BONNEVILLE POWER ADMINISTRATION

Non-Industrial Custom and Energy Smart Reserve Power Impact Evaluation Final Report

Submitted by Evergreen Economics
In partnership with SBW Consulting and Apex Analytics, LLC

Sept. 30, 2025

1 EXECUTIVE SUMMARY

This document presents the results of an impact evaluation of Bonneville Power Administration's (BPA's) non-industrial custom portfolio and Energy Smart Reserved Power (ESRP) program. A team led by Evergreen Economics (the evaluation team, the Evergreen team, or the team), which includes SBW Consulting and Apex Analytics, conducted this research.

The primary objectives of this evaluation were to:

- Estimate first-year kWh savings and cost-effectiveness for the non-industrial custom portfolio, including the ESRP program, to understand the savings performance.
- Develop recommendations on measurement and verification (M&V) procedures, including when savings can be reliably estimated, for custom measures, using the BPA M&V Protocol Selection Guide, including the protocol called Engineering Calculations with Verification (ECwV).

This evaluation covered three domains: Option 1 and Option 2 utilities, as well as ESRP projects outside of utility service territories.

1.1 METHODOLOGY

This evaluation represents Option 1 and Option 2 utilities' non-industrial custom projects as well as custom projects for facilities with no utility that are served directly by the Federal Columbia River Power System. The sample design targeted a 90/10 confidence level and precision and was developed based on BPA tracking data for projects that completed invoicing in the most recent and complete year (fiscal year 2023).

The evaluation team conducted the sampling using Dalenius-Hodges stratified random design based on domain and reported kWh savings for each project. The sample size was 32 sites.

The data collection approach utilized a combination of sources, including file review, site visits, and time of use metering. The evaluation team conducted site-specific measure savings analysis using a multistep process starting with a review of the project energy model, collecting supplemental data where needed, running the model, and estimating savings.

The team completed data collection and analysis for the sample, then compiled project-level results to estimate the electric savings and cost-effectiveness for the non-industrial custom program portfolio and the ESRP program using a ratio analysis.

1.2 SUMMARY OF FINDINGS AND RECOMMENDATIONS

The overall results for this study showed that the non-industrial custom portfolio realized 181 percent of reported savings, and ESRP realized 101 percent of savings, as the final column of Table 1 shows.

There were substantial differences between BPA-reported savings and savings documented in the sample project files within Option 2. When this set of differences was removed, the ratio of Option 2 evaluated savings was 109 percent of documented savings. ESRP evaluated savings were similar to BPA project file savings, with a 105 percent realization rate. This second set of realization rates is shown in the "realization rate excluding tracking data errors" column in Table 1.

Table 1: Non-industrial custom evaluated realization rates by domain

Domain	BPA-Reported Savings (kWh)	BPA Project File Savings (kWh)	Evaluated Savings (kWh)	Realization Rate Excluding Tracking Data Errors	Overall Realization Rate
Option 1	13,671,838	13,670,559	10,497,240	77%	77%
Option 2	9,163,681	28,360,352	30,841,631	109%	337%
Subtotal	**22,835,519**	**40,030,910**	**41,338,870**	**98%**	**181%**
ESRP	15,731,260	15,090,730	15,871,446	105%	101%

The overall sampling precision totaled 9 percent for a 90 percent two-tailed confidence interval, which was more precise than the target design of 10 percent (at 90 percent).[1]

Option 1 results (77 percent) were lower than the last evaluation conducted on Option 1 non-industrial non-lighting projects (115 percent), while Option 2 results (109 percent) were comparable to the previous evaluation of Option 2 non-industrial, non-lighting projects (110 percent). ESRP results (105 percent) were significantly higher than the last BPA impact evaluation of ESRP (49 percent).

The portfolio's evaluation results excluding the tracking data issues showed realization rates very close to 100 percent, indicating that overall, the non-industrial custom portfolio is realizing the savings reported in the BPA project files. However, at the individual project level, the high degree of variability in realization rates leads to lower reliability of savings for individual projects than utilities and end-use customers may be expecting. The portfolio is highly cost-effective, generating approximately $2 in benefits for every $1 spent.

The key findings from this study are summarized below.

- **Discrepancies between project file savings and BPA-reported savings were common within Option 2 utility projects.** Upon investigation, there were no clear trends or explanations for these differences. The team suspects that the issues encountered during this evaluation were associated with BPA's transition of its tracking systems.
- **Multiple projects did not accurately calibrate savings models to consumption or accurately normalize usage to typical usage,** leading to inaccuracies in project file savings.
- **Small- and medium-sized projects showed similar results in evaluated savings using the BPA Engineering Calculations with Verification (ECwV) protocol or high-rigor measurement and verification (M&V) methods.** While there was variability from site to site and the ECwV results were less precise, ECwV results were similar to

[1] When tracking data errors are included, the sampling precision values are much higher, especially for Option 2. Evergreen recommends focusing on results excluding tracking data errors.

evaluated results overall. This finding is consistent with the results from a similar analysis conducted on BPA's Option 1 and Option 2 custom industrial portfolio.

The evaluation includes recommendations to address these issues to improve the accuracy of non-industrial custom and ESRP program savings reporting (see Section 5 for more detail):

- BPA should review project savings compared to BPA-reported savings to identify and correct any major discrepancies. Going forward, BPA should review its reporting procedures to ensure that the data entered into its reporting system reflect the final project documented total savings. Making these improvements would improve the reliability of expected project savings for BPA, utilities, and end-use customers.
- Where feasible, for projects where the ECwV protocol is not being applied, BPA should calibrate savings estimation models to actual pre-installation and/or post-installation consumption data, in order to more closely estimate and report savings that the program, utility, and end-use customer will actually realize.
- BPA should continue applying ECwV to the same size range of projects that it is using now since the protocol is producing reliable results. This practice saves resources for BPA and utilities, and our results show that this protocol is producing accurate savings. However, BPA should be cautious of applying ECwV to larger projects.

2 INTRODUCTION

This document presents the results of an impact evaluation of BPA's non-industrial custom portfolio and ESRP program. A team led by Evergreen Economics (the evaluation team, the Evergreen team, or the team), which includes SBW Consulting and Apex Analytics, conducted this research.

2.1 DEFINITIONS AND KEY TERMS

This report uses the following terms throughout:

- **Option 1 Utilities:** Utilities that have opted to have BPA engineers manage and estimate the savings for custom projects at sites where site specific characteristics will determine energy savings estimates.
- **Option 2 Utilities:** Utilities that have opted to have their own engineers do the site-specific savings estimates and otherwise manage the implementation of custom projects with a BPA-approved M&V plan.
- **ESRP Program:** Includes projects at Federally Chartered Irrigation Districts served by BPA Reserved Power, and BPA and other federal agency infrastructure projects served by Station Service. Projects are programmatically managed by BPA. BPA engineers estimate the savings for custom projects at sites where site-specific characteristics will determine energy savings estimates.

See Appendix A for definitions of key terms in this report, such as reported savings, measure, and realization rate.

2.2 BACKGROUND

Consistent with Regional Technical Forum (RTF) guidelines, BPA aims to achieve 90 percent coverage of the energy efficiency portfolio through impact evaluation in a four-year period.[2] When selecting which programs to evaluate in a given year, BPA balances the objectives of portfolio coverage, strategic research needs, timely feedback, annual budgets, and the cost and effort required.

BPA conducted impact evaluation planning in 2019-2020 to determine what evaluation activities had occurred previously and what evaluation needed to occur in the next four year-period to address its evaluation policy. The outcome of this effort was the 2020-2021 evaluation plan,[3] which categorized the portfolio into unique domains grouped by similar delivery approaches (utility type, measure type, and sector). The evaluation plan proposed conducting the impact evaluations on a consistent rolling basis, with one domain-specific study approximately every six months across the four-year period. This evaluation is aligned with the priorities identified in the 2020-2021 evaluation plan, as the last study in the sequence (since 2020, BPA has completed evaluations of its custom industrial and of non-residential lighting portfolios).

---

[2] Regional Technical Forum. 2020. *Regional Technical Forum Operative Guidelines for the Assessment of Energy Efficiency Measures*: https://nwcouncil.app.box.com/v/2020RTFGuidelines (see Section 5.2.1).
[3] Evergreen Economics. 2020. *Bonneville Power Administration 2020-2021 Evaluation Plan.* https://www.bpa.gov/-/media/Aep/energy-efficiency/evaluation-projects-studies/bpa-2020-21-impact-evaluation-plan.pdf

2.3 STUDY OBJECTIVES

The primary objectives of this evaluation were to:

- Estimate first-year kWh savings and cost-effectiveness for the non-industrial custom portfolio, including the ESRP program, to understand the savings performance.
- Develop recommendations on M&V procedures for custom measures, including when savings can be reliably estimated, using the BPA M&V Protocol Selection Guide, including the ECwV protocol.

This evaluation covered both Option 1 and Option 2 utilities, as well as ESRP projects outside of utility service territories.

3 METHODOLOGY

This section summarizes the methods used by the Evergreen team to conduct this evaluation. Appendix C provides additional detail on the study methods.

