FILED
10-01-15
04:59 PM

BEFORE THE PUBLIC UTILITIES COMMISSION
OF THE STATE OF CALIFORNIA

Application of California-American Water Company (U210W) for Approval of the Monterey Peninsula Water Supply Project and Authorization to Recover All Present and Future Costs in Rates.	A.12-04-019 (Filed April 23, 2012)

MOTION TO DISMISS THE PROCEEDING
ON THE MONTEREY PENINSULA WATER SUPPLY PROJECT
BECAUSE OF DATA TAMPERING

RON WEITZMAN

23910 Fairfield Place
Carmel, CA 93923
Telephone: (831) 375-8439
Facsimile: (none)
Email: ronweitzman@redshift.com
President, Water Plus

Date: October 1, 2015

I. INTRODUCTION

Pursuant to Rule 11.2 of the Rules of Practice and Procedure and with the upcoming pre-hearing conference ("PHC") scheduled by ALJ Gary Weatherford in his ruling on September 24, 2015, in mind, I submit this motion on behalf of Water Plus to dismiss the proceeding on the Monterey Peninsula Water Supply Project, which has now become questionable as the result of not only a conflict of interest compromising the project's draft Environmental Impact Report ("DEIR") but also evidence, provided here, of data tampering in the project's evaluation.

Involving slant wells never before used for desalination, the project is an extremely costly and risky experiment funded by ratepayers who are neither entrepreneurs nor venture capitalists. In the PHC, the Commission must take a serious new look at regulating a utility when it has moved out of the monopolistic market of water purveyance and into the competitive market of water supply. Regulation in a competitive market is not the appropriate business of the Commission and is likely the source of all the problems incurred by the project since the inception of this proceeding.

Since the conflict of interest is well-documented, I am going to focus on the demonstration of data tampering to improve the fit of a project-proposed model to measurements of water elevation showing the effect of pumping in aquifers accessed by slant wells proposed by the project for feed-water intake.

II. DATA TAMPERING

Conflict of interest is one thing, data tampering as a manifestation of that conflict quite another. If conflict of interest is sufficient reason to terminate a project, as it appears to have been in the case of the Regional Desalination Project, then certainly tampering with data used to evaluate the MPWSP must lead to the same end for that project. This section will cite only one instance of data tampering in the MPWSP. Although others may exist, one should constitute sufficient grounds to terminate the project.

Because of a cease-and-desist-order deadline imposed by the state Water Resources Control Board, insufficient time exists to evaluate the technical feasibility of slant wells at the proposed CEMEX site. For that reason, hydrogeologists involved in the design and evaluation of the slant wells have proposed a model to predict the results of using slant wells over a long period of time. Substituting for actual experimentation, the model is critical to the prediction of the project's long-term success or failure.

In the DEIR, Geoscience Support Services, at the center of the conflict of interest, described the model's evaluation in Appendix E2 (p. 28), in reference particularly to the scatterplot presented in Figure 37, presented here in

Appendix A. The scatterplot represents 1,573 measurements (observations) and matched (calibrated) model predictions of water elevation taken monthly over 32 years from three aquifers in 17 wells scattered throughout the Salinas Valley. Calibration effectively equalized the means of the predicted and observed measurements.

A. Evaluation of Overall Fit. The scatterplot shows the extent of correspondence, or correlation, between the individual predicted and observed measurements for the complete data set. To assess this correspondence:, Geoscience used a descriptive statistic called "relative error" and defined as the ratio of (a) the standard deviation of differences ("residuals" or "errors") between the matched predicted and observed measurements to (b) the range (difference between largest and smallest) of observed measurements. The industry standard for a "good fit" is a value of this statistic, expressed as a percentage, no larger than 10.0. The relative error for the scatterplot in Figure 37 being 9.5 percent, Geoscience concluded the fit for its model to be good.

The use of relative error defined in this way appears to be unique to hydrogeology. Generally, in statistics, the correlation coefficient or its square is used to evaluate goodness of fit. Geoscience did not report the correlation coefficient expressing the relationship between predicted and observed measurements in Appendix E2. Through the auspices of the Commission, I received a spreadsheet of the data described by the scatterplot in Figure 37. From those data, I computed the correlation coefficient to be equal to .80. Its square, .64, indicates the proportion of variation in observed measurements that is predictable from the model. One minus that square, .36, indicates the proportion of variation in observed measurements that is not predictable by the model, meaning that 36 percent of the variation in the observed measurements represents error.

