Strategic Energy Management Program & Savings Review

National Grid Rhode Island

December 29, 2020

Table of contents

List of tables

1 EXECUTIVE SUMMARY

National Grid Rhode Island contracted DNV GL to review and assess the methodology and calculations for estimating electric energy savings from the program year 2019 (PY2019) industrial strategic energy management (SEM) demonstration initiative administered by Cascade Energy[1]. This report presents DNV GL's findings and recommendations based on our review of the measurement and verification (M&V) methods used to estimate electric savings at the seven non-wastewater treatment sites participating in the SEM demonstration. This report does not provide an independent evaluation of the savings estimated.

Continuous improvement programs such as SEM typically determine claimed savings based on analysis of metered whole-facility consumption data, using a regression model from prior to SEM initiation to predict what consumption would have been in a later, reporting period, absent the SEM. This SEM forecast modeling approach was used for all seven SEM projects. Developing SEM forecast models and savings requires identifying the drivers of consumption for each site, as well as identifying "non-routine events" (NRE) that change consumption in either direction and are not part of the SEM savings to be measured. This review assessed the model selection and treatment of NRE to inform savings risks carried by National Grid.

National Grid's 2019 Plan for Commercial and Industrial (C&I) Energy Efficiency Solutions and Programs (2019 Plan) defined SEM as "a set of processes for business energy management" whose main goals are to:

- Activate industrial and manufacturing customers, through a multiplicity of interventions including individual and group coaching
- Address O&M measures in the short term
- Pursue capital measures in the medium term
- Establish a culture of continuous improvement in its energy performance over a longer-term period[2]

1.1 Approach

DNV GL used the following methods to assess the program design and savings estimates:

File Review

- DNV GL engineering staff reviewed site documentation to assess evaluability, identify areas of savings risk, and identify opportunities to improve documentation.

Baseline Calculation Review

- DNV GL compared program regression methodology to industry best practice.
- DNV GL completed a statistical review of the current electric baselines used by the program to verify that models used can be replicated using the data provided.

1 Cascade Energy: https://cascadeenergy.com/company/
2 National Grid's 2019 Plan for Commercial and Industrial (C&I) Energy Efficiency Solutions and Programs: http://rieermc.ri.gov/wp-content/uploads/2018/09/2019-eepp-attachment-2-commercial-programs-final-draft.pdf#page=62&zoom=100,116,604

1.2 Key Findings and Recommendations

This SEM program documentation and M&V methodology review produced the following key findings:

- Overall, the program is well documented and follows industry best practice for measuring achieved savings during the measurement period. DNV GL did not identify any significant gaps in the documentation and was able to re-create the baseline regression models and associated savings estimates using the data provided in program documentation. The provided data covered just shy of 12 months of pre-period data for five sites and 8 months for two. All sites had data for 8 post-period months. These period durations were sufficient for assessing the validity of the savings methods.
- The National Grid demonstration program measurement methodology follows Bonneville Power Administration's Monitoring, Targeting and Reporting (MT&R) Reference Guide.[3] Cascade Energy was one of the manual's authors. The manual aligns with industry best practice for site level measurement of energy savings based on utility meter data for SEM type programs. The manual was originally written for BPA's Energy Smart Industrial SEM program.[4] An independent evaluation of this program was completed in February 2017.[5]
- The site reports and methods are consistent with the BPA SEM M&V guidelines, which are consistent with the International Performance Measurement and Verification Protocol (IPMVP) [6] and ASHRAE 14[7]. Both are internationally recognized protocols for the measurement and verification of energy savings. The BPA Manual adheres to widely accepted principles of M&V and produces verifiable energy savings calculations. Therefore, the methodology used is consistent with current industry best practices.
- National Grid is appropriately managing savings risk through its use of best practice methods to estimate savings and the documentation available to support savings claims. DNV GL has identified additional opportunities to further reduce program savings risk.
- Current program documentation is sufficient to support future evaluation, however the improvements discussed in this report are recommended by DNV GL. One process improvement identified is to improve the identification and tracking of non-routine events and associated adjustments, including measures rebated through other National Grid energy efficiency programs.
- DNV GL has identified several areas which could be improved upon and where the BPA M&V guidelines and related industry standards are less prescriptive. Most of these are areas that the industry in general has begun to grapple with in recent years with the increase in use of daily and hourly models, and the general expansion of SEM programs. Best practices are still in flux for these following areas of suggested improvement:
  - A standardized approach for reporting and addressing cases of possible multicollinearity
  - A standard approach of correcting for autocorrelation
  - Potential improved calculation of savings uncertainty (standard error or confidence bounds)
  - Additional guidance on non-routine event identification and adjustment methodology, together with more thorough documentation and explanation for such adjustments
  - Model acceptance guidelines when using data at weekly, daily, or finer time intervals.

3 BPA guide available here: https://www.bpa.gov/EE/Policy/IManual/Pages/IM-Document-Library.aspx
4 BPA Energy Smart Industrial SEM: https://www.bpa.gov/EE/Sectors/Industrial/Pages/Strategic-Energy-Management.aspx
5 SBW and Cadmus Group. 2017. "Industrial Strategic Energy Management (SEM) Impact Evaluation Report". Bonneville Power Administration: https://www.bpa.gov/EE/Utility/research-archive/Documents/Evaluation/170222_BPA_Industrial_SEM_Impact_Evaluation_Report.pdf
6 https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp
7 Measurement of Energy, Demand, and Water Savings, ASHRAE Guideline 14-2014, www.ashrae.org, 2014.

2 INTRODUCTION

National Grid's 2019 Plan for Commercial and Industrial (C&I) Energy Efficiency Solutions and Programs (2019 Plan) included the Strategic Energy Management (SEM) Demonstration Initiative.[8] The 2019 Plan defined SEM as follows:

> *Strategic energy management (SEM) is a set of processes for business energy management. The main goal of SEM is to activate industrial and manufacturing customers, through a multiplicity of interventions including individual and group coaching, to address O&M measures in the short term, pursue capital measures in the medium term and establish a culture of continuous improvement in its energy performance over a longer-term period. The energy benefits of SEM include reduced energy consumption through improved energy efficiency and energy conservation, improved demand management and the potential for reduced demand charges, decreased overall energy cost, and reduced greenhouse gas (GHG) emissions.*

2.1 Key Research Questions

National Grid contracted DNV GL to review all relevant program documents and available individual project data with the objective of developing a thorough understanding of the program approach, operations, and energy savings. National Grid requested responses to the following five key research questions specific to savings estimation and evaluability:

1. Can National Grid expect the proposed electric savings calculation methods and algorithms to provide defensible estimates of the electric energy savings acquired by the program?
2. Does the methodology or algorithm align or deviate from current SEM savings estimation best practice?
3. What significant risks to electric savings is the program carrying using the proposed methodology?
4. Does the program include documentation sufficient to support future energy savings measurements and independent evaluation?
5. What adjustments to the methodology and algorithm does the contractor recommend that will likely improve the accuracy of the savings estimation and mitigate savings risk?

In addition, National Grid asked DNV GL to provide responses to the following program comparison questions:

- How does National Grid's SEM program incentive structure compare to SEM programs serving similar customers in the Pacific Northwest?
- What key challenges exist for Pacific Northwest SEM program administrators (PAs) specific to offering multiple programs to a single customer?
- How do these SEM program administrators design their programs to manage these challenges?

This report first presents DNV GL's review of the program's savings estimation methods and associated documentation. The report then provides answers to the requested program comparison questions. The report concludes by answering the five key research questions and providing recommendations for National Grid's consideration. The appendix provides our site-specific review findings.

8 National Grid's 2019 Plan for Commercial and Industrial (C&I) Energy Efficiency Solutions and Programs: http://rieermc.ri.gov/wp-content/uploads/2018/09/2019-eepp-attachment-2-commercial-programs-final-draft.pdf#page=62&zoom=100,116,604

3 GENERAL FINDINGS ON SAVINGS METHODOLOGIES

3.1 Overview of M&V Process

Although not explicitly stated in the Energy Savings Reports and Model Reports developed as part of National Grid's Strategic Energy Management (SEM) program, it is clear that the program vendor is following Bonneville Power Administration's Monitoring, Targeting and Reporting (MT&R) Reference Guide (BPA M&V guidelines).[9] The BPA M&V guideline is a based on a regression model of whole-facility consumption that is generally consistent with the International Performance Measurement and Verification Protocol (IPMVP) for Option C (whole building).[10] The BPA guidelines have been developed over many years with engagement of a broad set of stakeholders, and are consistent with industry best practices for SEM savings quantification. This study focused on electric savings methods and did not include gas.

3.1.1 Measurement Boundary

The initial step in the BPA M&V process is to identify the measurement boundary. The measurement boundary is important because a full accounting must be made of all energy that crosses this border to accurately quantify the energy savings. The measurement boundary defines what is in scope for analyzing improvement in energy performance. Particularly for manufacturing facilities, there may be multiple meters, processes, and affected systems. The goal is to ensure that all systems potentially affected by the energy efficiency improvements are accounted for, either directly in the consumption data analyzed or if necessary, in a side calculation for interactive effects. In addition, it's important to avoid quantifying savings in one part of the facility that's achieved by shifting some function to another part outside the boundary.

