MUNI-LCC: An Open Toolkit Linking Measured Energy Savings, Life-Cycle Cost, and Federal Performance Reporting for Municipal Energy Projects

Stephen Olumide Akindipe [1]

[1]*Independent Researcher, Albany, NY, USA*
Correspondence: oluakins1108@gmail.com

Energy upgrades and retrofits in small and mid-sized municipalities are funded through federal and state awards. Recipients of federal awards must report project accomplishments against award objectives, provide reasons for objectives not met, and account for any cost overruns or high unit costs pursuant to 2 CFR 200.329. The existing methods for evaluating savings over the project's lifespan are standardized and mostly prepared through the support of a consultant or a licensed software. This financial implication creates a burden for municipalities with budget constraints. MUNI-LCC addresses this challenge: It is an open-source python toolkit that implements Options A and C of the International Performance Measurement and Verification Protocol as defined in the U.S. Department of Energy FEMP M&V Guidelines Version 5.0. It implements life-cycle cost analysis consistent with ASTM E917 through the NIST Handbook 135 framework and integrate the content elements of 2 CFR 200.329 to ensure accurate reporting. MUNI-LCC introduces a distinguishing feature whereby verified quantities directly feed the life-cycle cost study to ensure the net savings, savings-to-investment ratio, and unit cost reported to the sponsor is computed from the metered savings.
Validated using a synthetic worked example with a known answer, a 3PH change-point baseline (CV(RMSE) 7.9%, $R^2 = 0.988$) yields 6,879 therms of avoided gas use against a true 6,629 therms, inside the 90% uncertainty of $\pm$2,352 therms. In 2,000 Monte Carlo replications of the same data-generating process, the ASHRAE Guideline 14 uncertainty approximation applied for Option C attains empirical coverage of 90.0% at a nominal 90%, and mean estimation error is -17 therms. The outcome of these tests led to two changes implemented in version 0.1.1: information-criterion model selection and an uncertainty reported for any sign of savings. These changes do not alter the published result of the worked example. Sensitivity scans demonstrate how measured-savings uncertainty and the discount rate propagate to net savings. The toolkit is released under the MIT license and archived on Zenodo (DOI 10.5281/zenodo.23825887).

Keywords: measurement and verification, IPMVP, life-cycle cost analysis, change-point regression, savings uncertainty, Uniform Guidance, Monte Carlo simulation, energy efficiency retrofits

I. INTRODUCTION

Energy efficiency and heating-system upgrades conducted on public or municipal buildings are often funded through various financial mechanisms, including local capital budgets, utility programs, and grants from federal and state agencies. For projects funded through federal awards, recipients are required to provide a program performance report to maintain compliance with the Uniform Guidance established in 2 CFR 200.329. The program performance report is a comparison of the project's actual accomplishments based on the award's objectives. Therefore, recipients are expected to provide reasons why certain objectives were not met, provide cost analysis and reasons for excess expenditure or high unit costs. For compliance purposes, this report must be submitted by the stated due dates [1]. For such energy innovation projects, the accomplishment is an estimate of the amount of energy not consumed due to the technological implementation. This accomplishment cannot be computed directly from a meter. It is the calculated difference between the metered energy use after the retrofit and the estimate without the retrofit.

The International Performance Measurement and Verification Protocol (IPMVP) standardizes these estimates. The most recent edition is the IPMVP Core Concepts 2022 which defines four core measurement and verification options [2]. The implementation of this protocol is administered by The U.S. Department of Energy Federal Energy Management Program (FEMP) M&V Guidelines Version 5.0, prepared by Pacific Northwest National Laboratory and issued in September 2024. This is a functional application of the IPMVP for monitoring federal performance-based contracts and provides quantitative acceptance values for regression baselines [3]. Model statistics and approximation for the fractional uncertainty of savings is sourced from ASHRAE Guideline 14 [4], and the Efficiency Valuation Organization publishes separate uncertainty guidance [5]. For the economic component, ASTM E917 defines the life-cycle cost (LCC) analysis for buildings [6]; NIST Handbook 135 provides the public federal formulation [7], and the NIST updates and publishes discount rates and energy price indices every fiscal year [8].

Although these methods are credible and widely applicable, their application requires statistical modeling, present-value accounting, and familiarity with grant reporting. In most cases, large agencies satisfy these requirements by using energy service contracts, consultants, and licensed software. However, recipients in small or mid-sized municipalities with budget constraints or limited staff capacity lack measurement and verification (M&V) staff or software. The best practical alternative to measure savings is to provide an engineering estimate which does not reflect the actual savings in the project. Hence, the reported accomplishments are considered a projection rather than an actual result.

Several open-access tools address some of these workflow limitations. For example, NIST's Building Life-Cycle Cost program (BLCC) and the Economic Evaluation Engine (E3) calculate life-cycle and benefit-cost measures under the FEMP rules [9, 10]. OpenEEmeter implements the CalTRACK daily,

billing and hourly methods [11]; nmecr, developed by kW Engineering for Normalized Metered Energy Consumption (NMEC) implements change-point, time-of-week, and temperature (TOWT) models based on ASHRAE Guideline 14 uncertainty [12]. LBNL's RMV2.0 supports baseline modeling and non-routine event screening on interval data [13]. Although these tools offer significant contributions, none of these tools supports the integration of verified savings into life-cycle cost metrics or to the content of a federal performance report.

This paper introduces MUNI-LCC, an open-source Python toolkit built to provide the integration between savings and LCC estimates. Contributions in this paper include: (i) Implements IPMVP Options A and C as established in FEMP v5.0, with every acceptance threshold stored alongside the section of the source document. (ii) Computes LCC, net savings (NS), savings-to-investment ratio (SIR), adjusted internal rate of return (AIRR), and simple and discounted payback consistent with ASTM E917 framework. It can also compute actual annual energy quantities from verified M&V results rather than projections. (iii) For effective grant reporting, it organizes the regulatory performance-report content established in 2 CFR 200.329(c)(1), (c)(2), and (e), automates and computes the due date, and lists every explanation required by the regulation that the user has not supplied to ensure accurate reporting. (iv) This toolkit is validated through a test against a synthetic worked example with a known injected savings value, and every metric in its documentation is rendered from a live statistics file regenerated based on continuous integration. Beyond the software description, this paper reports a Monte Carlo evaluation of the Option C uncertainty approximation based on the worked example's data-generating process, the two modifications influenced by the evaluation, and the sensitivity scans of the life-cycle cost results.

The paper is structured as follows. Sec. II sets out the methodological framework and equations implemented by the toolkit. Sec. III explains the implementation. Sec. IV discusses the worked example and results. Sec. V reports the Monte Carlo validation and sensitivity analysis. Sec. VI outlines the limitations, and Sec. VII is the conclusion.

II. METHODOLOGICAL FRAMEWORK

A. Applied Standards

Table I contains the frameworks applied for the development toolkit and their respective functions. Of these frameworks, ASTM E917 and ASHRAE Guideline 14, are copyrighted and were not reproduced. The life-cycle cost (LCC) equation is consistent with NIST Handbook 135 and the public federal formulation of the same practice [7]. In accordance with ASHRAE Guideline 14, the published model statistics and the fractional savings uncertainty approximation were applied in this toolkit (Sec. II D).

