NREL
NATIONAL RENEWABLE ENERGY LABORATORY

# Evaluation of Automated Model Calibration Techniques for Residential Building Energy Simulation

Joseph Robertson and Ben Polly
*National Renewable Energy Laboratory*

Jon Collis
*Colorado School of Mines*

**NREL is a national laboratory of the U.S. Department of Energy**
**Office of Energy Efficiency & Renewable Energy**
**Operated by the Alliance for Sustainable Energy, LLC.**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

**Technical Report**
NREL/TP-5500-60127
September 2013

Contract No. DE-AC36-08GO28308

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

# Evaluation of Automated Model Calibration Techniques for Residential Building Energy Simulation

Joseph Robertson and Ben Polly
*National Renewable Energy Laboratory*

Jon Collis
*Colorado School of Mines*

Prepared under Task No. BE13.0102



National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

**Technical Report**
NREL/TP-5500-60127
September 2013

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

*Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.*

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

# Acknowledgments

This work was funded by the U.S. Department of Energy Building Technologies Program. The authors wish to thank David Lee (U.S. Department of Energy Team Leader, Residential Buildings) for his continued support. The authors also wish to thank Craig Christensen, Mike Gestwick, and Ron Judkoff for their insightful comments related to this study. An additional thanks to Ron Judkoff for his guidance in pointing us toward this interesting line of inquiry.

# Nomenclature

| | |
|---|---|
| $\bar{y}$ | Arithmetic mean of the reference utility data vector components |
| $\boldsymbol{\beta}$ | Matrix whose columns contain $\zeta_k$ polynomial term regression coefficients |
| $\chi_i^2$ | Chi-squared value corresponding to input $i$ for sensitivity analysis in Procedure 3.1 |
| $\Delta E$ | Objective function value difference between successive perturbations |
| $\delta_{TOTAL}$ | Consolidated index for measuring overall goodness-of-fit |
| $\eta$ | Number of random samples from each triangular probability distribution in the sensitivity analysis for Section 2.4 and Procedures 3.2–3.3 |
| $\Gamma_{ref}$ | Pre-retrofit annual energy use (synthetic billing data from reference model) |
| $\Gamma_{uncal}$ | Predicted pre-retrofit annual energy use from uncalibrated model |
| $\lambda$ | Number of approximate inputs considered in the study |
| $\mu_j$ | Mean of the $\eta$ annual output values corresponding to input $j$ in the sensitivity analysis for Section 2.4 and Procedures 3.2–3.3 |
| $\nu_i$ | Explicit input value randomly selected from triangular probability distribution for approximate input $i$ |
| $\varphi^c$ | Annual calibrated model predicted savings as a percent of the pre-retrofit reference utility data |
| $\varphi^r$ | Annual reference savings as a percent of the pre-retrofit reference utility data |
| $\psi^c$ | Calibrated model's post-retrofit annual energy savings prediction |
| $\psi^r$ | Reference model's post-retrofit annual energy savings prediction |
| $\psi^u$ | Uncalibrated model's post-retrofit annual energy savings prediction |
| $\rho_{exp}$ | Expected number of occurrences on levels $x_i^{low}, x_i^{mid}, x_i^{high}$ for sensitivity analysis in Procedure 3.1 |
| $\rho_{obs,s,i}$ | Observed number of occurrences on levels $x_i^{low}, x_i^{mid}, x_i^{high}$ for sensitivity analysis in Procedure 3.1 |
| $\sigma_j$ | Standard deviation of the $\eta$ annual output values corresponding to input $j$ in the sensitivity analysis for Section 2.4 and Procedures 3.2–3.3 |
| $\tau$ | Number of strong parameters identified from sensitivity analyses |

| | |
|---|---|
| $\mathbf{X}$ | Design matrix used in central composite design |
| $\mathbf{Y}$ | Response matrix used in central composite design |
| $\varepsilon(\psi^c)$ | Absolute error in savings predictions for the calibrated model |
| $\varepsilon(\psi^u)$ | Absolute error in savings predictions for the uncalibrated model |
| $\xi_j$ | Sensitivity analysis coefficient corresponding to input $j$ for Section 2.4 and Procedures 3.2–3.3 |
| $\zeta_k$ | Quadratic polynomials that best fit each of the sets of simulation data in the least squares sense |
| $BoC$ | Difference in absolute error of savings predictions between calibrated and uncalibrated models |
| $E_{new}$ | Objective function value of new (successive) perturbations |
| $E_{old}$ | Objective function value of old (previous) perturbations |
| $N$ | Number of iterations per parameter for simulated annealing algorithm |
| $n$ | Number of utility data points used in calibration (12, 365, or 8760) |
| $p$ | Number of parameters in the model |
| $T_0$ | Initial temperature for simulated annealing algorithm |
| $TBoC$ | Sum of the $BoC$ across all retrofit measures |
| $x_i^{high}$ | High level median value in LHMC discretization corresponding to input $i$ |
| $x_i^{low}$ | Low level median value in LHMC discretization corresponding to input $i$ |
| $x_i^{max}$ | Triangular probability distribution's maximum value corresponding to input $i$ |
| $x_i^{mid}$ | Middle level median value in LHMC discretization corresponding to input $i$ |
| $x_i^{min}$ | Triangular probability distribution's minimum value corresponding to input $i$ |
| $x_i^{nom}$ | Triangular probability distribution's nominal value corresponding to input $i$ |
| $\hat{\boldsymbol{y}}$ | Vector of simulation-predicted utility data |
| $\boldsymbol{y}$ | Vector of reference utility data |

# Acronyms

| | |
|---|---|
| AC | Air conditioner |
| ACH | Air changes per hour |
| AH | Air handler |
| BEopt | Building Energy Optimization program |
| BESTEST-EX | Building Energy Simulation Test for Existing Homes |
| CV | Coefficient of variation |
| ELA | Effective leakage area |
| LHMC | Latin Hypercube Monte Carlo |
| MEL | Miscellaneous electric loads |
| MGL | Miscellaneous gas loads |
| NMBE | Normalized mean base error |
| OSB | Oriented strand board |
| RA | Return air |
| RMSE | Root mean square error |
| SA | Supply air |
| SEER | Seasonal energy efficiency ratio |
| SHGC | Solar heat gain coefficient |
| SLA | Specific leakage area |
| UA | Unfinished attic |
| XML | Extensible markup language |

# Executive Summary

This simulation study adapts and applies the general framework described in Building Energy Simulation Test for Existing Homes (Judkoff et al. 2010) for self-testing residential building energy model calibration methods. BEopt™/DOE-2.2[1] is used to evaluate four mathematical calibration methods in the context of monthly, daily, and hourly synthetic utility data for a 1960s-era existing home in a cooling-dominated climate. The home's model inputs are assigned probability distributions representing uncertainty ranges, pseudo-random selections are made from the uncertainty ranges to define "explicit" input values, and synthetic utility billing data are generated using the explicit input values. The four calibration methods evaluated in this study are: (1) an ASHRAE 1051-RP-based approach (Reddy and Maor 2006); (2) a simplified simulated annealing optimization approach; (3) a regression metamodeling optimization approach; and (4) a simple output ratio calibration approach. The calibration methods are evaluated for monthly, daily, and hourly cases; various retrofit measures are applied to the calibrated models and the methods are evaluated based on the accuracy of predicted savings, computational cost, repeatability, automation, and ease of implementation.

Two utility billing scenarios were investigated: one in which the uncalibrated model overpredicts ("Scenario 1") the reference billing data and another in which the uncalibrated model underpredicts ("Scenario 2") the reference billing data. Figure ES1 and Figure ES2 illustrate the differences in annual energy savings prediction accuracy for retrofit measures when using the calibrated versus uncalibrated building models. The various retrofit measures considered in this study were first applied to the reference models (models with "explicit" input values) to develop reference energy savings. The retrofit measures were then applied to the uncalibrated model ("best-guess" building model composed of nominal input values) and each calibrated model obtained by performing the three input calibration methods. Annual energy savings predictions using the uncalibrated and calibrated models are displayed in the figures as percent errors in savings predictions relative to the reference energy savings; values appearing closer to the red vertical dashed line represent better agreement with reference savings than values farther from the red vertical dashed line.[2]

[1] Throughout the document, both "BEopt" and "BEopt/DOE-2.2" refer to the Building Energy Optimization software BEopt used in conjunction with the DOE-2.2 simulation engine.

[2] Relative errors may be high because reference energy savings for some retrofit measures are low; e.g., the Low Solar Absorptance Roof measure reduces reference energy use by 65 kWh (0.3%) for Scenario 1 and 302 kWh (1.1%) for Scenario 2.

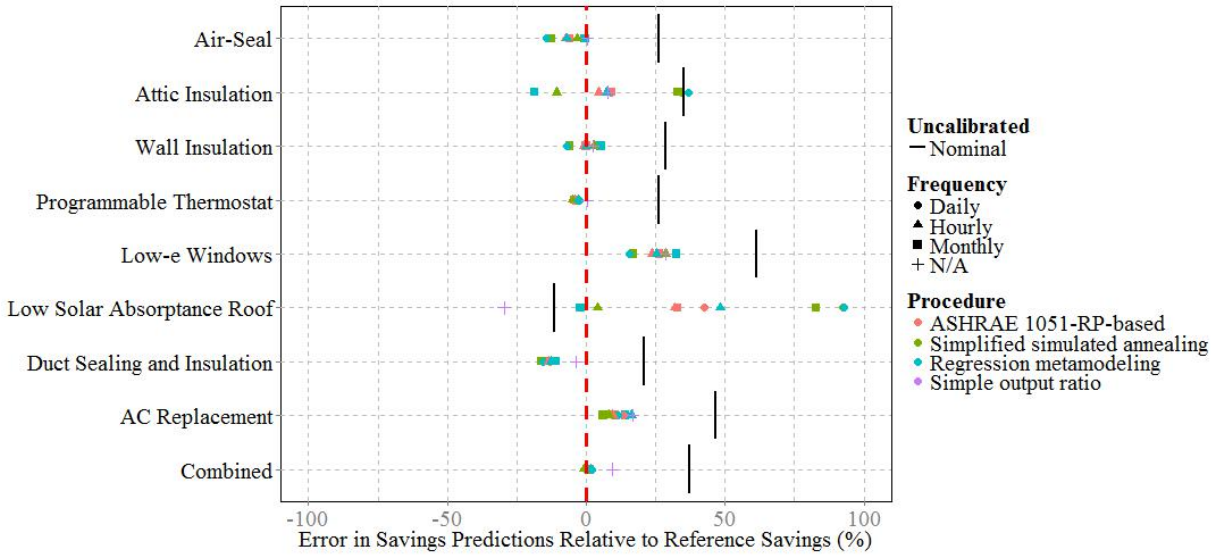


**Figure ES1. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 1**

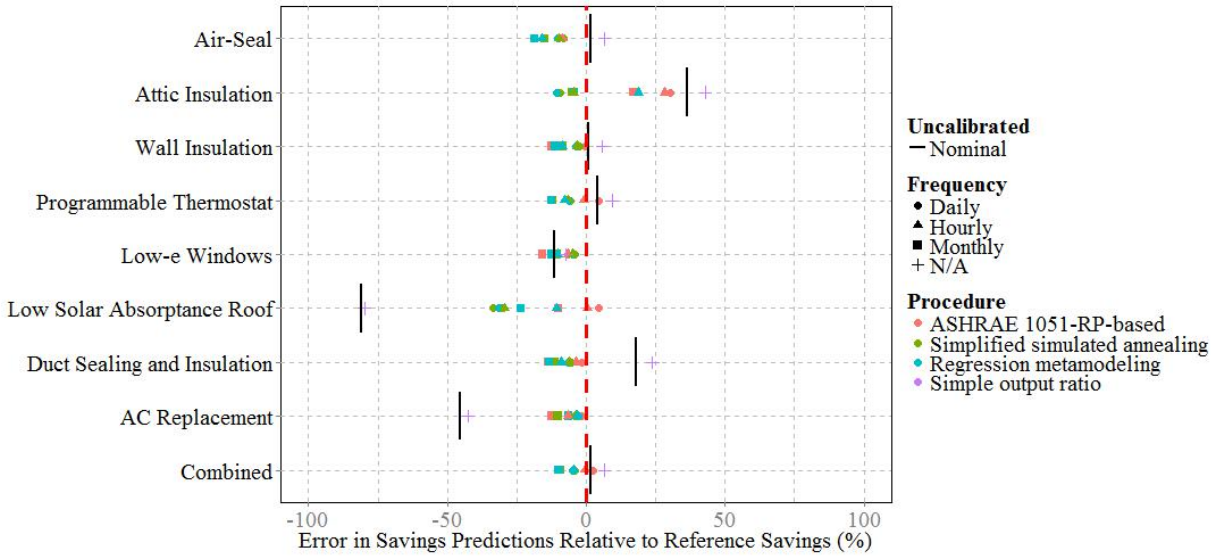


**Figure ES2. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 2**

These figures, as well as results shown throughout the report, demonstrate that calibration generally improves the accuracy of savings predictions, but that significant errors in calibrated models can exist even when pre-retrofit model predictions match well with the billing data (i.e., billing data can be matched for the wrong reasons). In terms of accurately predicting retrofit energy savings, the most computationally expensive method investigated in this study (ASHRAE 1051-RP based) performed slightly better than the two other less expensive input calibration methods. The ASHRAE 1051-RP-based method required additional simulations to develop "uncertainty" ranges for savings predictions, but for most predictions in this study the ranges did

not contain the "true" savings value. Calibrations to the hourly and daily billing data generally produced more accurate savings predictions than calibrations to monthly data. The added benefit of higher frequency data was more apparent for the calibration scenario (Scenario 2) where the uncalibrated model annual energy use prediction was fairly close (–5%) to the billing data (due to compensating errors) than for the scenario (Scenario 1) where the uncalibrated model significantly overpredicted the billing data (+26%).

Overall, the results suggest that:

1. The optimization problem is still significantly underdetermined when calibrating to monthly, daily, and hourly data using the approaches investigated in this study, which are largely based on previous studies performed mostly in the context of commercial buildings.
2. Additional research is needed to develop improved or alternate approaches that take full advantage of the additional informational content contained in the high-frequency residential billing data.