To comply with the BPA M&V process, the implementer must identify the utility meters or submeters that exist within the measurement boundary. The implementer must also review available data and determine if a single model for energy use will suffice or if multiple models would have been more appropriate.

DNV GL found the National Grid implementer defined the electric measurement boundary for all seven sites reviewed. In all cases, the boundary was a single participant site. In multiple cases, a single participant site includes multiple electric meters. For all sites with multiple meters, the meters were combined to develop a single measurement model for the site.

3.1.2 Energy Driver Identification

To be compliant with BPA methods, the implementer also needs to identify all energy drivers in the measurement boundary. At the facilities reviewed, the primary energy driver is production, so an understanding of the production processes is also necessary to construct reliable models of energy consumption. The implementer also needs to identify all other energy drivers for potential inclusion in the measurement model. Examples of such "other" energy drivers are ambient weather conditions, raw material properties, and day of the week (weekend/weekday). These variables are often important drivers in any model developed for energy consumption. Table 3-1 shows the count of final independent variables used across all seven sites, including the number of independent production variables used in the regression in the bottom third of the table.

9 BPA guide available here: https://www.bpa.gov/EE/Policy/IManual/Pages/IM-Document-Library.aspx
10 International Performance Measurement and Verification Protocol. Efficiency Evaluation Organization. 10000-1.2012. www.evo-world.org

Table 3-1. Count of independent variables used across final models for seven sites

Production Schedule Indicator Variables	Number of Sites
Plant Shut Down Indicator	1
Sunday or Holiday	1
Non-Holiday Saturday	2
Weekend	1
Weekend or Holiday	1
Production Yes/No	5
Weather Variables	
CDD_db	5
HDD_db	5
CDD_wb	1
Dry Bulb Temperature [°F]	1
Production Volume, Number of variables	
1 x Production Variable	3
2 x Production Variables	1
3 x Production Variables	1
4 x Production Variables	1
7 x Production Variables	1

The National Grid implementer defined the identified energy drivers in the Model Report for each site. In general, the drivers identified are production schedule, production volume, and weather.

3.1.3 Baseline Data and Hypothesis Model

The BPA M&V guidelines list the next step of the M&V process as the determination of the baseline period. Specifically, the baseline period should encompass two or more cycles and include a wide range of the hypothesized primary and secondary energy drivers. The end of the baseline period should be as close as possible to the start of the reporting/intervention starting date. The guidelines also provide guidance on the minimum number of baseline data points that should be used, stating that the minimum number of baseline data points is 6 times the number of coefficients. This is to avoid problems with model over-fit and the resulting model performance deterioration. A summary of the baseline data used in this study is provided in Table 3-2. This amount of data was sufficient for the study purpose of examining savings methods and risks.

Table 3-2. Summary of baseline observations

Site	Observations	
	Frequency	**#**
A	Daily	239
B	Daily	330
C	Daily	353
D	Daily	231
E	Weekly	51
F	Daily	354
G	Daily	361

The BPA M&V guidelines also recommend a visual inspection of the data after it is collected. This visual inspection can help identify any areas of potential outliers and/or anomalous data points which may would require further investigation. The guidelines also provide methods to avoid double counting of savings of incentivized or non-incentivized energy projects that the facility may have participated. When examining the data, the BPA guidelines repeatedly underscore the importance of working closely with the facility energy manager or other personal who can speak to the actual operation of the plant and who can elucidate any physical reasons behind data anomalies.

The National Grid implementer appears to default to a 12-month baseline period but shortened many baseline periods due to identified non-routine events that occurred such as completed capital projects. The baseline period length is typically a subjective decision made by the modeler. The modeler assesses the benefits and risks then selects the baseline period to use. Although use of less than 12 months of data in the calculation of weather sensitive measures carries additional statistical bias risk, the decision and reasoning for it are documented in the Model Report.

The Model Report also documents the data inspection completed by the implementer. The implementer's process first flags outlier residuals during the baseline period, then reviews why the outlier occurred, and then assesses if adjustments to the data or model are necessary. This process occurs during model development. In many cases, this process identified periods with missing electric meter data. In other cases, the implementer excluded periods of data from the baseline model due to data issues. If no reason for the outlier could be identified the data was kept in the model. DNV GL supports the use of the residual outlier analysis for model review. The process both helps identify potential data issues but can also show periods when the baseline model isn't estimating actual consumption well.

3.1.4 Modeling

Constructing a hypothesis model is the next step in the BPA M&V process. This model is informed by the visual inspection of the facility's data, knowledge of the workings of the facility, and knowledge of the variables that drive energy usage. Eight months of post-period data were available for modeling the seven sites. The guidelines specifically warn about the possibility of missing variables and non-linear relationships. The guidelines also mention the need to check for correlations between the regressor variables. The resulting multicollinearity is cited as a potential issue and the reader is referred to the NW Industrial Strategic Energy Management (SEM) Collaborative[11].

A regression-based baseline model is then created. It was assumed based on the BPA guidelines that a multivariate, ordinary least squares (OLS) regression estimation technique is used to create the baseline model. The guidelines first suggest examining the statistical significance of the coefficients on the independent variables and using a combination of a t-statistic value of greater than or equal to 2.0 and a p-value of less 0.10 to determine its worthiness of inclusion. The guidelines also recommend using the IPMVP threshold recommendation of a R-squared greater than 0.75 to determine if the baseline model fits the data well. The guidelines make several recommendations about using the Durbin-Watson test for autocorrelation but provide no further guidance on how to proceed if autocorrelation is detected. Table 3-3 later in this section shows the regression information for the seven sites reviewed compared to the suggested range from the BPA guidelines.

11Tools and Methods for Addressing Multicollinearity in Energy Modeling. NW Industrial Strategic Energy Management (SEM) Collaborative. 2013.

Table 3-3. Site baseline model information compared to guideline suggested range

Regression Information	Suggested Range	Site A	Site B	Site C	Site D	Site E	Site F	Site G
Data Frequency	NA	Daily	Daily	Daily	Daily	Weekly	Daily	Daily
Number of Observations =	>= 54	239	330	353	231	51	354	361
R-squared =	> 75%	76.9%	84.3%	96.0%	80.7%	66.9%	90.5%	98.0%
Adjusted R-squared =	> 75%	76.1%	84.1%	95.9%	80.3%	62.4%	90.4%	97.9%
Standard Error =	NA	4,028.0	1,996.1	628.3	2,138.5	8,058.0	3,883.8	2,831.1
Coeff. of Variation (RMSE) =	< 20%	6.3%	8.3%	5.3%	17.2%	5.9%	11.0%	6.1%
F Statistic =	NA	95.5	434.9	1,665.3	188.7	14.8	833.3	1,529.6
Autocorrelation Coefficient =	NA	25.6%	15.5%	47.8%	14.3%	44.5%	17.9%	16.0%
Fractional Savings Uncertainty (68% Conf.,.6% sav.,241 smpls.) =	< 50%	112.8%	33.1%	-840.2%	44.0%	-37.0%	19.6%	146.5%
Net Determination Bias =	< 0.00005	-6.57E-14	4.26E-16	5.52E-15	1.70E-15	2.11E-15	3.66E-15	-1.10E-15
Overall P-Value =	< 0.1	3.90E-89	4.70E-203	0.00E+00	1.04E-117	4.25E-13	3.11E-261	0.00E+00

3.2 Review of Rhode Island Vendor Reports

3.2.1 Review Summary

This review of methods used to establish the baseline began with reading and going through all the documentation and material provided for each site. This included a site report detailing the vendor's work and an excel file showing the data used to construct the predicted energy usage model along with actual energy usage.

Using the BPA M&V guidelines, the vendor reports were examined. These key points summarize the reports and models used for 2019.

- The measurement boundary was typically the whole facility and was clearly described in each report.
- A description of the energy drivers was also provided, along with the description of the sources of any data that came from outside of the facility (i.e., weather data).
- The beginning and end dates of the baseline period were included along with any dates that were excluded from the baseline period. If any dates were excluded, an explanation for the exclusion was provided, however no information on efforts to diagnose and mitigate the data issues that led to the need to exclude the data was provided, which might be important evidence for illustrating the rigor taken as part of the savings process. The baseline period was typically for at least one full calendar year; if a full year was decided to be unnecessary, or not available, the reasoning why was provided.
- The treatment and modifications of the data was also documented. The most common modification was related to non-routine adjustments that resulted from incentivized energy programs, both in the baseline and measurement periods. While these adjustments represent best practice, no further information was provided on the source of the adjustment, rationale for application, and details needed to ensure it was made correctly (e.g., lifetime factors, timing of program participation).
- A detailed description of the baseline model and its associated statistics was also provided. Using the data provided, the baseline model was independently replicated by DNV GL and compared to the reported model. For each site, the DNV GL analysis exactly replicated the vendor's model using the data provided and the same independent variables.

