Road Sandy Springs, GA 30350 consultrfi@aol.com	

BEFORE THE PUBLIC UTILITY COMMISSION

OF THE STATE OF OREGON

UM 1355

Investigation into Forecasting Forced Outage Rates

Portland General Electric Company

PGE

Portland General Electric

July 16, 2010

UM 1355 / PGE / 300
Tinker - Weitzel

BEFORE THE PUBLIC UTILITY COMMISSION

OF THE STATE OF OREGON

Forecasting Forced Outage Rates

Opening Testimony

Portland General Electric Company

Jay Tinker
David Weitzel

July 16, 2010

Table of Contents

I. Introduction

Q. Please state your names and positions with PGE.

A. My name is Jay Tinker. I am a Project Manager for PGE. My qualifications are on Page 20 of PGE Exhibit 100.

My name is David Weitzel. I am a Project Manager for PGE. My qualifications appear at the end of this testimony.

Q. What is the purpose of your testimony?

A. In Order No. 09-479, the Commission proposed a modification of Staff's methodology for addressing extraordinary forced outage rates (FORs) for particular years in the calculation of forced outage rates for coal-fired generating facilities:

> "The methodology for calculating the forced outage rate shall be as set forth in Staff/200, Brown/8-15, except that, instead of adjusting the FOR to the 10th or 90th percentile values for the calendar year, the mean annual FOR from the unit's entire historical data shall be substituted."

A similar modification to Staff's proposal had been advocated by ICNU witness Falkenberg in ICNU/300. In Order No. 10-157, the Commission granted PGE permission to file additional testimony to address ICNU's FOR proposal. The purpose of our testimony is to address the analysis that ICNU witness Falkenberg provided in his testimony in support of the modification of Staff's proposal.

Q. What is the principal conclusion of your testimony?

A. We demonstrate that Mr. Falkenberg's analysis does not support his assertion that the replacement of outliers by long-run averages results in superior forecasts of forced outage rates. His analysis suffers from at least two serious methodological errors, one of which is a fatal error that renders his analysis meaningless. Further, even if we accept his analysis and ignore these serious methodological errors, the "advantage" he claims for his method does not approach any standard level of statistical significance. When his analysis is replicated with the methodological errors corrected, the forecast performance of the "mean replacement" strategy and the "90/10" replacement strategy are virtually identical. Consequently, he has provided no evidence that supports a modification of Staff's proposal.

Q. How is your testimony organized?

A. We begin with brief summaries of the major points in Section II. We first explain (in brief) the nature of the two principal methodological errors in Mr. Falkenberg's analysis. Following this, we discuss the lack of statistical significance in Mr. Falkenberg's results, then summarize the results of an analysis with the methodological errors corrected. Next we observe that the approach to outlier replacement in the forecasting literature resembles Staff's approach (and not Mr. Falkenberg's suggested approach). Finally, we describe additional problems with Mr. Falkenberg's discussion of mean reversion and the FOR data used in the analysis. These summary discussions are followed by detailed explanations of the major points.

II. Overview

Q. Please explain the nature of the principal methodological errors in Mr. Falkenberg's analysis.

A. We summarize the two methodological errors below. We will expand on these points later in our testimony in Section III.

1. A very basic and intuitive principle of forecasting is that you cannot use information to generate forecasts that is not available at the time the forecast is being made. Forecasters refer to the information that *is* available at forecast time as the *information set.*

   > "Before forming a forecast, it is necessary to assemble the set of information that is to be used to make the forecast. For example, if times series forecasting, the forecast might be formed from the ***observed past and present values of the series."*** [*emphasis added*][1]

   Specifically, Mr. Falkenberg errs in his analysis by generating forecasts using information that is not available at the time the forecast is being generated. Using the 20-year mean (or the mean for all available data) to replace values in the four-year averages effectively uses the future to explain the future. This is a fatal flaw in his analysis.

2. In comparing the relative forecasting performance of the "mean replacement" strategy and the "90/10" replacement strategy, Mr. Falkenberg compares the sum of squared errors across the group of PacifiCorp plants resulting from the "mean replacement" strategy with the sum of squared errors resulting from the "90/10" replacement strategy. It has been recognized for some time in the forecasting literature that there are serious problems with using root mean square error to aggregate across different time series:

[1] C. W. J. Granger, *Forecasting in Business and Economics*, p. 11.

