Tools and Methods for Addressing Multicollinearity in Energy Modeling

The NW Industrial Strategic Energy Management (SEM) Collaborative was formed in 2012 by the Bonneville Power Administration (BPA), Energy Trust of Oregon (ETO), and the Northwest Energy Efficiency Alliance (NEEA) for the purpose of identifying and addressing market barriers to SEM adoption in the industrial sector. Three teams were subsequently formed to focus on: 1) Small-to-Medium Industrial Solutions, 2) Market Analysis and Planning, and 3) Energy Tracking and Savings Protocols (ETSP).

The members of the ETSP team include representatives from BPA, ETO, NEEA, Idaho Power, U.S. Department of Energy, the Consortium for Energy Efficiency, Puget Sound Energy, BC Hydro, and the Northwest Food Processors Association. This paper is the initial work product of the ETSP team, which was tasked with the identification of consistent and defensible methodologies for measuring and verifying SEM energy savings. The examples in this paper were drawn from recent modeling efforts in BPA's Energy Smart Industrial and the Energy Trust of Oregon's Production Efficiency programs.

Introduction

After reviewing SEM measurement and verification (M&V) protocols from different programs, the ETSP team identified multicollinearity as a common statistical issue in industrial data sets. Understanding that multicollinearity has the potential to affect the specification of regression-based energy models used to determine adjusted baselines, more consistent treatment of this issue may help improve confidence in SEM-based savings, and thereby address a potential market barrier. Therefore, the team compiled this paper for the purposes of outlining the implications of multicollinearity in the context of SEM measurement and verification, and providing examples of how program implementers have successfully identified and treated the presence of multicollinearity among a set of predictor variables.

Statistical Definition

Multicollinearity is present when two or more predictor variables in a regression model are correlated among themselves. When two independent variables tend to move together through time, including both variables may not add appreciably to the explanatory power of the model, compared to just having one or the other in the model. Essentially, the additional variable has little new information to provide. While perfect multicollinearity involves two variables that are linear translations of each other (R^2=1), partial-multicollinearity is more commonly observed in the context of regression-based energy models.

In the practice of developing predictive energy models, multicollinearity typically arises from one of the following:

1. There are two or more observed factors that each drive energy consumption and that trend together in fairly consistent ways.
2. There's an observed factor or factors that drive energy consumption, and these drivers also drive other observed variables.
3. There's an unobserved factor that drives energy and some other observed factors. Then the observed factors are all proxies for the true drivers, and the modeler is faced with developing the best possible predictive model from the available data.

Implication for Model Results

When multicollinearity is present, the coefficient of any one independent variable depends on which of the other correlated variables are included in the model. The ordinary least squares (OLS) results may indicate a good overall fit to the sample data in terms of R^2 and overall predictive capability. However, including or excluding other independent variables or using a different sample (or base period) may lead to large changes in the estimated coefficients when multicollinearity exists. Hence, the coefficients must be interpreted given the other explanatory variables in the model. The reliability of the coefficients will be reflected appropriately in the OLS-based standard errors.

The presence of correlated variables should serve as a warning that the statistical significance of a variable in a particular regression does not by itself indicate how closely that variable is correlated with energy consumption. Therefore, when faced with multicollinear variables, the modeler should exercise caution in excluding any variables that might actually be significant drivers of energy use.

Example #1: Municipal Wastewater Treatment Facility

Approximately 3-4% of the electricity consumed in the U.S. is used for the treatment, conveyance, and disposal of water and wastewater. The vast majority of electric utilities have a wastewater treatment (WWT) facility within their service areas, and there is a growing awareness of the benefits of SEM among WWT professionals. In the practice of establishing meter-level, regression-based energy models for wastewater treatment facilities, SEM practitioners are often confronted with two or more variables that exhibit varying degrees of multicollinearity. While plant influent is likely a primary energy driver, a range of other factors may affect the energy use within the facility. Some of these are physically related to the influent flow.

This example provides a short case study in the treatment of multicollinearity in a wastewater treatment facility. The program participant is a municipal operator of a 10-40 million gallon per day (MGD) open-system treatment facility in the Pacific Northwest. Figure 1 provides an illustration of the main process steps in the plant's operations. The electrical energy measurement boundary was defined by the utility's revenue meter, which covered both the water treatment and solids handling systems.

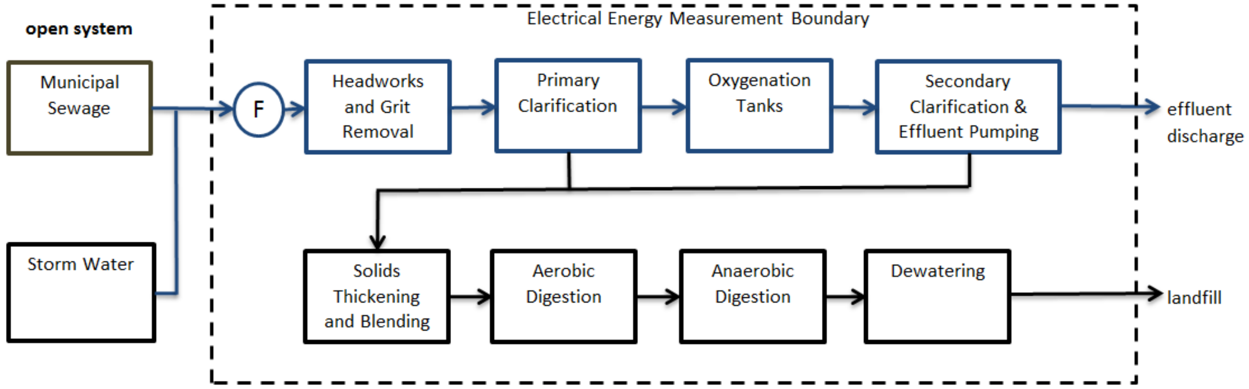


Figure 1. Process Flow Illustration

Review of Hypothesis Variables

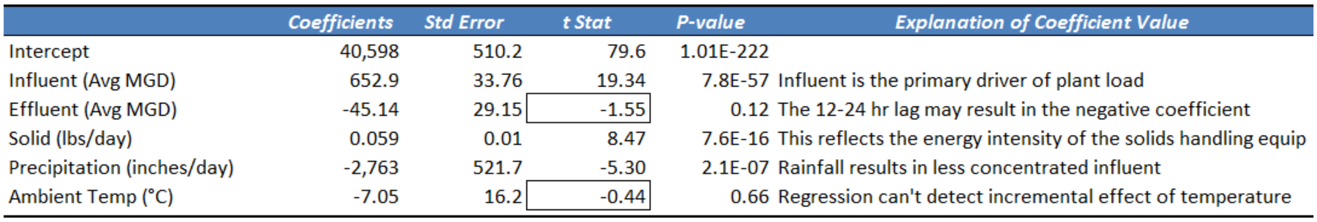
Equipped with an informed understanding of the process steps and electrical loads, the model developer begins by establishing a set of hypothesis predictor variables. Table 1 provides a summary of the variables selected at the outset of this effort, with a brief explanation of each variable’s association with energy use and the potential mechanisms for multicollinearity with other variables.