Vendor Statistical Summary Information

The vendor provided an autocorrelation coefficient with each regression that was performed in the savings reports. The savings report provided no additional information on the presence of autocorrelation or its effect on the baseline model. DNV GL considered the likelihood that the associated error terms of the models deviated from the assumed "independent and identically distributed" (i.i.d.) structure was high. In each replication of the models performed by DNV GL a Durbin-Watson test was performed, in many of the cases the detected autocorrelation was significant enough to warrant further consideration under the BPA M&V guidelines. The vendor appeared aware of these potential issues and made a note of this on several occasions, but no statistical test was performed, and no additional action was taken.

The vendor also made note of the potential for multicollinearity in several occasions. The vendor typically detected the possibility of multicollinearity by looking at the correlation coefficients between the independent variables. The vendor did not state if any diagnostic test was used to determine if or to what extent the multicollinearity existed. In order to more closely examine this potential issue, DNV GL looked at the variance inflation factor (VIF) for each model. This factor can be used to help determine if the variance for each coefficient estimate was impacted by multicollinearity and if corrective measures should be taken. Only one model was observed where the VIF was greater than 5 (a threshold value that is sometimes used to determine if additional action is needed). In this instance, corrective action appears to be unwarranted. While the vendor appears aware of the potential for problems with this issue there is no indication that any further diagnostics were considered or how the determination that no multicollinearity problem existed was made.

Finally, the limitations of this review also need to be considered. Because only the raw data was provided, replicating the results of the vendor's regression was a relatively trivial task. Judgements on the inclusion of variables, statistical test, and statistical diagnostics used have been made, but this review cannot speak to the methods used to determine the exclusion of data points or possible energy drivers. Because these decisions were made in the processing steps between the raw and the final modeling data, they were outside of the scope of this review. The statements provided about the exclusion of data points are consistent with the BPA M&V guidelines and, if taken at face value, would be appropriate exclusions. With additional documentation these exclusions could be easily verified. However, the suite of energy driving variables that were considered by the vendor was not provided. Therefore, this analysis was limited to considering model over-fit (the inclusion of variables that were not statistically significant) but is mute on model under-fit. For a site by site review please see Appendix A.

3.2.2 Findings from Savings Review

Overall the site reports and methods are consistent with the BPA M&V guidelines, which are consistent with the International Performance Measurement and Verification Protocol (IPMVP) [13] and ASHRAE 14[14]. Both are internationally recognized protocols for the measurement and verification of energy savings using whole-facility consumption analysis. The BPA Manual adheres to widely accepted principles of M&V and using it can produce verifiable energy savings calculations. Therefore, the methodology used is consistent with current industry best practices.

DNV GL has identified several areas which could be improved upon and where the BPA M&V guidelines and related industry standards are less prescriptive. Most of these are areas that the industry in general has begun to grapple with in recent years with the increase in use of daily and hourly models, and the general expansion of SEM programs. The BPA Manual refers to some guidance documents available on the NW Industrial Strategic Energy Management (SEM) hub[15]. However, best practices are still in flux for these areas which are listed below:

1. A standardized approach for reporting and addressing cases of possible multicollinearity
2. A standard approach of correcting for autocorrelation
3. Potential improved calculation of savings uncertainty (standard error or confidence bounds)
4. Additional guidance on non-routine adjustments, together with more through documentation and explanation for such adjustments
5. Model acceptance guidelines when using data at weekly, daily, or finer time intervals.

13 https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp
14 Measurement of Energy, Demand, and Water Savings, ASHRAE Guideline 14-2014, www.ashrae.org, 2014.
15 https://semhub.com/resources

Multicollinearity

In several instances the vendor reported that multicollinearity was a possible issue, though no further detail or comment was provided. It was left unclear if there were any issues, if any corrective action was taken, or further insight into the vendors decision making process. The BPA manual states (p. 10) that "when multicollinearity is present, the modeler should clearly explain the rationale for both the inclusion and exclusion of variables in the energy model." While there is some explanation of the included variables for each model, there is no explanation of inclusion or exclusion of variables in relation to observed multi-collinearity.

The manual also states, "The presence of collinear variables can affect the precision of individual coefficients and can understate the statistical significance of individual predicator [sic] variables." This statement may be misleading. Multi-collinearity is a fact of life. When two or more drivers that affect energy consumption are correlated with one another, the ability to separate their effects statistically is reduced. When the collinear variables are included in a model together, the statistical significance of each coefficient is weaker than if each were included alone. On the other hand, excluding other drivers leads to omitted variable bias of two kinds. First, the estimated coefficient of the included variable incorporates effects of the omitted correlated variables and does not reflect only the direct effects of the included variable. Second, predicted energy savings based on the model will not reflect any incremental effects of the omitted variables beyond the portion correlated with the included variable.

There is no one right way to address multi-collinearity in model development. In the context of savings estimation, if the predicted savings have small standard errors regardless of the presence of multi-collinearity, and the predicted savings are at conditions well within the combination of values used to fit the model, the savings estimates can be considered reliable. On the other hand, if the multi-collinearity is sufficient to make the savings estimates poorly determined, it may be better to eliminate one or more of the correlated variables. However, omitting variables that are physically expected to be important drivers of consumption is an added source of uncertainty in the savings estimate. This uncertainty is greater if the reporting period conditions are different from the baseline period used to fit the model.

For example, suppose that production level and moisture content of a feedstock are both drivers of energy consumption, and that in the baseline period these are closely correlated as a result of seasonal patterns affecting both. If it's not possible to separate their effects but a good fit can be obtained with one, and they always move together in both the baseline and reporting period, it may be acceptable to drop one from the model. On the other hand, if they are less correlated in the reporting period than in the baseline, the accuracy of the prediction may suffer. It's also possible than an alternative model specification could allow both effects to be included in some form.

The BPA Manual refers to guidance in a paper on "Tools and Methods for Addressing Multi-Collinearity" published by the NW Industrial Strategic Energy Management (SEM) Collaborative.[16]

DNV GL used a diagnostic statistical variable called the Variance Inflation Factor (VIF) to examine the effect any multicollinearity may have had on the variance estimates for each coefficient. The VIF statistic is not the only method of detecting multicollinearity and a number of other methods would be appropriate. In several of the reports the vendor raised the specter of multicollinearity but never provided their reasoning as to the conclusion that any correlation between independent variables was benign. A stronger defense of the

16 Available here: https://conduitnw.org/Handlers/conduit/FileHandler.ashx?rid=1762

savings calculations could be made if the vendor provided additional detail that describes their reasoning. An example of such reasoning could look like the following:

"In DNV GL's replication of the results for the location at Site 4, a VIF of 5.65 was recorded for the independent variable "Knitting_Yards". A threshold of 5 is frequently used for the VIF to determine if further investigation is warranted. The t and p-values for the associated parameter estimate on "Kitting_Yards" indicate that the coefficient is strongly statistically significant. The estimated savings using the model has an acceptable standard error. The correlation observed between this variable and the others is similar in the reporting period to that in the baseline period. Given the significances of the variable, the importance of this factor in the production process, the well-determined savings even with the multi-collinearity, and the similarity of the correlations between the reporting and baseline periods, it is reasonable to conclude that multicollinearity had minimal effects on the inference related to this variable. Therefore, in this case the effects of any multicollinearity did not appear to warrant further model modifications."

As stated above it is unclear from the vendor report the reasoning surrounding the possible cases of multicollinearity. If any action was taken, then providing supporting documentation and explanation would be appropriate. If not, then providing additional detail to justify inaction could give stakeholders greater confidence in the models used for savings calculations.

Autocorrelation

A data collection frequency of once per day was typically used to form the data set from which the baseline models were derived. Given the nature of energy consumption and this rate of data collection, it is common to expect a certain degree of autocorrelation in the data. Simply put, autocorrelation is the tendency of observations in a series to be more similar to those which occur temporally closer versus those that are further away. The vendor appears to be aware of such issues and presented the autocorrelation coefficient in the report. However, no statistical tests were performed, and no comments were provided. At issue here might be that the BPA M&V guidelines provide little guidance on any model corrections if the presence of autocorrelation is detected. Thus, the vendor is conducting and reporting on calculations suggested in the Manual, but any model adjustment in response to these results is not documented.

The BPA Manual recommends a Durbin-Watson test and residuals plots to identify autocorrelation and suggests that "high autocorrelation may indicate the omission of a key variable, or the occurrence of an event that changed energy consumption characteristics during the baseline." The Manual also refers to the same paper on *Tools and Methods for Addressing Autocorrelation in Energy Modeling,* produced by the NW Industrial Strategic Energy Management (SEM) Collaborative.[17] This document provides further examples of these same points.

