Exhibit No.:	
Issues:	Weather Normalization and Water Utilization Trend Estimates
Witness:	Edward L. Spitznagel, Jr.
Exhibit Type:	Surrebuttal
Sponsoring Party:	Missouri-American Water Company
Case No.:	WR-2008-0311 SR-2008-0312
Date:	October 16, 2008

MISSOURI PUBLIC SERVICE COMMISSION

CASE NO. WR-2008-0311
SR-2008-0312

SURREBUTTAL TESTIMONY

OF

EDWARD L. SPITZNAGEL, JR.

ON BEHALF OF

MISSOURI-AMERICAN WATER COMPANY

BEFORE THE PUBLIC SERVICE COMMISSION

OF THE STATE OF MISSOURI

IN THE MATTER OF MISSOURI-AMERICAN WATER COMPANY FOR AUTHORITY TO FILE TARIFFS REFLECTING INCREASED RATES FOR WATER AND SEWER SERVICE	)))))	**CASE NO. WR-2008-0311** **CASE NO. SR-2008-0312**

AFFIDAVIT OF EDWARD L. SPITZNAGEL, JR.

Edward L. Spitznagel, Jr., being first duly sworn, deposes and says that he is the witness who sponsors the accompanying testimony entitled "Surrebuttal Testimony of Edward L. Spitznagel, Jr."; that said testimony and schedules were prepared by him and/or under his direction and supervision; that if inquires were made as to the facts in said testimony and schedules, he would respond as therein set forth; and that the aforesaid testimony and schedules are true and correct to the best of his knowledge.

Edward L. Spitznagel, Jr.

Edward L. Spitznagel, Jr.

State of Missouri
County of St. Louis
SUBSCRIBED and sworn to
Before me this 13th **day of** October **2008.**

Lisa Caress

Notary Public

My commission expires: 04/11/2010

SURREBUTTAL TESTIMONY
EDWARD L. SPITZNAGEL
MISSOURI-AMERICAN WATER COMPANY
CASE NO.WR-2008-0311
CASE NO.SR-2008-0312

TABLE OF CONTENTS

SURREBUTTAL TESTIMONY

EDWARD L. SPITZNAGEL, JR.

WITNESS INTRODUCTION

Q. PLEASE STATE YOUR NAME, BUSINESS ADDRESS, AND EMPLOYER.

A. My name is Edward L. Spitznagel, Jr., and my business address is Campus Box 1146, One Brookings Drive, St Louis, Missouri 63130. I am employed by Washington University.

Q. WHAT IS YOUR PRESENT POSITION?

A. I am Professor of Mathematics in the College of Arts and Sciences at Washington University. I also hold a joint appointment in the Division of Biostatistics of the Washington University School of Medicine.

Q. ARE YOU THE SAME EDWARD L. SPITZNAGEL, JR WHO FILED DIRECT AND REBUTTAL TESTIMONY IN THIS CASE?

A. Yes, I am.

PURPOSE AND SCOPE

Q. WHAT IS THE PURPOSE OF YOUR SURREBUTTAL TESTIMONY?

A. I will respond to the Staff member Jerry Scheible's rebuttal testimony in which it is claimed there is no evidence for weather affecting water usage. I will also respond to Mr. Scheible's claim that "omission of the 2006 data amplifies any significant change in usage between 2005 and 2007, causing any predicted value to be skewed artificially high or low, accordingly."

ERRORS IN STAFF'S ANALYSIS

Q. PLEASE DESCRIBE ERRORS MADE BY MR. SCHEIBLE.

A. There are three errors. The first error is the attempt to use precipitation, i.e., monthly rainfall, as the weather variable that drives water consumption. The second error is to have plotted rainfall and consumption versus time on the same graph, in an attempt to match a pattern of rainfall with consumption. The third error is making no attempt to adjust for different consumer behavior in different months.

Q. PLEASE DISCUSS MR. SCHEIBLE'S ERROR IN USING MONTHLY RAINFALL AS THE WEATHER VARIABLE DRIVING CONSUMPTION.

A. Until this present case, Staff had traditionally used a soil moisture index as a predictor of consumption. The difference between rainfall and soil moisture lies in the fact that the soil has a finite capacity to store water. Once this capacity is reached, the soil cannot hold any additional moisture until it gradually dries out over subsequent rain-free days. Staff's soil moisture index used for ratemaking purposes in the 1990's allowed for the soil to have a one-inch capacity. Thus, for example, 3 inches of rain falling on a single day could contribute, at most, one inch of moisture to the soil, and usually less depending on how much moisture remained in the soil from the last rainfall. Since consumers will water lawns and gardens based on their grass and plantings needing water contained in the soil, rainfall itself is a very poor driver of water consumption.