In the nearer term, calibration methods similar to those described in this study could be implemented in residential simulation tools and tested in the field for automated calibrations to monthly billing data. They could be implemented in the context of emerging industry standards for residential model calibration (such as Building Performance Institute Standard 2400 [BPI 2011]). Software developers who have the capability to run batch simulations in parallel (e.g., through cloud-computing) could employ methods similar to the regression metamodeling optimization approach to reduce the time required for automated calibration.

# Table of Contents

**Acknowledgments** ..... **iii**
**Nomenclature** ..... **iv**
**Acronyms** ..... **vi**
**Executive Summary** ..... **vii**
**1 Introduction** ..... **1**
**2 Approach for Comparing Calibration Techniques** ..... **3**
2.1 Define the House ..... 3
2.2 Define *Approximate* Inputs ..... 4
2.3 Generate Utility Data ..... 4
2.4 Identify the Most *Influential* Inputs ..... 5
2.5 Perform Calibrations ..... 6
2.6 Apply Retrofit Measures ..... 7
2.7 Compare Calibration Procedures ..... 8
**3 Calibration Techniques** ..... **9**
3.1 ASHRAE 1051-RP-Based Approach ..... 9
3.1.1 Define Influential Input Parameters ..... 9
3.1.2 Perform a Coarse Grid Calibration ..... 9
3.1.3 Perform a Refined Grid Calibration ..... 12
3.1.4 Predict Ranges of Energy Savings for Retrofit Measures ..... 14
3.2 Simplified Simulated Annealing Optimization Approach ..... 17
3.2.1 Sensitivity Analysis ..... 18
3.2.2 Iterative Search ..... 18
3.3 Regression Metamodeling Optimization Approach ..... 26
3.3.1 Sensitivity Analysis ..... 27
3.3.2 Central Composite Design ..... 27
3.3.3 Optimization ..... 27
3.4 Simple Output Ratio Calibration Approach ..... 35
**4 Discussion** ..... **37**
4.1 Accuracy of Predicted Energy Savings ..... 37
4.2 Computational Cost ..... 39
4.3 Automation ..... 41
4.4 Repeatability ..... 41
4.5 Ease of Implementation ..... 41
**5 Conclusions and Future Work** ..... **43**
5.1 Conclusions ..... 43
5.2 Future Work ..... 44
**Glossary** ..... **46**
**References** ..... **47**
**Appendix A Approximate Inputs, Uncertainty Ranges, and Explicit Input Values** ..... **50**
**Appendix B Sensitivity Analysis Results** ..... **55**
**Appendix C Process Flow** ..... **57**
**Appendix D Latin Hypercube Monte Carlo** ..... **58**
**Appendix E Central Composite Design** ..... **59**
**Appendix F Goodness-of-Fit Indices** ..... **60**
**Appendix G Comparing Wall and Ceiling R-Values to BESTEST-EX** ..... **61**
**Appendix H Additional Figures** ..... **64**

# List of Figures

Figure ES1. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 1 ..... viii
Figure ES2. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 2 ..... viii
Figure 1. Triangular probability distribution ..... 4
Figure 2. The 24 influential inputs ..... 6
Figure 3. Procedure 3.1 refined grid results for calibration to monthly data, Scenario 1 ..... 13
Figure 4. Procedure 3.1 refined grid results for calibration to daily data, Scenario 1 ..... 13
Figure 5. Procedure 3.1 refined grid results for calibration to hourly data, Scenario 1 ..... 13
Figure 6. Procedure 3.1 refined grid results for calibration to monthly data, Scenario 2 ..... 14
Figure 7. Procedure 3.1 refined grid results for calibration to daily data, Scenario 2 ..... 14
Figure 8. Procedure 3.1 refined grid results for calibration to hourly data, Scenario 2 ..... 14
Figure 9. Procedure 3.1 BoC, Scenario 1 ..... 15
Figure 10. Procedure 3.1 BoC, Scenario 2 ..... 15
Figure 11. Procedure 3.1 energy savings predictions for combined retrofit, Scenario 1 ..... 16
Figure 12. Procedure 3.1 energy savings predictions for combined retrofit, Scenario 2 ..... 16
Figure 13. Estimated ranges for energy savings predictions using Procedure 3.1, Scenario 1 ..... 17
Figure 14. Estimated ranges for energy savings predictions using Procedure 3.1, Scenario 2 ..... 17
Figure 15. Procedure 3.2 $\xi$ sensitivity analysis results ..... 18
Figure 16. Procedure 3.2 optimization results, Scenario 1 ..... 19
Figure 17. Procedure 3.2 optimization results, Scenario 2 ..... 19
Figure 18. Procedure 3.2 parameter convergence for calibrations to monthly data, Scenario 1 ..... 20
Figure 19. Procedure 3.2 residual convergence for calibrations to monthly data, Scenario 1 ..... 20
Figure 20. Procedure 3.2 parameter convergence for calibrations to daily data, Scenario 1 ..... 21
Figure 21. Procedure 3.2 residual convergence for calibrations to daily data, Scenario 1 ..... 21
Figure 22. Procedure 3.2 parameter convergence for calibrations to hourly data, Scenario 1 ..... 22
Figure 23. Procedure 3.2 residual convergence for calibrations to hourly data, Scenario 1 ..... 22
Figure 24. Procedure 3.2 parameter convergence for calibrations to monthly data, Scenario 2 ..... 23
Figure 25. Procedure 3.2 residual convergence for calibrations to monthly data, Scenario 2 ..... 23
Figure 26. Procedure 3.2 parameter convergence for calibrations to daily data, Scenario 2 ..... 24
Figure 27. Procedure 3.2 residual convergence for calibrations to daily data, Scenario 2 ..... 24
Figure 28. Procedure 3.2 parameter convergence for calibrations to hourly data, Scenario 2 ..... 25
Figure 29. Procedure 3.2 residual convergence for calibrations to hourly data, Scenario 2 ..... 25
Figure 30. Procedure 3.2 BoC, Scenario 1 ..... 26
Figure 31. Procedure 3.2 BoC, Scenario 2 ..... 26
Figure 32. Procedure 3.3 optimization results, Scenario 1 ..... 28
Figure 33. Procedure 3.3 optimization results, Scenario 2 ..... 28
Figure 34. Procedure 3.3 parameter convergence for calibrations to monthly data, Scenario 1 ..... 29
Figure 35. Procedure 3.3 residual convergence for calibrations to monthly data, Scenario 1 ..... 29
Figure 36. Procedure 3.3 parameter convergence for calibrations to daily data, Scenario 1 ..... 30
Figure 37. Procedure 3.3 residual convergence for calibrations to daily data, Scenario 1 ..... 30
Figure 38. Procedure 3.3 parameter convergence for calibrations to hourly data, Scenario 1 ..... 31
Figure 39. Procedure 3.3 residual convergence for calibrations to hourly data, Scenario 1 ..... 31
Figure 40. Procedure 3.3 parameter convergence for calibrations to monthly data, Scenario 2 ..... 32
Figure 41. Procedure 3.3 residual convergence for calibrations to monthly data, Scenario 2 ..... 32
Figure 42. Procedure 3.3 parameter convergence for calibrations to daily data, Scenario 2 ..... 33
Figure 43. Procedure 3.3 residual convergence for calibrations to daily data, Scenario 2 ..... 33
Figure 44. Procedure 3.3 parameter convergence for calibrations to hourly data, Scenario 2 ..... 34
Figure 45. Procedure 3.3 residual convergence for calibrations to hourly data, Scenario 2 ..... 34
Figure 46. Procedure 3.3 BoC, Scenario 1 ..... 35
Figure 47. Procedure 3.3 BoC, Scenario 2 ..... 35
Figure 48. Procedure 3.4 BoC, Scenario 1 ..... 36
Figure 49. Procedure 3.4 BoC, Scenario 2 ..... 36
Figure 50. Predicted annual percent energy savings error relative to reference savings for each

retrofit measure, Scenario 1 ..... 38
Figure 51. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 2 ..... 38
Figure 52. Process flow for input calibration methods ..... 57
Figure 53. Histogram of equivalent wall assembly R-values resulting from 50,000 sets of random input selections from ranges specified in this study ..... 61
Figure 54. Histogram of equivalent wall assembly R-values resulting from 50,000 sets of random input selections from ranges specified in BESTEST-EX ..... 62
Figure 55. Histogram of equivalent ceiling assembly R-values resulting from 50,000 sets of random input selections from ranges specified in this study ..... 62
Figure 56. Histogram of equivalent ceiling assembly R-values resulting from 50,000 sets of random input selections from ranges specified in BESTEST-EX ..... 63
Figure 57. Procedure 3.1 energy savings predictions for air-seal retrofit, Scenario 1 ..... 64
Figure 58. Procedure 3.1 energy savings predictions for air-seal retrofit, Scenario 2 ..... 64
Figure 59. Procedure 3.1 energy savings predictions for attic insulation retrofit, Scenario 1 ..... 65
Figure 60. Procedure 3.1 energy savings predictions for attic insulation retrofit, Scenario 2 ..... 65
Figure 61. Procedure 3.1 energy savings predictions for wall insulation retrofit, Scenario 1 ..... 66
Figure 62. Procedure 3.1 energy savings predictions for wall insulation retrofit, Scenario 2 ..... 66
Figure 63. Procedure 3.1 energy savings predictions for programmable thermostat retrofit, Scenario 1 ..... 67
Figure 64. Procedure 3.1 energy savings predictions for programmable thermostat retrofit, Scenario 2 ..... 67
Figure 65. Procedure 3.1 energy savings predictions for low-e windows retrofit, Scenario 1 ..... 68
Figure 66. Procedure 3.1 energy savings predictions for low-e windows retrofit, Scenario 2 ..... 68
Figure 67. Procedure 3.1 energy savings predictions for low solar absorptance roof retrofit, Scenario 1 ..... 69
Figure 68. Procedure 3.1 energy savings predictions for low solar absorptance roof retrofit, Scenario 2 ..... 69
Figure 69. Procedure 3.1 energy savings predictions for duct sealing and insulation retrofit, Scenario 1 ..... 70
Figure 70. Procedure 3.1 energy savings predictions for duct sealing and insulation retrofit, Scenario 2 ..... 70
Figure 71. Procedure 3.1 energy savings predictions for AC replacement retrofit, Scenario 1 ..... 71
Figure 72. Procedure 3.1 energy savings predictions for AC replacement retrofit, Scenario 2 ..... 71
Figure 73. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.2, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 72
Figure 74. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.2, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 72
Figure 75. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.2, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 73
Figure 76. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.2, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 73
Figure 77. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.2, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 74
Figure 78. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.2, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 74
Figure 79. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.3, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 75
Figure 80. Illustration of agreement between calibrated model predictions and reference utility

data for calibration to daily data using Procedure 3.3, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 75
Figure 81. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.3, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 76
Figure 82. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.3, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data..... 76
Figure 83. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.3, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 77
Figure 84. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.3, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data. ..... 77

# List of Tables

Table 1. Characteristics of the Modeled House ..... 3
Table 2. Reference Utility Data ..... 5
Table 3. Retrofit Measures Applied in the Analysis ..... 7
Table 4. Procedure 3.1 Results of $\chi 2$ Sensitivity Analysis, Scenario 1..... 11
Table 5. Procedure 3.1 Results of $\chi 2$ Sensitivity Analysis, Scenario 2..... 12
Table 6. TBoC, Scenario 1 ..... 37
Table 7. TBoC, Scenario 2 ..... 37
Table 8. Computational Cost..... 40
Table 9. Operational Inputs and Uncertainty Ranges ..... 50
Table 10. Asset Inputs and Uncertainty Ranges..... 51
Table 11. Reference Model Input Values for Scenario 1 ..... 53
Table 12. Reference Model Input Values for Scenario 2 ..... 54
Table 13. The 24 Influential Inputs ..... 55
Table 14. Procedure 3.2 and 3.3 Results of $\xi$ Sensitivity Analysis ..... 56

# 1 Introduction

Building energy simulation programs are often used to model the thermal performance of commercial and residential buildings, and more specifically to recommend energy efficiency upgrade packages and operational strategies for existing buildings. These simulation programs may require hundreds of model inputs, many of which have high levels of uncertainty because they rely on software default assumptions and imperfect field data collection procedures. Therefore, accurate model predictions are not guaranteed, even if the underlying physical algorithms are accurate, because uncertainty in model inputs propagates uncertainty in model output. That is, perfect agreement between model predictions and measured output data cannot be expected if the input parameter values have significant errors. Modelers often employ calibration procedures for pre-retrofit building models to reconcile software predictions and measured energy uses. The building model is said to be "calibrated" once a specified level of agreement is achieved between model-predicted and measured data. The general assumption is that calibrating the building model increases the accuracy of energy savings predictions for retrofit measures.

Model calibration methods reported in the literature range in complexity from manual calibration based on user judgment to automated calibration based on analytical, numerical, and statistical methods (Kissock et al. 2003; Lee and Claridge 2002; New et al. 2012; "Treat: high-performance software for building energy analysis" 2010; Subbarao et al. 1988; Sun and Reddy 2006). Most methods pertain to commercial buildings with high-stake retrofit measure considerations (Bertagnolio et al. 2010; Carroll and Hitchcock 1993; Norford et al. 1994; Pappas and Reilly 2011; Pedrini et al. 2002; Raftery et al. 2009, 2011; Westphal and Lamberts 2005; Yoon et al. 2003). ASHRAE 1051-RP (Reddy and Maor 2006) recognized the need for consistency in model calibration techniques, stating that model calibration generally has been regarded as more of an art than a science. They established detailed guidelines for calibrating commercial building models. Specifically, ASHRAE 1051-RP investigates systematic and mathematical approaches for commercial building model calibration that incorporate ASHRAE Guideline 14 (ASHRAE 2002) calibrated-simulation criteria. Posing the search for a calibrated model as an underdetermined optimization problem, the methodology uses multiple calibration solutions to encourage an automated calibration and a consideration of uncertainty in retrofit savings predictions. ASHRAE 1051-RP states that its methodology may not be cost effective for residential buildings.