Table 1. Common Energy-Driver Variables in WWT Facilities

Variable	Association with Energy	Potential for Collinearity
Influent (untreated wastewater into plant)	Key driver of system load in most process steps, particularly the energy-intensive secondary treatment step.	Typically, plant influent will always be included, and thereby becomes the reference variable for assessing multicollinearity.
Effluent (treated, clean water leaving plant)	Also closely associated with plant load.	Effluent volume is typically closely associated with influent, but offset by the lead time of the treatment process.
Solids Processed	Indicative of the load on the solids-handling portion of the plant.	If the plant is run continuously, seven days per week, this parameter may be strongly correlated with influent.
Precipitation	In an open system that collects storm water, higher precipitation leads to more influent, with a lower concentration of solids and biological oxygen demand.	Influent and precipitation typically follow similar trends, depending on the system design.
Ambient Temperature	The solubility of oxygen in water varies with temperature, which impacts aeration load.	Ambient temperature and precipitation may be loosely correlated.

Before assessing multicollinearity, the model developer may perform a regression that includes all the hypothesis variables, for the purpose of gaining a cursory assessment of the relative statistical significance each variable. Table 2 shows the regression output for the five variables outlined in the preceding table.

Table 2. Regression Results When all Five Predictor Variables are Included



	Coefficients	*Std Error*	*t Stat*	*P-value*	*Explanation of Coefficient Value*
Intercept	40,598	510.2	79.6	1.01E-222	
Influent (Avg MGD)	652.9	33.76	19.34	7.8E-57	Influent is the primary driver of plant load
Effluent (Avg MGD)	-45.14	29.15	-1.55	0.12	The 12-24 hr lag may result in the negative coefficient
Solid (lbs/day)	0.059	0.01	8.47	7.6E-16	This reflects the energy intensity of the solids handling equip
Precipitation (inches/day)	-2,763	521.7	-5.30	2.1E-07	Rainfall results in less concentrated influent
Ambient Temp (°C)	-7.05	16.2	-0.44	0.66	Regression can't detect incremental effect of temperature

A commonly-applied rule of thumb would guide the modeler to carefully consider the statistical significance of variables with absolute value T-statistics of less than 2.0. For example, the IPMVP references a T-statistic of greater than 2.0 as providing a reasonable degree of confidence of a variable’s impact on energy use. On this basis, effluent and ambient temperature should not be treated as statistically significant variables, while the other four variables should be retained for further analysis. However, the T-statistic for effluent indicates moderate statistical significance, and the model developer can’t dismiss the possibility that its impact on energy is understated due to the inability of the regression analysis to empirically separate the effect of influent volume from the effect of effluent volume over the observed range of conditions. A basic examination of the scatter diagrams between these two variables and energy use, shown in Figures 2 and 3, leads to one of the following three conclusions:

- Influent and effluent are both significant drivers of energy use,
- Influent is a driver of both energy and effluent, resulting in a strong correlation between effluent and energy (or the reverse),
- Influent, effluent, and energy are all driven by some other factor, resulting in correlation among the three variables. Identifying the true "driver" may require a more detailed understanding of the mechanism at play.

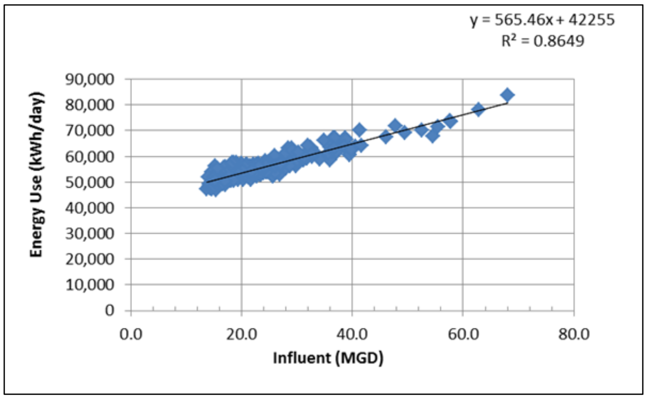


Figure 2. Influent versus Energy Use

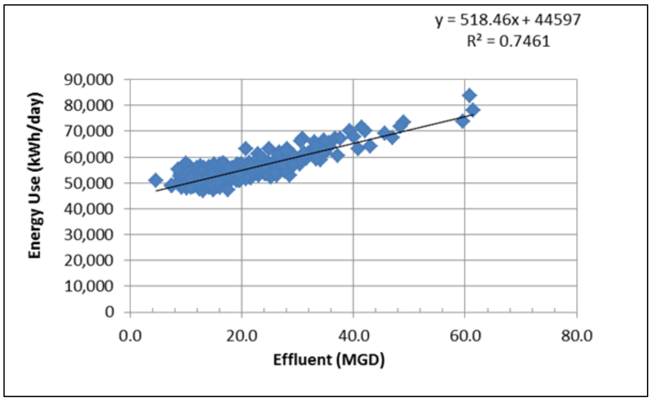


Figure 3. Effluent versus Energy Use

In the practice of multivariable regression, a general assessment for multicollinearity can be performed by regressing each predictor variable against the other hypothesis variables, and examining the coefficient of determination (R^2) of each relationship. As a rule of thumb, any bivariate correlation with $R^2>0.7$ is an indication that multicollinearity needs to be carefully considered in the variable selection process. The generation of a matrix of x-y scatter diagram, as shown in Figure 4, can provide a reasonable assessment of correlation among two variables. However, it must be recognized that multicollinearity is a multivariate problem, and while a simple matrix of correlation coefficients and scatter diagrams can identify two independent variables that are highly correlated, this exercise has limited ability to detect an independent variable that is highly correlated to a combination of predictor variables. A more complete assessment regresses each independent variable on all of the others.