> "The RMSE [root-mean-squared error] is unreliable. Related to this is its poor protection against outliers. We do not recommend the RMSE for assessing the level of accuracy."[2]

In our corrected version of Mr. Falkenberg's analysis, we employ an accuracy measure that is superior to RMSE for summing across plants[3].

Q. What is the problem with the (lack of) statistical significance of Mr. Falkenberg's results?

A. As indicated above, Mr. Falkenberg's claim for the superiority of the "mean replacement" strategy over the "90/10" replacement strategy is based on a comparison of the sums of squared (forecasting) errors resulting from the two approaches. Even if we take his analysis at face value (and ignore the methodological problems cited above), his claim of superiority for the "mean replacement" strategy is unsupported because the two sums of squares are statistically indistinguishable. In our detailed analysis, we employ a two-sample permutation test and show that the observed difference in the sums of squared errors could have occurred with a 0.4 probability when there is no underlying difference in the two methods. However, a typical probability ("p-value") accepted as "statistically significant" is 0.05 or less. Thus, the probability associated with the observed difference is far from a conventional significance level.

Q. What was the result when you corrected Mr. Falkenberg's analysis?

A. We modified Mr. Falkenberg's analysis so that the forecasts used only information that was available on or before the forecast date. We also used a better method for aggregating

---

[2] J. Scott Armstrong and Fred Collopy, "Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons", *International Journal of Forecasting*, 8 (1992), 69-80.

[3] Note that the problem here is adding up across plants; this does not imply a general problem with the use of RMSE. For example, PGE 100 (Pages 12-13) discusses RMSE results of various approaches to FOR forecasting for a <u>single</u> plant.

forecast errors across plants. When we made these changes, the forecast results from the two methods were virtually indistinguishable.

Q. Is Mr. Falkenberg's "mean replacement" strategy consistent with practice in the forecasting literature?

A. We were not able to locate any papers supporting a mean replacement strategy for outliers. We *were* able to find an example that closely matches the outlier replacement strategy that Staff has proposed. This is discussed further in Section III.

Q. Mr. Falkenberg claims that "reversion to the mean" supports the use of the "mean replacement" strategy. Is this correct?

A. Mr. Falkenberg's arguments are misleading and logically inconsistent. As explained in Section III, the history of annual FORs for a given plant can best be viewed as a *non-stationary* time series. For a time series to be mean-reverting, it has to have a stationary mean. In any event, the two pieces of Mr. Falkenberg's forecasting procedure (four-year average and "mean replacement") are logically inconsistent if the series is mean-reverting. If the FOR series *is* mean-reverting, then the use of a four-year average to forecast makes no sense. Conversely, if the use of a four-year average to forecast makes sense, the FOR series *cannot be* mean-reverting. You can't have it both ways. As a reminder, the Commission found (Order No. 09-479) that the use of the four-year average generally is appropriate. This suggests that plant FORs are not mean-reverting.

Q. Are there issues with the data that Mr. Falkenberg employed in his analysis that make it difficult to generalize his results to the application of the Commission's procedure to other utilities?

A. Yes. Two problems stand out. First, the durations of individual outages in the PacifiCorp data set have an upper bound of 28 days. However, there is no upper bound on the duration of actual individual outages for PGE plants. Second, Mr. Falkenberg used collar values that are based on individual plant histories rather than NERC data.[4]

Given the available data, we were not able to explore the consequences of the upper bound on outage durations. We have examined the effect of substituting NERC data in Mr. Falkenberg's analysis to collar outlying values. The results of this analysis are presented in the detailed analysis in Section III.

[4] Mr. Falkenberg uses data for all years to establish a plant's collar values. These same collar values are then applied in all years. Similar to Mr. Falkenberg's mean replacement method, this procedure is problematic because it uses future data to establish collar values. However, using only data to date for any particular year would result in having to calculate 90th and 10th percentile values from very small data sets. Also, the problem of using future data is more serious in the calculation of means for use in the replacement strategy than in the establishment of collar values.

III. Detailed Discussion

Q. Why do you need to discuss the statistical content of Mr. Falkengerg's procedure?

A. We believe that most statisticians, confronted with the problem of forecasting forced outage rates with the data available to us, would agree on the importance of a few basic elements of our forecasting problem. Evaluation of these elements leads to specific conclusions about how to forecast.

Q. Please expand on Mr. Falkenberg's methodological error in using the twenty-year (or available data) average to generate forecasts.