It is true that autocorrelation can be the result of an omitted variable or break point, and these effects are worth investigation. It is also true that with daily or even weekly data, whatever disturbances affect consumption up or down in one period of observation are likely to persist across multiple observations. This is the (typically unobserved) physical source of the statistical phenomenon of serial correlation. These effects can include, for example, particular workers being on or off the job, machines drifting out of calibration for a while before being corrected, or unusual weather effects not accounted for in the models. It may not be possible or practical to identify specific variables to account for these disturbances.

17 Available here: https://conduitnw.org/Handlers/conduit/FileHandler.ashx?rid=1762

From a technical perspective, autocorrelation does not bias the point estimates for the coefficients produced by OLS. The estimate with autocorrelation correction will not match the simple Ordinary Least Squares estimate, but both can be considered unbiased estimates of the same quantity. However, autocorrelation does bias the error estimates. Thus, in the presence of autocorrelation, results that appear to be statistically significant when autocorrelation is ignored might in fact be much less strongly significant when the autocorrelation is taken into account. Hence, if autocorrelation is observed, a recommended practice would be to explore the possibility that autocorrelation has affected the predicted energy usage as well as to apply corrective measures to avoid inferring that a savings estimate is strongly determined when in fact it may not be. In the savings calculator workbooks, the auto-correlation factor is incorporated into the Fractional Savings Uncertainty (FSU) calculation, to avoid understating the uncertainty in the presence of auto-correlation. However, as noted below, the accuracy of the FSU formula for the types of models reviewed here is itself uncertain.

DNV GL used a Durbin-Watson test as part of the review of each project's modeled energy usage. This test looks specifically for autocorrelation in the first lagged observation, that is, the observation from the immediately prior time period (written as AR1). DNV GL also examined a partial autocorrelation plot to determine if correlations between more distant lags existed. Using these diagnostic tools, it was determined that autocorrelation was a common issue seen in the data provided.

To see if the presence of autocorrelation affected the estimated energy savings, DNV GL re-estimated the predicted energy usage, using the same model specified by the vendor but further specifying an AR1 error structure. Table 3-4 below compares the preliminary savings to the savings calculated using the predicted energy usage from the model with the AR1 error structure. Also included are the +/- error bands for the corresponding 80 percent confidence intervals. The table shows that for 5 of the 7 sites the estimates from the AR1 model were similar to those from the OLS model. However, for one site the (negative) savings changed by a factor close to 5, and for one the savings changed from positive to negative. For both these sites the savings was not statistically significant at the 80% confidence level, so it's not surprising that an adjustment to the specification substantially changes the result.

These re-estimated savings should not be viewed as definitive, in part due to availability of only 8 months of post period data. It is important to note that continuous improvement savings can show up slowly, turning apparent negative savings to statistically significant positive over time. This analysis was done to illustrate the effects that autocorrelation might have on the savings point estimate calculations.

Table 3-4. Preliminary savings estimates vs. savings with AR1 regression

	Ordinary Least Squares Performance Period Savings kWh		AR1 Regression Performance Period Savings kWh		Ratio AR1/OLS	
Site	Estimate	80% confidence interval +/-	Estimate	80% confidence interval +/-	Savings Estimate	Confidence Interval Width
A	97,725	117,526	-330,105	233,657	-3.4	2.0
B	134,310	40,032	146,086	48,944	1.1	1.2
C	-2,398	11,814	-11,728	20,976	4.9	1.8
D	102,196	53,293	132,997	70,677	1.3	1.3
E	-269,389	60,573	-324,250	119,830	1.2	2.0
F	501,433	70,749	487,635	88,175	1.0	1.2
G	41,257	59,040	38,164	68,499	0.9	1.2

There are various methods the vendor could explore when adjusting for autocorrelation. These include but are not limited to: specifying the error structure (for example, with the AR1 structure applied in the above analysis), modeling first differences, or using feasible general least squares (FGLS) estimating methods. There may also be cases were the vendor has a compelling reason to make no adjustments and use OLS. A thoughtful justification should be provided so that the program can appropriately defend the savings calculations to stakeholders.

Savings Uncertainty

The ASHRAE approximation used to calculate the Fractional Savings Uncertainty (FSU) incorporates an empirically determined factor, based on previously studied regressions using weather and other variables. How well this approximation and its empirical factor apply to regressions involving multiple production variables is unclear, and no guidance is offered by ASHRAE 14 or the BPA Manual for assessing that.

The full calculation of confidence interval bounds or standard error of the predicted annual usage depends on several factors:

- The regression coefficient of variation,
- the reporting-period and baseline-period sample sizes,
- the correlations among the predictors, and
- the difference in the average values of the predictors between the baseline and reporting periods.

The first two of the these are included in the approximation formula, but the last two are not. Thus, we would expect the approximation to be worse when one or both of these two factors is very different from those for the original cases used to develop the empirical approximation, but we don’t know what those conditions were.

ASHRAE describes the approximation as avoiding the need to perform matrix algebra calculations. As part of the line fitting, the internal Excel regression function already calculate all the elements that would be needed for the package to be able to produce the standard error or confidence interval for the predicted value. However, Excel does not appear to easily provide these values. For purposes of this review, DNV GL calculated the regressions and correct confidence intervals in SAS.

Table 3-5 compares the uncertainty of the estimated savings using the FSU approximation with the full standard error calculation. Both calculations are designed to express the standard error of the estimated savings as a percent of the savings estimate. Two forms of the full calculation are shown. The first is based on the Ordinary Least Squares regression used in the workbooks. The second is for the AR1 regressions summarized in Table 3-4.

DNV GL also checked the Fractional Savings Uncertainty calculations from the tool and determined that the calculation is consistent with the formula in the Manual. Two minor anomalies were identified:

1. The Statistical Summary in the workbook includes the item “Fractional Savings Uncertainty (68% Conf.,.6% sav.,241 smpls.).” The indicated .6% savings and 241 sample points are artefacts of a particular analysis and not applicable to other cases.
2. It appears that the t-statistic used to calculate the FSU is based on degrees of freedom ignoring the intercept term in the count of regression parameters. With large numbers of observations in the model, this makes no material difference to the results.

For the 68% confidence level reported, with the sample sizes and numbers of parameters for the projects studied, the t-statistic in the FSU calculation is very close to 1. This means that the FSU approximates the standard error of the estimated savings, expressed as a fraction of the estimate. The ratio of the standard error of estimated savings to the savings estimate itself is the *relative standard error* (RSE) of the savings estimate. Thus, we can assess the accuracy of the FSU approximation for the reviewed cases by comparing the calculated FSU approximation with the RSE based on the full OLS calculation. The comparison is shown in Table 3-5.

Table 3-5. Comparison of FSU via BPA Manual Approximation vs Relative Standard Error via full calculation

	OLS			AR1 AutoCorr Adjusted		
Site	**FSU**	**RSE = SE/Est**	**FSU/RSE**	**FSU**	**RSE = SE/Est**	**FSU/RSE**
A	1.12	0.94	1.2	-0.18	-0.55	0.3
B	0.33	0.23	1.4	0.26	0.26	1.0
C	-8.37	-3.85	2.2	-0.83	-1.4	0.6
D	0.44	0.41	1.1	0.29	0.42	0.7
E	-0.36	-0.17	2.1	-0.17	-0.28	0.6
F	0.2	0.11	1.8	0.17	0.14	1.2
G	1.46	1.12	1.3	1.29	1.4	0.9

For the OLS models, the calculated FSU is higher than the actual ratio of OLS standard error to OLS estimate, by up to a factor of 2. This is not surprising, because the calculated FSU includes an adjustment for auto-correlation, while the OLS standard error does not. Thus, if OLS models are used to estimate savings despite indications of auto-correlation, the FSU approximation including the correction for auto-correlation provides a more realistic assessment of the savings uncertainty than the OLS relative standard error ignoring the auto-correlation.

For the AR1 model, the calculated FSU tends to understate the actual relative standard error, but not for all the projects. The FSU approximation was not developed for AR1 models using daily data, so it's not surprising its performance is mixed for these models. If AR1 models or other approaches are used to address auto-correlation, the software tools that fit these models should be used to calculate correct relative standard errors for the resulting savings estimates, rather than relying on the FSU approximation.

Non-Routine Event/Adjustment Documentation

The introduction to the BPA manual states that "Specific focus is given to methodologies for … adjusting the baseline model for non-routine changes to plants or systems." However, the guidance does not refer to "non-routine" adjustments elsewhere. There are helpful charts in Manual Appendices A and B on how to adjust for unrelated capital projects in the middle of the baseline (pre) or reporting (post) period. There does not appear to be guidance on how to handle other types of non-routine events (NREs). As noted, limited guidance on treatment of non-routine events is recognized as a gap in industry practice guidelines overall.

In particular, if a consumption data point outlier is identified, calculated savings will be affected by whether that is assumed to be a data error, a freak event unlike anything likely to occur again, or a kind of occasional high or low day that occurs at some rate every year—possibly with no firm explanation. If it's bad data or truly exceptional, it can make sense to omit it from the baseline or reporting period. If its nature is such that it happens now and then, it needs to be accounted for in both.