The 36 percent representing error corresponds to the 9.5 percent "relative error" as an evaluation of goodness of fit. The relative error makes the fit appear to be much better than the traditional statistical evaluation. The reason for this appearance is that the numerator of the relative-error ratio is one single standard-deviation unit while the denominator, being a range, is at least six standard-deviation units, in fact 6.36 such units computed from the data. Although the overall fit of the model to the data may be good, the relative-error ratio considerably exaggerates how good it is. To a layman, "relative error" is misleading.

Appendix B illustrates the traditional, or correlational, method of assessing goodness of fit. Examination of this illustration is not just a technical exercise. It provides the foundation for understanding the data tampering described in the next section. The right triangle in the illustration shows a unit (1.0) of observed-measurement spread as the hypotenuse and the predicted (.80) and error (.60) contributions to that spread as the other two sides. Being the sides of a right triangle, their squares (.64 and .36) add up to the square of the hypotenuse (1.0). The important thing to note here is that the triangle is a right triangle: Being equal to 90 degrees, the angle between the sides representing the predicted and error components indicates zero correlation between them. Changes in one are not even partially predictable from changes in the other. The predicted and error components are independent of each other, as they should be.

Although this is true, or approximately true, for the complete data set, the computed correlation coefficient between predicted and error measurement components being very close to zero (0.06), that is not the case for the portion of the data set representing the 180-foot aquifer.

B. Evaluation of Fit for the 180-foot Aquifer. The complete data set consists of three separate portions representing different aquifers. Two are the 400- and 900-foot aquifers. The test well at CEMEX does not directly access those aquifers. The third aquifer is the 180-foot aquifer, which the test well does directly access. The portion of the scatterplot for the data subset representing this aquifer is indicated by filled yellow circles in Figure 37. Although the trend of the complete data set, going more or less compactly from lower left to upper right, indicates a rather good correlation between the predicted and observed measurements, the visible circles representing the 993 data points for the 180-foot aquifer show no such trend. Yet, the computed correlation coefficient for this data subset, .82, was even greater than that for the data set as a whole. That result puzzled me.

To investigate what might be responsible for that result, I added the squared correlation coefficients between the observed and the predicted measurements and between the observed and the error measurements reported for the 180-foot aquifer. If the predicted and error measurements were uncorrelated (the angle between triangle sides representing them being 90 degrees), then the sum I computed should be equal to 1.00, as in the triangle of Appendix B. Instead, the sum was equal to 0.70, far from 1.00, indicating an angle different from 90 degrees and reflecting a correlation different from zero between predicted and error measurements. So, I computed that correlation and found it in fact to be equal to -0.45, far from zero. For the predicted and error components of a measurement, that means decreasing one tends to increase the other.

How could that happen? Suppose you wanted to shrink a large error component closer to zero by 3 units without changing the value of the measurement. The error component could equally likely be positive or negative. If positive, you would subtract 3 from it while adding 3 to the predicted component; if negative, you would add 3 to it while subtracting 3 from the predicted component. Carried out intentionally and systematically, such action, increasing one component while decreasing the other, would tend to produce the observed negative correlation of -0.45. That correlation is solid evidence of data tampering to improve the goodness of fit of the model to the water elevation measurements obtained in the 180-foot aquifer.

Ironically, despite that data manipulation, the relative error for the 180-foot-aquifer portion of the complete data set was equal to 11 percent, above the 10 percent or lower industry standard for a good model fit. The 180-foot aquifer is one of two directly accessed by the test well. The other is the Dune Sand aquifer. Of all the 5,273 data points represented in Figure 37, none applies to the Dune Sand aquifer. So, for the two aquifers directly accessed by the test well, the evidence shows a poor model fit for the one while being non-existent for the other. What investor would take a risk on a project built on such an infirm foundation?