Rainfall and the Palmer Drought Severity Index are two very different concepts and they are only weakly correlated with each other. In my Schedule 1, which is based on the St. Louis County residential data used by Mr. Scheible in his rebuttal testimony, the correlation between rainfall and usage was −0.1000, very small in magnitude and not statistically significant. However, when I add to Schedule 1 the Palmer Drought Severity Index (PDSI), which is the moisture index I use in weather normalization, it can be seen that the PDSI is very different from total rainfall, with only 0.3224 correlation between the two. That is, rainfall accounts for only 0.3224^2, or 10.4% of the variation in PDSI. Rainfall and PDSI are thus very different measurements of moisture.

Q. PLEASE DISCUSS MR. SCHEIBLE'S ERROR PLOTTING RAINFALL AND CONSUMPTION VERSUS TIME ON THE SAME GRAPH.

A. in an attempt to match a pattern of rainfall with consumption, Mr. Scheible plotted both on a common y-axis, with time represented on the x-axis. Only if there were a very strong relationship of both to time would one be able to see the connection between the two. A much better approach is to produce an XY Scatterplot, with the moisture variable on the x-axis, and consumption on the y-axis. However, even this plotting method will not show a relationship between moisture and consumption until the third of Mr. Scheible's errors is accounted for.

Q. PLEASE DISCUSS MR. SCHEIBLE'S THIRD ERROR, WHICH IS NOT ACCOUNTING FOR MONTHLY DIFFERENCES IN WATER USAGE PATTERNS.

A. In the hot summer months, considerably more water is used than in the cooler months of November through March. Unless this variation is accounted for, it becomes statistical "noise," obscuring the relationship between moisture and water consumption. The correct way to assess the effect of soil moisture on water usage is to control for the calendar month. The best way to control for calendar month is to include it as a categorical covariate directly in a multiple regression model, as I did in Schedule 1 to my rebuttal testimony (which in turn is taken from Page 2 of Schedule ELS-2, ELS Direct Testimony). In that analysis, the model R-square is 88.04%, out of a possible 100%. However, a reasonable alternative for demonstration purposes is to predict the deviation in consumption from the monthly average consumption. I have calculated those deviations and illustrated their relationship to the PDSI in three different ways.

The first is the correlation coefficient between PDSI and deviation from monthly average, −0.2946, marked in bold in Schedule 1 of my surrebuttal testimony. As will be seen in the regression table described below, this correlation is statistically significant. By comparison, the correlation coefficient between rainfall and deviation from monthly average is only −0.1008, which is not statistically significant.

The second is the Scatterplot of those deviations on the y-axis versus the PDSI on the x-axis. Superimposed on the Scatterplot is the best-fitting (in the sense of least squares) regression line, whose equation $y = -0.1024\,x + 0.0439$ appears in a legend in the upper right corner of the plot. The legend also contains the value of the squared correlation coefficient, 0.0868 (= (-0.2946^2)). The squared correlation is the percent of variation, 8.68%, in usage that is accounted for by variation in soil moisture. The coefficient of x, −0.1024, is the decrease in water consumption, in thousands of gallons per month, corresponding to a one-unit change in PDSI. As PDSI (i.e., soil moisture) increases, say, from −1 to 0, residential customers on average decrease their usage by 102.4 gallons each month. Multiplied by the number of residential customers, this would amount to a very substantial change in revenue to the Company. This correlation between PDSI and usage is visually demonstrated by the line which slopes downward from left to right. For example, when the PDSI (i.e., soil moisture) is negative, or drier than normal, consumption is greater than the average. Conversely, when the PDSI is positive, or wetter than normal, consumption is less than average.

The third is a regression calculation in which PDSI is used to predict deviation of monthly consumption from its average. The regression calculation shows the fit of the least squares line to be statistically significant, with P-value of 0.0120. The line equation is the same as that shown in the XY scatter plot. The additional information in the tabular output contains what is called a P-value for the slope coefficient. The value of 0.0120 is the probability of two equivalent events, under the null hypothesis that PDSI and consumption are unrelated. The first is obtaining a correlation of magnitude (positive or negative) at least as large as 0.2946. The second, equivalent to it, is obtaining a line slope of magnitude (positive or negative) at least as large as 0.1024. A P-value of 0.05 (=1/20) or less is considered to be statistically significant. In this instance, the P-value of 0.0120 (≈ 1 chance out of 83) is therefore strong evidence that PDSI and consumption are indeed related.

Because these regression models are in Excel format and are not based on optimal statistical methods, but rather are done to illustrate the problems with Mr. Scheible's calculations, they do not give the same degree of precision to weather normalization that my original general linear models do. Given this additional support from these Excel models, I believe the Commission should accept my original estimates.