Residential building modelers are currently employing calibration techniques in the field. The National Renewable Energy Laboratory, along with a working group of industry experts and residential audit tool providers, developed the Building Energy Simulation Test for Existing Homes (BESTEST-EX) to provide a method for software developers to test the accuracy of building energy audit software and calibration procedures (Judkoff et al. 2010). The BESTEST-EX Working Group field trials of preliminary test cases showed a variety of model calibration methods and a general benefit to calibrating residential building models. Although emerging standards (BPI 2011) are attempting to define calibration requirements for residential buildings (e.g., goodness-of-fit acceptance criteria for pre-retrofit predicted and measured energy uses and upper and lower bounds for input parameter adjustments), no widely accepted systematic calibration guidelines are available for residential building applications. Research is needed to explore automated calibration approaches in the context of residential buildings; methods

developed primarily for commercial buildings (such as those described in ASHRAE 1051-RP) may serve as a starting point, but modified and alternative methods tailored to the informational content of residential billing data may be needed.

When developing guidelines for residential model calibration, it is important to consider the specific needs of the residential sector. To achieve the national energy savings goals set forth by the U.S. Department of Energy for the residential sector, millions of homes must be retrofitted. The time and training required for analysts to perform accurate manual calibrations could be barriers to achieving retrofits at this scale. Allowing manual adjustments also makes it difficult to certify analysis software and procedures because human judgment reduces transparency and repeatability. Thus, the residential sector needs repeatable and automated model calibration techniques. To be cost effective, these techniques must be computationally less expensive than those used for commercial buildings. Furthermore, "smart-meters," which typically collect subhourly electricity use data, are becoming more common in residential buildings. Research is needed to understand and evaluate repeatable and automated residential building calibration methods in the context of higher frequency, smart-meter data.

BESTEST-EX uses a single software tool[3] to define a general framework for testing calibration methods; that framework is adapted and applied for this simulation study. BEopt/DOE-2.2 is used to evaluate four mathematical calibration methods in the context of monthly, daily, and hourly utility data for a 1960s-era ranch-style home in a cooling-dominated climate.[4] The four calibration methods implemented in this study are:

- An ASHRAE 1051-RP-based approach
- A simplified simulated annealing optimization approach
- A regression metamodeling optimization approach
- A simple output ratio calibration approach.

The calibration methods are evaluated for monthly, daily, and hourly cases; various retrofit measures defined in BESTEST-EX are applied to the calibrated models and the methods are evaluated based on the accuracy of predicted savings, computational cost, repeatability, automation, and ease of implementation. The purpose of this study is to investigate whether current automated calibration methods can be adapted, applied, and streamlined for residential building applications. This represents an initial step in the overall effort to develop optimized, automated calibration procedures for residential buildings.

[3] See "Performing Calibration Tests Without Using Reference Programs" (Appendix B) of Judkoff et al. 2011.
[4] See beopt.nrel.gov and Christensen et al. 2006 for more information on BEopt.

# 2 Approach for Comparing Calibration Techniques

The following sections describe the approach used to evaluate model calibration techniques. The approach is based on the self-testing procedure described in BESTEST-EX (Judkoff et al. 2010).

## 2.1 Define the House

One 1960s-era all-electric ranch-style home, partly based on BESTEST-EX (Judkoff et al. 2010) Case L200EX-P, is considered in the analysis.[5] The key pre-retrofit characteristics of the modeled house are given in Table 1.

**Table 1. Characteristics of the Modeled House**

| Category | Pre-Retrofit Characteristics |
|---|---|
| Location | Las Vegas, Nevada |
| Orientation | Front of house faces south |
| Dimensions | North/south = 57 ft<br>East/west = 27 ft<br>1 story (8 ft) |
| Garage | None |
| Neighbors | At 15 ft east/west |
| Eave Depth | 2 ft |
| Vented Crawlspace | 2.0 ACH, R-19 between joists |
| Exterior Walls | 2 × 4, 16 in. on center, wood siding, no cavity insulation (R-1 air gap) |
| Unfinished Attic | 2 × 6 joists, R-11 insulation, 3.5 in. thickness |
| Window Type | Single pane<br>Aluminum frame with thermal break<br>National Fenestration Research Council ratings for standard-size: U = 0.774 Btu/h·ft$^2$·°F, solar heat gain coefficient (SHGC) = 0.679 |
| Window Area/Distribution | 20% of exterior wall area<br>33.3% of window area each on north/south<br>16.7% of window area each on east/west |
| Furnace | Electric |
| Air Conditioner (AC) | Seasonal energy efficiency ratio (SEER) 10 |
| Ducts | Uninsulated, in crawlspace, leakage fraction = 0.30 |
| Living Space Specific Leakage Area (SLA) | Ratio = 0.000886 |
| Mechanical Ventilation | Spot vent only (bathroom, kitchen) |
| Water Heater | Electric, energy factor = 0.92, in living space |
| Major Appliances | Refrigerator, electric range, dishwasher, clothes washer, electric clothes dryer |
| Thermostat Set Points | Heating 68°F<br>Cooling 78°F |

[5] Including detailed heating and cooling systems (which were not considered in BESTEST-EX) not only allowed easier modeling with BEopt but permitted the testing of equipment-related retrofits.

## 2.2 Define *Approximate* Inputs

Probability distributions were assigned for each of $\lambda$ *approximate* BEopt inputs to model the uncertainty in these inputs. Probability distributions describe the likelihood of random variables taking certain values. *Triangular probability distributions* were used for this analysis. This type of distribution is characterized by having the greatest probability of selection at the "best-guess," or nominal, value, with linearly decreasing probability to zero at the range extrema (Judkoff et al. 2010; Kotz and van Dorp 2004; Reddy and Maor 2006). An asymmetric triangular probability distribution is shown in Figure 1, where "Nominal" refers to the nominal ("best-guess") value, "Min" the minimum value, and "Max" the maximum value.

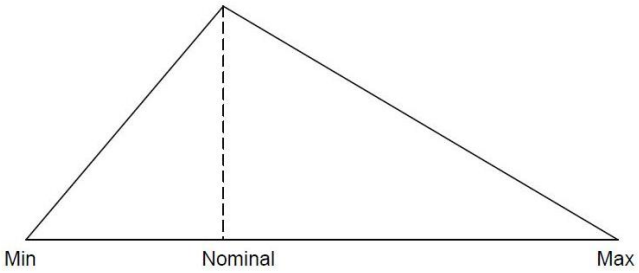


**Figure 1. Triangular probability distribution**

Table 9 of Appendix A lists the approximate inputs primarily related to the operation of the modeled home ("operational inputs"), including minimum, maximum, and nominal values, as well as units of measure. Table 10 of Appendix A lists the approximate inputs primarily related to the physical features of the modeled home ("asset inputs"), including minimum, maximum, and nominal values, as well as units of measure. Some ranges were specified based on BESTEST-EX (Judkoff et al. 2010) and limits set forth in Building Performance Institute Standard 2400 (BPI 2011). Other ranges were estimated using engineering judgment.[6] Appendix G compares the ranges for wall and ceiling assembly R-values in this study to those in BESTEST-EX.

## 2.3 Generate Utility Data

The first step in generating synthetic electric utility data[7] (kWh) was to randomly select an *explicit input value* $v_i$ from each approximate input's triangular probability distribution. Each probability distribution function is continuous, so randomly selected values could fall anywhere between the minimum and maximum values shown in Table 9 and Table 10. Equations for this selection scheme are shown in Appendix A. Explicit input values were then substituted into the model's building description file to create the *reference model*, and the reference model was simulated in BEopt/DOE-2.2 for one year using Typical Meteorological Year 3 weather data for Las Vegas, Nevada ("National Solar Radiation Data Base" 2008). The set of $n$ = 8,760 hourly total site electricity use data points were extracted from the simulation output. This set of hourly data became the *reference utility data* for hourly calibrations; reference utility data for monthly

[6] More research is needed to characterize uncertainty ranges for building input parameters.
[7] This study considers only electric utility data since the modeled home consumes no other fuel type.

and daily calibrations were obtained by summing the appropriate hourly reference data into $n$ = 12 monthly and $n$ = 365 daily data points, respectively.

To obtain two distinct calibration scenarios, this process was performed for 100 sets of randomly selected explicit input values $v_i$ (reference models) to generate 100 sets of reference utility data. The 100 sets of reference utility data were sorted on increasing annual cooling electricity, and the scenarios corresponding to the 5th and 95th sets of reference utility data were chosen as the overprediction and underprediction calibration scenarios investigated in this study. The uncalibrated model (based on the nominal inputs) that overpredicted the reference utility data (5th set) is referred to as "Scenario 1" and the uncalibrated model that underpredicted the reference utility data (95th set) is referred to as "Scenario 2." Scenario 1 and Scenario 2 are such that the uncalibrated model overpredicts and underpredicts the reference electricity consumption data by + 25.6% and – 4.7%, respectively.[8] See Table 2 for a summary of this synthetic utility bill data selection. The selected sets of $v_i$ are given in Table 11 and Table 12.

**Table 2. Reference Utility Data**

| | Cooling Electricity (kWh) | *Percent Difference* | Electricity Consumption (kWh) | *Percent Difference* |
|---|---|---|---|---|
| Uncalibrated | 6,145 | | 25,710 | |
| Scenario 1 | 4,619 | +33.1% | 20,473 | +25.6% |
| Scenario 2 | 9,225 | –33.4% | 26,965 | –4.7%* |

* Scenario 2 had compensating errors in which the uncalibrated model overpredicted heating energy but underpredicted cooling energy, which resulted in better agreement with annual reference electricity consumption compared to Scenario 1.

## 2.4 Identify the Most *Influential* Inputs

We performed a preliminary sensitivity analysis to compare calibration methods and to identify the 24 most influential inputs from the approximate inputs listed in Table 9 and Table 10. These inputs were considered as starting points for each applicable calibration procedure.[9] The sensitivity analysis was a Monte Carlo procedure in which $\eta = 100$ random samples were made from each approximate input's triangular probability distribution; the selected input values were substituted into the building description file one at a time (holding the other inputs at their nominal values), and then all files were simulated in automated batch mode (2,400 total simulations). A dimensionless sensitivity coefficient, $\xi_j$, was calculated for each of the $\lambda$ approximate inputs perturbed in the analysis according to:

$$\xi_j = \frac{\sigma_j}{\mu_j}, \tag{1}$$

[8] Although the calibration scenarios are characterized in terms of direction (i.e., overprediction and underprediction), the relative degree of difference between the uncalibrated and reference models may be a more influential factor for calibration success.

[9] ASHRAE 1051-RP suggests that 20–24 influential inputs be identified using walk-through audits and heuristics (Reddy and Maor 2006). To allow for fair comparison across calibration procedures, the same set of 24 influential inputs was considered for each applicable calibration procedure.

where:
$\sigma_j, \mu_j$ = the standard deviation and mean of the $\eta$ annual output values corresponding to each input $j$ (Hamby 1994).

Each $\xi_j$, therefore, is the output coefficient of variation. A descending sort of the $\xi_j$ values revealed the 24 most sensitive parameters, which we call "influential inputs." The sensitivity coefficients of the influential inputs are graphically displayed in Figure 2. See Table 13 for these coefficients. Also, Figure 52 in Appendix C shows the process flow for developing approximate input ranges and identifying influential inputs.

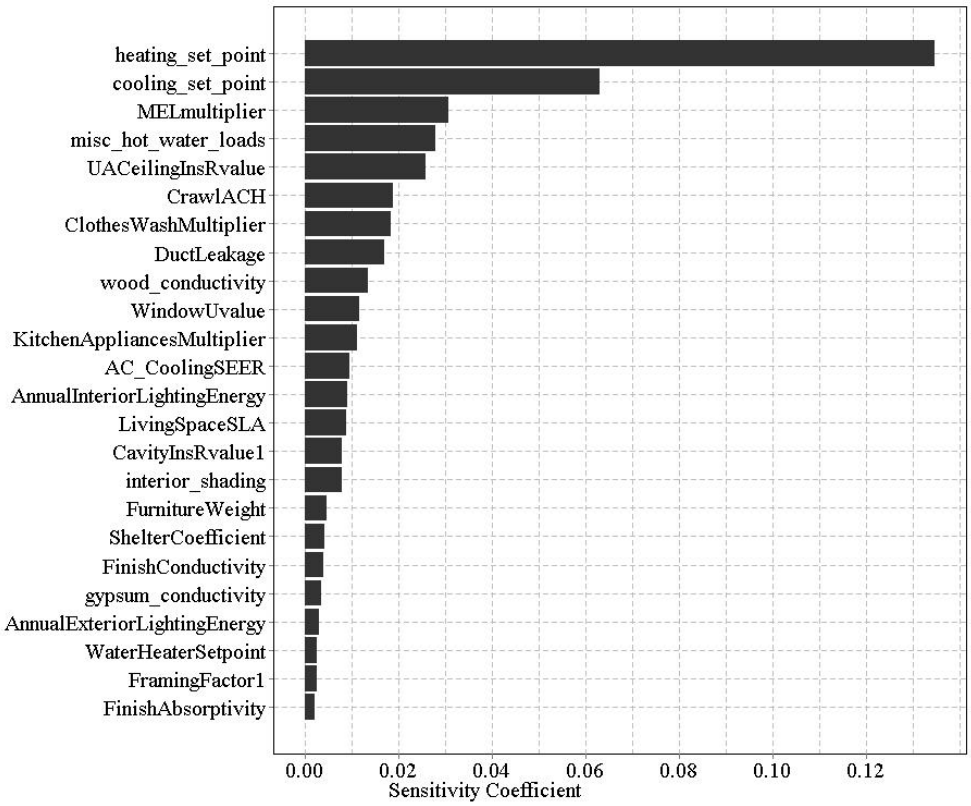


**Figure 2. The 24 influential inputs**

## 2.5 Perform Calibrations

Four calibration techniques were implemented and compared for this analysis:

1. An ASHRAE 1051-RP-based approach
2. A simplified simulated annealing optimization approach
3. A regression metamodeling optimization approach
4. A simple output ratio calibration approach.

The first three of these techniques require sensitivity analyses to reduce the dimensionality of the search space. These analyses were performed on the subset of 24 influential inputs shown in

Figure 2. Detailed descriptions and results for the four calibration techniques are presented in Section 3.

## 2.6 Apply Retrofit Measures

After the four calibration techniques listed in Section 2.5 were performed and their calibrated models recovered (if applicable), they were evaluated based on their ability to recover the reference input values (explicit input values) and improve the accuracy of energy savings predictions for retrofit measures. The four techniques, along with various retrofit measure applications, will be referred to as Procedure 3.1, Procedure 3.2, Procedure 3.3, and Procedure 3.4, respectively. Table 3 describes various retrofit measures, most of which are based on BESTEST-EX (Judkoff et al. 2010) that were applied to:

- The reference model (applying the explicit inputs)
- The nominal ("best-guess") model, called the *uncalibrated model*
- Each *calibrated model*.