A. It is relatively easy to find passages in the forecasting literature that make it clear that forecasts should only use information available at the point in time that the forecast is being made. For example:

> "Given the availability of a set of observations up to, and including, y_T, the optimal predictor l step ahead is the expected value of Y_{T+l} ***conditional on the information at time T.***"[5] [*emphasis added*]

In this excerpt, "T" is the point in time at which the forecast is being made.

Q. Can you provide a simple example?

A. Yes. Let's suppose that you were asked to estimate the US GDP for 2011. Now, if you were omniscient and knew the average level of GDP for the three years 2010-2012, you would likely be able to make a better forecast of 2011's GDP than you would in the absence of this information. However, you are not omniscient, so you don't know the average for these three years. In fact, no one knows the average of future GDP. But when Mr. Falkenberg uses the twenty-year average as a replacement value, because he's forecasting within his sample, he is acting as if he were omniscient.

---

[5] Andrew C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, p. 34.

Q. Can you provide a more formal demonstration of the problem with Mr. Falkenberg's procedure?

A. Yes. There are two parts to Mr. Falkenberg's method: (1) average forced outage rates over the entire period for which they are available and (2) whenever an individual observation is "collared", replace that observation in the four-year average with the average from (1) in the calculation of the four-year average used to forecast the next year's FOR. In Appendix 1, we show that if this type of replacement procedure is legitimate, then it is possible to generate "perfect" forecasts, that is, forecasts with zero forecast errors. Clearly this result makes no sense. In our corrected version of Mr. Falkenberg's analysis, we use only information that is available on or before the forecast date.

Q. What is the problem with simply comparing the sums of squared errors resulting from the two approaches as Mr. Falkenberg has done in his analysis?

A. Mr. Falkenberg's analysis compares the sums of squared errors resulting from the <u>different</u> outlier replacement strategies. This is equivalent to comparing the root-mean square errors (RMSE) resulting from the different outlier replacement strategies. The forecasting literature advises against comparing the sums of squared errors across forecasting methods:

> "A particularly sobering finding is the poor performance of the RMSE, the most commonly used error measure. Its use is related to its popularity with statisticians and its interpretability in relation to business decisions ***and not to its efficiency in choosing accurate forecasting methods***. [*emphasis added*] The reliability of RMSE is poor and it is scale dependent."[6]

This passage and the prior passage cited above in our introduction have emphasized three limitations of summing up squared errors across series: scale dependency, reliability, and sensitivity to outliers. Scale dependency means that the measure does not control for

[6] Dennis Ahlburg, "Error Measures and the Choice of a Forecast Method", *International Journal of Forecasting*, 8 (1992), p. 69.

differences in the size of errors across plants. "Reliability addresses the question of whether repeated application of a procedure will produce similar results."[7] Sensitivity to outliers is important because the results may be driven by a few observations and the conclusion may not accurately reflect the relative performance of the methods. In our reworking of Mr. Falkenberg's results, we employ an alternative accuracy measure called "Relative Geometric Root Mean Square Error – RGRMSE," that is more resistant to outliers[8]. However, we still face potential problems with scale dependency and reliability.

Q. Why is it important to consider the lack of statistical significance of Mr. Falkenberg's results?

A. Mr. Falkenberg's testimony implies that his method holds a clear and unambiguous advantage over Staff's proposed method. The results of statistical testing indicate, to the contrary, that there is a high degree of ambiguity as to his "advantage". We emphasize again that this is true even if we ignore the serious methodological problems with his analysis and take his results at face value. To assess the statistical significance of his results, we have simply taken the sums of squares that he produced with his methodology and performed a two-sample permutation test for a difference in the means of the squared errors resulting from the two methods.

Q. Why did you use a permutation test?

A. Permutation tests have undergone in-depth review by telecommunications regulators and have been used extensively to test for parity of telephone service in local competition.