3.1 SAMPLE DESIGN

Table 2 shows the number of non-industrial custom measures and custom ESRP program measures with their associated savings in our sample frame by domain and size strata.[4]

Table 2: Non-industrial custom and ESRP sample design

Utility Type	Strata	Reported Savings (kWh)		Number of Reported Measures	Sample Size (Measures)
		Average	Total		
Option 1	0	7,029	168,688	24	0
	1	85,392	5,635,898	66	7
	2	343,866	5,845,718	17	7
	Certainty	2,021,533	2,021,533	1	1
	Subtotal	**126,591**	**13,671,838**	**108**	**15**
Option 2	0	10,646	212,920	20	0
	1	114,326	4,344,379	38	6
	2	511,820	4,606,383	9	6
	Subtotal	**136,771**	**9,163,681**	**67**	**12**
Federal/ESRP Program	1	849,292	1,698,583	2	2
	Certainty	4,677,559	14,032,677	3	3
	Subtotal	**3,146,252**	**15,731,260**	**5**	**5**
Total		**214,260**	**38,566,779**	**180**	**32**

*Stratum *0* denotes the excluded projects (based on very small *ex ante* savings). The *certainty* projects represent a significant portion of total reported energy savings within the domain and are considered as necessary for the evaluation and therefore are not subject to random selection.

3.2 DATA COLLECTION

The evaluation team's general approach to evaluation data collection was to fully leverage the data collected by BPA, project engineers, and utility program staff throughout the process of developing each project and to collect additional data from end users to achieve reliable estimates of savings for the sampled projects. The evaluation team collected the necessary data using a combination of the following approaches (each of which is described in more detail in Appendix C):

- File review
- Telephone/email discussion with project engineers

[4] Note that the sample design was based on the tracking data extract provided by BPA from its tracking system (BEETS). The sample provided robust, statistically valid results for the accuracy of documented savings compared to evaluated savings, but the overall accuracy of tracked savings for the population is highly uncertain.

- Telephone/email discussion with end users
- Site visits
- Affected system trend metering
- Supplemental weather data gathering
- Cost effectiveness parameter data collection

3.3 MEASURE SAVINGS ANALYSIS

The evaluation team estimated savings for sampled lighting projects and measures using the following steps (each of which is described in more detail in Appendix C):

1. Review project energy models
2. Assess determinant reliability and collect supplemental data
3. Run the model and estimate evaluated savings
4. Assess the impact of interactive measures
5. Determine the time-based value of savings and cost effectiveness

3.4 STUDY AND DOMAIN ANALYSIS

Once the evaluation team completed data collection and analysis for the sample, the team compiled a workbook containing all individual measure-level findings about key drivers for deviations between evaluated savings and original savings estimates. The team used the measure-level results to estimate the electric savings and cost-effectiveness for the non-industrial and ESRP portfolios and by this study's domain category (which is utility type) using a ratio analysis. This report presents details of this approach in Appendix C.

4 FINDINGS

This section presents impact evaluation results for BPA's non-industrial custom portfolio and ESRP program.

The section is organized as follows:

- Overall results
- Measure-specific findings
- Key drivers of savings
- Lifetime savings
- ECwV savings estimation efficacy
- Cost-effectiveness

Appendix B provides site-specific savings estimation details.

4.1 OVERALL RESULTS

This subsection provides the overall results for this impact evaluation of custom non-industrial and ESRP projects installed by subdomain (Option 1 and Option 2 utility customers) with completed reporting in the most recent and complete year (fiscal year 2023).

This section presents two distinct sets of findings. First, as a result of tracking data issues, there were discrepancies between the project file savings obtained for the sample and BPA's reported savings. Those issues are primarily present in Option 2 projects.[5] Three out of twelve Option 2 projects reviewed had documented savings that were at least 70 percent higher than was found in BPA's Energy Efficiency Tracking System (BEETS). One project had documented savings over 10 times the reported savings. There were also differences between savings documented in individual project files compared to evaluated savings across all subdomains. This section presents results of these two sets of findings separately (two sets of interim realization rates) and also in combination (the overall realization rate) so that the tracking issues are separated from the site-level realized savings. To account for BPA tracking data discrepancies, the evaluation team calculated three realization rates to address individual steps in the analysis:

- Realization Rate 1 (RR1): Ratio of project file savings to BPA-reported savings.

$$RR1 = \frac{Project\ file\ savings}{BPA\ reported\ savings}$$

- Realization Rate 2 (RR2): Ratio of evaluated savings to project file savings.

$$RR2 = \frac{Evaluated\ savings}{Project\ file\ savings}$$

- Realization Rate 3 (RR3): Ratio of evaluated savings to BPA-reported savings, and the product of RRs 1 and 2.

[5] Note that the evaluation team obtained BPA project file savings only for sampled projects, not for the population. The team extrapolated the difference between BPA-reported and project file savings from the sample to the population by sample stratum.

$$RR3 = RR1 * RR2 = \frac{Project\ file\ savings}{BPA\ reported\ savings} * \frac{Evaluated\ savings}{Project\ file\ savings} = \frac{Evaluated\ savings}{BPA\ reported\ savings}$$

RR2 is typically the focus of custom project evaluations. RR1 is usually not calculated or reported, with BPA-reported savings in previously evaluated domains typically being very close or equal to what the project files contain. RR1 shows the impact of a tracking data issue that primarily impacts Option 2 projects, while RR2 is based on sampled project measure-specific issues identified by the evaluation.

The overall realization rate (RR3) is the product of RRs 1 and 2. The evaluation team estimated the overall realization rates for Option 1, Option 2, and ESRP measures to be 77 percent, 337 percent, and 101 percent, respectively. Across Option 1 and Option 2 measures, the team estimated the overall realization rate to be 181 percent.

FIRST-YEAR SAVINGS

As Figure 1 shows, evaluated savings were slightly lower than BPA-reported savings and BPA project file savings for Option 1 measures.

Evaluated savings were substantially higher than BPA-reported and BPA project file savings for sampled Option 2 measures. Evaluated savings for ESRP were slightly lower than BPA project file savings and approximately the same as BPA-reported savings. Across all domains, total BPA project file savings were much closer than BPA-reported savings to total evaluated savings for sampled projects.

Figure 1: Reported, project file, and evaluated first-year savings by domain (for the population)

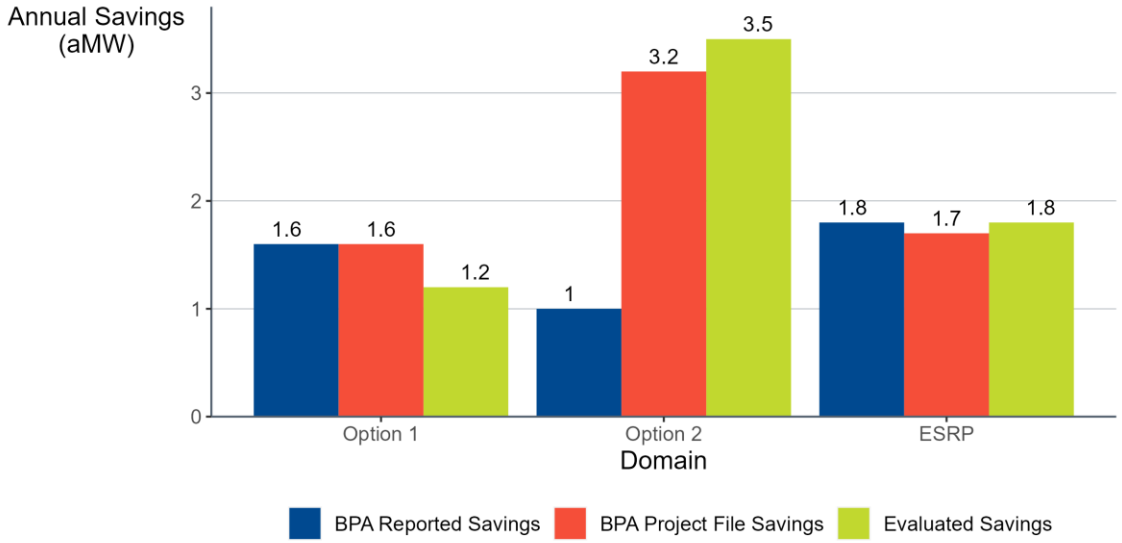


Table 3 shows the final results in two ways, first with RR2, which represents each stratum's realization rate as the ratio of the evaluated savings to project file savings, which excludes

tracking data errors and is most comparable to prior year evaluations. RR3 is the ratio of evaluated savings to BPA-reported savings and includes tracking data errors. In order to allow for a clean comparison, all savings have been adjusted to be savings at the generator, inclusive of busbar factors and water savings generation credit.

Table 3: Non-industrial custom evaluated realization rates by domain

Domain	Savings			Realization Rates		
	(a) BPA-Reported Savings (kWh)	(b) BPA Project File Savings (kWh)	(c) Evaluated Savings (kWh)	RR1 Ratio of Documented Savings to Tracking Data	RR2 Realization Rate Excluding Tracking Data Errors	RR3 Realization Rate Including Tracking Data Errors
				b/a	(c/b)	(c/a)
Option 1	13,671,838	13,670,559	10,497,240	100%	77%	77%
Option 2	9,163,681	28,360,352	30,841,631	309%	109%	337%
Subtotal	**22,835,519**	**42,030,910**	**41,338,870**	**184%**	**98%**	**181%**
ESRP	**15,731,260**	**15,090,730**	**15,871,446**	**96%**	**105%**	**101%**

Table 4 shows BPA-reported savings and evaluated savings, by stratum, for the population. The Evergreen team estimated evaluated savings (value c in the table) by taking the product of RR1 and RR2 (RR3) times tracked savings (value a in the table) at the stratum level. Option 2 had measures with large outliers in the comparison of tracked and project file savings, which drove the RR3 results.