**Table 3. Retrofit Measures Applied in the Analysis**

| Retrofit Measure | BEopt Input | Change[a] | Units |
|---|---|---|---|
| Air-Seal | LivingSpaceSLA | – 0.000451 | $in.^2/ft^2$ |
| Attic Insulation | UACeilingInsRvalue | + 28.9 | $h{\cdot}ft^2{\cdot}°F/Btu$ |
| | UACeilingInsThickness | + 8.0 | in |
| Wall Insulation | CavityInsRvalue1 | 13.0 | $h{\cdot}ft^2{\cdot}°F/Btu$ |
| Programmable Thermostat | heating_set_point | – 6.0 (10PM–6AM) | °F |
| | cooling_set_point | + 6.0 (8AM–5PM) | °F |
| Low-e Windows | WindowUvalue | 0.279 | $Btu/h{\cdot}ft^2{\cdot}°F$ |
| | WindowSHGC | 0.346 | – |
| Low Solar Absorptance Roof | RoofingMaterialAbsorptivity | 0.2 | – |
| Duct Sealing and Insulation | UnconditionedDuctRvalue | 8.0 | $h{\cdot}ft^2{\cdot}°F/Btu$ |
| | ReturnLeak | 0.01 | frac of AH[b] fan flow |
| | SupplyLeak | 0.09 | frac of AH fan flow |
| | AHLeakRA | 0.04 | frac of AH fan flow |
| | AHLeakSA | 0.01 | frac of AH fan flow |
| AC Replacement | AC_CoolingSEER | 16.0 | kBtu/kWh |
| | HeatingSizeMethod | autosize | kBtu/kWh |
| | CoolingSizeMethod | autosize | tons |
| Combined | … | … | … |

[a] Leading +/− denotes a change relative to the pre-retrofit building; all other values are absolute (replace the value for the pre-retrofit building). Pre-retrofit values can be found in Table 1.
[b] Air handler.

After each retrofit measure was applied, we labeled post-retrofit annual results:

- $\psi^r$ for the reference model energy savings prediction
- $\psi^u$ for the uncalibrated model energy savings prediction
- $\psi^c$ for the calibrated model energy savings prediction.

## 2.7 Compare Calibration Procedures

The absolute errors $\varepsilon(\psi^c)$ and $\varepsilon(\psi^u)$ were then calculated for the calibrated and uncalibrated model energy savings predictions using:

$$\varepsilon(\psi^c) = |\psi^c - \psi^r|, \tag{2}$$

$$\varepsilon(\psi^u) = |\psi^u - \psi^r|, \tag{3}$$

and the benefit of calibration (BoC) (Judkoff et al. 2010) was calculated using:

$$BoC = \varepsilon(\psi^u) - \varepsilon(\psi^c). \tag{4}$$

If BoC $>$ 0, then energy savings predictions are improved by calibration. The BoC has the same units as the energy savings predictions (kWh) and therefore can be compared across retrofit measures. BoC can also be converted to monetary values by assuming average utility prices (Judkoff et al. 2010), which clarifies the BoC with respect to the overall cost of the audit and home improvements.

We can express the annual calibrated model predicted energy savings, $\psi^c$, as a percent of the pre-retrofit reference utility data, $\Gamma_{ref}$, using:

$$\varphi^c = \frac{\psi^c}{\Gamma_{ref}} \times 100\%, \tag{5}$$

and the annual reference energy savings, $\psi^r$, as a percent of the pre-retrofit reference utility data using:

$$\varphi^r = \frac{\psi^r}{\Gamma_{ref}} \times 100\%. \tag{6}$$

Equations (5) and (6) are used in Sections 3.1.4 and 4.1 to produce graphical representations of energy savings prediction accuracy for the calibration cases.

# 3 Calibration Techniques

This section describes the application of Procedures 3.1–3.4 to the over- and underprediction calibration scenarios presented in Section 2.3. Overall results comparisons between procedures are presented in Section 4. Procedures 3.1–3.3 each contain steps selecting adjusted inputs and inverting to recover calibrated inputs. Figure 52 of Appendix C shows the process flow for these steps.

## 3.1 ASHRAE 1051-RP-Based Approach

The calibration procedure applied in this section (Procedure 3.1) is based on the four-step methodology described in ASHRAE 1051-RP (Reddy and Maor 2006). The four general steps are:

1. Define influential parameters.
2. Perform a coarse grid calibration.
3. Perform a refined grid calibration.
4. Predict ranges of energy savings for retrofit measures.

A unique feature of this calibration procedure is that it considers multiple calibration "solutions" for each calibration scenario (because there are multiple ways to "match" utility billing data by adjusting input values). Each calibration solution leads to a separate "calibrated" pre-retrofit model; these models are used to predict a range of potential energy savings for each retrofit. The steps of Procedure 3.1 are described in Sections 3.1.1–3.1.4.

### *3.1.1 Define Influential Input Parameters*

This initial stage includes gathering and quality checking energy use data and building specifications, as well as defining a *heuristic template* containing the building's most influential input parameters. The building's heuristic template isolates influential parameters from non-influential parameters. Included in the heuristic template are the influential parameters' assigned "best-guess" estimates and ranges of variation characterized by minimum and maximum values; ASHRAE 1051-RP suggests that 20–24 influential parameters be identified and assembled into the building's heuristic template. Influential parameter identifications and ranges of variation assignments are made by building professionals with extensive field experience using a walk-thru audit and practical past experience (Reddy and Maor 2006). For this analysis, we considered the 24 influential inputs based on parametric sensitivity tests as identified in Section 2.4 and presented in Figure 2.

### *3.1.2 Perform a Coarse Grid Calibration*

This step of the methodology identifies feasible combinations of input parameters by employing a Latin Hypercube Monte Carlo (LHMC) stratified sampling procedure. ASHRAE 1051-RP suggests that 5,000–10,000 three-level LHMC realizations are adequate for actual buildings, and that substantially fewer realizations are adequate for synthetic buildings. For this study, 2,500 LHMC realizations were simulated to limit computational cost. The procedure for performing a three-level LHMC is described in Appendix D.

See Eqs. (18) and (19) of Appendix F for coefficient of variation (CV) root mean square error (RSME) and nominal mean base error (NMBE) goodness-of-fit index definitions. ASHRAE 1051-RP adopts ASHRAE Guideline 14 "calibrated-simulation" criteria and refers to calibration solutions that meet these criteria as *feasible calibration solutions*. See criteria in Section 5.3.2.4 of (ASHRAE 2002) for acceptable calibration solutions:

1. Monthly ($n = 12$) utility data: CV(RMSE) ≤ 15% and –5% ≤ NMBE ≤ 5%, or
2. Hourly ($n$ = 8,760) utility data: CV(RMSE) ≤ 30% and –10% ≤ NMBE ≤ 10%.[10]

In this analysis, for Scenario 1, the monthly calibration case identified 73 feasible calibration solutions, the daily calibration case identified 147 feasible calibration solutions, and the hourly calibration case identified 245 feasible calibration solutions. For Scenario 2, the monthly calibration case identified 142 feasible calibration solutions, the daily calibration case identified 185 feasible calibration solutions, and the hourly calibration case identified 478 feasible calibration solutions.

Next, ASHRAE 1051-RP employs a sensitivity analysis to help "alleviate the curse of dimensionality" (Reddy and Maor 2006). Specifically, it describes a $\chi^2$ application for determining parameter sensitivity. The $\chi^2$ formula for three-level sampling of each influential parameter's probability distribution is given by

$$\chi_i^2 = \sum_{s=1}^{3} \frac{\left(\rho_{obs,s,i} - \rho_{exp}\right)^2}{\rho_{exp}}, \tag{7}$$

for $i = 1, \ldots, 24$, where:
$\rho_{obs,s,i}$ = observed number of occurrences on level $s$ (i.e., number of samples from one of $x_i^{low}, x_i^{mid}, x_i^{high}$), and
$\rho_{exp}$ = expected number of occurrences on each of the three levels (i.e., total feasible solutions divided by three).

Equation (7) was applied to the subset of realizations resulting in feasible calibration solutions, which were sorted in descending order of $\chi_i^2$. In ASHRAE 1051-RP, parameters with a $\chi_i^2$ value of at least 9.21 are considered strong ($\alpha = 0.01$); all other parameters are considered *weak*. In this analysis, we investigated a streamlined approach of using a constant number of strong parameters across calibration Procedures 3.1–3.3 and across monthly, daily, and hourly cases.[11] Preliminary results showed that the six greatest $\chi_i^2$ values were approximately equal to or greater than ASHRAE 1051-RP's prescribed 9.21 for the daily calibration scenario (the middle of the three frequency levels investigated). Therefore, the top $\tau = 6$ parameters of the sorted $\chi_i^2$ were considered strong parameters and are included in Section 3.1.3 as *adjustable* parameters for the refined grid calibration. Results for the sensitivity analyses are given in Table 4 and Table 5.

[10] ASHRAE Guideline 14 does not report criteria for acceptable calibration solutions using daily utility data; for daily utility data, this analysis assumes CV(RSME) ≤ 18% and –6% ≤ NMBE ≤ 6% (obtained by interpolating monthly and hourly criteria).
[11] $\chi^2$ tests are not performed in Procedures 3.2 and 3.3, so the streamlined approach was necessary for these two procedures and therefore was also used for Procedure 3.1 to enable simpler comparisons across all procedures.

**Table 4. Procedure 3.1 Results of $\chi^2$ Sensitivity Analysis, Scenario 1**

| $n$ | Strong Parameter | Level Frequency $x_i^{low}$ | $x_i^{mid}$ | $x_i^{high}$ | $\chi_i^2$ |
|---|---|---|---|---|---|
| 12 | heating_set_point | 72 | 1 | 0 | 140.1 |
| | cooling_set_point | 0 | 2 | 71 | 134.3 |
| | MELmultiplier | 48 | 22 | 3 | 42.0 |
| | misc_hot_water_loads | 31 | 33 | 9 | 14.6 |
| | LivingSpaceSLA | 33 | 15 | 25 | 6.7 |
| | gypsum_conductivity | 17 | 33 | 23 | 5.4 |
| 365 | heating_set_point | 146 | 1 | 0 | 288.0 |
| | cooling_set_point | 0 | 11 | 136 | 232.9 |
| | MELmultiplier | 89 | 50 | 8 | 67.0 |
| | misc_hot_water_loads | 73 | 53 | 21 | 28.1 |
| | AC_CoolingSEER | 28 | 56 | 63 | 14.0 |
| | LivingSpaceSLA | 67 | 38 | 42 | 10.1 |
| 8,760 | heating_set_point | 231 | 14 | 0 | 410.8 |
| | cooling_set_point | 4 | 53 | 188 | 222.4 |
| | MELmultiplier | 140 | 79 | 26 | 79.7 |
| | ClothesWashMultiplier | 108 | 89 | 48 | 23.0 |
| | misc_hot_water_loads | 107 | 85 | 53 | 18.1 |
| | UACeilingInsRvalue | 55 | 85 | 105 | 15.5 |

**Table 5. Procedure 3.1 Results of $\chi^2$ Sensitivity Analysis, Scenario 2**

| $n$ | Strong Parameter | Level Frequency $x_i^{low}$ | $x_i^{mid}$ | $x_i^{high}$ | $\chi_i^2$ |
|---|---|---|---|---|---|
| 12 | cooling_set_point | 130 | 11 | 1 | 217.6 |
| | heating_set_point | 115 | 27 | 0 | 152.8 |
| | MELmultiplier | 17 | 37 | 88 | 56.6 |
| | AC_CoolingSEER | 68 | 46 | 28 | 17.0 |
| | interior_shading | 42 | 36 | 64 | 9.2 |
| | ClothesWashMultiplier | 35 | 44 | 63 | 8.6 |
| 365 | cooling_set_point | 158 | 26 | 1 | 230.8 |
| | heating_set_point | 132 | 53 | 0 | 143.1 |
| | MELmultiplier | 28 | 49 | 108 | 55.8 |
| | interior_shading | 32 | 70 | 83 | 22.8 |
| | misc_hot_water_loads | 55 | 41 | 89 | 19.8 |
| | gypsum_conductivity | 43 | 69 | 73 | 8.6 |
| 8,760 | heating_set_point | 310 | 168 | 0 | 302.3 |
| | cooling_set_point | 315 | 133 | 30 | 261.4 |
| | MELmultiplier | 76 | 154 | 248 | 93.1 |
| | interior_shading | 133 | 131 | 214 | 28.1 |
| | misc_hot_water_loads | 121 | 146 | 211 | 27.1 |
| | AnnualInteriorLightingEnergy | 135 | 137 | 206 | 20.5 |

Another objective of Section 3.1.2 is to determine the top 10 feasible calibration solutions. These are determined by sorting in ascending order the "total goodness-of-fit" values, $\delta_{TOTAL}$, of feasible calibration solutions and subsequently selecting the corresponding top 10 calibration solutions. The formula for "total goodness-of-fit" is given by Eq. (21) of Appendix F.

### 3.1.3 Perform a Refined Grid Calibration

For this step, ASHRAE 1051-RP suggests a manual or mathematical optimization approach. The latter is adopted in this study, specifically a gradient-based nonlinear optimization technique. We chose to apply a simulated annealing algorithm, using Eq. (18) from Appendix F as its objective function. A detailed description of the simulated annealing algorithm is provided in Section 3.2.2. A separate optimization was performed for each of the top 10 feasible calibration solutions. For each optimization, the values of the strong parameters identified in Section 3.1.2 were the initial guesses for the simulated annealing algorithm. All other weak inputs were initially set at their uncalibrated (nominal) values and were not adjusted during the optimization. Some optimization algorithm parameters must be prescribed before executing the calibration; we chose an initial simulated annealing temperature $T_0 = 10$ and the number of iterations per parameter $N = 100$.[12] The main objective was to refine the top 10 feasible calibration solutions to recover 10 *refined* calibration solutions. The 10 refined calibration solutions are referred to as

[12] Simulated annealing algorithm parameter values for $T_0$ and $N$ were chosen based on the authors' experience and the convergence observed during preliminary simulations; algorithm performance depends on these chosen values. More information describing the simulated annealing implementation is provided in Section 3.2.2.