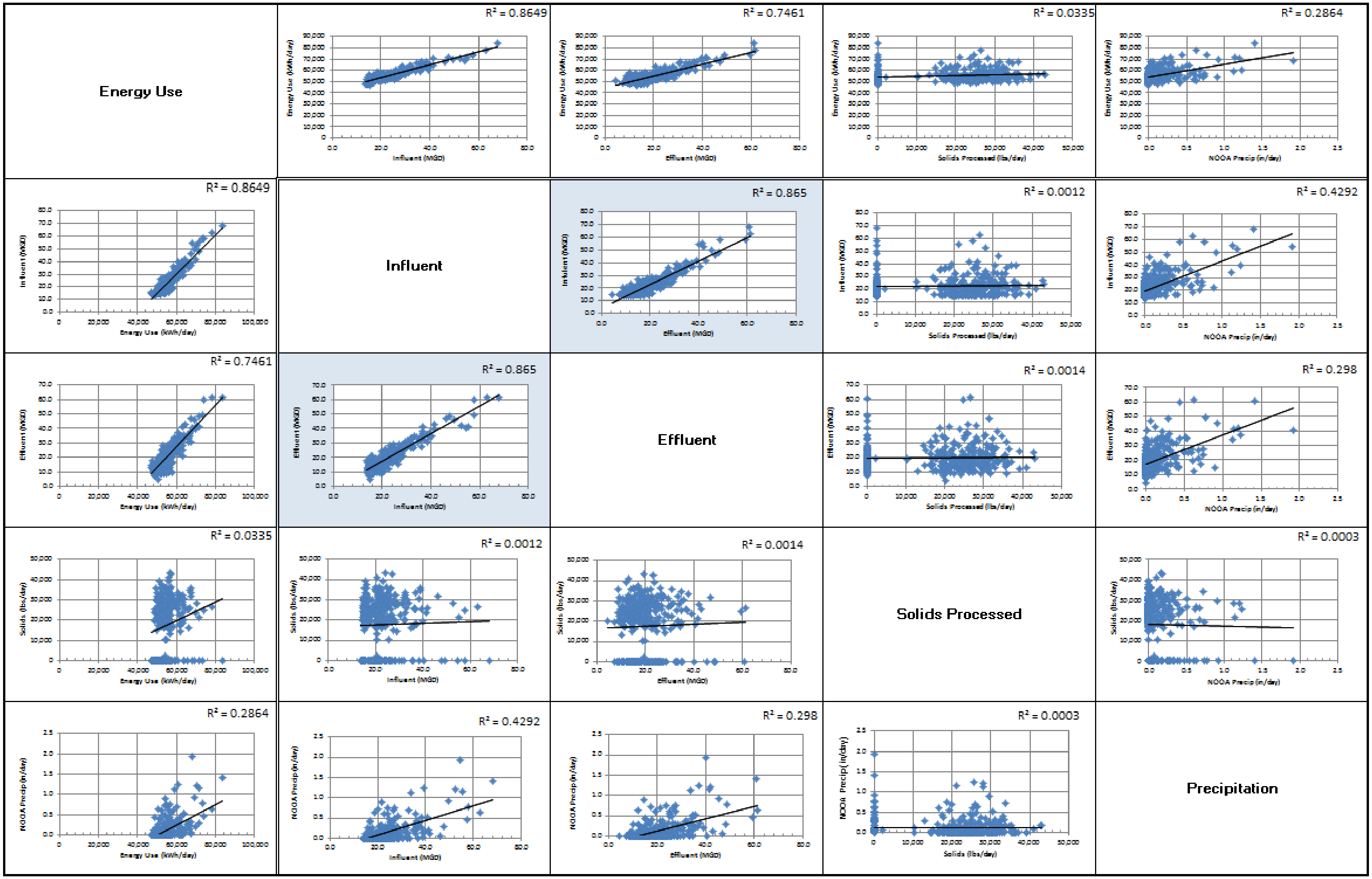


Figure 4. Scatter Diagram Matrix

Casual inspection of the scatter plots confirms the presence of collinearity between the influent and effluent variables, which is validated by their respective coefficient of determination (R^2=0.87). As one might expect, influent and precipitation exhibit a modest degree of correlation, but the R^2 of 0.43 doesn't violate common statistical guidelines.

Variance Inflation Factor

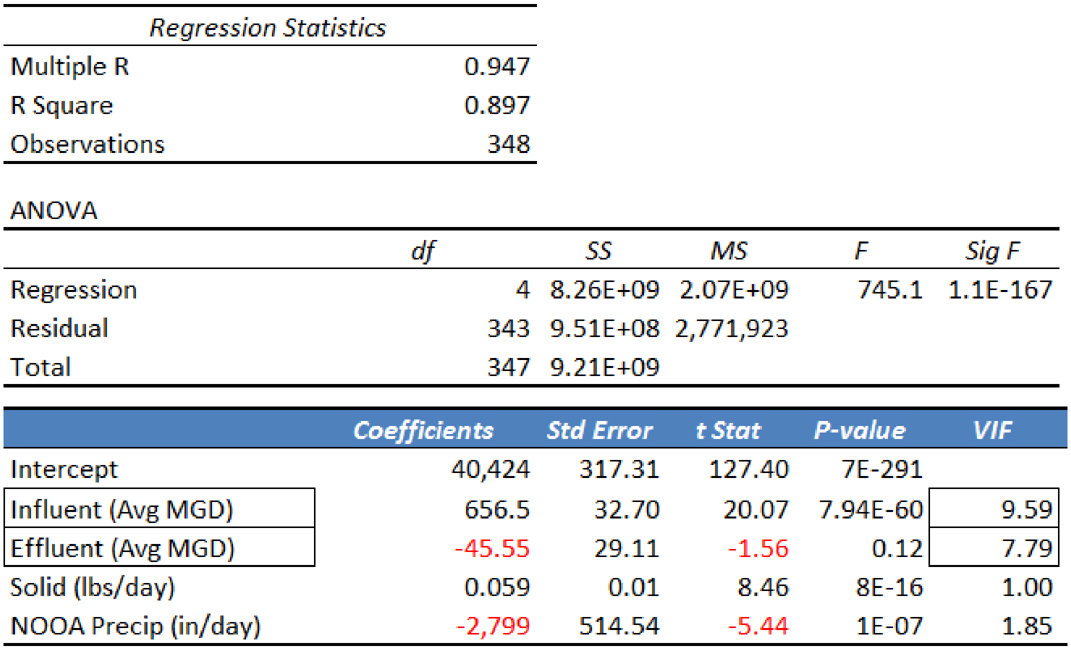
At this point, the model developer may decide to further examine the severity of multicollinearity using the Variance Inflation Factor (VIF). While this exercise isn't normally required in the development of energy models, it may provide additional insight when one encounters counterintuitive results from a regression output (e.g. beta coefficient signs or magnitudes, low t-stats). The VIF provides a reliable index of how much the variance of an estimated regression coefficient is increased because of the collinearity, when compared to having uncorrelated predictors. The VIF is useful because it provides an indication of the inter-correlation effects among all the variables, not just two at a time.