The evaluation team reviewed BPA-reported and project file savings for the sample and noted discrepancies in some of the Option 2 measures. There was no clear trend in the Option 2 discrepancies. BPA project file savings as a fraction of BPA-reported savings were 100 percent for Option 1 measures, 309 percent for Option 2, and 96 percent for ESRP.

Table 4: Non-industrial custom evaluated first-year savings (population) and combined realization rates by domain

		Savings			Realization Rates		
		(a)	(b)	(c)	RR1	RR2	RR3
Domain	Strata	BPA-Reported Savings (kWh)	BPA Adjusted Savings (kWh)	Evaluated Savings (kWh)	(b/a)	(c/b)	(RR1* RR2 at the Stratum Level)
Option 1	0*	168,688	168,673	129,519	100%	77%	77%
	1	5,635,898	5,634,635	3,637,466	100%	65%	65%
	2	5,845,718	5,845,718	4,876,590	100%	83%	83%
	99	2,021,533	2,021,533	1,853,665	100%	92%	92%
	Subtotal	**13,671,838**	**13,670,559**	**10,497,240**	**100%**	**77%**	**77%**
Option 2	0*	212,920	658,958	716,611	309%	109%	337%
	1	4,344,379	22,741,293	23,193,524	523%	102%	534%
	2	4,606,383	4,960,100	6,931,495	108%	140%	150%
	Subtotal	**9,163,681**	**28,360,352**	**30,841,631**	**309%**	**109%**	**337%**
Subtotal		**22,835,519**	**42,030,910**	**41,338,870**	**184%**	**98%**	**181%**
ESRP	1	1,698,583	1,620,710	2,353,305	95%	145%	139%
	99	14,032,677	13,470,020	13,518,141	96%	100%	96%
	Subtotal	**15,731,260**	**15,090,730**	**15,871,446**	**96%**	**105%**	**101%**

*The Evergreen team assigned the average domain-level realization rate to these strata, as the sample did not include any sites from these strata.

SAMPLING ERROR

The actual sampling relative precision totaled 9 percent for a 90 percent two-tailed confidence interval (*Table 5*). This precision is slightly better than predicted during the sample design development (10 percent precision). Totals for increasing portions of the population tend to have improved precision compared to their components, because the propagation of error results in a tightening distribution as one adds together uncertain values. This causes the total of Option 1 and Option 2 to have better precision than either of Option 1 and Option 2 by themselves.

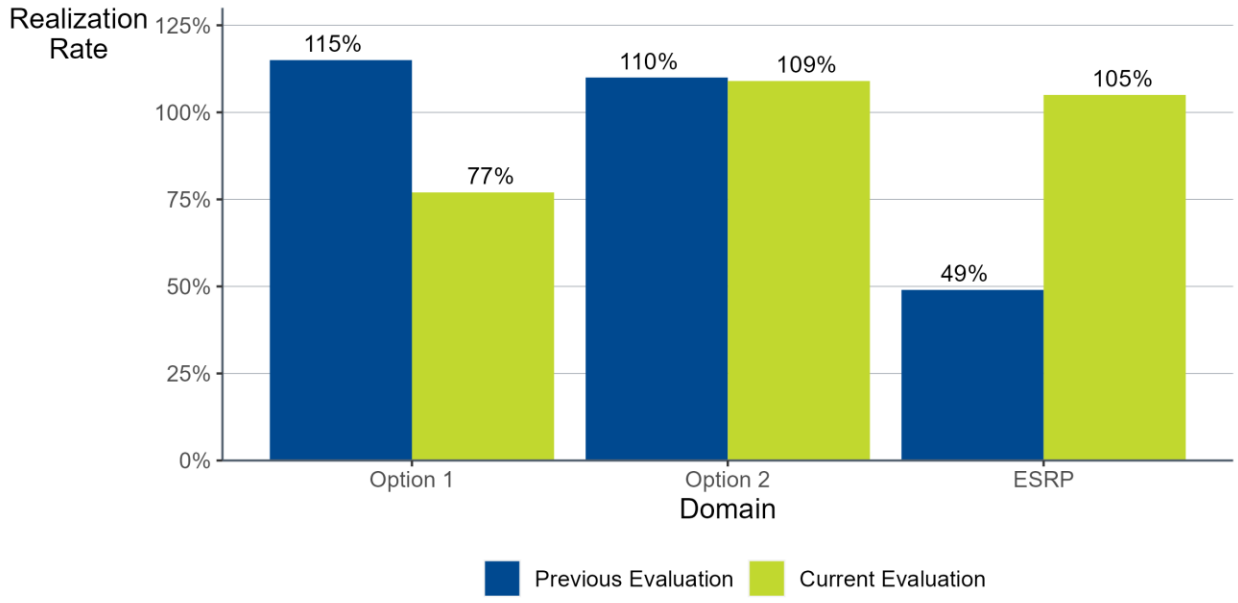
Table 5: Non-industrial custom evaluated sampling precision

Domain	Sampling Relative Precision (90% Two-Tailed)
Option 1	22%
Option 2	15%
Subtotal Option 1 and Option 2	12%
ESRP	0%
Total	**9%**

COMPARISON TO PRIOR EVALUATION

Option 1 results (77 percent) were lower than the last evaluation conducted on Option 1 non-industrial non-lighting projects (115 percent), while Option 2 results (109 percent) were comparable to the previous evaluation of Option 2 non-industrial non-lighting projects (110 percent). ESRP results (105 percent) were significantly higher than the last BPA impact evaluation of ESRP (49 percent). This comparison is based on RR2 results, excluding the tracking data issues.

Figure 2: Realization rate comparison across evaluations



4.2 MEASURE-SPECIFIC FINDINGS

Measure-level results (as presented in Section 4.1, RR2 results) based on the evaluation sample of 32 measures were highly variable, with realization rates ranging from 0 to 2.7 (Figure 3), with each measure in the sample represented by a single point. Measures are arranged along the x-axis by domain, with blue representing ESRP measures, orange representing Option 1, and teal representing Option 2. The size of each point corresponds to its evaluated savings value (expressed in average megawatts, or aMW). Option 1 realization

rates showed a negatively skewed distribution, with a few sites demonstrating realization rates below 1.0, bringing the weighted average realization rate of this group down to 77 percent. Option 2 had a few large measures with realization rates substantially higher than 1.0, resulting in an average realization rate of 109 percent. One relatively small ESRP measure had a large realization rate, but all other ESRP measures had realization rates close to 1.0, leading to a domain level realization rate of 105 percent.