*calibrated models*. The results of the refined grid calibrations are given in Figure 3 through Figure 8. Recovered values are the averages across the 10 refined calibration solutions.

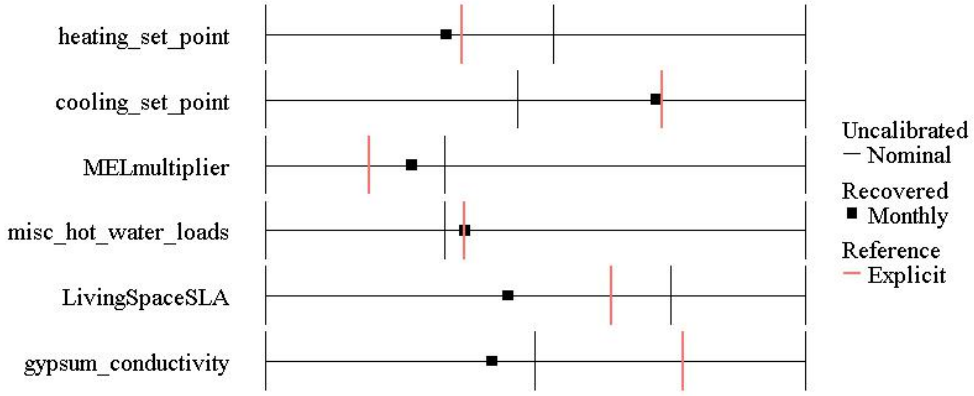


**Figure 3. Procedure 3.1 refined grid results for calibration to monthly data, Scenario 1**

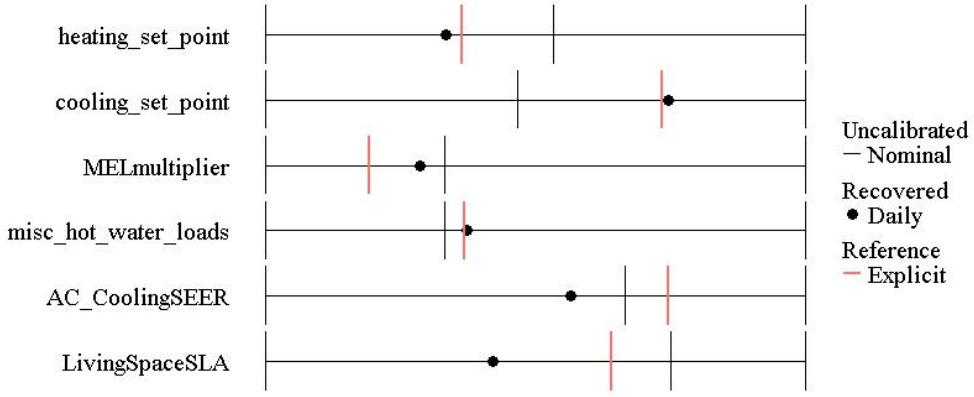


**Figure 4. Procedure 3.1 refined grid results for calibration to daily data, Scenario 1**

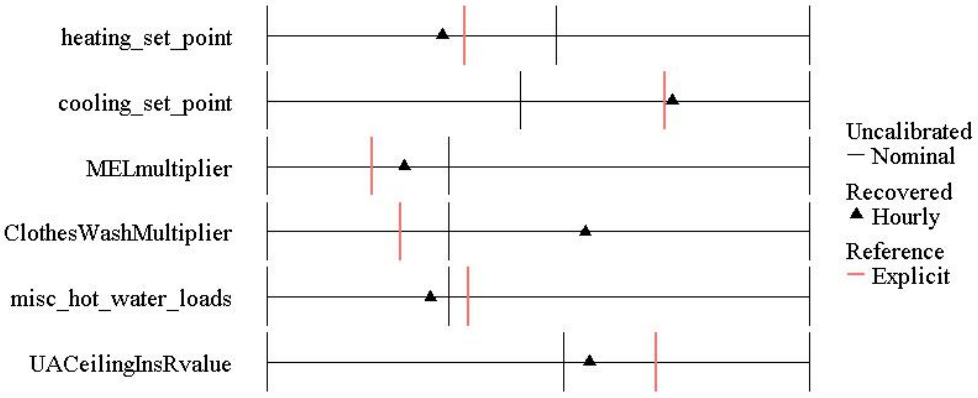


**Figure 5. Procedure 3.1 refined grid results for calibration to hourly data, Scenario 1**

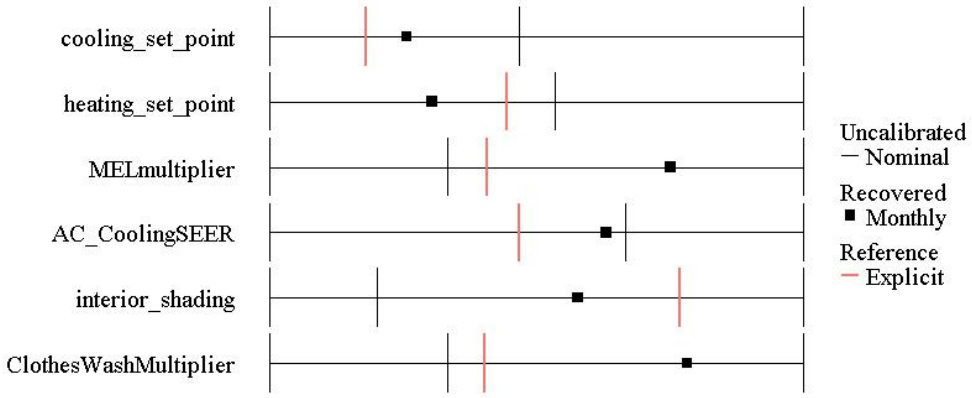


**Figure 6. Procedure 3.1 refined grid results for calibration to monthly data, Scenario 2**

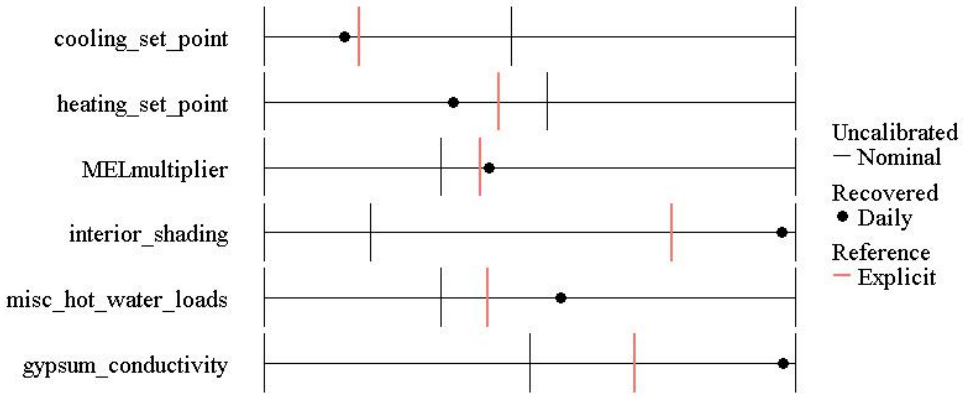


**Figure 7. Procedure 3.1 refined grid results for calibration to daily data, Scenario 2**

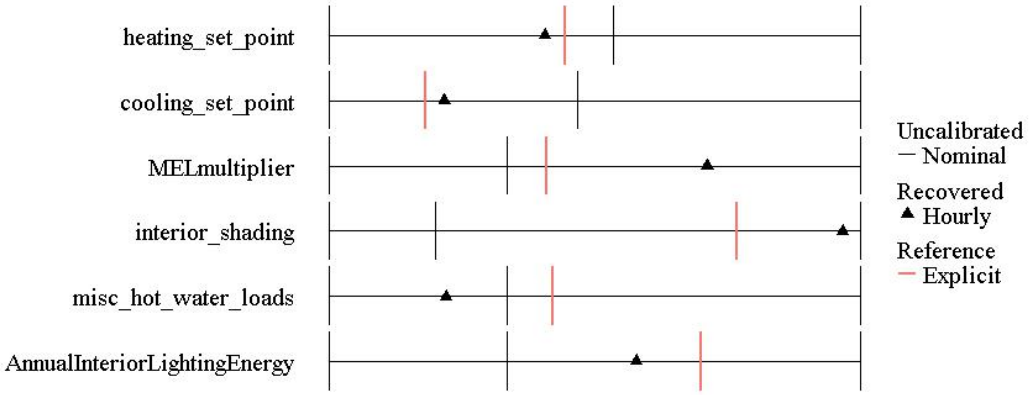


**Figure 8. Procedure 3.1 refined grid results for calibration to hourly data, Scenario 2**

### *3.1.4 Predict Ranges of Energy Savings for Retrofit Measures*

Using the 10 calibrated models, this step attempts to predict ranges of energy savings for the retrofit measures defined in Table 3. Each retrofit measure was individually applied to each of

the 10 calibrated models identified in Section 3.1.3, and calibrated model energy savings predictions were calculated. The mean savings across the calibrated model energy savings predictions were calculated and used to compute the BoC. These results are summarized in Figure 9 and Figure 10.

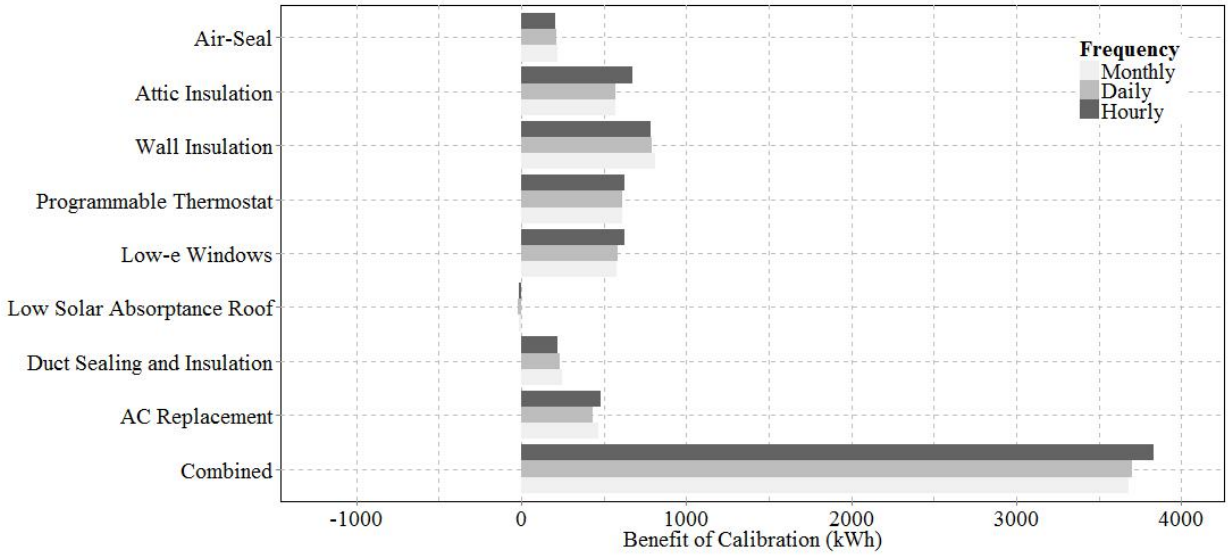


**Figure 9. Procedure 3.1 BoC, Scenario 1**

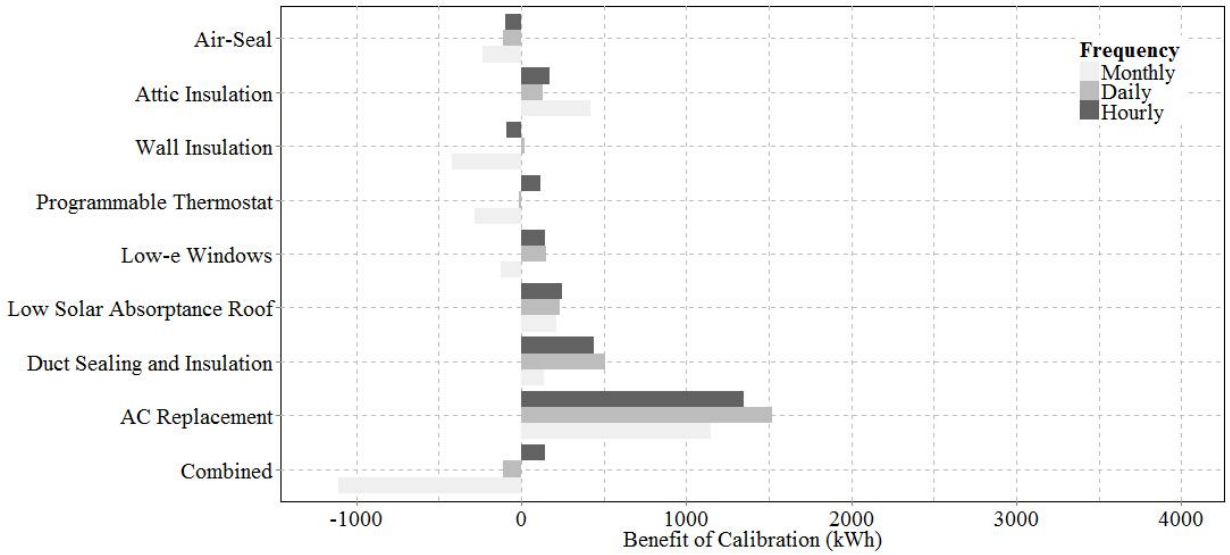


**Figure 10. Procedure 3.1 BoC, Scenario 2**

Equations (5), (6), and (21) were used to produce graphical representations (see Figure 11 and Figure 12) for the accuracy of Procedure 3.1 in predicting energy savings for the retrofit measures. For visual clarity, energy savings predicted by the uncalibrated model are not plotted but are noted in the top right of each figure. The scale varies on each figure so that the distribution of savings predictions can be seen, even if they occur over a narrow range of

savings. Additional plots are given in Appendix H. For each plot, $\delta_{TOTAL}$ on the y-axis represents the "total-goodness-of-fit" calculated using Eq. (21) of Appendix F. Smaller $\delta_{TOTAL}$ values represent "better" agreement (according to Eq. (21)) between predicted pre-retrofit energy use and synthetic utility billing data. In Figure 11, the reference savings line covers the line representing savings using the hourly calibrated model.