Equation 1 provides a useful formula for calculating VIFs for multiple predictor variables:

$$Equation\ (1)\text{: } Variable\ Inflation\ Factor\ \left(VIF_j\right) = \frac{S_j^2(n-1)SE_j^2}{MSE_{residuals}}$$

The following table shows the standard Excel regression output, with an additional column for each variable's standard deviation, along with the VIF calculated by Equation (1). Note that the temperature variable was determined to be statistically insignificant, and was removed from the set of hypothesis variables. Variance Inflation Factor values range from one, indicating that a variable isn't correlated with any other predictors, to infinity, for near perfect correlation in which there exists no unique solution for the regression coefficient. As a rule of thumb, if any of the VIF values is greater than five, the modeler should consider taking steps to address multicollinearity.

Table 3. Regression Output with VIF for Five Variables



Regression Statistics	
Multiple R	0.947
R Square	0.897
Observations	348

ANOVA

	df	*SS*	*MS*	*F*	*Sig F*
Regression	4	8.26E+09	2.07E+09	745.1	1.1E-167
Residual	343	9.51E+08	2,771,923		
Total	347	9.21E+09			

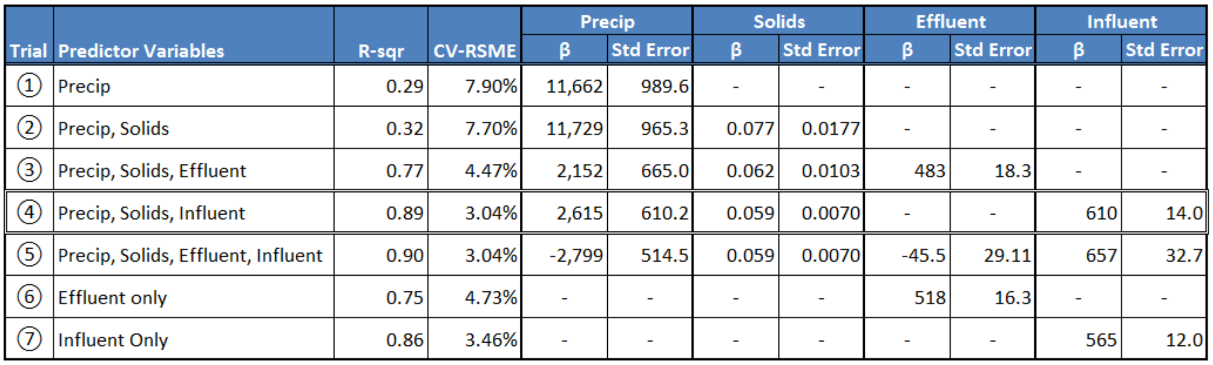
	Coefficients	*Std Error*	*t Stat*	*P-value*	*VIF*
Intercept	40,424	317.31	127.40	7E-291	
Influent (Avg MGD)	656.5	32.70	20.07	7.94E-60	9.59
Effluent (Avg MGD)	-45.55	29.11	-1.56	0.12	7.79
Solid (lbs/day)	0.059	0.01	8.46	8E-16	1.00
NOOA Precip (in/day)	-2,799	514.54	-5.44	1E-07	1.85

To put the Variance Inflation Factor figure in context with other regression statistics, the VIF of 9.59 for influent indicates that the standard error term associated with its beta coefficient is approximately three times as large as is would be if it were uncorrelated with other predictor variables ($\sqrt{9.59} = 3.1$). In this case, the model developer considered the implications of eliminating either the influent or effluent variable from the model specification.

Stepwise Regression and Model Specification

A typical approach in this decision process involves performing the regression with and without each variable. For illustrative purposes, Table 4 shows the results of a stepwise regression process, beginning with precipitation as the single variable.

Table 4. Stepwise Regression Summary



				Precip		Solids		Effluent		Influent	
Trial	Predictor Variables	R-sqr	CV-RSME	β	Std Error	β	Std Error	β	Std Error	β	Std Error
①	Precip	0.29	7.90%	11,662	989.6	-	-	-	-	-	-
②	Precip, Solids	0.32	7.70%	11,729	965.3	0.077	0.0177	-	-	-	-
③	Precip, Solids, Effluent	0.77	4.47%	2,152	665.0	0.062	0.0103	483	18.3	-	-
④	Precip, Solids, Influent	0.89	3.04%	2,615	610.2	0.059	0.0070	-	-	610	14.0
⑤	Precip, Solids, Effluent, Influent	0.90	3.04%	-2,799	514.5	0.059	0.0070	-45.5	29.11	657	32.7
⑥	Effluent only	0.75	4.73%	-	-	-	-	518	16.3	-	-
⑦	Influent Only	0.86	3.46%	-	-	-	-	-	-	565	12.0

A comparison of the model fitness statistics, coefficient of determination (R^2), and the coefficient of variation (CV-RSME), shows incremental improvement in model resolution provided by each variable. A comparison of trials demonstrates that retaining influent as the construct of plant load (Trial ④) provides a higher level of model resolution versus the alternative option of retaining the effluent variable (Trial ③).

Trial ⑤ captures the original hypothesis model, without ambient temperature. While the presence of multicollinearity didn't affect the predictive capability of the model, including both variables in the model produces a beta coefficient for effluent that can't be interpreted as an accurate representation of effluent's independent incremental impact on plant energy use. Thus, one can observe instability in the beta results for effluent among different trials (Trial ⑤ versus ③, and Trial ⑤ versus ⑥). If an accurate characterization of a variable's independent influence on energy use is a desired outcome of the regression exercise, then this phenomenon should be recognized. Also, the presence of multicollinearity in Trial ⑤ may have led the modeler to exclude the effluent variable due to the marginally significant T-statistic, without further consideration. While that decision would have led to the same outcome in this example, in other instances it could result in the erroneous exclusion of a significant predictor variable, resulting in omitted variable bias.