One of the benefits of SEM should be the reduction in anomalous unexplained consumption and increasingly predictable consumption due to better process controls. If such anomalies are screened out of the analysis just because they're anomalies, we lose the ability to observe these expected improvements. While the effects on savings of the occasional anomalous day may be small in general, SEM is typically looking for improvements on the order of a couple percent per year. An excluded unplanned shut-down week or a couple days of uncontrolled operations could easily swing the measured savings by a meaningful amount on that scale.

The BPA Manual states that "The modeler must provide a supporting explanation when removing statistical outliers," but does not describe what such explanation should entail. We recommend that the justification specifically address whether this type of anomaly is or is not to be expected occasionally over the course of a year, and what that implies for savings.

In addition, we recommend more complete documentation of adjustments made for measures installed via other programs. In general, the SEM documentation provided little information on these adjustments. This was exemplified by the vendor stating that a non-routine adjustment was needed between two dates because the facility participated in an incentivized energy program, but no supporting documents or information on how that event was tracked by National Grid was provided. National Grid should prioritize a process improvement that results in the inclusion of measure tracking information (i.e. Project or Measure Number) within SEM program documentation. This link will help ensure that no savings at a single site are double counted.

DNV GL did not verify the other program activity and found no "red flags" in our review that would raise doubt as to the validity of these claims. However, the inclusion of such documentation could provide an additional level of assurance to all stakeholders. DNV GL also recommends that National Grid review its tracking data to confirm that both SEM savings and savings from other measures are being tracked under the same customer or site identification number.

Model Acceptance Guidelines

The BPA manual follows guidance from IPMVP and ASHRAE for statistical model fit, including a threshold R^2 value, the Net determination bias, and t-statistics for individual coefficients. Most of these guidelines are carry-overs from experience with monthly data models.

For R^2 in particular, daily or weekly data can be exhibited to be noisier, leading to lower R^2, yet still produce tighter savings estimates (even with autocorrelation appropriately accounted for). In general, R^2 is a less helpful indicator of the quality of model estimates than the standard errors or confidence intervals for the estimates of interest. Over time, we anticipate that the industry will develop guidelines more specific to analysis of data at daily or hourly time scales.

For a linear model, if the net determination bias is calculated for the same data used to fit the model, the net determination bias is zero, apart from machine calculation noise. The reported values were all quite small, and it's unlikely that this statistic is used to guide model selection or acceptance except as a cross-check.

4 PROGRAM COMPARISON

This section first compares National Grid's SEM program incentives and eligibility requirements to programs serving similar customers in the Pacific Northwest. The section then describes the key challenges that exist for the same Pacific Northwest SEM program administrators (PAs) and how they manage these challenges through their program design or delivery.

4.1 Incentive Comparison

Table 4-1 compares the incentive structures of for programs administered by Energy Trust of Oregon, Puget Sound Energy (PSE), and the Bonneville Power Administration (BPA) to the National Grid offering. The average retail cost of energy was taken from The Energy Information Administration (EIA). Incentive and eligibility information was taken from publicly available program documentation. Hyperlinks to the sources of information are provided in Table 4-2.

National Grid's SEM incentives, as a portion of energy cost, are notably lower than the others examined. Akin to many program decisions, these generally lower incentive rates may still be sufficient to draw customers into the initiative and achieve program savings goals.

Table 4-1: Program incentive structure comparison

Program Administrator	SEM eligibility hurdles	Avg. Cost of Energy[18]	SEM Incentive Rate
National Grid RI, Industrial SEM Program	None	$0.175 per kWh $1.00 per therm	$0.03 per kWh $0.30 per therm
Energy Trust of Oregon, Industrial SEM	$50,000 per year of energy spend	$0.070 per kWh $0.48 per therm	$0.04 per kWh $0.40 per therm
Puget Sound Energy, Industrial System Optimization (ISOP)		$0.093 per kWh $0.68 per therm	$0.05 per kWh $0.80 per therm
Puget Sound Energy, Industrial Strategic Energy Management (ISEM)	Buildings must collectively use greater than 3,000,000 kWh or 150,000 therms (or a combination of both) annually.	$0.093 per kWh $0.68 per therm	$0.02 per kWh $0.32 per therm $25,000 per year MAX
Bonneville Power Administration (BPA), Energy Smart Industrial, Strategic Energy Management		$0.051 per kWh BPA does not supply natural gas.	$0.025 per kWh, No documentation of cost If documentation of cost provided, then lesser of the following for years 1 and 2: • $0.075 per kWh of SEM Verified Savings • 70 percent of documented action-item costs

18 DNV GL used the following U.S. EIA tables as sources of the cost of energy for program participants. DNV GL used information on the "Industrial" sector (non-transport) across both electricity and gas. If a program administrator serves customers across multiple utilities, DNV GL calculated a weighted average cost using the data provided.
Electricity: EIA-861- schedules 4A & 4D and EIA-861, https://www.eia.gov/electricity/data.php
Gas: Annual company level data from Form EIA-176, https://www.eia.gov/naturalgas/data.php

Table 4-2: Comparison program and rate information

Program Administrator/ Source	Sources
Energy Trust of Oregon, Industrial SEM	Program Information: https://www.energytrust.org/wp-content/uploads/2020/04/IND_SEM_FS_2005.pdf Participant average cost of energy was calculated as a weighted average from EIA tables for IOU customers in Oregon as Energy Trust administers programs for IOU customers. Actual customer tariffs can be viewed at the following locations. • PGE All Tariffs: https://www.portlandgeneral.com/our-company/regulatory-documents/tariff OR https://www.portlandgeneral.com/-/media/public/documents/rate-schedules/all_tariffs48.pdf • Pacific Power Rates: https://www.pacificpower.net/about/rates-regulation/oregon-rates-tariffs.html • NW Natural Rates: https://www.nwnatural.com/uploadedFiles/OR_Summary_Monthly_2019.pdf
Puget Sound Energy, Industrial System Optimization	Program Information: https://www.pse.com/rebates/business-incentives/energy-management-programs/industrial-system-optimization-program Participant average cost of energy was taken from EIA tables. Actual customer tariffs can be viewed at the following locations. • Elec Rate Pamphlet: https://www.pse.com/-/media/Project/PSE/Portal/Rate-documents/summ_elec_7399_ncmp_res_comm_2020_05_01.pdf • Elec Rate Information: https://www.pse.com/-/media/Project/PSE/Portal/Rate-documents/summ_elec_prices_2020_07_03.pdf • Gas Rate Pamphlet: https://www.pse.com/-/media/Project/PSE/Portal/Rate-documents/summ_gas_7401_ncmp_res_comm_2020-05-01.pdf • Gas Rate Information: https://www.pse.com/-/media/Project/PSE/Portal/Rate-documents/summ_gas_prices_2020_05_01.pdf
Puget Sound Energy, Industrial Strategic Energy Management	Program Information: https://www.pse.com/rebates/business-incentives/energy-management-programs/industrial-strategic-energy-management Rate information shown in row above.
Bonneville Power Authority, Industrial	Program Information: https://www.bpa.gov/EE/Sectors/Industrial/Pages/Strategic-Energy-Management.aspx Program Manual: https://www.bpa.gov/EE/Policy/IManual/Documents/2020-2021_IM_Updated_3-20.pdf Umbrella Program Factsheet 1: https://www.bpa.gov/EE/Sectors/Industrial/Documents/ESI_Fact_Sheet_for_Utilities.pdf Umbrella Program Factsheet 2:https://www.bpa.gov/EE/Sectors/Industrial/Documents/ESI_FactSheet.pdf Participant average cost of energy was calculated as a weighted average from EIA tables for largest non-IOU utilities in Washington state. Actual customer tariffs vary based on location.

4.2 Multiple Program Offering Challenges

DNV GL received responses to National Grid's questions from all three program administrators: PSE, BPA, and Energy Trust. It is important to note that only PSE is both the customer's utility provider and energy efficiency program administrator. Energy Trust and BPA are only program administrators.

What key challenges exist for Pacific Northwest SEM program administrators specific to offering multiple programs to a single customer?

Program administrators responded with the following key challenges.

1. Avoiding customer confusion due to a new program offering that requires a different level of customer participation and functions very differently from traditional rebate offerings.
2. Ensuring that the customers who enroll in the program are good candidates for the program and will remain engaged throughout the program experience.
3. Ensuring appropriate accounting of program savings due to participation in other energy efficiency programs or other non-routine events that occur at a participant facility.

How do these SEM program administrators design their programs to manage these challenges?

While each program administrator designs their own portfolio and structures their programs differently, two central themes were provided in the responses.