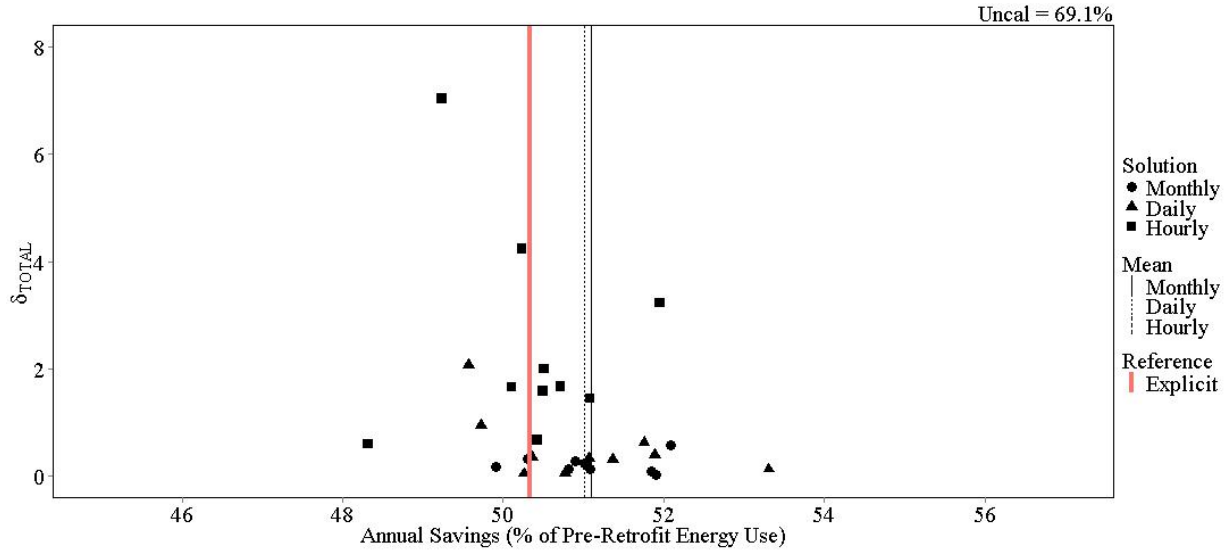


**Figure 11. Procedure 3.1 energy savings predictions for combined retrofit, Scenario 1**

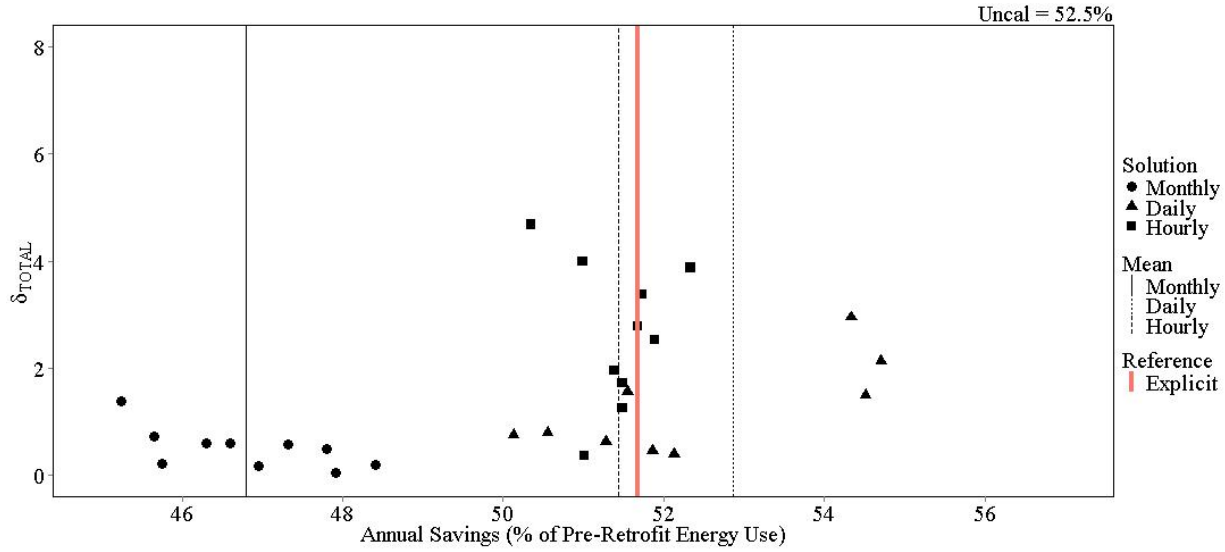


**Figure 12. Procedure 3.1 energy savings predictions for combined retrofit, Scenario 2**

To estimate ranges of energy savings, we used the 10 predicted savings values to calculate the 99% tolerance limits ($\gamma = 0.01$) between which 95% ($\alpha = 0.05$) of the energy savings

predictions would fall, assuming an approximate normal distribution (Walpole and Myers 1989). These tolerance intervals are shown in Figure 13 and Figure 14.

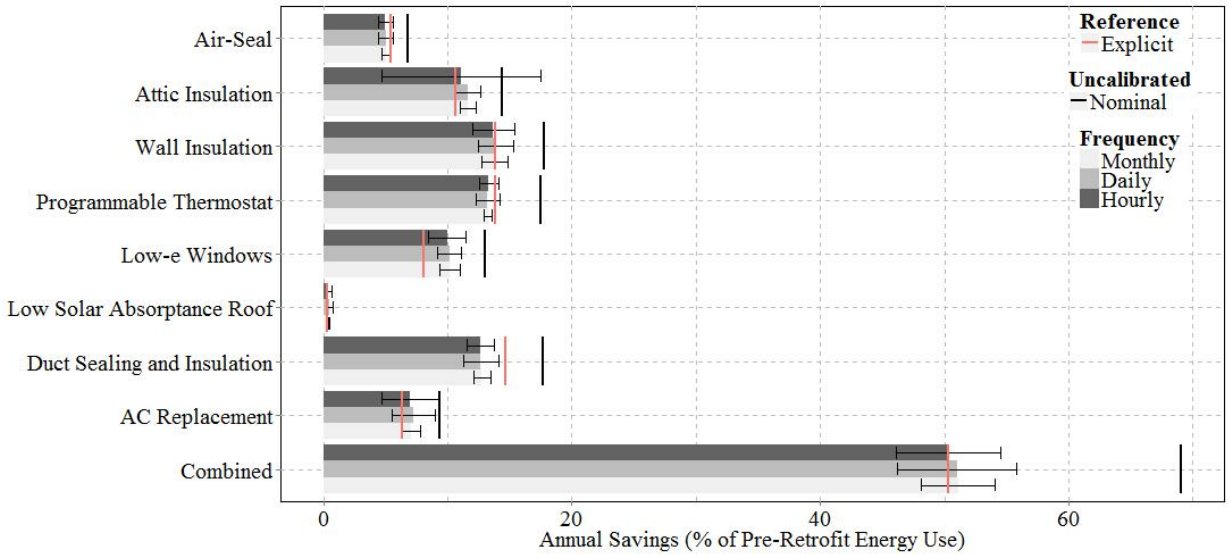


**Figure 13. Estimated ranges for energy savings predictions using Procedure 3.1, Scenario 1**

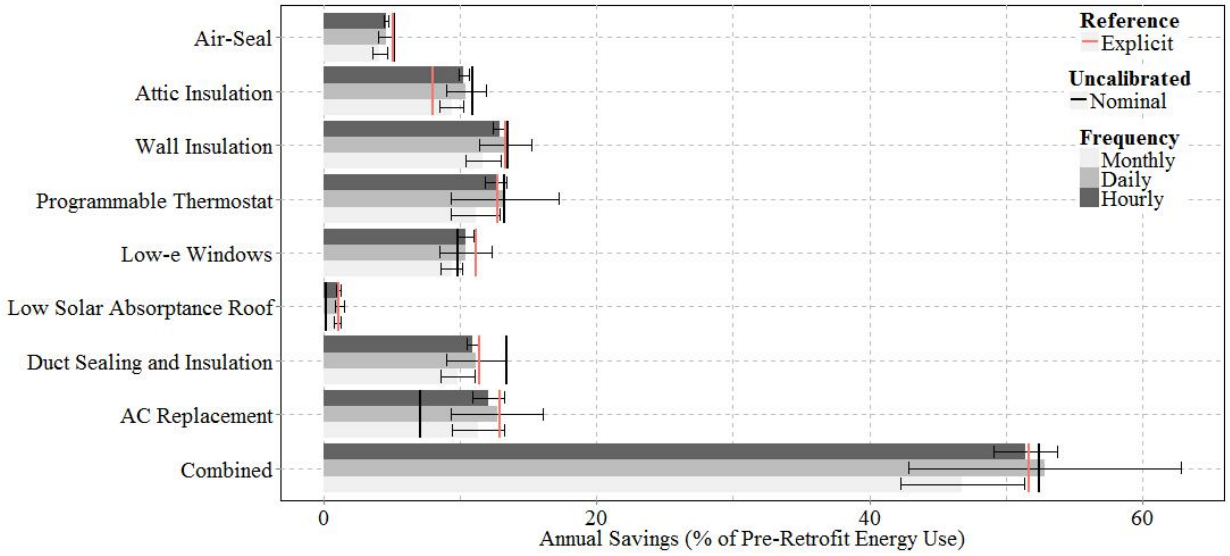


**Figure 14. Estimated ranges for energy savings predictions using Procedure 3.1, Scenario 2**

## 3.2 Simplified Simulated Annealing Optimization Approach

Simulated annealing is an efficient optimization algorithm for nonlinear inversion (Kirkpatrick et al. 1983; Kuperman et al. 1990). The steps of the simulated annealing algorithm include initial gradient calculations, randomly perturbing the parameters, evaluating the change $\Delta E$ in the

objective function, deciding whether to accept perturbations resulting in positive $\Delta E$, and lowering the simulated annealing temperature $T$ (Collins et al. 1992). For this calibration technique, the simulated annealing implementation follows an initial sensitivity analysis for reducing parameter search space dimensionality.

### 3.2.1 Sensitivity Analysis

The sensitivity analysis for this procedure follows the method described in Section 2.4, except that we assume the analyst preselects 24 influential inputs to be considered in the analysis. To provide a fair comparison across calibration procedures, we began with 24 influential inputs from Figure 2. The most sensitive $\tau = 6$ of these were selected as strong parameters to be adjusted in the simulated annealing algorithm.[13] All other input parameters were frozen at their uncalibrated (nominal) values during the optimization. The results for the top $\tau = 6$ parameters of this sensitivity analysis, with $\eta = 100$ random samples, are shown in Figure 15. Note that we expect the results to be very similar to those shown in Figure 2, but not necessarily identical, because the analyses are based on different sets of randomly selected values. See Table 14 for these coefficients in tabular form.

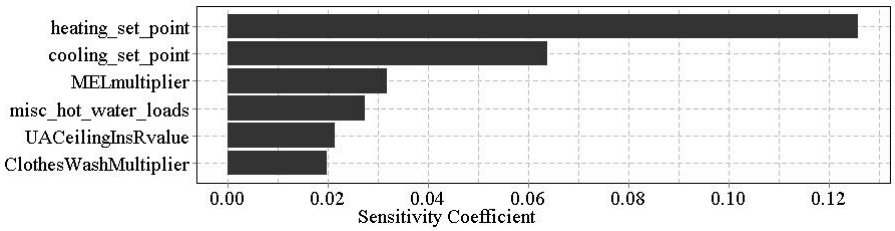


**Figure 15. Procedure 3.2 $\xi$ sensitivity analysis results**

### 3.2.2 Iterative Search

We made the strong parameters listed in Table 14 adjustable parameters for the simulated annealing algorithm and began the iterative search at nominal values, choosing the initial system temperature $T_0 = 10$ and the number of iterations per parameter $N = 100$, as in Procedure 3.1. Each adjusted parameter was bounded by the minimum and maximum values of its probability distribution. We defined the probability $P(\Delta E)$ that a perturbation results in an acceptable energy increase (i.e., objective function value increase) by the commonly used Boltzmann probability factor (Kirkpatrick et al. 1983):

$$P(\Delta E) = \exp\left(\frac{-\Delta E}{T}\right), \tag{8}$$

where:
$\Delta E = E_{new} - E_{old}$,
$E_{new}$ = objective function value of perturbed guess,
$E_{old}$ = objective function value of previous guess, and
$T$ = simulated annealing temperature.

[13] This number of adjustable parameters was chosen to be equal to that in Section 3.1.3.

The temperature parameter $T$ decreases after each iteration according to a "fast simulated annealing cooling schedule," given by:

$$T = \frac{T_0}{i}, \tag{9}$$

for $i = 1, \ldots, N$ (Collins et al. 1992; Szu and Hartley 1987). For larger values of $T$, the algorithm is more likely to accept perturbations that have caused energy use to increase (less agreement between predicted energy use and billing data). The iterative search is thus permitted to escape local minima as it investigates the parameter space (Kirkpatrick et al. 1983). Figure 16 and Figure 17 give the results of the simulated annealing optimization in terms of the calibrated input values.