The final model specification for this waste treatment facility, as outlined in Trial ④, is given by Equation (2):

$$Equation\ (2){:}\ \ Energy\ Use\ \frac{kWh}{day} = 40{,}533 + 610.2\ (Influent) + 0.059\ (Solids) - 2{,}615\ (Precipitation)$$

This model meets the basic criteria of the program administrator's SEM M&V guidelines. Figures 5 and 6 illustrate the strong agreement between actual and predicated energy consumption values across the observed range of energy consumption, and the unbiased nature of the model error (residuals) over the baseline period.

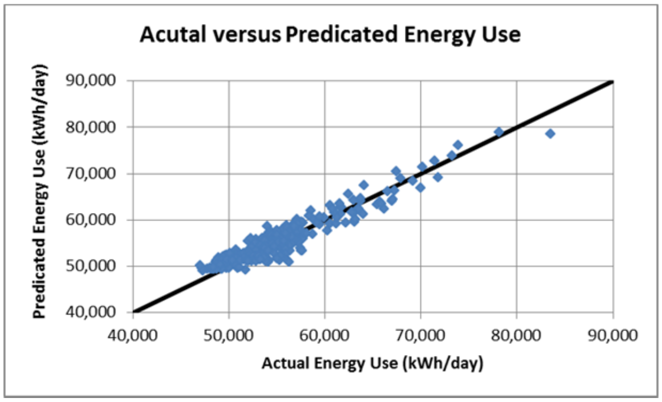


Figure 5. Actual versus Predicted Energy Use (Trial 4)

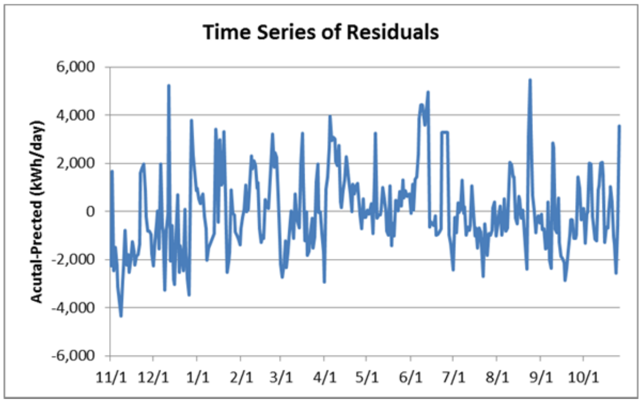


Figure 6. Time Series of Residuals (Trial 4)

Example #2 – Wood Products Company

When trying to identify important energy drivers, a company often provides many different production variables to add into the model. Many times, these variables are correlated to each other because all production tends to trend together. Depending how closely related to each other the variables are, a modeler much decide if multicollinearity in the model is causing an unstable model or if all variables are necessary to use to describe energy use at the site. One example of this is a wood products company that creates wood pieces of varying lengths. When providing the production energy driver data, the facility provided the following variables:

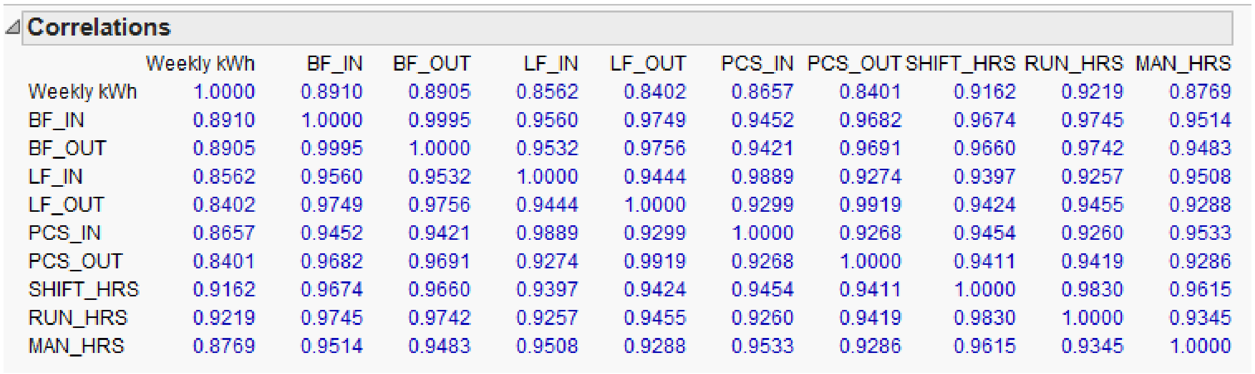
Table 5. Production Variables Provided by Plant

Production Variable	Description
Board-feet in	Board feet is the thickness of the board width times the length. This variable represents the incoming lumber into the plant.
Board-feet out	Total board-feet out is measured after the plant converts the incoming lumber into pieces of varying lengths and styles. The difference between board-feet in and out would be any scrap that is created throughout the process.
Linear-feet in	Linear-feet in is simply the total length of the incoming lumber.
Linear-feet out	Linear-feet out is the total length of the lumber after it has gone through the manufacturing process. The difference between linear feet in and linear feet out is any scrap created in the process.
Pieces in	Total number of pieces of incoming lumber.
Pieces out	Total number of pieces after manufacturing.
Shift Hours	Weekly hours worked by manufacturing employees.

Production Variable	Description
Run Hours	Weekly machine run time.
Man Hours	Weekly hours worked by all employees.

At first inspection, it is clear that all these variables are very closely related to each other. The employee hours data should increase if the production data increases and vice versa. When creating an energy model, the goal is to create the simplest model that best represents the energy usage at the facility. So, while the facility provided many variables for consideration, not all the variables are needed. Step one is determining whether any of the variables correlate to their energy use or to each other. Looking at the correlations of all the variables, it is clear that all nine variables provided are strongly correlated with one another.