1. Central account management. Each PA relies on account managers to understand a customer's current programming needs and determine if SEM is a good option for the customer.
    a. For BPA, an account manager is assigned to a local utility and its territory to develop an energy efficiency plan for the industrial customers in that utility's territory. If SEM is part of a utilities plan, then the account manager will include this option in their outreach to the industrial customers in that territory. The knowledge and experience of the account manager mitigates customer confusion and ensures that each customer participates in the program that best fits their needs.
    b. Energy Trust's account managers are contractors hired to deliver their industrial portfolio. These contractors are responsible for the long-term customer relationship on behalf of Energy Trust. The account manager has access to program materials designed to assist customer navigation of Energy Trust's offerings. Energy Trust has also aligned account managers with its SEM contractors to "improve communications, goal alignment, and limit customer transaction burden".
    c. PSE has internal Energy Management Engineers (EMEs) that are assigned to specific customers. These EMEs are not the utility's account managers, but specific to energy efficiency programs. The use of EMEs is intended to reduce customer confusion by creating a consistent energy efficiency experience for industrial customers.
2. Customer participation tracking. While the account managers discussed above mitigate customer confusion and customer engagement risk, coordinated central program tracking is required to ensure appropriate savings accounting. Both Energy Trust and PSE have taken steps to ensure that all program activity is recorded under the same customer or site number and that program processes are in place to review program tracking data before and during SEM engagement.

5 CONCLUSIONS AND RECOMMENDATIONS ON KEY RESEARCH QUESTIONS

The following table shows the six core research questions and the conclusion or recommendation that address each.

Research Question	Conclusion/ Recommendation
1. Can National Grid expect the proposed savings calculation methods and algorithms to provide defensible estimates of the energy savings acquired by the program?	Conclusion 1A
2. Does the methodology or algorithm align or deviate from current SEM best practice?	Conclusions 1A, 1B
3. What significant risks to savings is the program carrying using the proposed methodology?	Conclusion 2
4. Does the program include documentation sufficient to support future energy savings measurements and independent evaluation?	Conclusion 3
5. What adjustments to the methodology and algorithm does the contractor recommend that will likely improve the accuracy of the savings estimation and mitigate savings risk?	Recommendations 1, 3, & 4

Conclusion 1A: The savings calculations as described in the BPA Manual are consistent with industry best practices for SEM. This manual has helped to establish industry best practices. The workbook tool correctly implements the calculations described in the manual (with two minor corrections suggested here). Therefore, National Grid can expect the proposed savings calculation methods and algorithms to provide defensible estimates of the energy savings acquired by the program. Some improvements could be made to methods, guidance, and documentation as described under Recommendation 1. These relate to areas where industry best practices are still in development.

Conclusion 1B: As noted, the savings calculations are consistent with current industry best practices. None of the issues identified in this review represent major risks. Nonetheless, several issues were identified that could result in mis-stated savings, increased savings uncertainty, mis-stated savings uncertainty, or incomplete justification of analysis steps. These risks to savings and savings calculation accuracy come from the following sources:

- **Multi-collinearity.** In the presence of multi-collinearity, it may at times be difficult to construct models that provide savings estimates with tight standard errors. The BPA Manual provides little guidance on how to assess whether multi-collinearity is a problem and doesn’t recommend specific steps that should be taken. The project reports provide some assessment of inter-correlation among variables, but in most cases don’t describe how this information was used in model specification choices. For the projects reviewed, there was no evidence that multi-collinearity was affecting the quality of the savings estimates. However, for models with wide error bands for estimated savings, it may sometimes be possible to reduce the level of uncertainty by addressing multi-collinearity. This review did not explore this possibility for the specific projects studied.

- **Autocorrelation.** In the presence of autocorrelation, the standard errors of the model fit calculated using standard OLS formulas may be understated. That is, the accuracy with which savings are determined may be reported to be artificially good, if relying on OLS standard error calculations.

  The BPA Manual recommends testing for autocorrelation, and the analysis tool includes a calculated auto-correlation coefficient. This factor is incorporated into the Fractional Savings Uncertainty calculation, to avoid understating the uncertainty in the presence of autocorrelation. However, as noted below, the accuracy of the FSU formula for the types of models reviewed here is itself uncertain.

  Durbin-Watson statistics calculated by DNV GL indicated auto-correlation for all the models reviewed. Alternative models developed in this review to account for autocorrelation (AR1 models) provided similar savings estimates to the reported OLS models for 5 of the 7 models reviewed, and inconsistent results for 2 projects that had wide error bands around their savings estimates. In all cases the standard errors based on the OLS models were tighter than those with the alternative model. That is, the savings estimation accuracy with the OLS model would be estimated to be better than it really is. However, the OLS standard errors are not reported in the workbooks and site reports, only the calculated Fractional Savings Uncertainty.

- **Fractional Savings Uncertainty.** The reported Fractional Savings Uncertainty is consistent with an approximation provided in ASHRAE-14, which is based on empirical results for some models including weather and other independent variables. The applicability of this approximation to models of industrial facilities with process-related variables is unclear. Key determinants of savings uncertainty from a linear regression model include the correlation among the predictor variables, and the difference in average value of predictors between the baseline and reporting periods. Neither of these is accounted for in the ASHRAE approximation.

  The tool correctly calculates the formula as given in the BPA Manual, with two minor corrections suggested:

  1. The label on the output in the tool includes incorrect parenthetical details.
  2. The degrees of freedom for the t-statistic calculation appears to be off by 1 for regressions including an intercept term.

  For the projects reviewed, the Fractional Savings Uncertainty based on the approximation indicates wider uncertainty than the OLS standard error calculation. Since the OLS calculation understates uncertainty when autocorrelation is present, the FSU approximation is directionally doing the right thing. For the alternative AR1 models developed in this review, the FSU approximation had inconsistent performance.

- **Non-Routine Adjustments.** The BPA Manual includes guidance on how to adjust for other incentivized energy efficiency projects during the baseline or reporting period. However, the project reports provide only limited information on the other projects adjusted for and how the adjustments were made. In addition, the Manual does not provide guidance on how to treat anomalous data points. The reports noted outliers excluded, but do not justify why such occasional unusual behavior points should not be considered to be part of ongoing operations.

- **Model Acceptance Guidance.** For R^2, the longstanding rules of thumb derived from monthly data may not be as useful for higher frequency data. Standard errors or confidence intervals for the estimates of interest from the model are a more valuable indicator of model performance. Given that the models are all linear, the net determination bias is expected to be 0. It's not clear how much emphasis is given to R^2 or to net determination bias during model development.

Recommendation 1: There are several items we recommend be developed, as specified below.

- **Multi-collinearity.** Additional guidance on handling multi-collinearity could be developed. The guidance could include:
    - Conditions when multi-collinearity should or should not be considered a problem
    - Strategies for addressing multi-collinearity when it is a problem
    - Appropriate documentation for how identified multi-collinearity was addressed and why.
- **Auto-correlation.**
    - Consider incorporating more explicit guidance or check points for addressing auto-correlation when it is identified, and for documenting what was explored. Methods to address autocorrelation may include
        - Inclusion of a previously omitted variable that accounts for the autocorrelation
        - Inclusion of a break point in the model
        - Inclusion of auto-correlated error structure in the regression calculation.
    - Consider incorporating methods to account for autocorrelation via the error structure in the regression calculation, and directly calculating the Fractional Savings Uncertainty for the autocorrelation corrected regression.
- **Fractional Savings Uncertainty.**
    - Correct minor errors in the calculated Fractional Savings Uncertainty in the workbook.
        - Correct the label on the Summary Sheet.
        - Correct the degrees of freedom calculation for models including an intercept term
    - If OLS regression is used in the presence of autocorrelation, continue to use the FSU approximation rather than the full statistical formula for a standard error, unless a standard error calculation properly accounting for the autocorrelation is developed.
    - If a method to account for autocorrelation via the error structure in the regression calculation such as an AR1 model is applied, use the correct statistical formulas for this error structure to calculate the Fractional Savings Uncertainty, rather than the ASHRAE 14 approximation.

- Consider further exploring the accuracy of the FSU approximation or alternative approximations for models similar to those so far in this program—specifically, for models using daily or weekly data, with several production-related predictors.

- **Non-Routine Adjustments.** We recommend requiring more complete documentation of non-routine adjustments. This documentation would include:
  - Program tracking identification for other incentivized measures
  - More complete description of how their effects were removed
  - Discussion of the justification for deleting any outliers, other than clearly bad data. In particular explain why these occasional unusual points should not be considered to be part of ongoing operations.
- **Model Acceptance Guidance.** Limit attention to R^2 or net determination bias in model development. This recommendation may be consistent with existing practice.

Conclusion 2: The risk to savings currently carried by the program is that the measured savings are not a result of the program but are instead a result of factors outside of the program. DNV GL concludes that National Grid is appropriately managing this risk through the best practice methods used to estimate savings and the documentation available to support savings claims. DNV GL identified the following additional opportunities to further reduce program savings risk.