Figure 3: Measure-level realization rates by domain

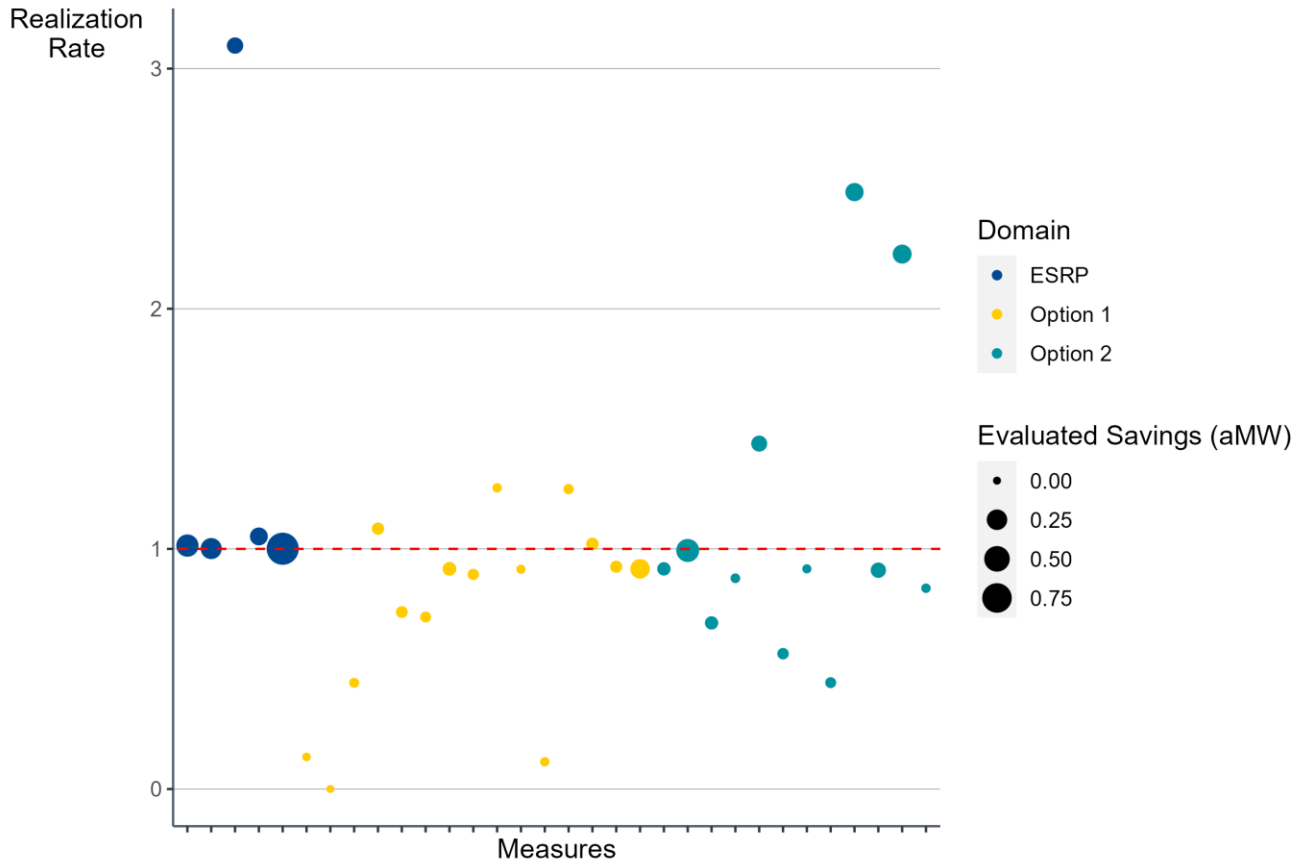
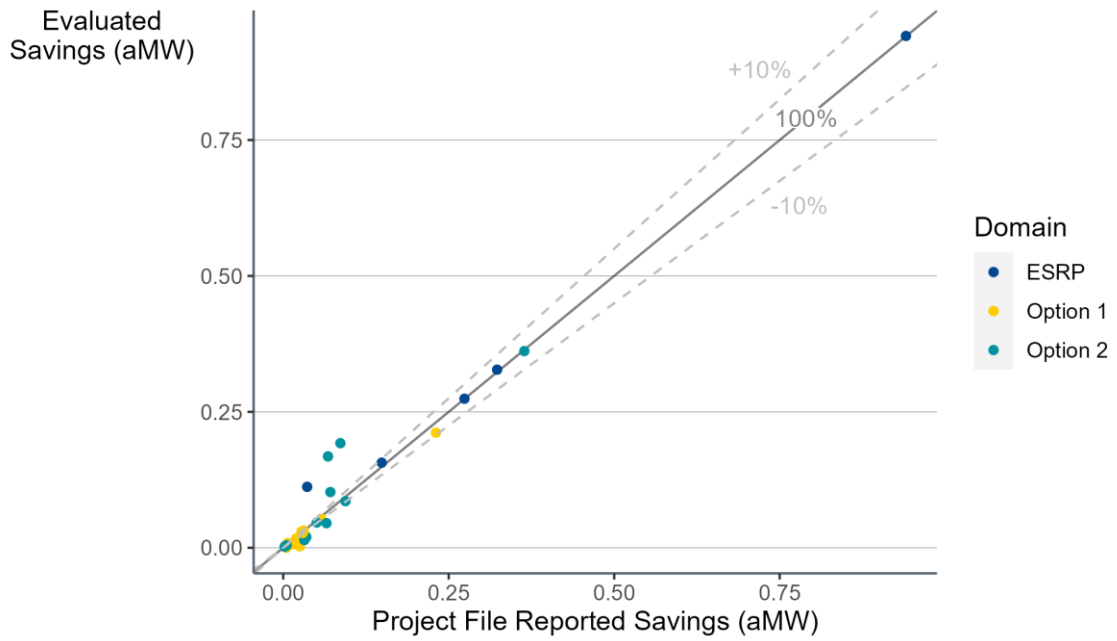


Figure 4 (on the next page) shows evaluated savings results by measure for the evaluation sample, expressed in aMW. Points lying above the gray diagonal line represent sites with evaluated savings higher than BPA project file savings, while those lying below the gray diagonal line represent sites with evaluated savings lower than project file savings. The dashed lines indicate +/- 10 percent of reported savings. Most Option 1 and ESRP projects have evaluated savings within 10 percent of reported savings. However, a handful of Option 2 measures have savings more than 10 percent above project file savings.

Figure 4: Evaluated savings versus project reported savings by measure

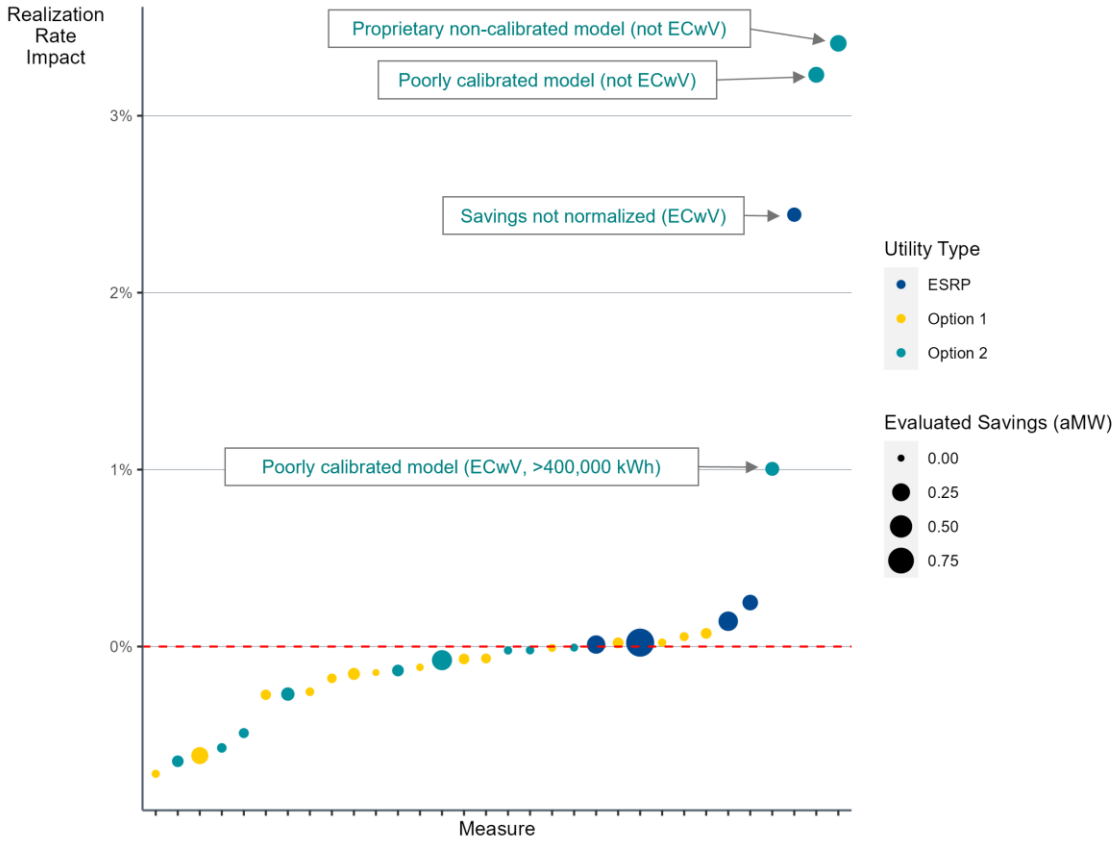


4.3 KEY DRIVERS OF SAVINGS DIFFERENCES

Figure 5 (on the next page) shows the impact of each reviewed project measure in the evaluation sample on the overall study realization rate. Project measures below the red dashed line are driving the realization rate below 1.0, while project measures above the line are driving the realization rate above 1.0. Most projects had little or no influence on the overall realization rate either because their realization rates were near 1.0 or their small size did not influence the total.

While most projects had realization rates near 1.0, there were several exceptions, with four relatively large measures with high realization rates raising the realization rate above 1.0, which drove the overall realization rate to be higher than 1.0. The callout boxes within the figure summarize the reasons for some of the most influential projects on the deviation in the realization rate from 1.0, which this report discusses in more detail after the figure.