The purpose of the model was to predict the long-term effect of pumping on water elevation in aquifers accessed by intake slant wells. The only slant well tested over time is the one at Dana Point, and the results there indicate a drop in pump efficiency of 43 percent over a six-year period (CONCUR, Inc., Final Report: Technical Feasibility of Subsurface Intake Designs for the Proposed Poseidon Water Desalination Facility at Huntington Beach, California, October 9, 2014, p.37). Pump efficiency is one of the variables in the model used to predict the effect of pumping on water elevation. As just demonstrated, the model is inadequate to predict that effect.. The only other option is long-term testing. Since the state's cease-and-desist order rules that option out, all the evidence to date indicates a poor prospect of success for a highly risky MPWSP.

III. CONCLUSION

Fort these reasons, particularly data tampering most likely produced by existing conflict of interest, Water Plus moves that the Commission dismiss the Monterey Peninsula Water Supply Project proceeding before the expenditure of any further ratepayer money. More generally, Water Plus recommends that the Commission exit the water-supply market because its presence there works only to decrease competition while increasing the opportunity for fraud.

Dated October 1, 2015

Respectfully submitted,

Ron Weitzman
President, Water Plus

APPENDIX A

California American Water and ESA
Monterey Peninsula Water Supply Project
Groundwater Modeling and Analysis

DRAFT

Comparison of Measured Versus Model-Calculated Groundwater Elevations
Transient Model Calibration (Water Years 1980-2011)