TABLE I: Frameworks implemented in MUNI-LCC toolkit

Source	Functions Applied
FEMP M&V Guidelines v5.0 (2024) [3]	Definitions of Options A and C; R^2, bias, and t acceptance values; whole-year baseline guidance
IPMVP Core Concepts 2022 [2]	Protocol on which FEMP v5.0 is based
ASTM E917-17(2023) [6]; NIST Handbook 135 [7]	LCC method, present-value factors, supplementary measures
NIST IR 85-3273-40 (2025) [8]	Discount rate (3.0%) and 40-year maximum study period
ASHRAE Guideline 14-2023 [4]	Model statistics; fractional savings uncertainty approximation
LBNL (2017) [14]	CV(RMSE) and FSU thresholds where FEMP states none
2 CFR 200.329 (2024) [1]	Report due dates and content elements

B. Life-cycle cost

All amounts applied in this equation are in constant base-year dollars discounted at a real rate d over a study period of n years, with costs at the end of each year by default and an optional mid-year convention for recurring costs. The present-value factors are:

$$\mathrm{SPV}(t,d) = (1+d)^{-t}, \tag{1}$$

$$\mathrm{UPV}(n,d) = \frac{1-(1+d)^{-n}}{d}, \tag{2}$$

$$\mathrm{UPV}^{*}(n,d,e) = \sum_{t=1}^{n} \left(\frac{1+e}{1+d}\right)^{t}, \tag{3}$$

where e is the constant real escalation rate. When a year-by-year price index I_t is supplied, as reported in the annual NIST supplement, the modified present-value factor is $\mathrm{UPV}^{*} = \sum_t I_t (1+d)^{-t}$. For the alternative LCC, the equation is defined

$$\mathrm{LCC} = C_0 + \mathrm{PV}_{\mathrm{repl}} - \mathrm{PV}_{\mathrm{res}} + \mathrm{PV}_{\mathrm{E}} + \mathrm{PV}_{\mathrm{W}} + \mathrm{PV}_{\mathrm{OM\&R}}, \tag{4}$$

where C_0 is the initial investment. Replacements and residual value discounted by Eq. (1). Energy, water and operation, maintenance and repair (OM&R) costs discounted by Eq. (3). According to NIST Handbook 135, investment-related costs are $I = C_0 + \mathrm{PV}_{\mathrm{repl}} - \mathrm{PV}_{\mathrm{res}}$. Operational costs are $O = \mathrm{PV}_{\mathrm{E}} + \mathrm{PV}_{\mathrm{W}} + \mathrm{PV}_{\mathrm{OM\&R}}$. Relative to a base case as the baseline,

$$\mathrm{NS} = \mathrm{LCC}_{\mathrm{base}} - \mathrm{LCC}_{\mathrm{alt}}, \tag{5}$$

$$\mathrm{SIR} = \frac{O_{\mathrm{base}} - O_{\mathrm{alt}}}{I_{\mathrm{alt}} - I_{\mathrm{base}}}, \tag{6}$$

$$\mathrm{AIRR} = (1+d)\,\mathrm{SIR}^{1/n} - 1. \tag{7}$$

These equations satisfy $\mathrm{NS} > 0 \Leftrightarrow \mathrm{SIR} > 1 \Leftrightarrow \mathrm{AIRR} > d$, which serves as the test suite checks. Simple and discounted payback are defined as the years until cumulative undiscounted or discounted operational savings equal the added initial investment cost, interpolated within the crossing year. Costs dated after the study period are excluded from this calculation. The study period is capped at 40 years, as the allowable maximum period enforced by FEMP legal guidelines.

TABLE II: Option C baseline model forms. E is energy use per day, T as the period mean temperature, c a change point, and p the number of parameters calculating change points.

Form	Model	p
mean	$E = \beta_0$	1
2P	$E = \beta_0 + \beta_1 T$	2
3PC	$E = \beta_0 + \beta_1 (T-c)^+, \beta_1 > 0$	3
3PH	$E = \beta_0 + \beta_1 (c-T)^+, \beta_1 > 0$	3
4P	$E = \beta_0 + \beta_1 (T-c)^- + \beta_2 (T-c)^+, \beta_1 < \beta_2$	4
5P	$E = \beta_0 + \beta_1 (c_H - T)^+ + \beta_2 (T-c_C)^+, \beta_{1,2} > 0$	5

C. IPMVP Option A

Option A isolates a retrofit and measures its key parameter or most critical performance parameter, while stipulating others that are less critical. For a lighting retrofit, the key parameter is the connected demand while the stipulated parameter is annual operating hours:

$$S_A = (\bar{P}_b - \bar{P}_r) H \phi, \tag{8}$$

where $\bar{P}_b$ and $\bar{P}_r$ are the means of field readings before and after the retrofit, H is the stipulated parameter, and ϕ is an optional interactive-effects factor (default 1). According to the standard, FEMP v5.0 requires the M&V plan to justify stipulated or estimated values [3]; the toolkit refuses a stipulation without a written basis. The uncertainty equation is

$$U_A = t_{1-\alpha/2,\nu} \sqrt{\mathrm{SE}_b^2 + \mathrm{SE}_r^2}\, H\phi, \qquad \nu = n_b + n_r - 2, \tag{9}$$

where SE is the standard error of each mean. The root-sum-square combination matches the FEMP v5.0 rule for independent error sources (Eq. 5-1). The stipulated parameter is treated as exact by agreement and the output reports the same.

D. IPMVP Option C

Option C uses the whole-facility meter where the savings are calculated from the utility data analysis of the entire facility. For each billing period i with length Δ_i days, total use y_i, and mean outdoor dry-bulb temperature T_i, the toolkit models use per day $E_i = y_i/\Delta_i$ with the change-point forms in Table II, consistent with the inverse-modeling family of the ASHRAE Research Project 1050 [15].

Here $(x)^+ = \max(x,0)$ and $(x)^- = \min(x,0)$. The coefficients are fitted by the least squares weighted by $w_i = \Delta_i$, which weights each period by its length. Change points are searched on a 1 F grid between the 10th and 90th percentiles of baseline temperatures, and fits with physically implausible slope signs are rejected. With $\hat{y}_i = \Delta_i \hat{E}_i$, n baseline periods and p parameters, the fit statistics are

$$\mathrm{CV(RMSE)} = \frac{1}{\bar{y}} \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-p}}, \tag{10}$$

$$\mathrm{NMBE} = \frac{\sum_i (y_i - \hat{y}_i)}{(n-p)\,\bar{y}}, \tag{11}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{12}$$

and the t-statistic of each slope is $t_j = \hat{\beta}_j / \mathrm{se}(\hat{\beta}_j)$ from the weighted covariance $s^2 (X^\top W X)^{-1}$ with $s^2 = \sum_i w_i r_i^2/(n-p)$, conditional on the selected change points. Because the weighted normal equations with an intercept give $\sum_i w_i (E_i - \hat{E}_i) = \sum_i (y_i - \hat{y}_i) = 0$, therefore NMBE is zero by construction up to rounding.