Figure 5: Overall realization rate deviation from individual projects



The most influential project measures impacting the realization rate had the following issues:

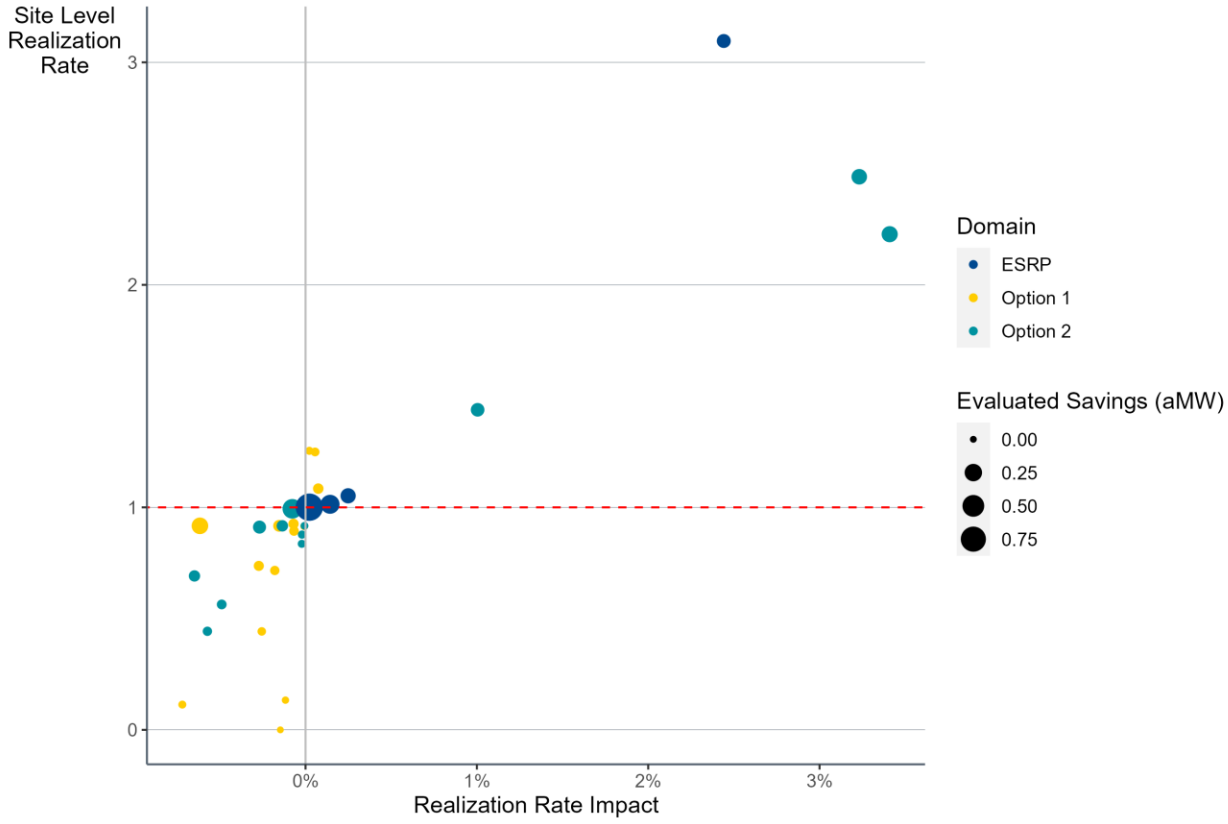
1. The largest contributor to deviations from a realization rate of 1.0 had **model calibration problems.** In the case of a new construction project where a building baseline was created with the DOE2 simulation engine but not provided to the evaluation team, a grocery store prototype model (EnergyPlus) was adapted instead. The evaluation team determined that the DOE grocery store prototype was an acceptable model to use as a baseline because its energy use intensity (EUI) falls within the range of average grocery store EUIs reported by the other investigated sources. Both the original model and the evaluation team's model vary considerably from the site's billing data.
2. The second most influential project **used an incorrect post-installation reporting period.** Savings for this project were calculated when the building was not yet back to normal operation following measure installation. Irregularities included fluctuations in occupancy, irregular HVAC scheduling due to the project, and incomplete commissioning. The evaluation team determined the date of return to normal operation and shifted the post period, leading to increased savings.
3. The third most influential project **calculated savings before installation was completed.** For this project, the claimed savings were estimated when the project was only at about one-third completion. The claim estimated savings using the ECwV protocol with the data that were available at the time, which only covered five months of operation. Once the project was completed fully, the losses observed during the first

few months were corrected. The evaluation calculated savings using additional post-install data which increased the savings achieved by the project.

4. The fourth most influential project **did not calibrate to bills.** One measure's savings had been calculated without baseline billing data, using the ECwV protocol, even though the project had higher than 400,000 kWh saved.[6] The evaluation team obtained billing data for both the baseline and post install cases and created an Energy Charting and Metrics (ECAM) model. The team used the weather normalized billing regression model to estimate savings that were greater than the savings claim calculated with Excel engineering calculations.

The measure impact map (Figure 6) shows the relationship between realization rate, size of measure savings, and resulting overall impact on the realization rate based on the evaluation sample results. The measure impact map combines all the information presented in the previous series of exhibits. The x-axis shows the impact on the overall realization rate, while the y-axis shows the measure-level realization rate. Measures in the lower left quadrant are driving the realization rate below 1.0, while sites in the upper right quadrant are driving the realization rate above 1.0. Larger dots represent larger measures, which generally increase their impact on overall realized energy savings.

Figure 6: Measure impact map



[6] ECwV is typically used for projects with savings less than 400,000 kWh.

4.4 LIFETIME SAVINGS

Table 6 shows the estimated evaluated lifetime savings for the sampled projects. The evaluated lifetime savings estimates are slightly higher than the reported lifetime savings. The Evergreen team did not make any adjustments to measure lifetimes because the team's review found that measure lifetimes for evaluated measures were exactly the same as stated for all projects. The team calculated evaluated lifetime savings based on the sample of projects, without weighting by population lifetime savings, because lifetime savings for the population were unavailable. As a result, the resulting lifetime savings realization rate of 105 percent for Option 1 and Option 2 measures is based on the sample, which is lower than the first-year savings realization rate (98 percent) that is weighted to the population.

Table 6: Lifetime savings for sampled projects

Utility Type	Evaluated Lifetime Savings (kWh)	Project File Lifetime Savings (kWh)	Lifetime Savings Realization Rate (RR2)
Option 1	48,246,962	59,707,403	81%
Option 2	80,001,192	62,875,972	127%
Subtotal	**128,248,153**	**122,583,375**	**105%**
ESRP	79,357,229	75,453,651	105%

4.5 ECWV SAVINGS ESTIMATION EFFICACY

Energy Calculation with Verification (ECwV) is a simplified approach to calculating energy savings for projects. Its use is supported for all projects with savings under 400,000 kWh and may be used on projects with savings greater than 400,000 kWh at the discretion of BPA. The evaluation looked at two issues with ECwV savings estimation methods:

1. Can ECwV provide an unbiased estimate of energy savings?
2. Did the Evaluator's implementation of BPA's ECwV M&V protocol provide an unbiased estimate of energy savings?

For the first question, the evaluation team estimated savings using an ECwV methodology and a regular high rigor evaluation method for all projects in the sample (Figure 7 on the next page).

Figure 7: Evaluated savings with and without ECwV

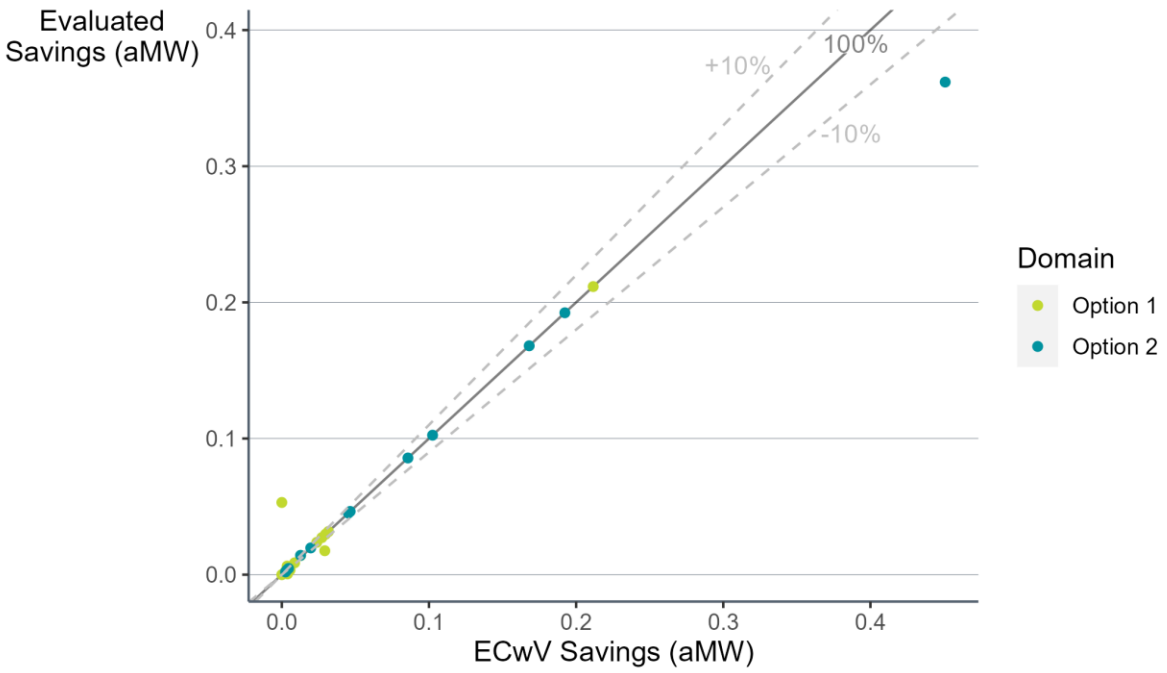


Table 7 (on the next page) compares evaluated savings with ECwV methodology to evaluated savings with high rigor methodology applied at the stratum level. The table shows the total savings of all projects in each stratum. If the ratio of the two is near 100 percent, it means that ECwV provided an unbiased estimate of the true savings. With an overall ratio of 104 percent between ECwV and evaluated savings, ECwV tends to show slightly higher savings than evaluated savings. The results below are for the sample only and are not weighted up to the population level.