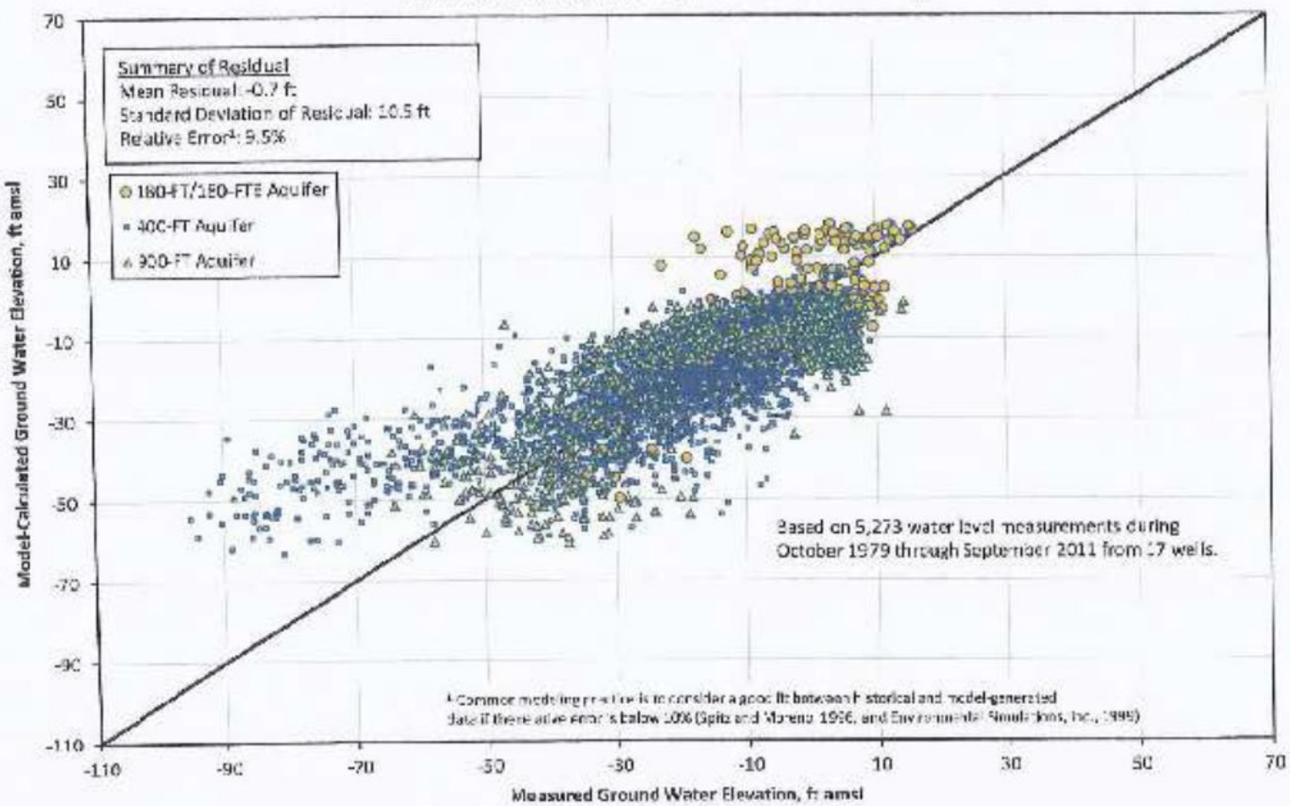
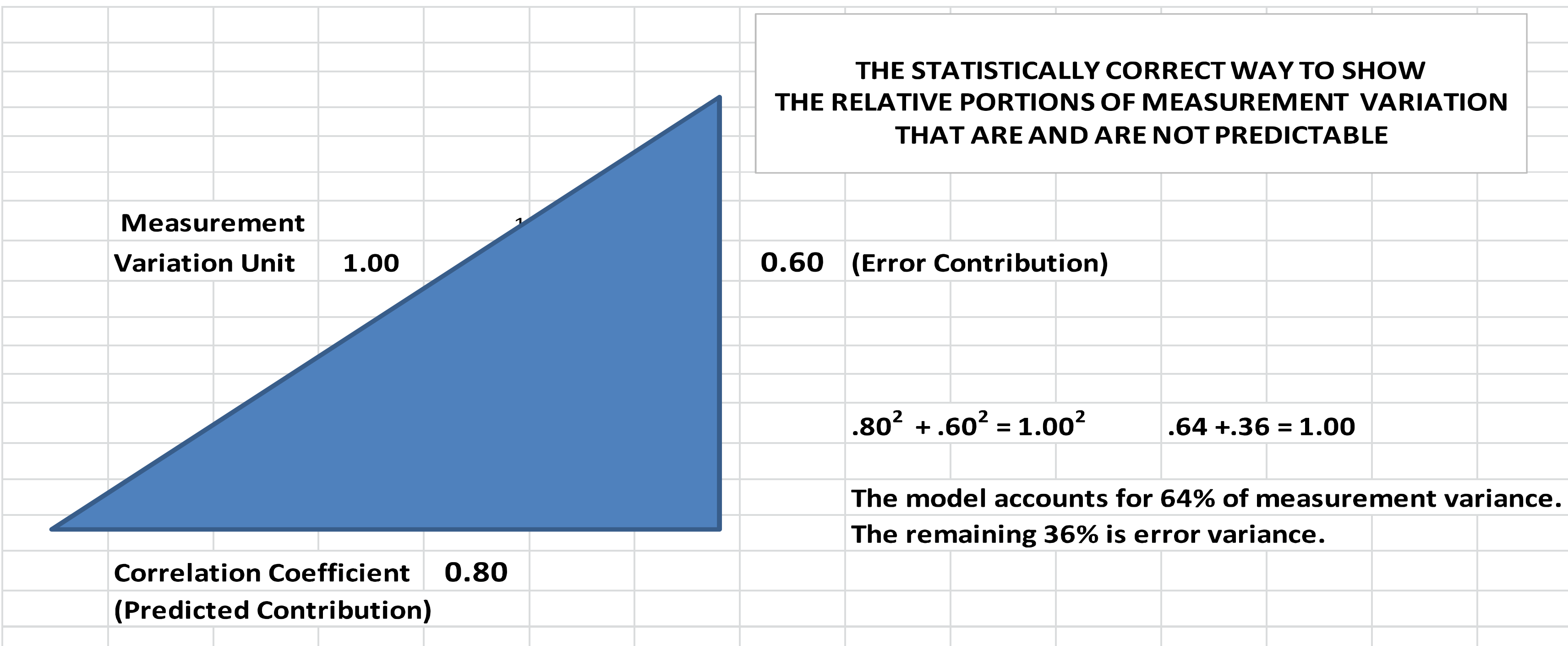


Figure 37

GEOSCIENCE Support Services, Inc.

17-Apr-15

APPENDIX B


THE STATISTICALLY CORRECT WAY TO SHOW
THE RELATIVE PORTIONS OF MEASUREMENT VARIATION
THAT ARE AND ARE NOT PREDICTABLE
Measurement
Variation Unit 1.00
0.60 (Error Contribution)
.80² + .60² = 1.00² .64 +.36 = 1.00
The model accounts for 64% of measurement variance.
The remaining 36% is error variance.
Correlation Coefficient 0.80
(Predicted Contribution)