Model selection. Among physically valid fits, models whose slopes all have $|t_j| > 2$ are preferred, and among this group, the one with the lowest Schwarz information criterion is selected,

$$\mathrm{BIC} = n \ln\left(\frac{\sum_i (y_i - \hat{y}_i)^2}{n}\right) + p \ln n, \tag{13}$$

with ties going to the model with fewer parameters. However, if no fit qualifies, the lowest-BIC valid fit is selected, and the acceptance check fails on the t-statistic. Version 0.1.0 of this toolkit selected the lowest CV(RMSE) instead; Sec. V B of this paper provides the reason for the modification in the toolkit design. With this implemented, the user can force a form when the M&V plan specifies one.

Avoided energy use. With the selected model applied to reporting-period temperatures and period lengths, the adjusted baseline is $\hat{Y}_{\mathrm{adj}} = \sum_{k=1}^{m} \Delta_k \hat{E}(T_k)$ over m reporting periods, and

$$S_C = \hat{Y}_{\mathrm{adj}} - \sum_{k=1}^{m} y_k, \qquad F = S_C / \hat{Y}_{\mathrm{adj}}. \tag{14}$$

This calculation is the avoided energy use of FEMP v5.0 and the IPMVP; normalized savings are not computed in v0.1 of this toolkit.

Savings uncertainty. For monthly models without autocorrelation correction, the toolkit adopts the ASHRAE Guideline 14 approximation for fractional savings uncertainty (FSU),

$$\mathrm{FSU} = \frac{t_{1-\alpha/2,n-p} \cdot 1.26 \cdot \mathrm{CV(RMSE)}}{F} \sqrt{\frac{n+2}{nm}}, \tag{15}$$

with absolute uncertainty $U_C = \mathrm{FSU} \cdot S_C$ [4, 16]. Eq. (15) is the expression of ASHRAE Guideline 14 Annex B without an autocorrelation adjustment. If the effective number of baseline points n' equal to n, therefore, the term $\sqrt{(n/n')(1+2/n')(1/m)}$ reduces to $\sqrt{(n+2)/(nm)}$. The constant 1.26 used in the equation was derived by Reddy and Claridge [16] and applied in the 2002 and 2014 editions of the guideline. Recent studies reported that the constant is retained in the 2023 edition [17]. In this study, the value has not been checked against a licensed copy of the current edition published in 2023. Furthermore, the approximation treats estimated change points as known, thereby excluding uncertainty in the change point of the 3PH, 4P, and 5P forms. In Sec. V A, the Monte Carlo evaluation re-estimates the change point in every replication and empirically tests this assumption. FEMP v5.0 cites EVO 10100-1:2018 for savings uncertainty [5], stating that federal projects do not require savings uncertainty calculations. For context, two properties of Eq. (15) are highlighted below. First, since $F = S_C/\hat{Y}_{\mathrm{adj}}$, the absolute uncertainty

$$U_C = 1.26\, t\, \mathrm{CV(RMSE)}\, \hat{Y}_{\mathrm{adj}} \sqrt{(n+2)/(nm)} \tag{16}$$

TABLE III: Default Option C acceptance criteria in `config/acceptance_criteria.json`.

Criterion	Default	Source
R^2	≥ 0.75	FEMP v5.0 §5.3.4.1
|NMBE|	$\leq 0.005\%$[a]	FEMP v5.0 §5.3.4.1
Slope $\|t\|$	> 2.0	FEMP v5.0 §5.3.4.1
CV(RMSE)	$\leq 25\%$	LBNL (2017) [14]
FSU	$\leq 50\%$	LBNL (2017) [14]

[a]NMBE is termed net determination bias in FEMP v5.0, which takes the 0.005% threshold from ASHRAE Guideline 14-2002. Satisfied by construction for days-weighted least squares with an intercept (Sec. II D). Retained for conformance with FEMP v5.0; it does not discriminate between models.

does not depend on the size of the savings; only FSU, its ratio to the savings, does. The toolkit therefore reports U_C for any sign of S_C and reports FSU only when $S_C > 0$; when it is not, the FSU acceptance check fails. Second, the approximation accounts for the baseline model error but not separately for noise in the reporting-period meter readings.

E. Acceptance criteria

The default acceptance criteria and their sources are documented in Table III. The R^2, NMBE, and slope-t thresholds are adapted from FEMP v5.0 §5.3.4.1, as reported from IPMVP and ASHRAE Guideline 14-2002. FEMP's calibration tolerances in its Table 4-2 (monthly CV(RMSE) 15%, mean bias $\pm 5\%$) apply to calibrated simulation under Option D and are not applied to Option C. A baseline that is not 12, 24, 36, or 48 months produces a warning according to FEMP v5.0 Section 4.4. All values can be configured because FEMP v5.0 requires the project M&V plan to state its own statistical validation targets.

F. Performance reporting under 2 CFR 200.329

The structure of the reporting module supports computation. Recipients must comply with the predefined structure. *Due date*, 200.329(c)(1): 90 calendar days after an annual reporting period ends; 30 days after a quarterly or semiannual period; 120 days after the period of performance ends for a recipient's final report, and 90 days for a subrecipient's. *Content*, 200.329(c)(2): for each objective, the target, the verified actual and its source (an Option A measure, Option C, an LCC result, or a reported value) and the reason when the objective is not met; budget against actual cost; and unit cost must be reported. Planned unit cost is $c_{\text{plan}} = B/Q_{\text{plan}}$ and actual unit cost is $c_{\text{act}} = C_{\text{act}}/Q_{\text{ver}}$, where Q_{ver} is a verified quantity. A cost overrun is $C_{\text{act}} > B$, and a high unit cost is $c_{\text{act}} > (1+\tau)c_{\text{plan}}$ with tolerance τ (default 0.10). *Significant developments*, 200.329(e): beneficial and adverse developments, with the corrective action plan required for adverse ones. A *completeness check* outlines all the required explanations that are missing; a report with open items is not ready for submission. The module produces content for the form prescribed by an awarding agency and does not replace that form.

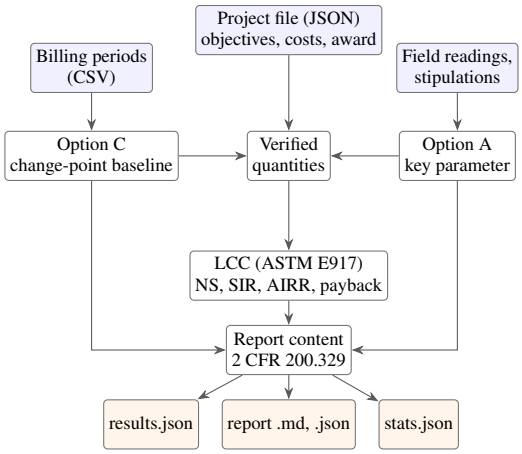


FIG. 1: Data architecture built into MUNI-LCC. One project file and one command (`muni-lcc run`) produce the verified results, the life-cycle cost study computed from verified quantities, and the performance report content. The statistics file feeds the rendered documentation, and continuous integration checks that it is reproduced.

G. Linking verification to cost

The link between the reporting modules is established as a quantity source. As built, any annual energy quantity in the LCC study can be declared as a reference to a verified result rather than a number. For Option C, it is linked to the adjusted baseline or reporting-period use, each annualized by $365/\sum_k \Delta_k$, or the Option A baseline or reporting-period annual quantity. The base case LCC then carries the adjusted baseline, and the alternative LCC carries metered use, to ensure that NS, SIR, and AIRR are ex post measures computed from metered savings.