- **Completed Action Documentation:** Current site reports include lists of the priority opportunities for the site and the full list of site opportunities and engagement actions. The program is following the best practice of recording when an opportunity was identified, who it was assigned to, and when it was completed. DNV GL observed that many of the priority opportunities were still in progress at the end of 2019 and many of the completed opportunities had completion dates at the end of the year, long after savings started to be measured. In these cases, the risk to savings is higher as the documentation supporting the measurement is weaker. National Grid can improve savings risk management by listing the actions that are believed to have resulted in the measured savings in the priority section and/or providing the date an opportunity starts to provide savings as well as the completed date (if different). Additionally, if measured savings are zero or negative, the summary report should state why the actions taken did not result in measurable savings. Addressing negative or zero savings will help customers understand the program's perspective on why no savings were achieved, especially if the actions were expected to achieve savings, and potentially help identify opportunities for program improvement.
- **Adjustments to Standard Operating Procedures.** Actions at industrial facilities are often determined based on the standard operating procedure (SOP). Discussion of a sites SOP was not observed in the provided documentation. The program could improve its management of savings risk by documenting what adjustments to the SOP were made during the year that are tied to recorded opportunities. If an action or set of actions is stated in the SOP, then it is likely to persist and continue to provide savings.
- **Agreements on modeling limitations and boundary conditions.** The program implementer is successfully managing savings risk by following a robust industry best practice modeling guideline. However, model selection and savings determination require multiple subjective decisions on the

baseline period length, non-routine adjustments to implement, independent variables to use, and model form. There is a scenario in which the implementer will estimate savings that National Grid believes are too risky based on the modelling indicators or site conditions. The current guidelines effectively warn modelers of increased risk, but do not tell them when to stop. Essentially, the guidelines identify the good green zone and the yellow warning zone, but not the red stop zone. National Grid could improve its savings risk management by establishing conditions under which the program will not claim savings using one of the available IPMVP Option C methods discussed in the guideline. As stated previously, the definition of the good green zone is expected change as more modeling is completed using higher frequency data.

Recommendation 2: Consider adjustments to the program agreement and documentation discussed in Conclusion 2 above.

Conclusion 3: The program contains documentation sufficient for independent evaluation. All production and consumption data is stored and accessible on the vendor's platform. The vendor also produces workbooks documenting model selection and annual savings estimation. National Grid should expect independent evaluators to request and review the source data used to develop models, all data from the measurement period, information supporting non-routine adjustments or non-standard baseline period lengths, records of a participant's engagement in the program, and records of the actions taken to reduce energy consumption at the facility. As noted in Conclusion 1B, project documentation is limited with respect to the following:

- Model selection rationale, particularly in the presence of multi-collinearity and auto-correlation.
- Details on non-routine adjustments made, including tracking information and project documentation for other incentivized measures, and outliers excluded.

Recommendation 3: We recommend improved documentation in all these areas, as described under Recommendation 1.

Recommendation 4: We recommend establishing program processes that ensure the information expected to be requested for evaluation is stored annually on a National Grid system and associated with the final energy savings estimate for the site in that year. Even if no savings are claimed for the site, National Grid should store the documentation and associate it with a zero-savings tracking record for the year. Implementing an annual documentation transfer will mitigate risks associated with potential future program delivery changes.

Appendix A: Site Level Findings

This appendix provides the results of DNV GL's replication of the program baseline models using the data provided by National Grid. These site-specific replication results are used to support the responses to National Grid's key research questions.

Site A

The model and modeling process for Site 1 followed the BPA M&V guidelines. Daily consumption data for 8 months in the pre period (7/1/2018 – 2/28/2019) and 8 months of post period were used in the analysis (3/1/2019-10/31/2019). Only one day was removed from the analysis (October 30, 2018) because of missing electric metering data. No documentation was provided to verify that this data point was truly missing. Non-routine adjustments were accounted for in accordance with the BPA M&V guidelines, however, there were no non-routine adjustments made to the study period.

One issue that the BPA M&V guidelines provide little guidance on is the presence of autocorrelation. The autocorrelation coefficient was presented, but no statistical test was performed and no comments were provided. In the replication of the results, a Durbin-Watson test was performed, which showed significant autocorrelation. However, in this case the effects of such autocorrelation may have little impact on the model's ability to accurately project energy usage, which is its main objective.

DNV GL also checks the variance inflation factor (VIF) in each replication, which can indicate the effects from multicollinearity.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	36616.00	2462.86	14.87	<.0001	0.00
Max(0, 65-DB)	1	-88.31	22.44	-3.93	0.0001	1.79
Max(0,DB-65)	1	653.04	75.91	8.60	<.0001	2.28
Holiday	1	-5094.90	1558.67	-3.27	0.0012	1.16
Phase_I	1	167.01	16.92	9.87	<.0001	1.35
Phase_II	1	152.61	14.11	10.81	<.0001	1.41
Phase_III	1	162.97	15.63	10.42	<.0001	1.11
Production	1	8100.95	2296.24	3.53	0.0005	1.59
ShutDown_Indicator	1	-17051.00	2449.05	-6.96	<.0001	1.45

Statistical Summary	
Number of Observations	239
R-Squared	0.769
Adjusted R-Squared	0.761
Root MSE	4027.951
Coeff of Var	6.35
F Statistic	95.53
Autocorrelation Coefficient	0.255
Durbin-Watson D	1.484

Site B

The model and modeling process for Site 2 has followed the BPA M&V guidelines. DNV GL was able to replicate the regression results with the data provided. The non-routine adjustments during the baseline period were in line with the BPA guidelines, however, there was no mention of any non-routine adjustments during the study period. Daily consumption data for 12 months in the pre period (1/1/2018 – 12/31/2018) and 8 months of post period were used in the analysis (3/1/2019-10/28/2019).

Electric consumption data were missing from March 29, 2018 through May 1, 2018, however there was no reason provided why data from the month of April was missing. There were two other dates with missing data; August 4, 2018 and August 21, 2018. The reason for both was "missing 7045 data," however, no explanation as of what that means was provided.

It was stated that all the linear regression assumptions are valid for the model, however, no detail on how these assumptions were checked was provided. In the replication of the regression results, tests for autocorrelated variances were performed and they showed significant autocorrelation. However, in this case the effects of autocorrelation may have little impact on the model's ability to accurately project energy usage. The report also stated that two of the independent variables had a high degree of correlation, however there was no mention of any check or test for multicollinearity.

In replicating the results, DNV GL checked the variance inflation factor (VIF), which showed minimal effects from multicollinearity.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	11500.00	460.92	24.95	<.0001	0.00
Max(0,50-DB)	1	-36.09	14.57	-2.48	0.0138	1.69
Max(0,DB-50)	1	156.59	13.88	11.28	<.0001	1.71
Production	1	5172.97	499.68	10.35	<.0001	1.50
Production_Quantity	1	0.05	0.00	24.97	<.0001	1.48

Statistical Summary	
Number of Observations	330
R-Squared	0.843
Adjusted R-Squared	0.841
Root MSE	1996.125
Coeff of Var	8.34
F Statistic	434.94
Autocorrelation Coefficient	0.153
Durbin-Watson D	1.663

Site C

The model and modeling process for Site 4 has followed the BPA M&V guidelines. DNV GL was able to replicate the regression model with the data provided. The non-routine adjustments during the baseline period were in line with the BPA guidelines and again there was no mention of any non-routine adjustments during the study period. The baseline period included a full year (March 1, 2018 – February 28, 2019) while the post period included 8 months (3/1/2019-10/31/2019). .

Data removed from the analysis were from May 18, 2018 through May 24, 2018 and on August 15, 2018. The reason stated was because of missing electric data, but no documentation was provided. The vendor states that there were three occasions of low usage with normal production (October 30, 2018, December 6, 2018, and January 30, 2019) and these data were removed from the analysis. No documentation was provided to support this and there did not appear to be any systematic removal of data.

The vendor once again asserted that all model assumptions were met but provided no detail on how this was verified. They did provide several graphs to show that the data conformed to the linearity assumption and a graph to demonstrate the normality of the residuals. The vendor calculated pairwise correlations and stated that multicollinearity would need to be considered during variable selection.

In DNV GL's replication of the results, a VIF of 5.65 was recorded (highlighted below) A VIF of 5 is sometimes used as a cutoff to indicate the presence of strong multicollinearity, however, the effects on the projected energy usage coming from the model is unknown. This model also had significant autocorrelation according to the Durbin-Watson test.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	6873.19	143.18	48.01	<.0001	0.00
Max(0, WB-54)	1	122.99	4.94	24.92	<.0001	1.02
Weekend_Holiday	1	-1535.69	121.87	-12.60	<.0001	2.92
Knitting_Yards	1	0.19	0.01	21.88	<.0001	5.65
Production_Ind	1	1837.63	154.12	11.92	<.0001	2.79
Sum_Warping	1	0.02	0.01	2.64	0.0086	1.97

Statistical Summary	
Number of Observations	353.00
R-Squared	0.96
Adjusted R-Squared	0.96
Root MSE	628.34
Coeff of Var	5.33
F Statistic	1665.34
Autocorrelation Coefficient	0.48
Durbin-Watson D	1.04

Site D

The model and modeling process for Site 3 has followed the BPA M&V guidelines. DNV GL was able to replicate the regression model with the data provided. The non-routine adjustments during the baseline period were in line with the BPA guidelines. However, again there was no mention of any non-routine adjustments during the study period. The baseline period did not include a full year and was only eight months in length (July 1, 2018 to February 28, 2019), as was the post period (3/1/2019 to 10/31/2019). The vendor provided the explanation that the first half of 2018 appeared to not be representative of current operations and therefore the baseline period was shortened. No other documentation was provided to support this claim.