Table 7: Evaluated savings with and without ECwV by size

Stratum	Savings (source=sample)				
	Evaluated ECwV Stratum Savings	Evaluated Stratum Savings	Ratio	Sample Size	Population
Option 1 Small	324,383	311,742	104%	7	66
Option 1 Medium	1,265,829	1,628,631	78%	7	17
Option 1 Certainty	1,853,665	1,853,665	100%	1	1
Subtotal	**3,444,077**	**3,794,038**	**91%**	**15**	**84**
Option 2 Small	5,050,757	4,279,373	118%	6	38
Option 2 Medium	4,883,902	4,883,902	100%	6	9
Subtotal	**9,884,659**	**9,163,275**	**108%**	**12**	**47**
Total	**13,378,536**	**12,957,313**	**103%**	**27**	**131**

4.6 COST-EFFECTIVENESS

Cost-effectiveness is calculated by summing up the net present value of costs (generally the incremental costs of the project) and comparing to the net present value of the benefits (including the value of energy and non-energy benefits (NEBs)). The custom non-industrial portfolio is strongly cost effective overall, based on the evaluation results, producing $1.99 in benefits for every $1 spent, as shown in Table 8.

Table 8: Benefit-cost results

Domain	Evaluated		
	Benefits ($)	Costs ($)	Benefit-Cost Ratio
Option 1	$3,002,577	$1,548,607	1.94
Option 2	$5,181,426	$2,914,383	1.78
ESRP	$4,317,239	$1,822,521	2.37
Total	**$12,501,243**	**$6,285,511**	**1.99**

5 KEY FINDINGS AND RECOMMENDATIONS

There are two distinct sets of findings. First, there were discrepancies between the project file savings and BPA's reported savings for Option 2 projects, as a result of tracking data issues. Three out of twelve Option 2 projects showed documented savings at least 70 percent higher than recorded in BPA's Energy Efficiency Tracking System (BEETS). There were also some differences in project file savings versus evaluated savings. The evaluation reports on these two sets of issues separately (two sets of interim realization rates) and also in combination (the overall realization rate) so that the tracking issues are separated from the site-level realized savings.

The population-level realization rates based on tracking data discrepancies (RR1) varied by domain: 100 percent for Option 1, 309 percent for Option 2, and 96 percent for ESRP. The realization rates based on the evaluated savings compared to BPA project file (RR2) savings (excluding tracking data discrepancies) also varied by domain: 77 percent for Option 1, 109 percent for Option 2, and 105 percent for ESRP. These RR2 results are most comparable to previous years' results.

The overall realization rate (RR3), including tracking data errors, is based on the product of the two interim realization rates with weights applied by stratum: 77 percent for Option 1, 337 percent for Option 2, and 101 percent for ESRP.

The portfolio's evaluation results excluding the tracking data issues showed realization rates very close to 100 percent, indicating that overall, the non-industrial custom portfolio is realizing the savings reported in the BPA project files. However, at the individual project level, the high degree of variability in realization rates leads to lower reliability of savings for individual projects than utilities and end-use customers may be expecting. The portfolio is highly cost-effective, generating approximately $2 in benefits for every $1 spent.

The remainder of this section offers suggestions for improving realized savings and the reliability of reported savings that, if taken, should lead to closer alignment of reported savings with evaluated savings for individual projects.

Key Finding: Discrepancies between project file savings and BPA-reported savings were common for Option 2 projects and contributed to a substantial portion of the differences between BPA-reported and evaluated savings. Upon investigation, there were no clear trends or explanations for these differences. The team suspects that the issues encountered during this evaluation were associated with BPA's transition of their tracking systems (i.e., from IS2 to BEETS).

Recommendation 1: The evaluation team recommends that BPA review Option 2 project savings, especially for larger projects, compared to BPA-reported savings to identify and correct any major discrepancies. Going forward, BPA should review its reporting procedures to ensure that the data entered into BEETS reflects the final documented project savings. Project documents should make the final total savings value by measure[7] clear. Making these improvements would improve the reliability of expected project savings for BPA, utilities, and

[7] A distinct measure is also referred to as a unique Technology/Activity/Practice (TAP) installed at a site.

end-use customers and improve impact evaluation realization rates (i.e., bring them closer to 100 percent).

Key Finding: Multiple projects did not accurately calibrate savings models to consumption or accurately normalize usage to typical usage.

Model calibration problems. BPA did not calibrate multiple projects to available consumption data for the pre-period or post-period. While this is expected for ECwV projects, there were problems observed with two projects that were not ECwV and a third project that had greater than 400,000 kWh savings but used an ECwV protocol.[8] A fourth project, which followed the ECwV protocol, was calibrated to post-installation data before the project had been commissioned. Of the 16 projects documented as ECwV during submission, 10 did not require calibration of post-energy consumption. Of the 16 measures that did not use ECwV for M&V, two measures were not calibrated prior to submission and required calibration during the evaluation. Both of these projects caused substantial changes in the overall realization rate.

Recommendation 2: Where feasible, for projects where the ECwV protocol is not being applied, BPA should calibrate savings to actual pre-installation and post-installation consumption data—including using post-installation consumption data for calibration after all commissioning is completed and any new equipment is in use, and normalizing savings estimates to typical weather, production, and other relevant variables—rather than relying on actual first-year usage. The guidelines should clarify the cases in which a model can or should be calibrated to billing data to maintain consistency across projects within the program.

Key Finding: Small- and medium-sized projects showed similar results in evaluated savings using the BPA ECwV protocol or high-rigor M&V methods. While there was variability from site to site and the ECwV results are less precise, ECwV results were similar to evaluated results overall. This finding is consistent with the results from a similar analysis conducted on BPA's Option 1 and Option 2 custom industrial portfolio. One project greater than 400,000 kWh was observed that used an ECwV protocol and would have benefited from additional rigor.

Recommendation 3: BPA should continue applying ECwV to the same size range of projects that it is using now. While the evaluation sample size is not large enough to provide a precise suggested size level, there is some evidence that ECwV is working relatively well for the projects to which it is currently being applied. This contrasts with the evidence that it does not work as well for larger projects.

[8] ECwV is typically used for projects with savings less than 400,000 kWh.

APPENDICES

APPENDIX A: DEFINITION OF KEY TERMS

This report relies on the following definitions of key terms.

Reporting System
BPA uses its reporting system to track projects completed by public power utilities under various programs and initiatives. For Option 1 utilities, BPA has detailed custom project proposals and completion reports in its system. Option 2 utilities and ESRP report high-level project information through a custom project completion report into the BPA system periodically.

Option 1 Utilities
Option 1 utilities are those that have opted to have BPA engineers manage and estimate the savings for custom projects at sites where site-specific characteristics will determine energy savings estimates.

Option 2 Utilities
Option 2 utilities are those that have opted to have their own engineers do the site-specific savings estimates and otherwise manage the implementation of custom projects with a BPA-approved measurement and verification plan.

Energy Smart Reserved Power (ESRP) Program
This program includes projects at Federally Chartered Irrigation Districts served by BPA Reserved Power, and BPA and other federal agency infrastructure projects served by Station Service. Projects are programmatically managed by BPA, and BPA engineers estimate the savings for custom projects at sites where site-specific characteristics will determine energy savings estimates.

Project
A project is a phase of work at an end user location that improves energy efficiency. An end user is the customer of a BPA utility. The project tracking data record a date when the project is complete. The data also contain information such as the name of the end user, the location where the work was carried out, and other data critical to this evaluation. End users may authorize the completion of many phases of work; BPA's reporting system tracks each of these phases as a separate project.

Measure
A measure is a distinct Technology/Activity/Practice (TAP) within a project. The BPA reporting system uses a standardized taxonomy (Technology/Activity/Practice) for classifying measures. For most projects, BPA or utility staff assign one of several possible TAP descriptions to each physical measure or change implemented as part of a project.

Project Engineers
Project engineers and program staff assist in the identification, development, savings estimation, cost-effectiveness analysis, M&V, and quality control review of projects. Project engineers may be BPA staff, utility staff, or staff of BPA or utility project implementation contractors. For the purposes of this evaluation, project engineers are not staff or contractors employed by the end users, even though the end user workforce may have played an important role in the development of a project.

M&V Model
Project engineers use this M&V model (an algorithm or calculation procedure) to estimate savings for the measures that comprise a project.

Reported Savings
Reported savings are the savings estimated by the project engineers and entered in the BPA reporting system based on the M&V model. Please note that the BPA system uses the term "estimated savings" for the savings estimated at the proposal stage and "actual savings" for the savings at the report completion stage. The BPA Implementation Manual does not require all projects to submit a formal proposal. Reported savings are the "actual savings" field in the reporting system. "Actual savings" is busbar savings (see the term busbar savings defined later in this section), equal to 1.09056 times site savings.

Evaluation Savings
Evaluation savings are the savings estimated by the evaluation team based on the evaluation model and rely on best practical data collection and savings estimation practices, as laid out in the RTF guidelines and informed by evaluator experience. The evaluation estimated the savings achieved during the first year of measure operation. If any of the evaluation data collection occurs more than one year after the measure was complete, it may indicate failures in the measure performance that are relevant to measure lifetime and not to the first-year savings. Evaluation savings estimates reflect the conditions of the measure during the first year of operation and are adjusted using the associated busbar factor.