III. IMPLEMENTATION

MUNI-LCC is built as a Python package (Python 3.9 or later) whose only runtime dependency is NumPy. The toolkit consists of six modules: `lcc`, `mv_option_a`, `mv_option_c`, `report`, `pipeline`, and a small `stats` helper that computes two-sided Student t critical values from exact closed forms for one and two degrees of freedom and a Cornish-Fisher expansion otherwise. The package source is about 1,090 lines including documentation strings. Fig. 1 illustrates the data architecture flow.

Inputs and interface. A project is one JSON file, and any section may be omitted. Option C reads two CSV files with defined columns `period`, `days`, `energy`, and `mean_temp`. The command `muni-lcc run project.json --out outputs` writes `results.json`, `performance_report.md`, and `performance_report.json`; `muni-lcc due-date` computes a report due date alone.

Traceability. Every acceptance threshold in the configuration file contains the exact text existing from its source. The Option A module contains the measured and stipulated parameters,the basis for each stipulation, and the report repeats both simultaneously. The FEMP discount rate file records the exact source of the NIST edition.

Testing and reproducibility. The test suite designed to validate the toolkit contains 37 tests. The tests are grouped into: closed-form checks of the factors in Eqs. (1) to (3) and of a hand-computed LCC; the NS, SIR, and AIRR identities; exact recovery of a known change-point model from noiseless data; recovery of a known injected savings fraction from noiseless and noisy data within the reported uncertainty; due dates for every reporting frequency; uncertainty reporting when savings are not positive; information-criterion selection; and an end-to-end test against the worked example's stored true answer. The worked example writes a statistics file from which the README and the software paper are rendered, and continuous integration fails if the example no longer reproduces those numbers. Every number in this study is likewise read from output files by a script (Sec. VII).

Versions. Version 0.1.0 was released on 14 October 2026 under the MIT license, with documentation under CC BY 4.0, and archived on Zenodo [18]. This paper describes version 0.1.1 [19], which adds the two changes discussed in Sec. V B and the scripts behind this paper. Every result of the worked example in Sec. IV is identical under both versions, and continuous integration is designed to ensure this is checked for every analysis.

IV. WORKED EXAMPLE

A. Synthetic building and data

The worked example built into this toolkit is a hypothetical 40,000 ft^2 town hall whose natural gas boiler and controls were replaced and the initial fluorescent lighting is converted to LED. The rationale for why synthetic data are used is to ensure that the true savings are known in advance, to ensure correctness of the toolkit. With seed 2026, monthly mean temperatures are drawn as rounded illustrative values plus Gaussian jitter with a standard deviation of 2.5 F. Based on this, the true daily gas usage is calculated as

$$E_{\text{true}}(T) = 40 + 6\,(58 - T)^{+} \text{ therms/day} \qquad (17)$$

The observed usage for each month adds a Gaussian noise with a standard deviation of $\sigma = 12$ therms/day. The baseline covers 24 months (2023 and 2024), and the reporting period covers 12 months (2025), where use is scaled by $1 - F_{\text{true}}$ with $F_{\text{true}} = 0.15$. The true avoided use is therefore computed as F_{true} times the noiseless baseline use at the reporting-period temperatures, $0.15 \times 44,191 = 6,629$ therms. For the LED lighting measure, five circuit-demand readings before and after the retrofit have estimated means of 14.6 and 6.2 kW, and annual operating hours are stipulated at 3,380 from the posted building schedule. All prices and costs used in the worked example are invented. The applied temperatures are not NOAA normals.

B. Option C baseline selection

All six forms presented in Table II were fitted. The 3PC form is rejected because its cooling slope contains the wrong sign for a heating meter, so Table IV compares the remaining five candidate fits. CV(RMSE) and BIC for the 4P and 5P forms are marginally higher than the 3PH form. Their added slopes are insignificant, so the selection rule chooses the 3PH model under either version,

$$\hat{E}(T) = 43.00 + 6.37\,(57 - T)^{+} \text{ therms/day}, \qquad (18)$$

which is relatively close to the generating model calculated in Eq. (17). The selected model satisfies all the acceptance criterion in Table III, and the 24-month baseline covers two whole years. Fig. 2 illustrates the fit.

TABLE IV: Candidate Option C baseline models for the worked example (24 months). The minimum slope $|t|$ is the smallest absolute t-statistic among a model's slopes; BIC is Eq. (13); n.s., not significant. 3PC is not listed in the table because no physically valid fit exists.

Form	CV %	R^2	min $\lvert t\rvert$	BIC	Status
mean	69.5	−0.025	n/a	380.9	valid
2P	17.3	0.939	18.8	316.3	valid
3PH	7.9	0.988	42.3	280.8	**selected**
4P	8.0	0.988	0.7	283.4	slope n.s.
5P	8.4	0.988	0.1	287.5	slope n.s.

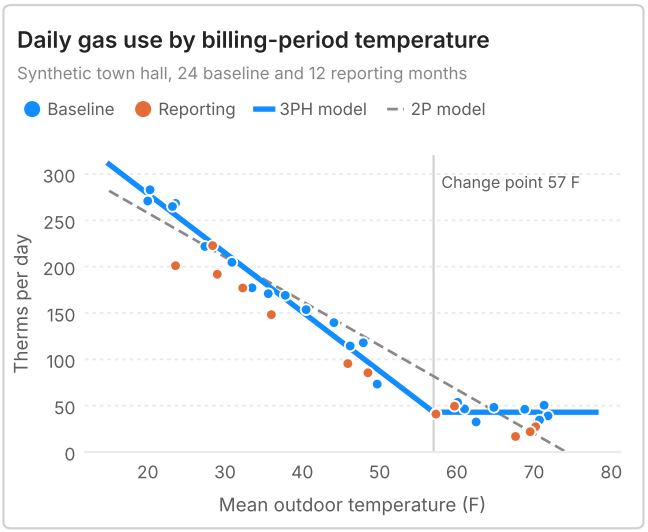


FIG. 2: Daily gas use against mean billing-period temperature for the synthetic town hall. The blue data-points represent the 24 baseline months while the orange data-points are the 12 reporting months. The selected 3PH baseline (solid) captures the flat base load above the change point that the 2P fit (dashed) misses. The vertical gap between the reporting months and the baseline is the avoided energy use.

C. Avoided energy use and its uncertainty

Table V compares the results with the already known truth. The adjusted baseline for the reporting year is 45,617 therms and metered use is 38,737 therms, so the avoided energy use is 6,879 therms, or 15.1% of the adjusted baseline (computed from unrounded values while the displayed figures are rounded independently). The true value is 6,629 therms; the error of 251 therms (3.8%) lies within the 90% uncertainty of ±2,352 therms (FSU 34.2%). Fig. 3 shows the monthly results. The monthly differences include meter noise, justi-

TABLE V: Worked example results as compared with the known truth. Option C uses a 90% confidence level. Each value is rounded independently from unrounded results.