TREND LINE ANALYSIS

Q. PLEASE DISCUSS YOUR CALCULATIONS IN SCHEDULE 2, WHICH SHOW THAT THE GAPS IN THE YEARLY DATA ACTUALLY IMPROVE THE STABILITY OF PROJECTIONS INTO THE FUTURE.

A. In his rebuttal testimony, Mr. Scheible claims that "omission of the 2006 data amplifies any significant change in usage between 2005 and 2007, causing any predicted value to be skewed artificially high or low, accordingly." This in fact is incorrect. While the data for 2006 was not used due to the conversion back from the 4-4-5 accounting method, a total of six years' data were employed by beginning with data from the year 2000. (The year 2003 was also not used because of anomalies associated with conversion to the 4-4-5 method.)

The greater spread of the x-values, consisting of the years 2000, 2001, 2002, 2004, 2005, and 2007, over an eight year period, causes the standard error of estimate for projections into both 2008 and 2009 to be smaller than if the consecutive six years from 2002 through 2007 had been used. These calculations can be found in the spreadsheet in Schedule 2, which implements a well-known formula, known as the Standard Error of the Estimated Mean Response, contained in virtually every introductory statistics book. For projection into 2008, the standard error of estimate using the six consecutive years 2002 through 2007 is the root mean square error S_e multiplied by 0.9309, whereas the standard error for the six non-consecutive years 2002 through 2007 omitting 2003 and 2006 is S_e multiplied by 0.9151. The smaller standard error corresponds to the better estimate. Similarly, for projections into 2009, changing the x-star value from 2008 to 2009 will yield standard errors of 1.1506 S_e and 1.0694 S_e respectively.

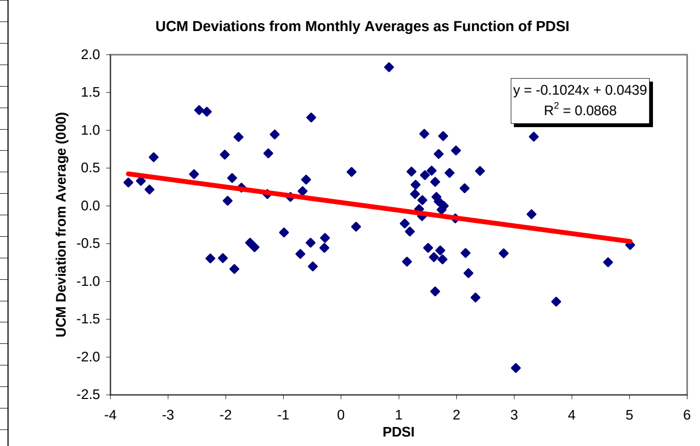
Q. DOES THIS CONCLUDE YOUR TESTIMONY?