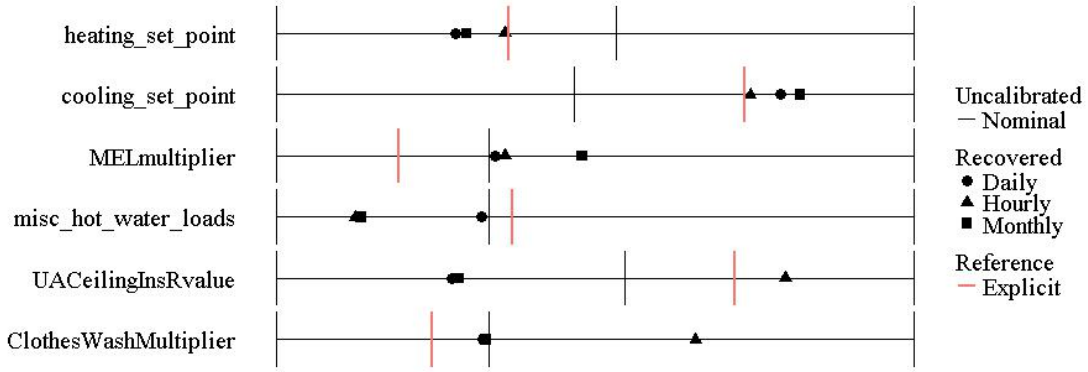


**Figure 16. Procedure 3.2 optimization results, Scenario 1**

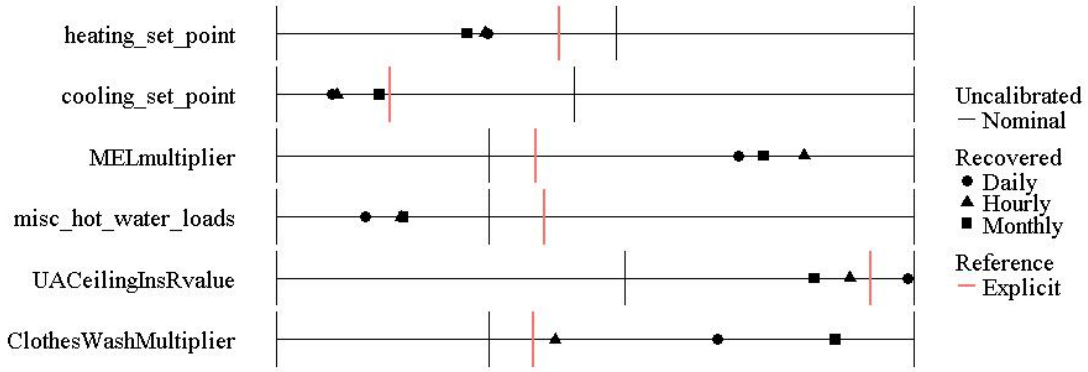


**Figure 17. Procedure 3.2 optimization results, Scenario 2**

Figure 18 through Figure 29 show the iterative search convergence processes for parameters relative to reference input values and for the objective function value (red lines are explicit reference values and dashed lines represent the number of function evaluations required to obtain

the residual minimum). The convergence plots for the objective function value depict that at higher temperatures $T$ the algorithm allows more energy increases and eventually, as $T$ cools according to the schedule, convergences on a least objective function value.