Table 6. Correlation Matrix of Predictor Variables



Correlations

	Weekly kWh	BF_IN	BF_OUT	LF_IN	LF_OUT	PCS_IN	PCS_OUT	SHIFT_HRS	RUN_HRS	MAN_HRS
Weekly kWh	1.0000	0.8910	0.8905	0.8562	0.8402	0.8657	0.8401	0.9162	0.9219	0.8769
BF_IN	0.8910	1.0000	0.9995	0.9560	0.9749	0.9452	0.9682	0.9674	0.9745	0.9514
BF_OUT	0.8905	0.9995	1.0000	0.9532	0.9756	0.9421	0.9691	0.9660	0.9742	0.9483
LF_IN	0.8562	0.9560	0.9532	1.0000	0.9444	0.9889	0.9274	0.9397	0.9257	0.9508
LF_OUT	0.8402	0.9749	0.9756	0.9444	1.0000	0.9299	0.9919	0.9424	0.9455	0.9288
PCS_IN	0.8657	0.9452	0.9421	0.9889	0.9299	1.0000	0.9268	0.9454	0.9260	0.9533
PCS_OUT	0.8401	0.9682	0.9691	0.9274	0.9919	0.9268	1.0000	0.9411	0.9419	0.9286
SHIFT_HRS	0.9162	0.9674	0.9660	0.9397	0.9424	0.9454	0.9411	1.0000	0.9830	0.9615
RUN_HRS	0.9219	0.9745	0.9742	0.9257	0.9455	0.9260	0.9419	0.9830	1.0000	0.9345
MAN_HRS	0.8769	0.9514	0.9483	0.9508	0.9288	0.9533	0.9286	0.9615	0.9345	1.0000

The scatter-plot matrix also shows these connections graphically:

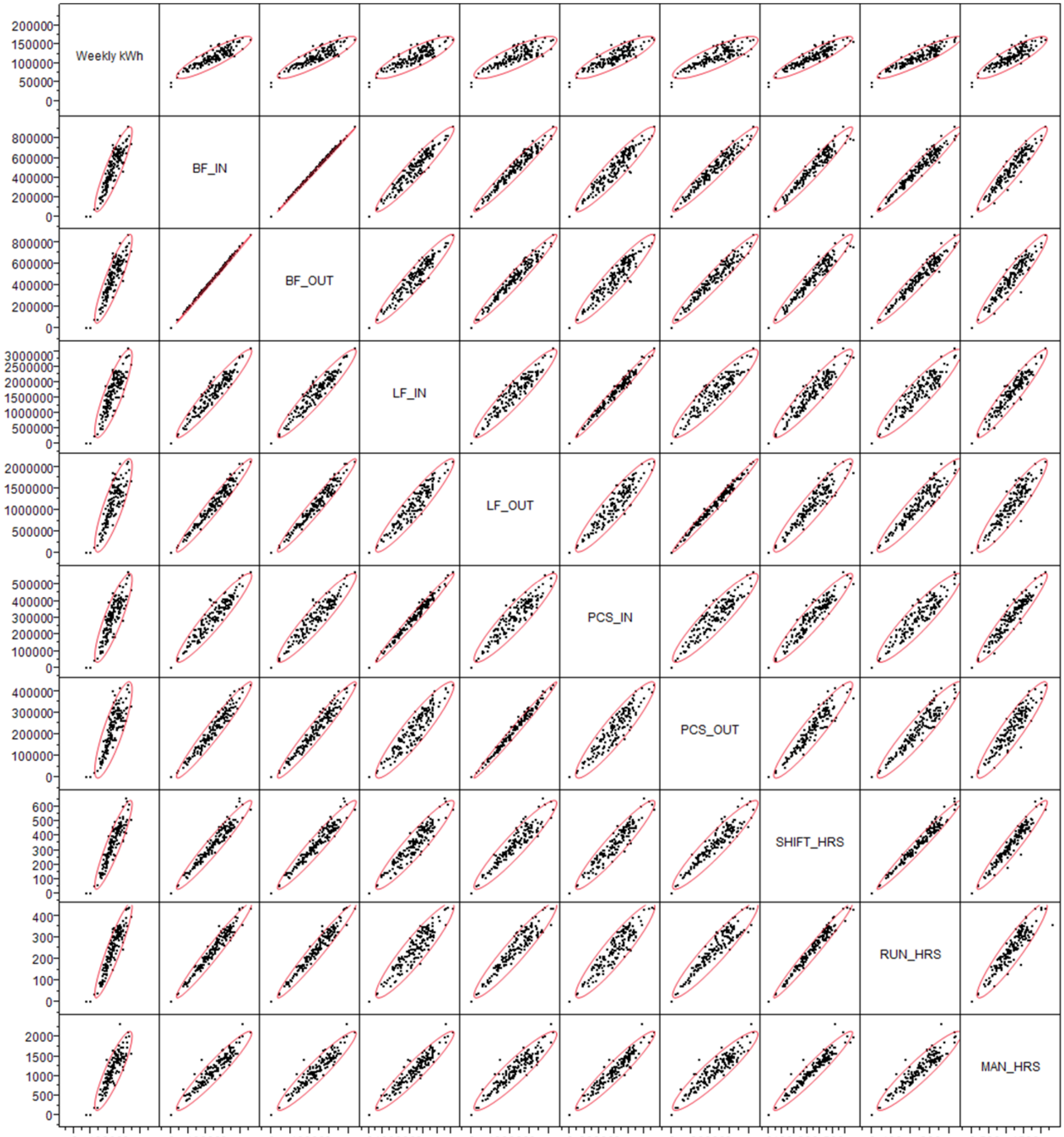
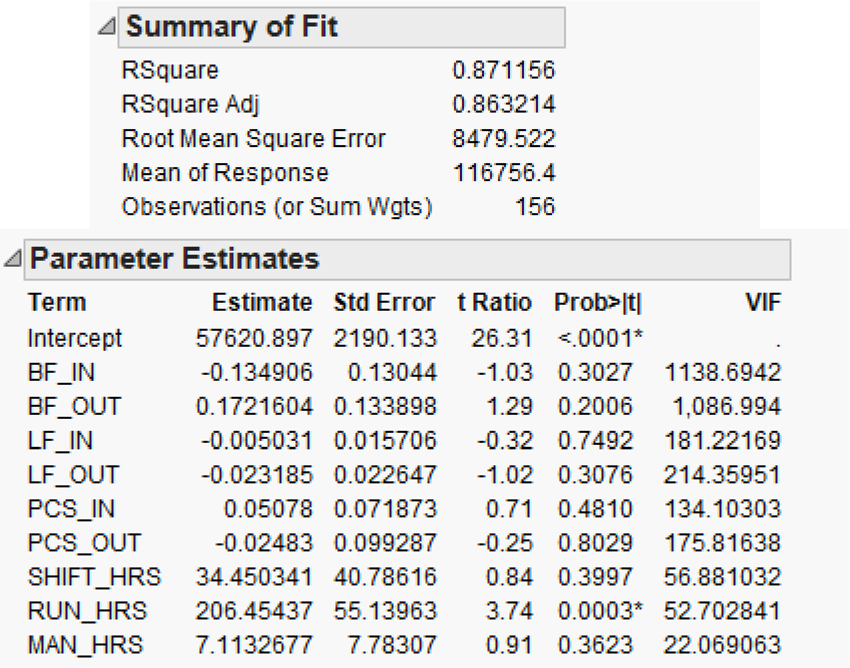