Quantity	MUNI-LCC	Truth
Option C model	3PH	3PH
Change point (F)	57	58
Adjusted baseline (therms)	45,617	44,191
Avoided energy use (therms)	6,879	6,629
Fractional savings (%)	15.1	15.0
90% uncertainty (therms)	±2,352	
FSU at 90% (%)	34.2	
Option A savings (kWh/yr)	28,392	28,392
Option A 90% uncertainty (kWh/yr)	±717	

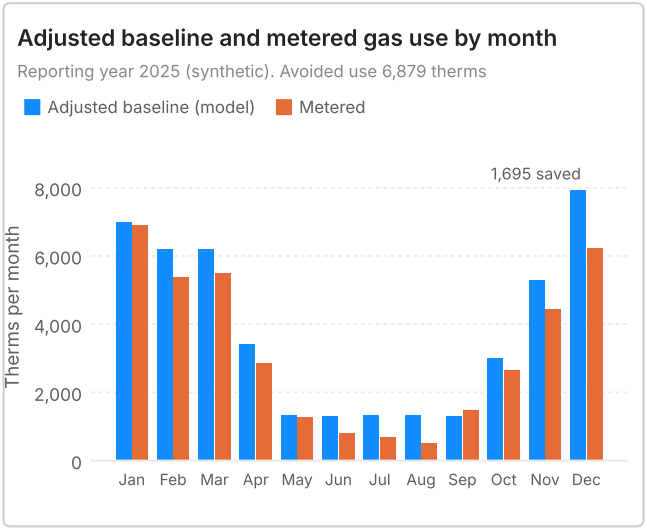


FIG. 3: Adjusted baseline (model applied to 2025 temperatures) and metered use by reporting month. The difference summed over the year is the avoided energy use, 6,879 therms. The label marks the month with the largest difference.

fying the wide variation between individual months although September shows slightly negative savings. The annual total is the quantity as described by the uncertainty.

D. Option A lighting savings

From Eq. (8), $(14.6 - 6.2)\,\text{kW} \times 3{,}380\,\text{h} = 28{,}392\,\text{kWh}$ per year, with a measurement uncertainty of ±717 kWh (2.5%) at 90% confidence. This band covers only the demand readings and does not account for the uncertainty in the stipulated operating hours. The estimate exactly equals the true value, but this is based on construction rather than evidence of accuracy because the synthetic readings were chosen with estimated means of exactly 14.6 and 6.2 kW, and the truth is defined from the same means and the same stipulated hours. However, Sec. V D tests the Option A uncertainty properly.

E. Life-cycle cost from verified quantities

The LCC study compares two scenarios between retaining the existing boiler and fluorescent lighting (base case) with the retrofit (alternative) over 25 years period at the FEMP 2025

TABLE VI: LCC breakdown for the worked example (present value, 2025 dollars; 25 years; 3.0% real discount rate). The totals and differences may differ by $1 from values computed in the table because the entries have been rounded independently from unrounded results.

Component	Base case	Retrofit
Initial investment	0	196,500
Replacements	74,994	7,702
Residual value (−)	0	9,552
Energy (verified quantities)	1,099,281	871,483
OM&R	113,185	69,653
LCC	1,287,460	1,135,786
Net savings		151,675
SIR; AIRR		2.27; 6.4%
Payback, simple; discounted		7.8 y; 9.0 y

real discount rate of 3.0%. Gas and lighting quantities are drawn from the verified results as described in Sec. II G, at invented prices of $1.20 per therm and $0.17 per kWh. For the base case, an end-of-life boiler replacement of $95,000 is expected in year 8. In the alternative case, an LED driver replacement of $12,000 is anticipated in year 15 and a residual boiler value of $20,000 in year 25. Table VI gives the breakdown. The retrofit has net savings of $151,675, an SIR of 2.27, an AIRR of 6.4%, a simple payback of 7.8 years, and a discounted payback of 9.0 years.

F. Performance report content

The award used in this worked example scenario has an annual reporting period ending 31 December 2025, which implies that the report submission is due on 2026-03-31 at 11:59PM. The award has three principal objectives guiding the accomplishments. Objective 1: 7,000 therms per year of avoided gas use - the results attains 98.3% of target and reported as target not met, with the user's explanation attached (the controls for one heating zone were commissioned in March). Objective 2: 25,000 kWh per year of lighting savings - the result attains 113.6% of target, an indicator that target is met and exceeded. Objective 3: positive net savings, is met. Project actual cost exceed the $185,000 budget by $11,500, so the report contains an explanation for the cost overrun. The actual unit cost on a verified basis is $28.56 per first-year therm avoided against a planned $26.43, 8.1% higher, which is within the 10% tolerance and is not flagged as a high unit cost. The completeness check finds no missing explanations.

V. VALIDATION AND SENSITIVITY

A single worked example applied in this study illustrates that the computation recovers a known answer once. However, the results do not establish the accuracy of the uncertainty stated. To ensure accuracy of results, this section performs multiple replications of the example's data-generating process, varies its input, and investigates how often the reported 90% band contains the true value of the avoided energy use. It also varies the life-cycle cost inputs. All analysis use the

latest version 0.1.1 with its default model forms, selection rule, and acceptance criteria; Sec. V B compares the results with version 0.1.0 on identical data. The scripts are described in Sec. VII.

A. Monte Carlo coverage of the Option C uncertainty

Design. Each replication draws a new 24-month baseline and 12-month reporting period from Eq. (17) with the reference settings of Sec. IV ($F_{\text{true}} = 0.15$, $\sigma = 12$ therms/day, temperature jitter 2.5 F). It runs the entire Option C procedure and records the estimate S_C, the reported half-width U_C at 80%, 90%, and 95% confidence, the selected form, and the acceptance outcome. The truth in each replication is calculated as F_{true} times the noiseless baseline use at that replication's reporting temperatures. Replication is covered when $|S_C - S_{\text{true}}| \le U_C$. For 2,000 replications, the Monte Carlo standard error of a coverage near 90% is about 0.7 percentage points. These conditions are favorable to the method: the generating model belongs to the fitted 3PH family, the errors are independent and Gaussian, and no non-routine events occur (Sec. VI).

Results. The estimator shows no detectable bias: the mean error is -17 therms compared to a mean true saving of 6,596 therms, and the root-mean-square error is 1,613 therms. Empirical coverage is 79.8%, 90.0%, and 94.6% at nominal 80%, 90%, and 95%, each within about one Monte Carlo standard error (Fig. 4). For monthly data generated without autocorrelation, the ASHRAE Guideline 14 approximation in Eq. (15) is therefore well calibrated in this setting. This is consistent with the position that the approximation's known shortfalls arise mainly from auto-correlated residuals in daily and hourly models [20], which this test does not include.

Model selection. The generating model is 3PH, and the selection rule chooses 3PH in 87.6% of replications, 4P in 11.6%, and 5P in 0.8%. The remaining over-parameterized choices occur when an extra slope passes the $|t| > 2$ screen and also lowers the BIC, which happens by chance in a minority of noisy baselines. Coverage is 90.8% in the replications where 3PH is selected and 84.9% where 4P is selected.

Acceptance outcomes. Every replication passes the R^2, NMBE, slope-t, and CV(RMSE) criteria (median CV(RMSE) 9.6%). The NMBE criterion is satisfied in every replication. As expected, its highest value is 1.5×10^{-13} percent, which is consistent with a value that is zero by construction under days-weighted least squares (Sec. II D). NMBE is retained to ensure conformance with FEMP v5.0 but does not discriminate between fits produced by the toolkit. The binding criterion is FSU: 72.7% of replications pass all criteria, and every failure is an FSU above 50%. The FSU of the worked example of 34.2% is therefore not unusual and the median over replications is 41.7%.