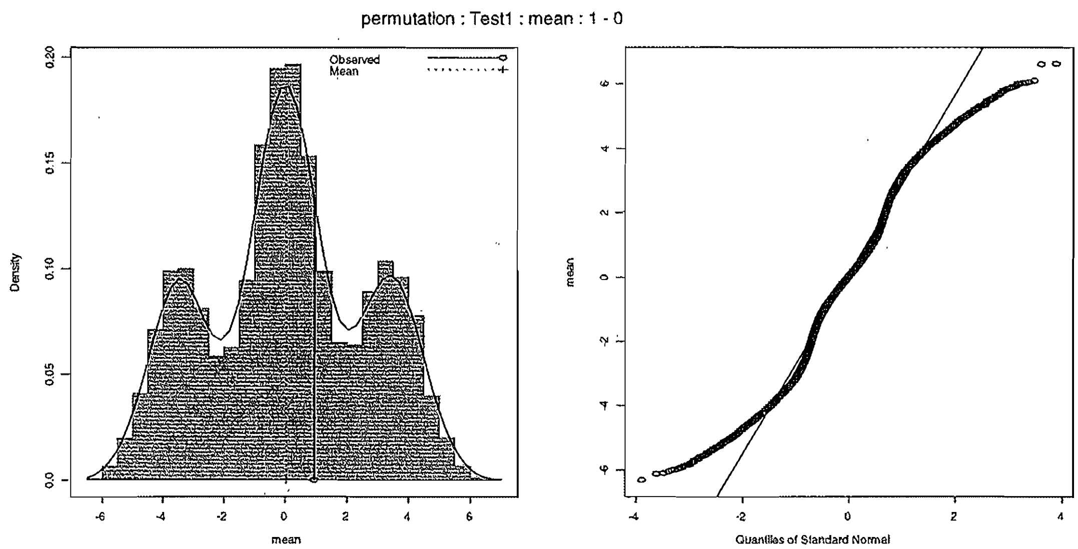
---

[7] J. Scott Armstrong and Fred Collopy, "Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons", *International Journal of Forecasting*, 8 (1992), 73.

[8] R. Fildes, "The Evaluation of Extrapolative Forecast Methods", *International Journal of Forecasting*, 8, (1992), pp. 81-98.

The alternatives to permutation tests, conventional two-sample "t-tests", are not robust to departures from normality of the sampling distribution. Since the permutation distribution approximates the sampling distribution, we can use the permutation distribution as a guide for the choice of the test. Figure 1 is a plot of the permutation distribution corresponding to Mr. Falkenberg's results. It is clear from inspection of the plot that the permutation distribution departs markedly from normal: the distribution is tri-modal and has "fatter tails" than a normal distribution. A permutation test is a better choice.

Figure 1



Q. What is the interpretation of the test results?

A. Test results for Mr. Falkenberg's original comparison are reported in the first row of Table 1 below. A permutation test indicates the probability of realizing a difference as large as (or larger than) the observed difference, given that the underlying distributions are the same. In other words, if the two forecasting methods are equivalent, what is the probability that we would observe a difference in sums of squares as large as (or larger than) the observed difference. If this probability is high, then the observed difference does not support the conclusion that there is an actual difference in the two methods. For Mr. Falkenberg's original results, this probability is very high. This probability is typically referred to as a "p-value". The p-value for Mr. Falkenberg's results is approximately 0.4. Conventionally, a p-value is not considered evidence of an actual difference unless it is 0.05 or less. The probability of observing the difference reported by Mr. Falkenberg when the two methods are equivalent is close to a coin flip.

Table 1
Test Results With Plant Data-Based Collars

Falkenberg Method	Staff Method	P-Value for Difference in Methods
Original Falkenberg Method (Squared Errors)*	Staff Method Using Plant Data for Collar (Squared Errors)	0.41
Falkenberg Method With Data to Date (Squared Errors)	Staff Method Using Plant Data for Collar (Squared Errors)*	0.50
Falkenberg Method With Data to Date (GRMSE)	Staff Method Using Plant Data for Collar (GRMSE)*	0.48

*Indicates the method that resulted in lowest sum of squared errors or lowest sum of natural logs of GRMSEs.

Q. Please describe the corrected versions of Mr. Falkenberg's analysis.

A. We have made two corrections to Mr. Falkenberg's analysis. The results are also reported in Table 1. First, as we have discussed, it is not legitimate to use the twenty-year average as a replacement value because that average uses information that post-dates the forecast date. In the corrected analysis that parallels Mr. Falkenberg's replacement strategy, we use a sample average (for replacing outliers) that uses data only up to the date that the forecast is being made. We first report (row 2 of Table 1) the results making only this single correction (in available information) for the sums of squared errors. Comparing Staff's proposal with Mr. Falkenberg's proposal, we find that the sums of squared errors are very close and the difference is far from statistically significant (as indicated by the high p-value of 0.5).

Second, to reduce the chance that results are unduly influenced by a few observations, we then incorporate geometric root mean square error (GRMSE) as the measure of forecast error. The GRMSE-based permutation test has a p-value of 0.5, indicating that any difference in forecast accuracy between methods is statistically insignificant. Finally, we calculated the relative geometric root mean square error (RGRMSE) of Staff's proposal and the corrected Falkenberg proposal. The result is 0.996, where a value of one indicates equality.