Existing Condition Baseline
For retrofit measures, the baseline would include the efficiency of existing equipment with remaining useful life.

Current Practice Baseline
BPA and the RTF have different definitions of current practice baseline for custom measures, based on BPA's M&V selection guide[9] and RTF guidelines,[10] respectively.

BPA: *"When the practitioner uses a current practice baseline, the efficiency level of the baseline equipment must be consistent with any state or local mandates for new equipment, which may vary from city to city and state to state."*

RTF: *"The practitioner needs to identify what would normally be done, based on prior experience with similar projects. The practitioner should start by using applicable codes and standards, or one of the following if they constitute a more energy efficient baseline for the measure and the information is practical to obtain and applicable to the delivered measure's location OR there is no applicable code or standard for the measure implemented.*

- *Recent similar purchases by the end user*
- *Documented end user plans or specifications*

---

[9] Bonneville Power Administration (BPA). 2018. *Measurement & Verification (M&V) Protocol Selection Guide and Example M&V Plan*. Page 7. https://www.bpa.gov/-/media/Aep/energy-efficiency/measurement-verification/1-bpa-mv-selection-guide.pdf (file will download automatically)
[10] Regional Technical Forum. 2020. *Guidelines for the Assessment of Energy Efficiency Measures*, Section 4.3.3, page 20. https://nwcouncil.app.box.com/v/2020RTFGuidelines

- *End user or vendor-developed alternative designs, considered as part of the measure selection process*
- *End user description of what was done in similar circumstances elsewhere in the facility or in another facility they operate*
- *Equipment vendor's description of what they would normally do for this end user"*

Realization Rate
Realization rate is the ratio of evaluation savings to reported savings. Realization rates greater than 1.0 mean that the evaluation savings were greater than reported savings.

Key Determinants
Key determinants influence the savings from a measure. The evaluation considered the following key determinants:

- **Connected load.** Baseline or efficient-case rated kW demand and/or the quantity of the equipment
- **Efficiency profile.** Part-load impacts on demand profile, including VFDs and HVAC interaction factors
- **Hours of operation.** Baseline or efficient-case schedule of operation for a measure
- **Load profile.** Facility occupancy rates and changes not captured by other categories of key determinants
- **Production.** Number of production units per unit of time
- **Weather.** Weather-based data used for weather-sensitive measures, such as dry and wet-bulb temperatures, or heating and cooling degree-days

Reasons for Difference
The reasons for difference are what was changed that caused a modification to one or more key determinants and ultimately savings. The evaluation team ranked impacts as causing a primary or secondary change in savings to give a sense of their scale, and assigned all reasons to one of the following categories:

- **Documentation error.** These include errors in calculations or values entered into models.
- **Other.** Commonly, a change in inputs due to a contradicting finding in the first year. This would indicate that the value for the key determinant in the project documentation was correct, but the value entered in the savings calculation did not match what was in the documentation. It could also indicate that the key determinant in the project documentation did not match what was found during the site visit or in trend data.

Measure Baseline
Measure savings must be determined against clearly defined baseline conditions. The RTF guidelines define two possible baseline conditions that were used in this evaluation:

- **Current practice.** A current practice baseline is used if the measure affects systems, equipment, or practices that are at the end of their useful life. The baseline is defined by the recent typical choices of the end user in purchasing new equipment and services. Current practice baseline is also used for new construction projects where there is no pre-existing systems, equipment, or practices.

- **Pre-conditions.** A pre-conditions baseline is used when the measure-affected equipment or practice still has remaining useful life. The baseline is defined by the existing condition at the end user site just prior to the delivery of the measure.

ProCost Model
ProCost is a spreadsheet tool developed by the Northwest Power and Conservation Council that computes regional measure lifecycle cost-effectiveness. ProCost uses regional economic and power system assumptions that are updated with each Council Power Plan.

Measure Lifetime
Measure lifetime, according to the RTF guideline for lifetime savings, is defined as the median number of years during which at least half the deliveries of a measure are in place and operable, i.e., producing savings. For example, consider the installation of 100 VFDs on pumps. If the VFDs were regularly inspected for many years, it would be possible to determine when each one became inoperable (failed mechanically or electrically or was removed from service). The lifetime for the measure would be the median number of years to measure failure, i.e., no longer producing savings. An estimate of measure lifetime is a required input to ProCost.

Incremental Costs and Benefits
When a measure is delivered, costs are incurred and benefits realized—e.g., the value of electricity savings and other nonelectric benefits, such as changes in operations and maintenance expenses. Only incremental costs and benefits are used in estimating life cycle costs and benefits.

A measure's incremental costs and benefits are those incurred in the efficient case delivery, beyond what is required to establish and maintain the baseline condition. For a precondition baseline, the baseline does not involve any change, and thus baseline costs and benefits are zero. In this case, incremental costs and benefits are equal to the efficient case costs and benefits. For measures with a current practice baseline, the baseline condition does require a change and therefore has costs and benefits. In this case, the incremental costs are the difference between the efficient case and the baseline case delivery.

NEBs (Nonelectric Benefits)
Nonelectric benefits are defined as any benefit, positive or negative, that the value of the electric savings or the measure incremental cost does not capture. NEBs include changes caused by the measure in the costs of operation and maintenance or other utilities such as gas, water, or wastewater. The RTF guidelines further explain these benefits (see the guideline for the estimation of incremental measure costs and benefits).

Total Resource Cost (TRC) Test
The TRC is one type of cost-effectiveness testing that includes all incremental cost and lifetime benefits of a measure, regardless of who pays for or receives them. BPA uses the definition of the TRC test consistent with the Northwest Power and Conservation Council.

Busbar Factor

Busbar factor is a term used to characterize transmission and distribution energy loss between a source of generation and the point of consumption. Busbar factors applicable to the year of evaluated projects are applied to savings calculations to represent additional savings due to electricity not needing to be transmitted.

APPENDIX B: SITE-SPECIFIC SAVING ESTIMATION

Table 9 provides the measure level results for the sample.

Table 9: Measure level results for the sample

Sample ID	Measure ID	Stratum	Measure Realization Rate	Measure Impact on Overall Realization Rate
103	16101	Option 1 Certainty	0.92	-0.62%
104	5901	Option 1 Medium	0.89	-0.07%
104	6001	Option 1 Small	1.25	0.02%
105	15901	Option 1 Medium	1.02	0.02%
106	5301	Option 1 Medium	0.92	-0.15%
107	16001	Option 1 Medium	0.93	-0.07%
108	3501	Option 1 Medium	1.08	0.07%
109	7201	Option 1 Medium	0.11	-0.72%
109	7202	Option 1 Small	1.25	0.06%
110	4401	Option 1 Medium	0.74	-0.27%
111	3001	Option 1 Small	0.44	-0.26%
112	701	Option 1 Small	0.00	-0.15%
113	201	Option 1 Small	0.13	-0.12%
114	5001	Option 1 Small	0.72	-0.18%
115	6501	Option 1 Small	0.92	-0.01%
118	13701	Option 2 Medium	2.23	3.41%
119	13601	Option 2 Medium	0.91	-0.27%
120	13501	Option 2 Medium	2.49	3.23%
121	11601	Option 2 Medium	0.69	-0.65%
122	13002	Option 2 Small	0.44	-0.57%
123	14701	Option 2 Small	0.84	-0.02%
124	12401	Option 2 Small	0.92	-0.01%
125	11901	Option 2 Small	1.44	1.00%
126	12201	Option 2 Medium	0.56	-0.49%
127	11801	Option 2 Small	0.88	-0.02%
128	11101	Option 2 Medium	0.92	-0.14%
129	11501	Option 2 Small	0.99	-0.08%
131	14403	ESRP Small	1.05	0.25%
132	9301	ESRP Small	3.10	2.44%
137	4101	ESRP Certainty	1.01	0.14%
138	9101	ESRP Certainty	1.00	0.01%
139	123309	ESRP Certainty	1.00	0.02%

APPENDIX C: DETAILED METHODOLOGY

This appendix provides more detail on the study methods (sample design, data collection, and analysis).

SAMPLE DESIGN

BPA's evaluation policies have established a target for impact evaluation, striving for evaluations that attain a relative error of 10 percent at the 90 percent confidence level, with a minimum acceptable level of 80/20. The sampling strategy targeted a 90/10 confidence level and precision for the non-industrial custom and ESRP evaluation.