B. Changes in version 0.1.1

The same tests were initially conducted on version 0.1.0 where two behaviors were identified and corrected in Version 0.1.1. Table VII compares both versions from identical data.

Model selection. Version 0.1.0 selected the significant-slope model using the lowest CV(RMSE). For the reference process, it chose a 4P or 5P model in 23.5% of replications. Since each slope's t-statistic is computed conditional on a change point that was chosen to minimize error, the $|t| > 2$ screen admits a spurious slope more often than its nominal level implies, and CV(RMSE) penalizes an added parameter only through $n - p$. The over-fitted models reported a lower CV(RMSE) and understated their own uncertainty. With selections by BIC, Eq. (13), penalizes each added parameter by $\ln n$. Reduced over-parameterized choices to 12.4% increased coverage to 90% from 89.5% to 90.0% and lowered the RMSE from 1,649 to 1,613 therms. To ascertain that the stricter rule does not underfit, the test was repeated on processes whose true model is 4P or 5P. For a true 4P model the two rules perform alike. For a true 5P model, BIC identifies the 5P form less often (30.7% against 56.0%), but the RMSE of the savings estimate changed by only 1.8% and coverage is 92.7% against 93.3%. For this process, the effect on the savings estimate is considerably small.

TABLE VII: Option C results based on identical synthetic data under versions 0.1.0 and 0.1.1. The reference process has a true 3PH model ($F_{\text{true}} = 15\%$, $\sigma = 12$ therms/day, 24-month baseline, 2,000 replications); the weak and noisy rows use 500 replications each.

	v0.1.0	v0.1.1
Reference process, true model 3PH		
Selects 3PH (%)	76.5	87.6
Selects 4P (%)	19.7	11.6
Selects 5P (%)	3.8	0.8
Coverage at 90% (%)	89.5	90.0
RMSE (therms)	1,649	1,613
Weak or noisy savings, coverage at 90% (%)		
$F_{\text{true}} = 5\%$	86.0	91.6
$\sigma = 36$	87.4	91.6
No uncertainty reported, $F_{\text{true}} = 5\%$ (%)	11.0	0.0
True model 4P, 300 replications		
Selects 4P (%)	70.0	72.0
Coverage at 90% (%)	92.3	92.7
RMSE (therms)	1,512	1,501
True model 5P, 300 replications		
Selects 5P (%)	56.0	30.7
Coverage at 90% (%)	93.3	92.7
RMSE (therms)	1,529	1,556

Uncertainty for non-positive savings. Version 0.1.0 computed the absolute uncertainty as FSU times savings and therefore reported none when estimated savings were zero or negative, which occurred in 11.0% of replications at 5% true savings. Version 0.1.1 computes U_C directly from Eq. (16), which does not involve savings, so a report can state that savings were not distinguishable from zero. Coverage at 5% true savings increases from 86.0% to 91.6%, and at $\sigma = 36$ from 87.4% to 91.6%.

Neither change affects the worked example: its candidate models rank the same under both rules highlighted in (Table IV), and its savings are positive.

C. Dependence on savings size, noise, and baseline length

Table VIII varies one setting at a time with 500 replications each (Monte Carlo standard error of about 1.3 points at 90%).

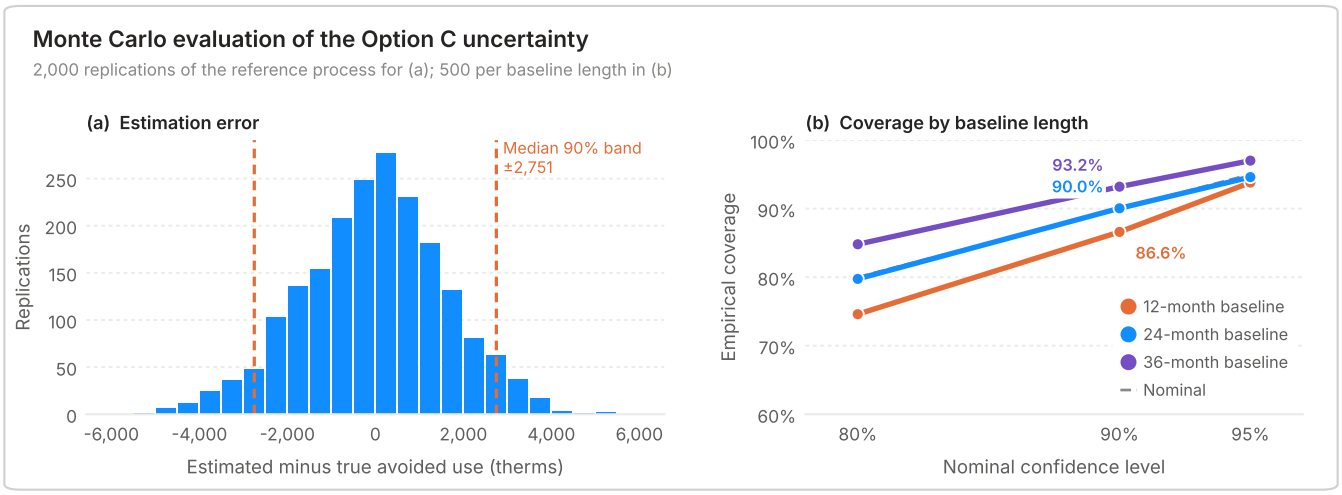


FIG. 4: Monte Carlo evaluation of Option C. (a) Distribution of estimation error over 2,000 replications of the reference process (bins of 500 therms); dashed lines mark the median reported 90% half-width. (b) Empirical coverage of the reported uncertainty against the nominal confidence level for 12-, 24-, and 36-month baselines. The 24-month reference uses 2,000 replications and the other two use 500 each.

Within each block, the same random draws are reused across settings, so differences between rows reflect the setting rather than sampling.

Savings fraction. Error and absolute uncertainty do not change with F_{true}, as Eq. (15) implies (Sec. II D): the RMSE is the same in every row of the block. The noticeable change in every block is the FSU, from a median of 106.3% at 5% savings to 20.9% at 30%, together with the acceptance rate, from 27.2% at 10% savings to 94.2% at 20%. In practical applications, the FSU criterion in Table III determines whether a project of this size and noise level can be verified at all from monthly bills. At 5% savings, the estimate is zero or negative in 10.6% of replications; the band is still reported, and coverage is 91.6%.

Noise. By halving the noise to $\sigma = 6$, the RMSE is instantaneously halved (781 therms) and every replication passes. At $\sigma = 36$ the RMSE is 4,709 therms and only 19.6% of replications meet the CV(RMSE) criterion, but the coverage stays at or above nominal in all noise settings: noticeably, the band widens with the noise as it should.

Baseline length. A 12-month baseline raises the RMSE to 2,365 therms and lowers coverage to 86.6% at nominal 90% and 74.6% at nominal 80%, an indicator that the approximation is somewhat optimistic for the shortest baseline acceptable by FEMP v5.0. A 36-month baseline reduces the RMSE to 1,499 therms, and coverage of 93.2% is slightly conservative. These results support the FEMP v5.0 guidance to use whole-year baselines, suggestive that one year baseline should be treated as a minimum rather than a target.