Q. Did you perform any additional tests?

A. Yes. Mr. Falkenberg's forecasts used distributions based on individual plant data to collar outliers. Paralleling the results just described, we have generated three sets of forecasts that use distributions based on NERC data to collar outliers. This procedure more closely resembles the procedure recommended by Commission Staff.[9]

Q. Please describe the results.

A. Test results using NERC collars to determine outliers are reported in Table 2 below. Test results for Mr. Falkenberg's original comparison (now with a NERC collar) are reported in the first row of Table 2. The high p-value, 0.4, indicates that, although in this case the Staff method appears to outperform Mr. Falkenberg's mean replacement method, the difference in performance is not statistically significant. Test results for a sample average (for replacing

[9] There are three limitations on our NERC collar-based work. 1) We only have NERC data relevant to 17 of the 26 plants used by Mr. Falkenberg in his analysis. Hence the data set for the NERC collar analysis is smaller. We also do not have NERC data for early years. To establish collars for these years, we pool all comparable NERC data that we do have. 2) It is our understanding that the PacifiCorp and NERC data are based on somewhat different definitions of forced outage rate; the former includes maintenance outages, whereas the latter does not. In the case of observations that are collared, i.e. are very high or very low, maintenance outages do not greatly affect the PacifiCorp FOR calculations, making the results of the two definitions very similar. 3) To facilitate comparisons with Mr. Falkenberg's work, we retain in all of our analyses his construction, under which the forecast for a particular year is the average of observations (collared, if appropriate) from the preceding four years. For example, data from 2001-2004 are used to forecast 2005. In practice, data from 2000-2003 were used to forecast 2005, as annual data for 2004 were not available late in 2004 when rates for 2005 were established.

outliers) that uses data only up to the date that the forecast is being made are reported in the second row of Table 2. Again the result, a p-value of 0.3, is not statistically significant. [It should be 0.05 or less.] Finally, results that incorporate the use of GRMSEs are reported in the third row of Table 2 and are not statistically significant, as the p-value is 0.5. In addition, we calculated the RGRMSE of Staff's NERC collar proposal and the corrected NERC collar and mean replacement Falkenberg proposal. The result is 1.008, where a value of one indicates equality.

Table 2
Test Results With NERC Data-Based Collars

Falkenberg Method	**Staff Method**	**P-Value for Difference in Methods**
Original Falkenberg Method (Squared Errors)	Staff Method Using NERC Data for Collar (Squared Errors)*	0.44
Falkenberg Method With Data to Date (Squared Errors)	Staff Method Using NERC Data for Collar (Squared Errors)*	0.35
Falkenberg Method With Data to Date (GRMSE)*	Staff Method Using NERC Data for Collar (GRMSE)	0.47

*Indicates the method that resulted in lowest sum of squared errors or lowest sum of natural logs of GRMSEs.

Q. Have you compared the biases of Staff's strategy and the strategy advocated by Mr. Falkenberg in ICNU 300?

A. Yes. We have computed the bias of Staff's NERC Collar (and replacement) strategy and that of a corrected version of Mr. Falkenberg's recommended strategy. The corrected version includes a collar based on NERC data and replaces outliers with a mean based on "data to date". The bias in each case was calculated as the sum of the forecast errors produced by the strategy, divided by the number of forecasts. The Staff approach has a bias of approximately -0.9 percent; the data to date mean replacement approach has a bias of approximately -1.5 percent. Both strategies appear to under-forecast forced outage rates, with the bias greater for the mean replacement approach. We did not examine the statistical significance of the biases or the difference between them.

Q. Can you cite a paper from the forecasting literature that supports Staff's approach to outlier replacement?

A. Yes. It was not surprising to discover that statisticians had addressed problems similar to the one we face in forecasting forced outage rates. We looked in particular for papers that focus on forecasting non-stationary time series that feature occasional outliers. Consider the following:

> "On the other hand, if the deviation between the predicted and the observed value is too large, the observation is considered an outlier and gets replaced by a ***boundary value.***"[10] [*emphasis added*]

We see that the methodology in the paper cited results in outliers being replaced with boundary values (much like the collar values in Staff's approach) and not with a sample mean.

Q. Mr. Falkenberg claims that plant forced outage rates are generated by a mean reverting process. Can you elaborate on the problems with this claim?