Figure 7. Scatter Diagram Matrix

If a model was created using all of these variables, multicollinearity would cause the coefficients and T-statistics to be highly unstable. The resulting parameter estimates are as follows:

Table 7. Regression Output with VIF for Nine Variables



Summary of Fit

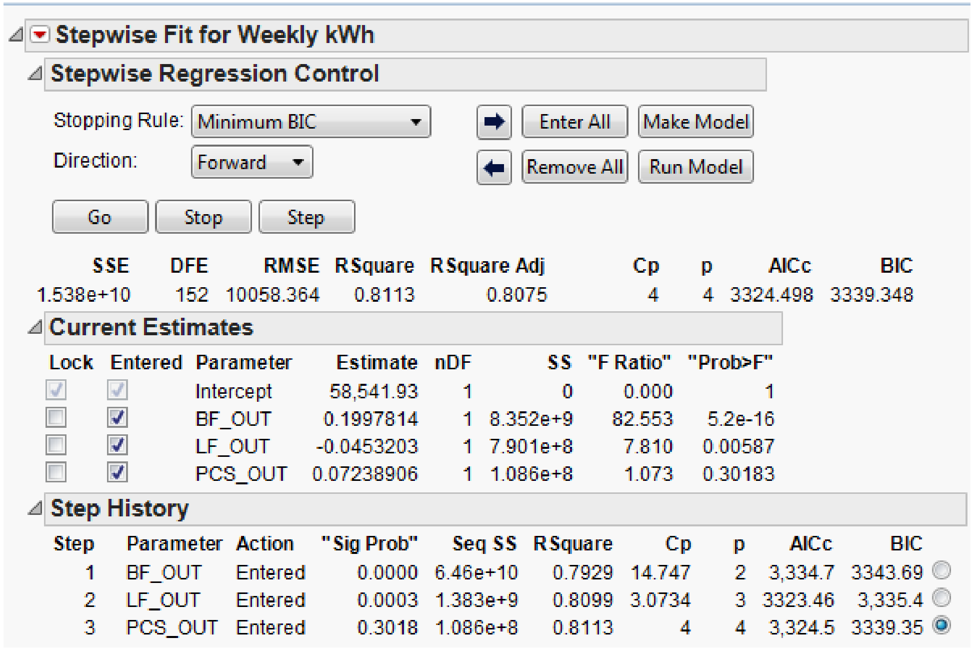
RSquare	0.871156
RSquare Adj	0.863214
Root Mean Square Error	8479.522
Mean of Response	116756.4
Observations (or Sum Wgts)	156

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	57620.897	2190.133	26.31	<.0001*	.
BF_IN	-0.134906	0.13044	-1.03	0.3027	1138.6942
BF_OUT	0.1721604	0.133898	1.29	0.2006	1,086.994
LF_IN	-0.005031	0.015706	-0.32	0.7492	181.22169
LF_OUT	-0.023185	0.022647	-1.02	0.3076	214.35951
PCS_IN	0.05078	0.071873	0.71	0.4810	134.10303
PCS_OUT	-0.02483	0.099287	-0.25	0.8029	175.81638
SHIFT_HRS	34.450341	40.78616	0.84	0.3997	56.881032
RUN_HRS	206.45437	55.13963	3.74	0.0003*	52.702841
MAN_HRS	7.1132677	7.78307	0.91	0.3623	22.069063

Drawing any conclusions from any one of these parameters in isolation would be erroneous because of multicollinearity. For instance, it appears from this regression that there is a negative relationship between linear feet in and electricity use. This is obviously not true (as can be seen in the scatter-plot matrix). What the regression actually is saying is that when linear feet is high *compared to what would be expected for a given combination of the other variables*, electricity use will be low *compared to what would be otherwise be expected for that combination.*

Because of the multicollinearity issue, the choice of variables was performed using a combination of judgment from the modeler and stepwise regression. First, all the “in” variables were rejected from the analysis, since manufacturing output is typically a more direct driver of energy consumption. Next, the “hours” variables were rejected because of the logistical implications of using employee hours in an energy model. Energy models should be tied to physical production whenever possible, so that process improvements that may use fewer or consistent employee hours but produce more would show up as energy savings. Finally, stepwise was used to determine the “most correlated” production output variable. From stepwise, it was apparent that adding additional production variables after the first really didn’t help the regression substantially, as seen below:



Stepwise Fit for Weekly kWh

Stepwise Regression Control

Stopping Rule: Minimum BIC

Direction: Forward

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
1.538e+10	152	10058.364	0.8113	0.8075	4	4	3324.498	3339.348

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	58,541.93	1	0	0.000	1
☐	☑	BF_OUT	0.1997814	1	8.352e+9	82.553	5.2e-16
☐	☑	LF_OUT	-0.0453203	1	7.901e+8	7.810	0.00587
☐	☑	PCS_OUT	0.07238906	1	1.086e+8	1.073	0.30183