A. Yes, it does.

	A	B	C	D	E	F
1	**St Louis Residential Quarterly**					
2		72 Months of Data				
3		Usage (000)	Rainfall	PDSI	Monthly Avg	Deviation
4	Jan-00	7.872	1.230	-3.47	7.5440	0.3285
5	Feb-00	8.168	3.110	-3.25	7.5230	0.6445
6	Mar-00	6.839	1.880	-3.32	6.6225	0.2164
7	Apr-00	6.364	1.840	-3.69	6.0536	0.3104
8	May-00	7.531	5.840	0.18	7.0835	0.4471
9	Jun-00	7.830	8.220	1.22	7.3779	0.4523
10	Jul-00	8.560	2.250	1.35	8.6035	-0.0431
11	Aug-00	10.949	3.640	1.99	10.2160	0.7333
12	Sep-00	9.534	2.620	1.63	10.6665	-1.1321
13	Oct-00	9.830	2.600	1.51	10.3850	-0.5546
14	Nov-00	9.279	2.790	1.45	8.8741	0.4048
15	Dec-00	7.796	1.350	1.10	8.0277	-0.2317
16	Jan-01	7.409	1.120	1.40	7.5440	-0.1354
17	Feb-01	7.983	2.480	2.41	7.5230	0.4599
18	Mar-01	6.571	1.450	1.74	6.6225	-0.0518
19	Apr-01	6.332	3.010	1.29	6.0536	0.2780
20	May-01	8.038	2.810	1.44	7.0835	0.9546
21	Jun-01	8.066	3.600	1.69	7.3779	0.6876
22	Jul-01	9.069	4.000	1.57	8.6035	0.4651
23	Aug-01	11.139	1.990	1.77	10.2160	0.9234
24	Sep-01	10.785	2.810	1.65	10.6665	0.1189
25	Oct-01	10.620	5.500	2.14	10.3850	0.2350
26	Nov-01	9.311	3.060	1.88	8.8741	0.4372
27	Dec-01	7.320	3.460	1.76	8.0277	-0.7074
28	Jan-02	6.919	3.160	2.16	7.5440	-0.6249
29	Feb-02	7.582	0.830	1.69	7.5230	0.0594
30	Mar-02	6.699	3.670	1.41	6.6225	0.0768
31	Apr-02	5.466	4.250	1.72	6.0536	-0.5875
32	May-02	6.974	7.810	3.30	7.0835	-0.1094
33	Jun-02	6.824	5.260	-0.29	7.3779	-0.5539
34	Jul-02	7.969	1.470	-0.71	8.6035	-0.6342
35	Aug-02	11.385	4.120	-0.52	10.2160	1.1691
36	Sep-02	11.612	2.440	-1.15	10.6665	0.9452
37	Oct-02	10.581	4.780	-0.67	10.3850	0.1960
38	Nov-02	9.569	1.140	-1.26	8.8741	0.6944
39	Dec-02	7.480	2.020	-1.50	8.0277	-0.5475
40	Jan-04	6.656	3.970	2.21	7.5440	-0.8882
41	Feb-04	7.840	0.850	1.63	7.5230	0.3165
42	Mar-04	6.458	4.360	1.98	6.6225	-0.1641
43	Apr-04	6.208	1.940	1.28	6.0536	0.1547
44	May-04	7.082	9.750	1.78	7.0835	-0.0011
45	Jun-04	7.038	0.830	1.19	7.3779	-0.3396
46	Jul-04	7.925	5.520	1.61	8.6035	-0.6785
47	Aug-04	8.071	4.100	3.03	10.2160	-2.1453
48	Sep-04	9.454	0.230	2.33	10.6665	-1.2124
49	Oct-04	9.759	3.210	2.82	10.3850	-0.6257
50	Nov-04	7.608	5.740	3.73	8.8741	-1.2663
51	Dec-04	8.944	1.770	3.34	8.0277	0.9162
52	Jan-05	7.028	9.010	5.01	7.5440	-0.5164
53	Feb-05	6.778	1.840	4.63	7.5230	-0.7450
54	Mar-05	6.970	1.470	-0.61	6.6225	0.3480
55	Apr-05	6.173	2.170	-0.88	6.0536	0.1198
56	May-05	6.595	0.780	-1.58	7.0835	-0.4883
57	Jun-05	7.621	5.100	-1.73	7.3779	0.2427
58	Jul-05	9.848	2.220	-2.33	8.6035	1.2443
59	Aug-05	9.380	3.870	-1.85	10.2160	-0.8356
60	Sep-05	11.576	5.300	-1.78	10.6665	0.9096
61	Oct-05	11.065	1.520	-2.02	10.3850	0.6801
62	Nov-05	8.185	3.350	-2.05	8.8741	-0.6896
63	Dec-05	9.295	1.220	-2.46	8.0277	1.2669
64	Jan-07	9.380	3.110	0.83	7.5440	1.8365
65	Feb-07	6.788	1.980	1.14	7.5230	-0.7354
66	Mar-07	6.197	2.800	-0.28	6.6225	-0.4253
67	Apr-07	5.778	3.180	0.26	6.0536	-0.2755
68	May-07	6.281	4.260	-0.49	7.0835	-0.8028
69	Jun-07	6.889	2.880	-0.53	7.3779	-0.4891
70	Jul-07	8.250	3.110	-0.99	8.6035	-0.3537
71	Aug-07	10.371	1.570	-1.28	10.2160	0.1550
72	Sep-07	11.037	1.710	-1.89	10.6665	0.3707
73	Oct-07	10.454	1.970	-1.97	10.3850	0.0692
74	Nov-07	9.293	1.250	-2.55	8.8741	0.4193
75	Dec-07	7.331	2.750	-2.27	8.0277	-0.6965

Correlations

	Usage (000)	Rainfall	PDSI	Monthly Avg
Rainfall	-0.1000			
PDSI	-0.1571	0.3224		
Monthly Avg	0.9025	-0.0627	-0.0334	
Deviation	0.4308	-0.1008	**-0.2946**	0.0000



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.2946334
R Square	0.0868088
Adjusted R S	0.0737632
Standard Err	0.6736637
Observations	72

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3.01985753	3.019857528	6.654267	0.011994
Residual	70	31.7675903	0.453822718		
Total	71	34.7874478			

	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.0439092	0.08119627	0.540778429	0.590377	-0.11803	0.205849921
PDSI	-0.1023789	0.03968812	-2.579586591	0.011994	-0.18153	-0.023223549

The accuracy of a future prediction as a function of gaps in the independent variable

No Gaps	Two Gaps	Three Gaps	x-star
2002	2000	1999	**2008**
2003	2001	2000	
2004	2002	2001	
2005	2004	2004	
2006	2005	2005	
2007	2007	2007	

Standard Error(y-hat) = Se times:

0.9309	0.9151	0.8621