D. Option A coverage

The exact match of Option A demonstrated using a worked example is built into the example (Sec. IV). To test Eq. (9), we drew 20,000 replications of five readings before and after the retrofit from normal distributions with means of 14.6 and 6.2 kW and a standard deviation of 0.20 kW, with the operating hours fixed at their stipulated value. Coverage is 80.1%, 90.1%, and 95.0% at nominal 80%, 90%, and 95%, and the median relative uncertainty at 90% is 2.7%. This result confirms the arithmetic of the interval for independent, normally distributed readings. It does not record any information about the uncertainty in stipulated hours, which the toolkit does not quantify by design.

E. Sensitivity of the life-cycle cost results

Propagation of measured-savings uncertainty. Each verified therm of gas savings calculated in this replication is worth $20.90 of net present value over the study period (rounded; the gas contribution of $143,751 is calculated from unrounded values). When both verified quantities are moved to the edges of their 90% bands (gas ±2,352 therms, lighting ±717 kWh), net savings increases from $100,413 to $202,937 and the SIR from 1.84 to 2.70. The cost-effectiveness verdict remains unchanged across the band, but its size is uncertain by roughly a third in either direction, which a report derived from a single projected savings would not have shown.

Discount rate, energy price, and study period. Fig. 5 varies one input at a time for each replication. Net savings fall with the real discount rate and reach zero at 9.2%, where the adjusted internal rate of return (AIRR) equals the discount rate indicated in Eq. (7). Net savings becomes positive in a study period of 10 years, then increases linearly with the gas price, and remains consistently positive ($7,924) even at a gas price of zero.

In this synthetic worked example, the avoided base-case boiler replacement and the lighting and maintenance savings exceed the added investment on their own. This further establishes that the verified gas savings worth about $143,751 in the present value determine the size of the net savings rather than its positive or negative sign.

VI. LIMITATIONS AND SCOPE

There are four core measurement and verification options defined by IPMVP. Of these options, the toolkit implements

TABLE VIII: Option C performance as one setting is varied from the reference ($F_{\text{true}} = 15\%$, $\sigma = 12$ therms/day, 24-month baseline). Errors are in therms. Coverage is the share of replications in which the reported band contains the truth. The last column is the share of replications where the estimated savings are zero or negative; the absolute uncertainty is still reported for them, but FSU is not, and they fail the FSU criterion.

Setting	Mean error	RMSE	Median FSU_{90} (%)	Cov. 80%	Cov. 90%	Cov. 95%	Pass all (%)	$S_C \leq 0$ (%)
Reference	−17	1,613	41.7	79.8	90.0	94.6	72.7	0.0
Savings fraction F_{true}								
5%	−18	1,686	106.3	80.2	91.6	95.6	1.4	10.6
10%	−18	1,686	61.4	80.2	91.6	95.6	27.2	0.8
20%	−18	1,686	31.4	80.2	91.6	95.6	94.2	0.0
30%	−18	1,686	20.9	80.2	91.6	95.6	100.0	0.0
Noise s.d. σ (therms/day)								
6	12	781	21.7	83.8	91.0	95.2	100.0	0.0
24	37	3,126	84.3	82.8	91.6	95.4	9.4	1.6
36	59	4,709	120.3	81.6	91.6	95.0	2.8	7.8
Baseline length (months)								
12	80	2,365	45.3	74.6	86.6	93.8	60.6	0.0
36	59	1,499	41.7	84.8	93.2	97.0	76.8	0.0

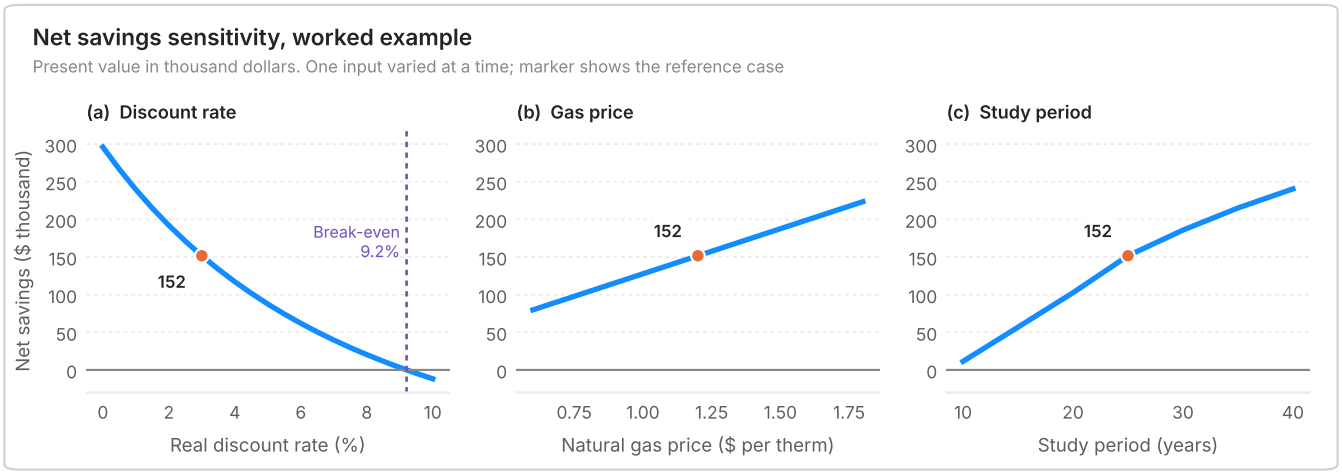


FIG. 5: Net savings of the worked example as one life-cycle cost input is varied: (a) real discount rate, (b) natural gas price, (c) study period. The orange marker is the reference case (3.0%, \$1.20 per therm, 25 years; net savings \$151,675); the dotted line in (a) marks the break-even rate.

two of the four IPMVP options: Options A (key parameter measurement) and C (whole facility). Option B (all parameters measured) and Option D (calibrated simulation) are not implemented in this toolkit. Option C uses temperature as the only independent variable; other variables such as occupancy, school calendars, and production are not modeled. Non-routine adjustments must be estimated outside the toolkit and documented in the M&V plan. Avoided energy use is computed, but the normalized savings are not computed. The uncertainty method applied is the ASHRAE Guideline 14 approximation for monthly data rather than the EVO 10100-1:2018 procedure cited in FEMP v5.0; it does not correct for autocorrelation and should not be applied to daily or hourly data without that correction [20, 21]. The 1.26 constant in the expression is attributed to Reddy and Claridge [16] and the 2002 and 2014 editions of the ASHRAE guideline. It has not been checked against a licensed copy of the current edition published in 2023 to check for any change in the value. Furthermore, the approximation treats estimated change points as known. The Monte Carlo results in Sec. V are based on a data-generating process that matches the 3PH model family, has independent Gaussian errors, and has no non-routine events, so they present calibration under favorable conditions. They do not constitute evidence of performance on real municipal billing data, which has not yet been tested in this toolkit. Life-cycle cost is calculated in constant dollars; taxes, financing, and incentives as cash flows are not modeled. Finally, the report module organizes the content as specified in 2 CFR 200.329 and checks its completeness. It is not a compliance determination procedure and does not replace the form prescribed by the awarding agency, and it does not model due-date extensions.