Data from October 30, 2018 to November 7, 2018 were removed from the analysis because of missing electric consumption data, but no documentation was given to support this claim. January 30, 2019 was also removed from the analysis because of missing electric data. December 6, 2018 and August 15, 2018 were also removed and the reason provided was that electric usage was unusually low even though production was normal. There did not appear to be any systematic removal of data. For the data points that were removed, a reason was given, but no supporting documentation was provided.

The vendor once again asserted that all model assumptions were met but provided no detail on how this was verified. The vendor did provide several graphs to show that the data confirmed to the linearity assumption and a graph to demonstrate the normality of the residuals. The graph of the residuals appeared to depart from normality, but we cannot determine if this departure was severe enough to influence parameter estimation. Again, the vendor calculated pairwise correlations and stated that multicollinearity could be an issue but provided no statistical check on this possible violation of assumptions.

In DNV GL's replication of the results, a Durbin-Watson test was performed again showing significant autocorrelation, though its effect on the parameter estimation could be limited. DNV GL also checked the VIF which showed minimal effects from multicollinearity.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	12173.00	485.58	25.07	<.0001	0.00
Max(0,55-DB)	1	69.30	18.02	3.84	0.0002	2.08
Max(0,DB-55)	1	11.63	23.35	0.50	0.6189	2.10
Total_Production	1	0.03	0.00	7.85	<.0001	1.48
NonHoliday_Sat	1	-4059.30	464.60	-8.74	<.0001	1.23
Sun_Holiday	1	-7989.95	399.05	-20.02	<.0001	1.42

Statistical Summary	
Number of Observations	231
R-Squared	0.808
Adjusted R-Squared	0.803
Root MSE	2138.509
Coeff of Var	17.19
F Statistic	188.74
Autocorrelation Coefficient	0.143
Durbin-Watson D	1.711

Site E

The model and modeling process for Site 5 has followed the BPA M&V guidelines. DNV GL was able to replicate the regression model with the data provided. There was no mention of any non-routine adjustments during either the baseline period or the study period. The baseline period included 51 weeks (March 10, 2018 – February 28, 2019) while the post period included 34 (3/2/2019 to 10/25/2019). The study planned to include a full year (52 weeks), but the first week of data had to be removed because of missing electric consumption data. The model was run using weekly consumption since the production data is mostly recorded weekly.

The only data that were removed from the analysis was from the first week of the baseline period. All other data were included in the analysis. The vendor again made the assertion that all standard assumptions on using linear regression were met. The model again showed significant autocorrelation, but in this model the R-squared value fell below the BPA guideline standards. The vendor acknowledged that the R-squared value fell outside of the suggested range (highlighted below) and states that, because the model uses weekly data, the statistic would be considered acceptable. DNV GL also acknowledges that the use of weekly data versus daily data could lead to a model whose fit fails to exceed the suggested R-squared value of 0.75.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	94675.00	5429.11	17.44	<.0001	0.00
Max(0, DB-56)	1	168.92	24.12	7.00	<.0001	1.23
Max(0,35-DB)	1	91.92	77.51	1.19	0.2421	1.66
CLX	1	0.07	0.04	1.72	0.0926	1.52
Insulating	1	0.01	0.00	2.79	0.0078	1.14
Cabling	1	0.05	0.03	1.59	0.1194	1.32
Jacketing	1	0.08	0.02	3.44	0.0013	1.68

Statistical Summary	
Number of Observations (weeks)	51
R-Squared	0.67
Adjusted R-Squared	0.62
Root MSE	8057.97
Coeff of Var	5.92
F Statistic	14.85
Autocorrelation Coefficient	0.40
Durbin-Watson D	1.02

Site F

The model and modeling process for Site 6 has followed the BPA M&V guidelines. DNV GL was able to replicate the regression model with the data provided. The non-routine adjustments during the baseline period were in line with the BPA guidelines. Again, there was no mention of any non-routine adjustments during the study period. The baseline period included a full year (January 10, 2018 – December 31, 2018) while the post period included 8 months (March 1, 2019-October 30 2019).

The baseline period was planned to begin on January 1, 2018, but data were removed from analysis for the first ten days of the baseline period because of missing electric data. No supporting documentation was provided. Two observations (May 29, 2018 and June 2, 2018) were removed from the data analysis and the vendor states the reason is that there was extremely high usage and no production.

The vendor once again asserted that all model assumptions were met but provided no detail on how this was verified. They did provide several graphs to show that the data confirmed to the linearity assumption and a graph to demonstrate the normality of the residuals. In this case, the residuals appeared to approximate normality, however the Durbin-Watson autocorrelation test showed significant first order autocorrelation was present.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	2901.16	915.45	3.17	0.0017	0.00
Dry Bulb Temp	1	110.10	12.12	9.09	<.0001	1.00
Non_Holiday_Saturday	1	3138.19	694.84	4.52	<.0001	1.35
Parts_Produced	1	0.01	0.00	35.71	<.0001	2.07
Production	1	7903.39	915.60	8.63	<.0001	1.71

Statistical Summary	
Number of Observations	354
R-Squared	0.91
Adjusted R-Squared	0.90
Root MSE	3883.81
Coeff of Var	10.99
F Statistic	833.31
Autocorrelation Coefficient	0.18
Durbin-Watson D	1.64

Site G

The model and modeling process for Site 7 has followed the BPA M&V guidelines. DNV GL was able to replicate the regression model with the data provided. The non-routine adjustments during the baseline period were in line with the BPA guidelines, again there was no mention of any non-routine adjustments during the study period. The baseline period included a full year (March 1, 2018 – February 28, 2019) while the post period included 8 months (March 1, 2019 to October 31, 2019).

Data were removed from analysis for August 15, 2018, October 30, 2018, and December 6, 2018 due to low electric usage and high production. The standard used for this decision was within the BPA M&V guidelines, but no documentation was provided. There was also missing data for January 30, 2019, so it was also excluded from the analysis.

The vendor once again asserted that all model assumptions were met but provided no detail on how this was verified. The vendor did provide several graphs to show that the data confirmed to the linearity assumption and a graph to demonstrate the normality of the residuals. In this case, the residuals appeared to approximate normality, however the Durbin-Watson autocorrelation test showed significant first order autocorrelation was present. This model also had variables for the pounds produced by production line, but for the most part these variables did not appear very strongly correlated. The vendor did not provide any analysis that was performed to demonstrate that there was no multicollinearity problem.

In DNV GL's analysis the VIF was calculated and one of the production variables had a slightly higher VIF than seen in most of the other regressions, however, it fell below 5 (4.23) and was not considered an issue. In this case the residuals appeared to approximate normality, however the Durbin-Watson autocorrelation test showed significant first order autocorrelation was present.

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|	VIF
Intercept	1	11856.00	626.94	18.91	<.0001	0.00
Max(0, DB-50)	1	215.16	18.82	11.43	<.0001	1.77
Max(0, 50-DB)	1	55.20	22.31	2.47	0.0138	1.82
Weekend Indicator	1	-5602.25	496.68	-11.28	<.0001	2.28
Production Indicator	1	5394.54	651.43	8.28	<.0001	2.09
Production Line 105	1	0.12	0.02	5.88	<.0001	1.55
Production Line 107	1	0.09	0.01	8.45	<.0001	2.57
Production Line 106	1	0.13	0.01	18.26	<.0001	4.23
Production Line 108	1	0.08	0.01	9.29	<.0001	2.39
Production Line 109	1	0.08	0.01	14.60	<.0001	3.04
Production Line 4_4	1	0.18	0.01	12.74	<.0001	1.81
Production Line 103_4_1	1	0.21	0.01	14.43	<.0001	1.74

Statistical Summary	
Number of Observations	361
R-Squared	0.98
Adjusted R-Squared	0.98
Root MSE	2831.07
Coeff of Var	6.14
F Statistic	1529.56
Autocorrelation Coefficient	0.16
Durbin-Watson D	1.68

About DNV GL

Driven by our purpose of safeguarding life, property and the environment, DNV GL enables organizations to advance the safety and sustainability of their business. We provide classification and technical assurance along with software and independent expert advisory services to the maritime, oil and gas, and energy industries. We also provide certification services to customers across a wide range of industries. Operating in more than 100 countries, our 16,000 professionals are dedicated to helping our customers make the world safer, smarter and greener.