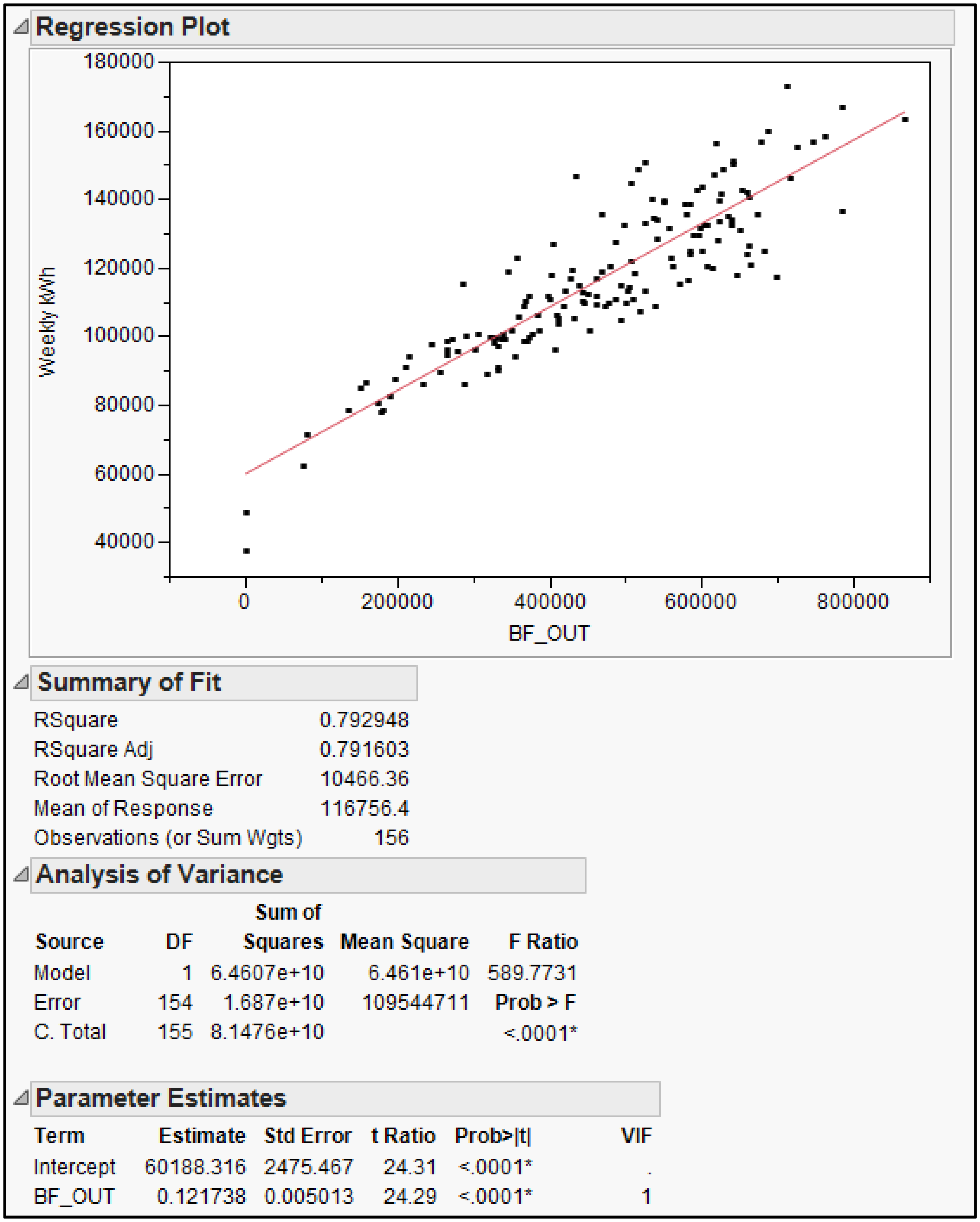
Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	BF_OUT	Entered	0.0000	6.46e+10	0.7929	14.747	2	3,334.7	3343.69	○
2	LF_OUT	Entered	0.0003	1.383e+9	0.8099	3.0734	3	3323.46	3,335.4	○
3	PCS_OUT	Entered	0.3018	1.086e+8	0.8113	4	4	3,324.5	3339.35	●

Figure 8. Stepwise Regression Output

Besides not explaining much additional variation, the addition of more variables (LF_Out and PCS_Out) after the first (BF_Out) creates counter-intuitive parameter estimates. While LF_Out was statistically significant (p=0.0003), on average the inclusion of this variable won't appreciably improve the accuracy of the predicted values from the model. The exception would arise if the model were consistently applied at the high or low end of the observed range of LF_Out, in which case the additional resolution provided by the inclusion of the variable might be important.

The modeling process started with nine possible production variables. Because of multicollinearity, a model specification that includes all the potential predictor variables results in individual coefficients that have poor p-values and don't make a lot of intuitive sense. A more meaningful model can be obtained by using stepwise regression, combined with knowledge of the process, to identify an initial set of primary energy drivers, and selectively adding variables to see how much the addition of other variables helps (or doesn't help) improve the model's fitness.



Summary of Fit

RSquare	0.792948
RSquare Adj	0.791603
Root Mean Square Error	10466.36
Mean of Response	116756.4
Observations (or Sum Wgts)	156

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6.4607e+10	6.461e+10	589.7731
Error	154	1.687e+10	109544711	Prob > F
C. Total	155	8.1476e+10		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	60188.316	2475.467	24.31	<.0001*	.
BF_OUT	0.121738	0.005013	24.29	<.0001*	1

Figure 9. Regression Output for Single-Variable Model

The final model outlined above could be further studied to see if weather variables or other non-correlated production variable could improve the amount of variation explained by the model.

Summary and Conclusions

Multicollinearity is a problem that often confronts program implementers in their efforts to select the appropriate predictor variables for regression-based energy models. While the presence of multicollinear variables has limited influence on the predictive capability of the final model, SEM practitioners should be mindful of the implications of multicollinearity on the ability to draw conclusions based on a cursory interpretation of the model coefficients.

The examples in the paper present methods of diagnosing and addressing multicollinearity. Diagnostics can include a simple matrix of correlation coefficients, but the variance inflation factor provides the most reliable method of testing for the presence of multicollinearity. In the context of top-down, regression-based energy modeling, the common solution involves dropping one of the offending variables from the regression analysis. However, this approach may induce omitted variable bias, which may negatively affect the predictive capability of the model. In practice, the model developer must interpret these statistical indicators with an informed understanding of how the process uses energy, while considering factors such as data availability and model complexity.

References

ASHRAE. 2002. *ASHRAE Guideline 14-2002 For the Measurement of Energy and Demand Savings.* American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. Atlanta: Ga.

Efficiency Evaluation Organization (EVO). *International Performance Measurement and Verification Protocol*. 2012. 10000-1.2012. www.evo-world.org

ESI Energy Performance Tracking (EPT) Team. 2012. *Monitoring, Targeting and Reporting (MT&R) Reference Guide – Revision 3.1.* http://www.bpa.gov/energy/n/pdf/MTR_Reference_Guide-Rev3_1.pdf. Portland, Ore: Bonneville Power Administration.

Neter, John and William Wasserman. 1974. *Applied Linear Statistical Models*. Pp 249-254, 339-347.

[SEP] Superior Energy Performance. 2012. *Superior Energy Performance Measurement and Verification Protocol for Industry*. http://www.superiorenergyperformance.net/pdfs/SEP_MV_Protocol.pdf. Oakland, Calif.: The Regents of the University of California.

Contributing Authors

Todd Amundson, Bonneville Power Administration
Mimi Goldberg, KEMA
Stephen King, Triple Point Energy
Aimie McKane, Lawrence Berkeley National Laboratory
Steve Martin, Energy Smart Industrial and Cascade Energy
Keri Macklin, Triple Point Energy
Mark Thompson, Northwest Energy Efficiency Alliance and Forefront Economics Inc.
John Thornton, Northwest Food Processors Association
Julia Vetromile, KEMA