VII. CONCLUDING REMARKS

MUNI-LCC is a compact open access toolkit designed to implement the performance reporting process for energy effi-

ciency upgrades, from metered data to a federal performance report based on IPMVP Options A and C as defined in FEMP v5.0, life-cycle cost consistent with ASTM E917, and content elements of 2 CFR 200.329. The primary essence for the development of the toolkit is to ensure that verified quantities, not projections, are calculated in the life-cycle cost study so that Net Savings (NS), Savings-to-Investment Ratio (SIR), Adjusted Internal Rate of Return (AIRR) and unit cost are accurately reported on a measured basis with every acceptance value referencing its source.

As observed from the synthetic example with a known answer, the toolkit recovered the injected gas savings within its stated 90% uncertainty. Across 2,000 replications of the same process, the Option C uncertainty attained coverage of 90.0% at a nominal 90% with no detectable bias, and Option A coverage is likewise close to the nominal value. These results demonstrate behavior under the simulated conditions of Sec. V and should not be assumed as evidence of performance on real municipal billing data. The findings from tests conducted in version 0.1.0 led to two changes that were implemented in version 0.1.1. The two changes include: model selection by information criterion, which cuts spurious change-point models from 23.5% to 12.4% of replications; and an absolute uncertainty reported for any sign of savings. Neither of these changes altered the worked example. The tests further demonstrate that the FSU criterion, not the goodness of fit, decides whether a modest saving can be verified from the monthly bills.

Future scope of work will include the integration of Option B (all parameter measurement), the EVO 10100-1:2018 uncertainty procedure alongside the current approximation, normalized savings, and additional independent variables, validated against real municipal billing data compared with an established open M&V implementation on the same data.

CODE AND DATA AVAILABILITY

The MUNI-LCC source code, documentation, synthetic example data, and tests are packaged in a repository publicly accessible at https://github.com/Olumide-Akindipe/muni-lcc under the MIT license (documentation CC BY 4.0). The manuscript source and the scripts supporting every table and figure in this paper (`validation.py`, `underfit_check.py`, `lcc_scans.py`, `figures.py`, and `make_numbers.py`, which writes every number in the text from the output files) are in the `preprint/` folder of the repository, together with their outputs. Version 0.1.1 is archived on Zenodo [19]; version 0.1.0 is archived separately [18], and all versions are available under the concept DOI 10.5281/zenodo.23362541. The package can be installed with `pip install muni-lcc`. Real utility data are not used.

COMPETING INTERESTS

The author declares no competing interests.

ACKNOWLEDGMENT

No financial support was received for this work.

REFERENCES

[1] 2 CFR 200.329, Monitoring and Reporting Program Performance, 2024. URL https://www.ecfr.gov/current/title-2/section-200.329. As amended at 89 FR 30136, April 22, 2024.

[2] Efficiency Valuation Organization. International Performance Measurement and Verification Protocol: Core Concepts 2022, 2022. URL https://evo-world.org.

[3] U.S. Department of Energy, Federal Energy Management Program. M&V Guidelines: Measurement and Verification for Performance-Based Contracts, Version 5.0. Technical report, U.S. Department of Energy, September 2024. URL https://www.energy.gov/sites/default/files/2024-10/mv_guide_5_0.pdf. Prepared by Pacific Northwest National Laboratory.

[4] ASHRAE. Guideline 14-2023, Measurement of Energy, Demand, and Water Savings, 2023.

[5] Efficiency Valuation Organization. Uncertainty Assessment for IPMVP, 2018. URL https://evo-world.org.

[6] ASTM International. Standard Practice for Measuring Life-Cycle Costs of Buildings and Building Systems, 2023. ASTM E917-17(2023).

[7] Sieglinde K. Fuller and Stephen R. Petersen. *Life-Cycle Costing Manual for the Federal Energy Management Program*. NIST Handbook 135. National Institute of Standards and Technology, 1996.

[8] National Institute of Standards and Technology. Energy Price Indices and Discount Factors for Life-Cycle Cost Analysis, 2025. Technical Report NIST IR 85-3273-40, National Institute of Standards and Technology, 2025. URL https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.85-3273-40.pdf.

[9] National Institute of Standards and Technology. Building Life Cycle Cost Programs (BLCC). URL https://www.nist.gov/node/1331206.

[10] National Institute of Standards and Technology. Economic Evaluation Engine (E3) Technical Manual. Technical Report NIST Technical Note 2225r1, National Institute of Standards and Technology, 2024.

[11] OpenEEmeter contributors. eemeter (OpenEEmeter). URL https://eemeter.readthedocs.io.

[12] kW Labs. nmecr: An implementation of peer-reviewed energy data analysis algorithms for site-specific M&V. URL https://github.com/kw-labs/nmecr.

[13] Lawrence Berkeley National Laboratory. RMV2.0: Advanced measurement and verification 2.0 (M&V 2.0) for commercial buildings. URL https://github.com/LBNL-ETA/RMV2.0.

[14] Lawrence Berkeley National Laboratory. Assessment of Automated Measurement and Verification (M&V) Methods: Acceptance Criteria and Documentation. Technical report, Lawrence Berkeley National Laboratory, 2017. URL https://bies.lbl.gov/sites/default/files/acceptcriteria-doc-2017.pdf.

[15] J. Kelly Kissock, Jeff S. Haberl, and David E. Claridge. Inverse modeling toolkit: Numerical algorithms. *ASHRAE Transactions*, 109(2), 2003.

[16] T. Agami Reddy and David E. Claridge. Uncertainty of "measured" energy savings from statistical baseline models. *HVAC&R Research*, 6(1):3–20, 2000. doi: 10.1080/10789669.2000.10391247.

[17] Léa Gondian and Thimothée Thiery. Validation of methods to estimate the uncertainty of buildings energy savings in a controlled numerical setting and Bayesian energy signature with autocorrelated errors, 2026. arXiv:2607.25382.

[18] Stephen Olumide Akindipe. MUNI-LCC: Open Life-Cycle Cost and Measurement-and-Verification Toolkit for Municipal Energy Projects, 2026. Version 0.1.0. DOI 10.5281/zenodo.23362542 (concept DOI 10.5281/zenodo.23362541). https://github.com/Olumide-Akindipe/muni-lcc.

[19] Stephen Olumide Akindipe. MUNI-LCC: Open Life-Cycle Cost and Measurement-and-Verification Toolkit for Municipal Energy Projects, 2026. Version 0.1.1. DOI 10.5281/zenodo.23825887 (concept DOI 10.5281/zenodo.23362541). https://github.com/Olumide-Akindipe/muni-lcc.

[20] Samir Touzani, Jessica Granderson, David A. Jump, and Derrick Rebello. Evaluation of methods to assess the uncertainty in estimated energy savings. *Energy and Buildings*, 193, 2019. doi: 10.1016/j.enbuild.2019.03.041.

[21] Travis Walter, Phillip N. Price, and Michael D. Sohn. Uncertainty estimation improves energy measurement and verification procedures. *Applied Energy*, 130:230–236, 2014. doi: 10.1016/j.apenergy.2014.05.030.