The Evergreen team prepared a detailed sample design for both Option 1 and Option 2 utilities as well as facilities with no utility that the Federal Columbia River Power System (FCRP) serves directly. The sample was based on the custom projects in IS2.0 data, pulled in mid-Jan. 2024 and comprised of projects that completed invoicing in the most recent and complete year (fiscal year 2023) at the time of the sample pull.[11] The Evergreen team excluded industrial sites and Option 1 and Option 2 utility lighting projects, as they were recently evaluated. The sampling unit of this study is a measure, defined as a unique TAP for a single site.[12] Evergreen defines each of these items below:

- **Measure:** A unique TAP installed at a site (i.e., three VFD upgrades at the same location invoiced at the same time would be one measure)
- **Project:** All TAPs invoiced at the same time at the same site, regardless of measure type (i.e., A VFD upgrade and new pump installed at the same site and invoiced for at the same time, would be one project, but two measures)
    - While not used for sampling, BPA enters data into IS2.0 as projects; therefore, the file review will request project files as part of the evaluation.
- **Site**: a distinct location where a measure is installed
    - Even if the locations are within the same address/facility, if transit time is required between the two locations, they are two sites (i.e., two buildings at one college campus may have the same address but are not within walking distance. These are two sites)

Evergreen developed a sample design that utilizes Dalenius-Hodges stratified random sampling based on domain and reported kWh savings for the measure. The evaluation team chose this method since it is the most appropriate for maximizing sampling precision across strata that are defined by measure size, given a fixed sample size.[13] The team selected a savings stratified random sample designed to achieve at least an expected precision of +/- 20 percent at an 80 percent level of confidence. This method reduces the variability in the sampling universe by segmenting the measures into more homogenous groups. Evergreen defined an excluded stratum (i.e., stratum 0) that contains very small measures; this is the group of measures that collectively account for less than 1 percent of the savings within each

[11] Going back more than a year is more difficult for utilities and end-use customers due to staff turnover, program changes, and misplaced documentation.

[12] For uniformity of the evaluation approach, evaluation and project resource management, and cost control, sampling is based on measure.

[13] Dalenius-Hodges is preferred to Neyman allocation when evaluators stratify on *ex ante* savings as opposed to end use (e.g., HVAC, refrigeration). Sample points are allocated evenly across strata, when possible. If a stratum is filled, any remaining points are allocated to the next largest stratum.

domain. The Evergreen team assigned measures that represent a significant portion (more than 10 percent of the domain) of total reported energy savings to a priority "certainty" stratum. The team considered these measures necessary for the evaluation; thus, they were not subject to random selection. The team then allocated moderately sized measures to probabilistic strata.

Table 10 shows the number of non-industrial custom measures and custom ESRP program measures with their associated savings in our sample frame by domain and size strata. The sample sizes shown in the table yielded a relative precision of +/- 12 percent at a 90 percent confidence level for the evaluation over the 12-month period. At the domain level, the samples for Option 1 yielded +/- 25 percent, and Option 2 yielded +/- 16 percent at a 90 percent confidence level. The evaluation conducted a census of ESRP sites.

The sample extract using this design contained 32 measures from 30 unique sites with 22 unique TAPs (i.e., distinct measure types). These values reflect the data contained in the sample measure project files.

Table 10: Non-industrial custom and ESRP sample design

Utility Type	Strata*	Reported Savings (kWh)		Number of Reported Measures	Sample Size (Measures)
		Average	Total		
Option 1	0	7,029	168,688	24	0
	1	85,392	5,635,898	66	7
	2	343,866	5,845,718	17	7
	Certainty	2,021,533	2,021,533	1	1
	Subtotal	**126,591**	**13,671,838**	**108**	**15**
Option 2	0	10,646	212,920	20	0
	1	114,326	4,344,379	38	6
	2	511,820	4,606,383	9	6
	Subtotal	**136,771**	**9,163,681**	**67**	**12**
Federal/ESRP Program	1	849,292	1,698,583	2	2
	Certainty	4,677,559	14,032,677	3	3
	Subtotal	**3,146,252**	**15,731,260**	**5**	**5**
Total		**214,260**	**38,566,779**	**180**	**32**

*Stratum *0* denotes the excluded projects (based on very small *ex ante* savings). The *certainty* projects represent a significant portion of total reported energy savings within the domain and are considered as necessary for the evaluation and therefore, the evaluation team did not subject them to random selection.

DATA COLLECTION

The Evergreen team developed procedures for data collection, adapting the already developed procedures for the recently completed evaluation of BPA's 2019-2021 custom industrial portfolio.

The team closely coordinated with BPA to notify utilities about projects selected for the study sample and provided them with the necessary information consistent with the communication protocols developed for this study (see Appendix B). The team developed materials for and hosted a data collection webinar to introduce sampled utilities to the study, notified them of data collection activities (including end-user contact protocols), and clarified roles and expectations.

The study included collecting data from at least 32 sites. The Evergreen team tracked and recorded dispositions for completed sites in an Excel-based project tracker. The team updated and shared the tracker with BPA on a weekly basis to record the status of each site and relevant information about the site (e.g., utility, assigned engineer, number of contacts made, level of complexity). The tracker supported any follow-up required by BPA to ensure response by end users and utilities.

In accordance with the research plan and approved communication protocols, the team collected data with the following methods:

- **File review.** The file review involved extracting all project information relevant to savings estimation including measure descriptions, baseline or efficient condition inputs, reported savings values, and the final version of the M&V model.
- **Telephone/email discussion with technical staff who were involved with the project.** Technical staff at the utility (for Option 1 utility customers) or BPA and/or vendors (for Option 2 utility and federal infrastructure/ESRP customers), were another possible source of data. As needed, the team contacted them by telephone or email to obtain information needed for the evaluation that the team did not find in the project files.
- **Telephone/email discussion with end users.** In some cases, the evaluation team needed to obtain information from the end user using telephone or email contacts.
- **Site visits.** Based on the file review and discussions with project staff, the contractor conducted in-person interviews with operation staff, reviewed electrical plans, and inspected control settings. They also reviewed manufacturers' specifications, took one-time measurements, and/or conducted short-term metering to gather more information from inspection of affected systems and equipment.
- **Affected system trend metering.** When necessary, the team collected additional metering data, such as billing or interval premise electricity usage data, if the initial metering data were not sufficient.

The data collection for project sites was a mix of low, medium, and high effort, based on project size and complexity. The evaluation team determined final sample size and quantity of each type once the team gathered and reviewed the project files. Assumptions regarding level of effort were as follows:

- **Low Effort:** Typically, did not require a new model or data logging. Onsite visits might have required only a nameplate verification and one-time reading, or an output of data from a control system.

- **Medium Effort:** May have required some trend logging for key equipment and/or a simple new model
- **High Effort:** Required logging or trend data for multiple parameters

Assumptions for level of effort within the initial sample erred on the side of more complex sites (higher effort).

SITE-LEVEL MEASURE ANALYSIS

The Evergreen team first estimated savings for the sample of custom and ESRP measures as described below.

- **Reviewed Project Energy Models:** The Evergreen team based the evaluation model on what it deemed to be reasonable and reliable. The team started by reviewing the existing energy model to confirm the model conformed to BPA's M&V protocols and assessed savings calculations to determine reliability of savings estimates. Evaluation versions of the energy models were updated to accurately represent the conditions observed during evaluation data collection. If needed, the team created a new energy model.
- **Assessed Determinant Reliability and Collected Supplemental Data:** The Evergreen team developed a list of critical determinants for each project, where critical was defined as having a significant (possibly 10 percent or more) impact on the calculated savings. The team then found and assessed the corresponding values used in the evaluation model and deemed whether they were reliable or not. For unreliable critical determinants, the evaluation team assessed what level of data collection involving the end user was necessary to obtain reliability for that determinant (telephone/email interview, site visit, or metering). The team gathered supplemental data as needed to support sufficiently reliable savings estimates.
- **Ran the Model and Estimated Evaluated Savings:** After the Evergreen team confirmed or obtained reliable determinant values through data collection, the team ran the evaluation model for each site and estimated site-level energy savings.
- **Treatment of Interactive Measures:** Savings achieved by one measure may have affected the savings of another measure—for example, a lighting upgrade that coincided with an HVAC upgrade affecting the same spaces within a building. If the two improvements occurred as part of separate projects that were completed at different times, this was not an issue for this evaluation. However, an issue could have arisen if one or more projects were completed at the same time. Using information collected from the reporting system, project engineers, and end users, the Evergreen team determined whether this occurred for any of the measures in the sample. If it did, the team obtained documentation for all the interactive measures to determine how the M&V models accounted for the interactions. To account for measure interaction in estimating the evaluation savings, the evaluation team identified and used the same measure order that was assumed when estimating each measure's savings.
- **Time-Based Value of Savings and Cost Effectiveness.** The Evergreen team assigned load shapes to individual measures using ProCost via BPA TAP reporting code. The team then calculated cost effectiveness and peak savings based on the generic calculator and project-specific ProCost analyses and reported on any differences.

The Evergreen team coordinated with BPA throughout the data collection and analysis process to ensure the data and information used to develop independent estimates of savings are accurate and complete.

STUDY DOMAIN ANALYSIS

Once the Evergreen team completed analysis for the sample, the team compiled site-level results to estimate the electric savings and cost-effectiveness for the portfolio using a ratio analysis. The team estimated typical first-year savings for Option 1, Option 2, and ESRP subdomains using the evaluation model results for the sample, weighted to reflect the proportion of the population represented by each stratum.

Prior to portfolio analysis, the team developed an analysis template workbook for BPA review. This workbook served as a template for conducting the subdomain and portfolio-level rollup calculations. The subdomain-level rollup analysis took the evaluation results for the sample of sites, extrapolated them to the stratum and subdomain levels, and ultimately calculated stratum and subdomain-level results.

After completing the analysis, the Evergreen team delivered a results workbook, including aggregated electricity savings and cost-effectiveness results for each subdomain.