A. No statistician would ever classify a time series as mean reverting if that series did not have a stable long-term mean. The entire history of annual FORs for a given plant can best be viewed as a *non-stationary* time series. For our purposes, this means that the expected annual FOR evolves over time, i.e., it is not constant. This view accords with common sense: A coal plant at the present time is the sum of repairs and upgrades that have occurred over its history; we don't expect the plant's expected FOR to be constant over time because the composition of the plant is not constant over time. This is why Parties have concluded

[10] Sarah Gelper, Roland Fried and Christophe Croux, *Robust Forecasting with Exponential and Holt-Winters Smoothing*, p. 6.

that using recent data (the most recent four-year average) is the best way to forecast next year's FOR.

Given the data, we can always calculate a twenty-year average of the historical FORs. It is far from clear that the resulting number corresponds to anything meaningful in a statistical sense. If the time series of forced outage rates were truly mean-reverting, then the basic forecasting technique (and not just the outlier replacement strategy) should properly make use of the long-term mean. The most recent four-average clearly does not capture all of the "information" contained in the historical average; it is *local*. So Mr. Falkenberg is combining one measure that is incompatible with mean reversion and stationarity (the most recent four-year average) with a replacement value (the historical average) that only makes sense if the time series is stationary. It is highly unlikely that a forecasting procedure based on a fundamental internal contradiction will produce optimal forecasts.

Q. Please summarize why you do not believe that Mr. Falkenberg has provided credible evidence in support of his proposed mean replacement method.

A. First, Mr. Falkenberg claims his recommendation will result in superior forecasting performance. However, he has failed to provide credible support for this assertion. His evaluation is based on a false premise, as it uses future data to predict the future. Second, either with or without corrections, his mean replacement procedure does not perform differently than Staff's collar procedure in any meaningful statistical sense. Finally, his claim that mean reversion provides support for his proposed procedure is inconsistent with the four-year average that the Commission generally accepts for forecasting forced outage rates.

Q. What should the Commission order in this proceeding?

A. The Commission should adopt the August 19, 2009 Stipulation Regarding All Issues for PGE, which uses Staff's collar approach.

IV. Qualifications

Q. Mr. Weitzel, please state your educational background and experience.

A. I received a PhD in Economics from the University of Washington in 1980 with a field in econometrics. In 1997, I obtained the Chartered Financial Analyst (CFA) designation. I have worked in the Rates and Regulatory Affairs department since 2009.

My forecasting work includes two projects for the Electric Power Research Institute; for one project I estimated the effects of time-of-use pricing on residential electricity demand and for a second project I estimated models to forecast industrial demand for energy. For Puget Power, I created statistical models to forecast energy savings from residential conservation programs. As a member of the GTE (and later Verizon) Demand Analysis and Forecasting Group, I was responsible for research design and for forecasting demand for telecommunication services. Also at Verizon, I participated in the development of statistical testing protocols to assess parity of service provision in local telecommunications markets. With Insightful Corporation, I developed models to forecast demand for consumer goods. Miscellaneous projects include forecasting the price of oil tanker services, forecasting water demand, and models to predict credit problems.

Q. Does this conclude your testimony?

A. Yes.

Appendix 1

In this appendix, we demonstrate that the assumed knowledge of future FORs in Mr. Falkenberg's mean replacement strategy would allow for perfect forecasts. For our formal demonstration, let the entire time period consist of T years, indexed as t = 1, T. Let t' be the point in time at which we are generating a forecast for year t'+1. Let F_t represent the FOR at time t. The average forced outage rate over the T years can be represented as:

$$(\Sigma_t F_t)/T$$

The four-year average corresponding to the forecast for period t'+1 is $(F_{t'}+F_{t'-1}+F_{t'-2}+F_{t'-3})/4$.

Mr. Falkenberg's method is to substitute the T-year average of the FOR values in the four-year average. The expression for the T-year average can be rewritten as a weighted average with equal weights, w_t, equal to 1/T:

$$\Sigma_t (1/T) F_t = \Sigma_t w_t F_t$$

If the basic procedure is legitimate, then there is no reason why we shouldn't "optimize" the weights in the weighted average. We claim that the optimal set of weights is $w_{t'+1} = 1$ and $w_t = 0$ for t not equal to t'+1. If we use these weights in our T-period average we get $F_{t'+1}$. If we substitute this for all four terms in our four-year average, we get $F_{t'+1}$ – a perfect forecast!

This example is an extreme case, but we believe it illustrates why the procedure is ill-advised. You simply don't know the extent to which you have merely "used the future to predict the future". Again, in our corrected version of Mr. Falkenberg's analysis, we use only information that is available on or before the forecast date.