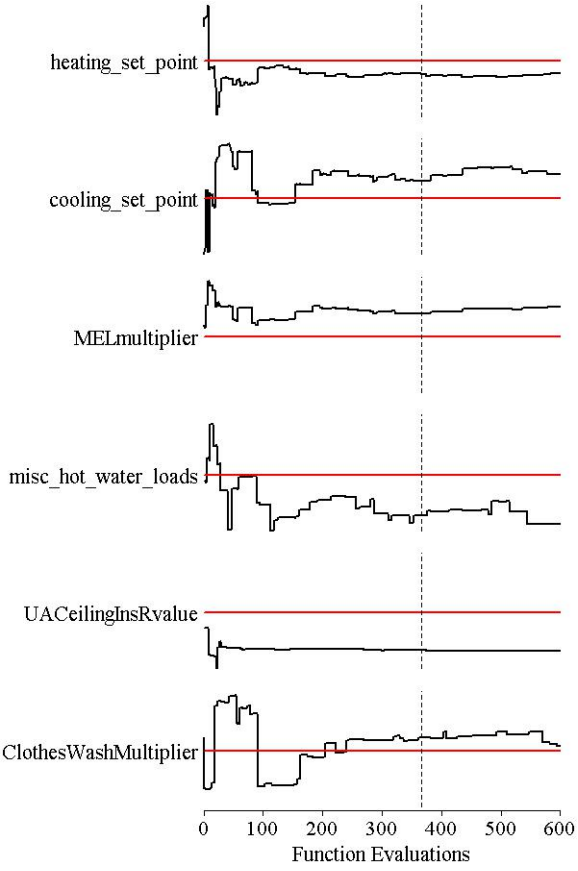


**Figure 18. Procedure 3.2 parameter convergence for calibrations to monthly data, Scenario 1**

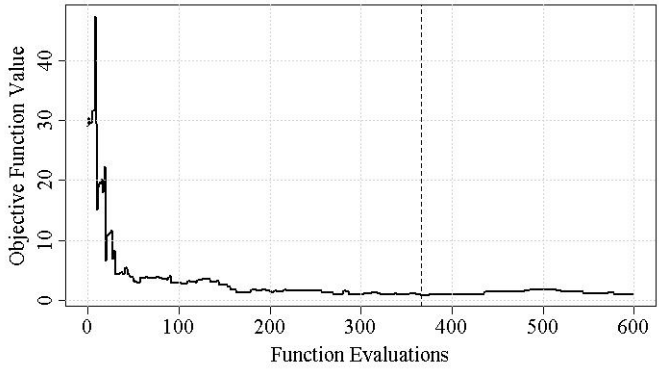


**Figure 19. Procedure 3.2 residual convergence for calibrations to monthly data, Scenario 1**

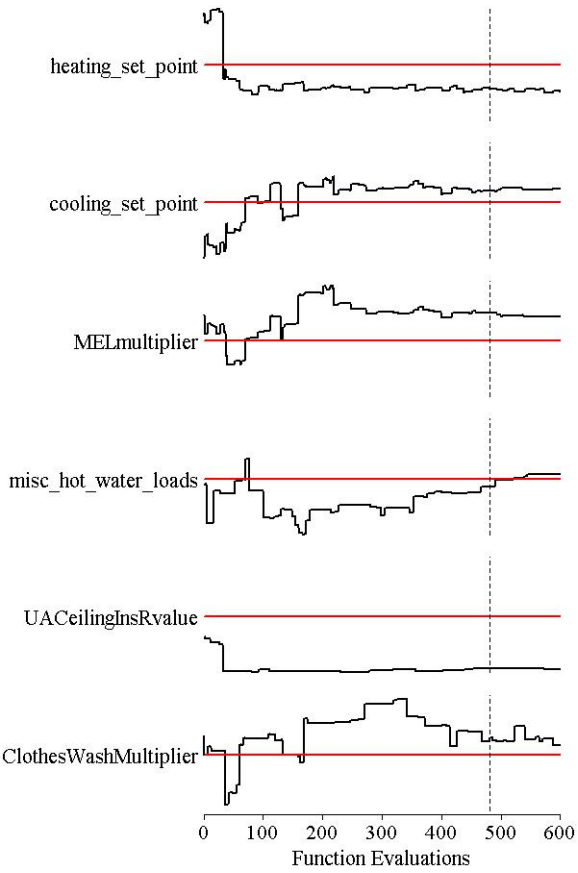


**Figure 20. Procedure 3.2 parameter convergence for calibrations to daily data, Scenario 1**

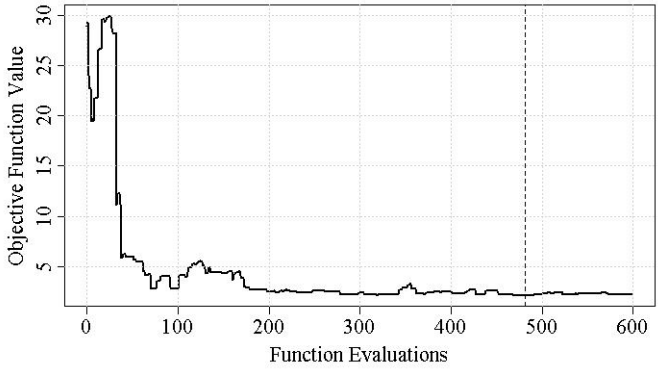


**Figure 21. Procedure 3.2 residual convergence for calibrations to daily data, Scenario 1**

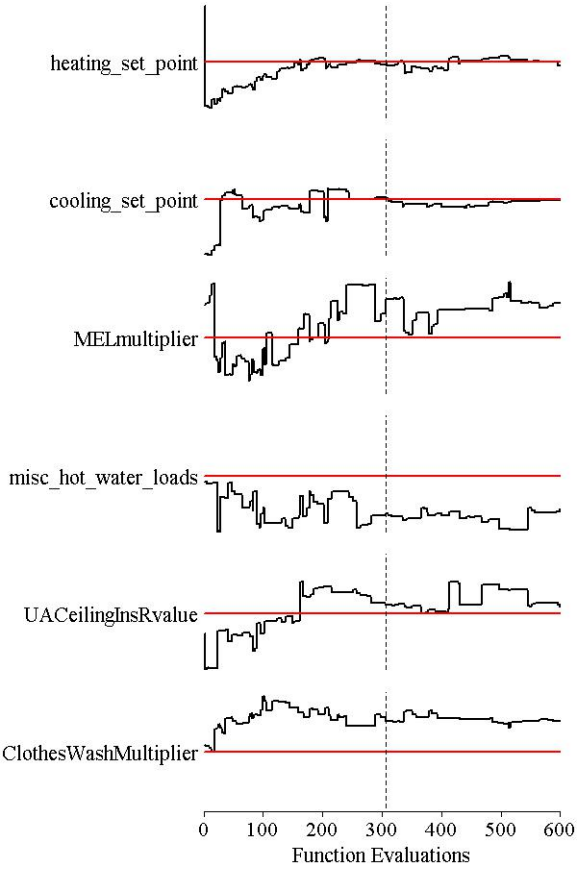


**Figure 22. Procedure 3.2 parameter convergence for calibrations to hourly data, Scenario 1**

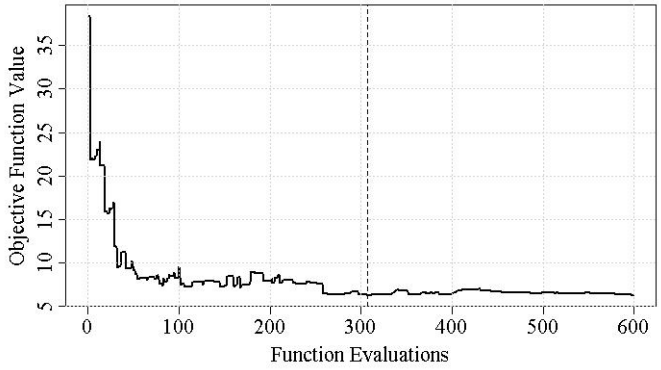


**Figure 23. Procedure 3.2 residual convergence for calibrations to hourly data, Scenario 1**

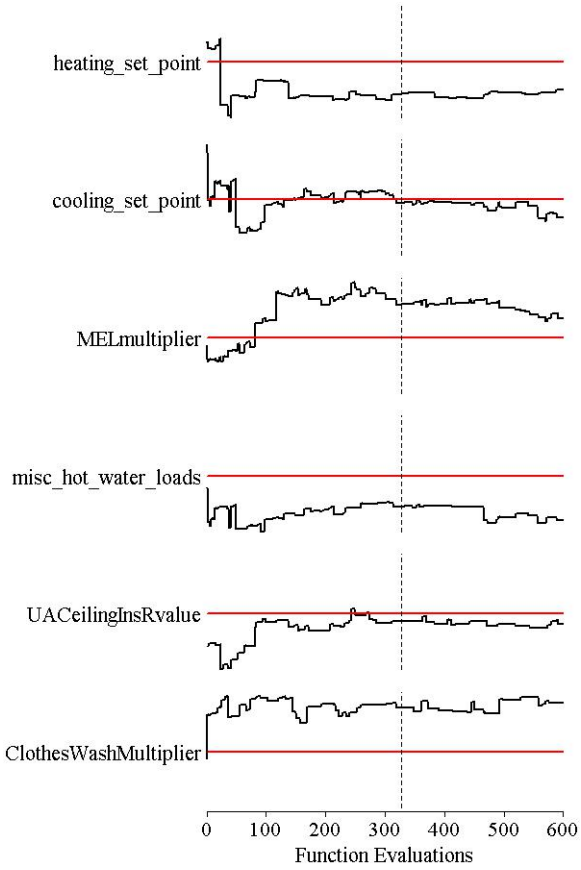


**Figure 24. Procedure 3.2 parameter convergence for calibrations to monthly data, Scenario 2**

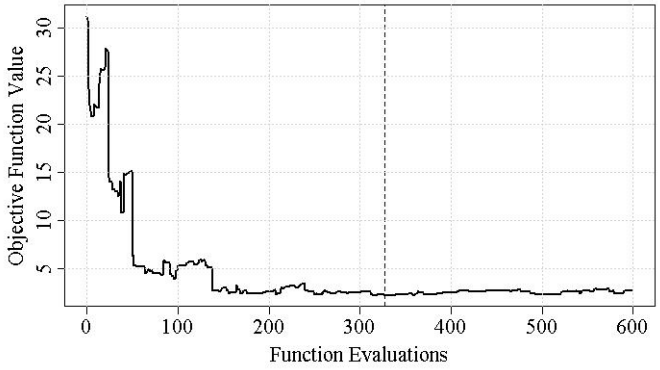


**Figure 25. Procedure 3.2 residual convergence for calibrations to monthly data, Scenario 2**

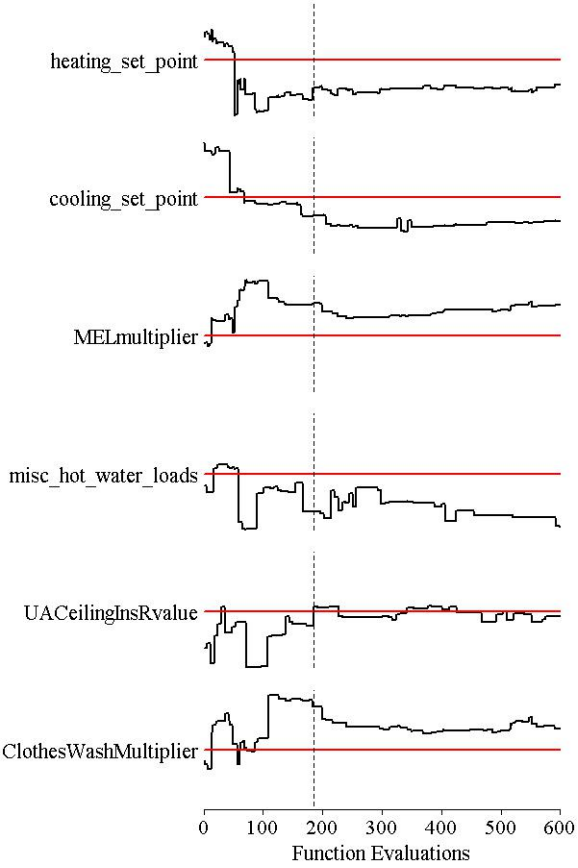


**Figure 26. Procedure 3.2 parameter convergence for calibrations to daily data, Scenario 2**

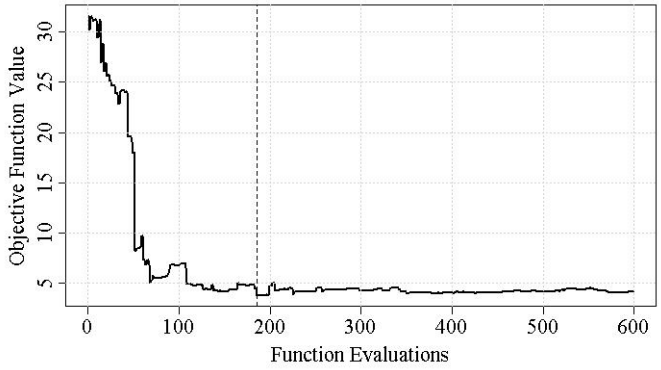


**Figure 27. Procedure 3.2 residual convergence for calibrations to daily data, Scenario 2**

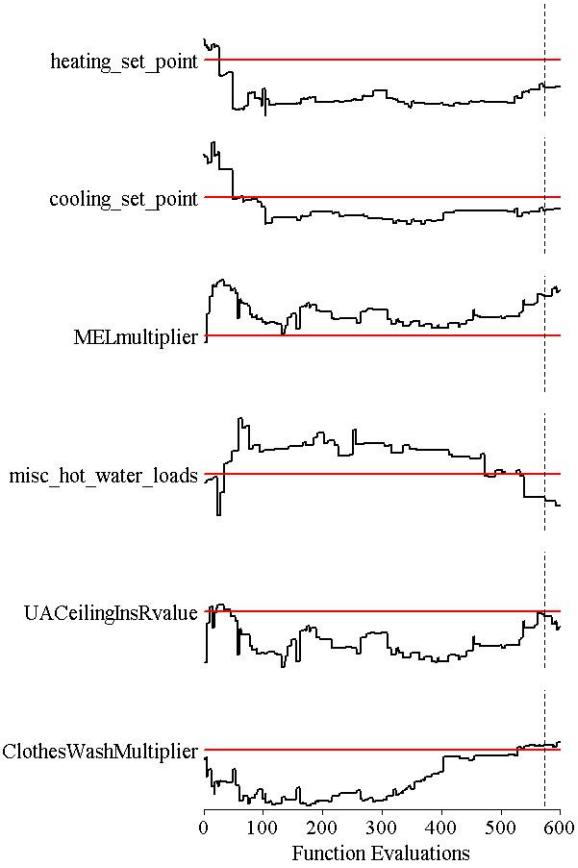


**Figure 28. Procedure 3.2 parameter convergence for calibrations to hourly data, Scenario 2**

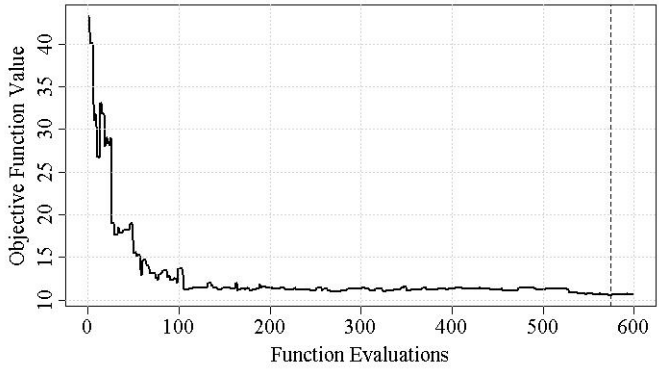


**Figure 29. Procedure 3.2 residual convergence for calibrations to hourly data, Scenario 2**

A graphical representation of output agreement is shown in Figure 73 through Figure 78 of Appendix H. The daily and hourly output was summed into monthly output so that a meaningful comparison between errors for the monthly, daily, and hourly calibrated models can be made.

Each retrofit measure described in Table 3 was applied to the calibrated model, and calibrated model energy savings predictions were calculated. Resulting BoC values were then calculated

using Eq. (4) to assess the benefit of calibration. These results are summarized in Figure 30 and Figure 31.

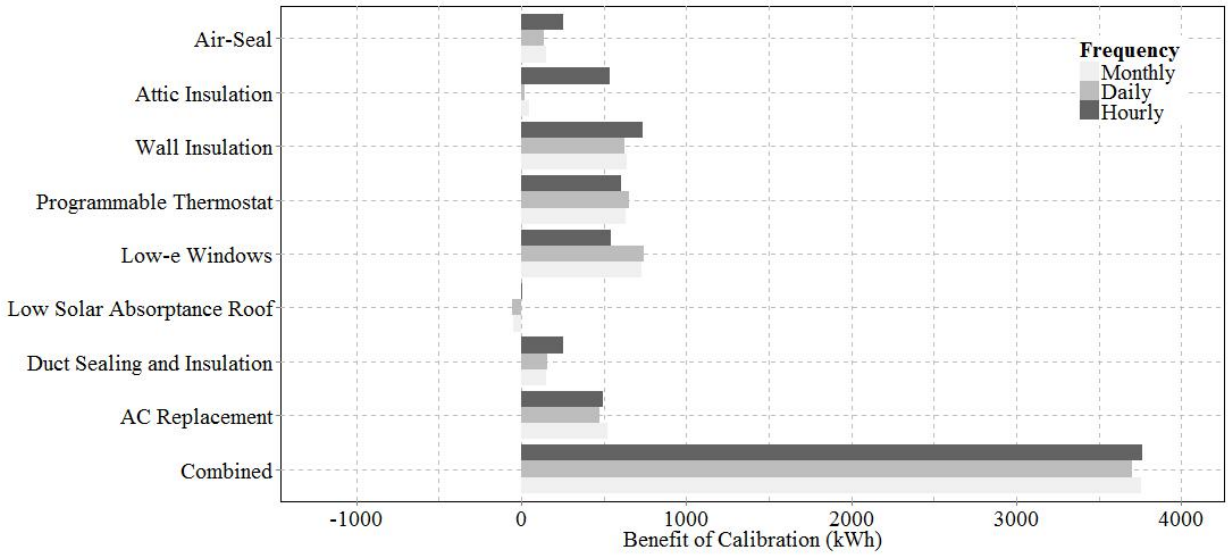


**Figure 30. Procedure 3.2 BoC, Scenario 1**

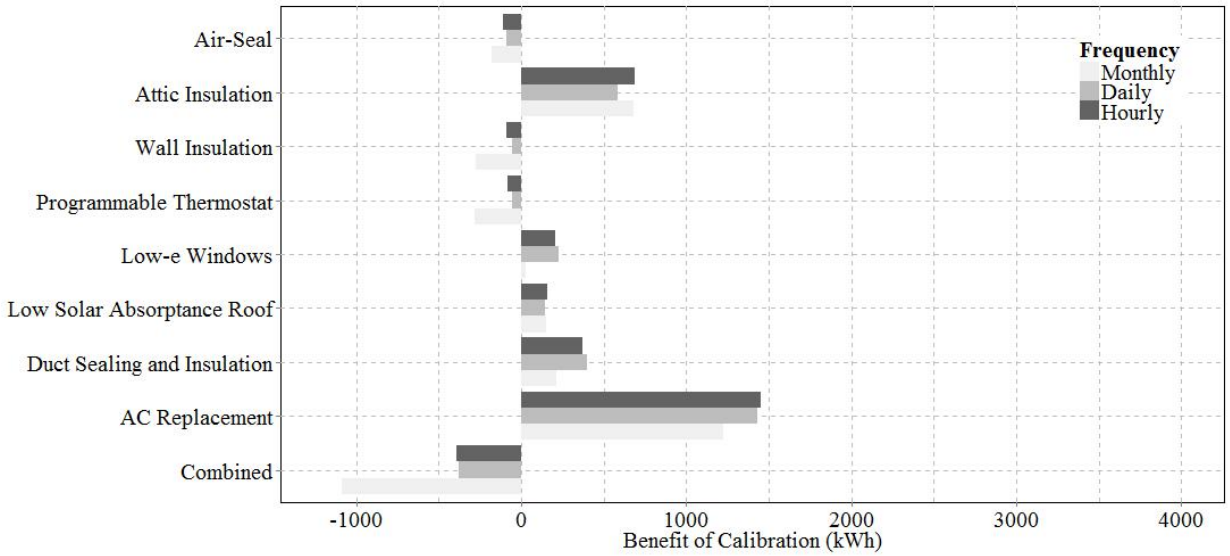


**Figure 31. Procedure 3.2 BoC, Scenario 2**

## 3.3 Regression Metamodeling Optimization Approach

This approach uses a central composite design and results from BEopt/DOE-2.2 simulations to build a statistical model. The reduced-order statistical model is then calibrated using the simulated annealing algorithm as described in Section 3.2.2. By using the simple approximation instead of the detailed model, it is easier to perform the analysis since computational effort is reduced (Manfren et al. 2013).

### 3.3.1 Sensitivity Analysis

The sensitivity analysis for this procedure is identical to that of Procedure 3.2 (see Section 3.2.1). Therefore, we considered the same parameters for adjustment as in Procedure 3.2 (see Figure 15).

### 3.3.2 Central Composite Design

This procedure uses a *response surface methodology*. We used a central composite design for the strong parameters identified in Section 3.3.1 to construct matrix $\mathbf{X}$. Central composite design matrix $\mathbf{X}$ is given in Appendix E along with a description of the response matrix $\mathbf{Y}$ and brief concept overview. We then performed a multiple linear regression by solving the system of normal equations, using the unique QR factorization[14] of $\mathbf{X}$ (Trefethen and Bau 1997), given in matrix-vector form by:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{Y}, \tag{10}$$

to obtain regression coefficient $\beta_0$, linear regression coefficients $\beta_1, \ldots, \beta_\tau$, first-order interaction regression coefficients $\beta_{12}, \ldots, \beta_{\tau-1\tau}$, and quadratic regression coefficients $\beta_{11}, \ldots, \beta_{\tau\tau}$ of second-order polynomial functions:

$$\zeta_k(x_1, \ldots, x_\tau) = \beta_0 + \sum_{i=1}^{\tau} \beta_i x_i + \sum_{i} \sum_{\substack{j \\ i<j}} \beta_{ij} x_i x_j + \sum_{i=1}^{\tau} \beta_{ii} x_i^2, \tag{11}$$

for $k = 1, \ldots, n$, where:
$n =$ number of utility data points used in the calibration (12, 365, or 8,760),
$\tau = 6$ parameters, and
$x_i =$ simulation input variables.

The resulting $\zeta_k$ provide the best fit to each of the sets of simulation data in the least squares sense, thereby attempting to predict response values (i.e., electricity consumption levels) at points other than those previously simulated according to the central composite design.

### 3.3.3 Optimization

Substituting the $\zeta_k$ for actual building simulations, we ran the simulated annealing algorithm as described in Section 3.2.2 with a starting temperature $T_0$ = 10 and the number of iterations per parameter $N$ = 100. Figure 32 and Figure 33 give the results of the simulated annealing optimization in terms of the calibrated input values.

[14] This factorization is the decomposition of $\mathbf{X}$ into the product $\mathbf{X} = \mathbf{QR}$.

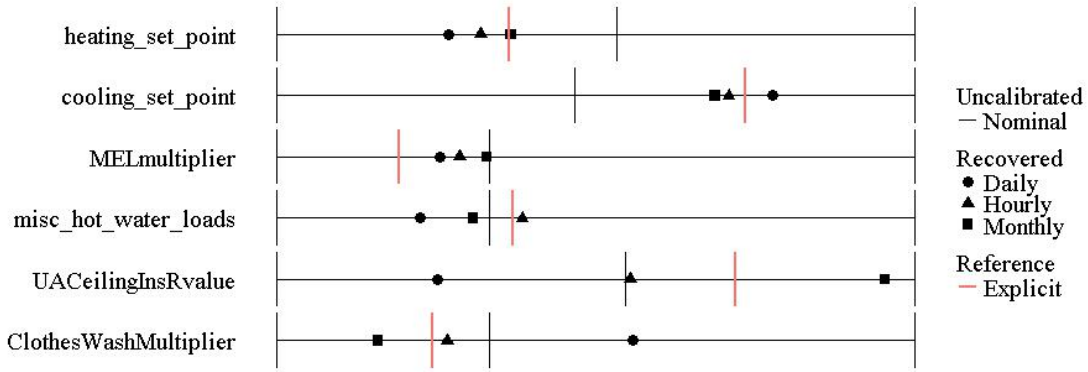


**Figure 32. Procedure 3.3 optimization results, Scenario 1**

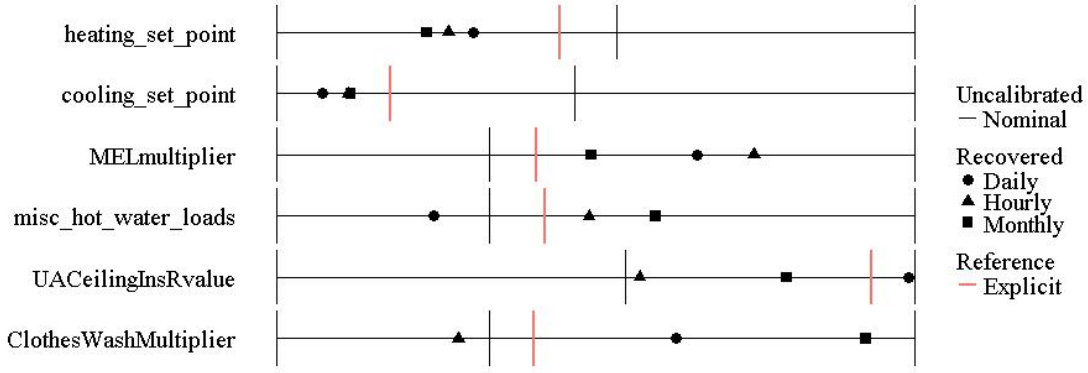


**Figure 33. Procedure 3.3 optimization results, Scenario 2**

Figure 34 through Figure 45 show the iterative search convergence processes for parameters relative to reference input values and for the objective function value (red lines are explicit reference values and dashed lines represent the number of function evaluations required to obtain the residual minimum). The convergence plots for the objective function value depict that at higher temperatures $T$ the algorithm allows more energy increases and eventually, as $T$ cools according to the schedule, convergences on a least objective function value.

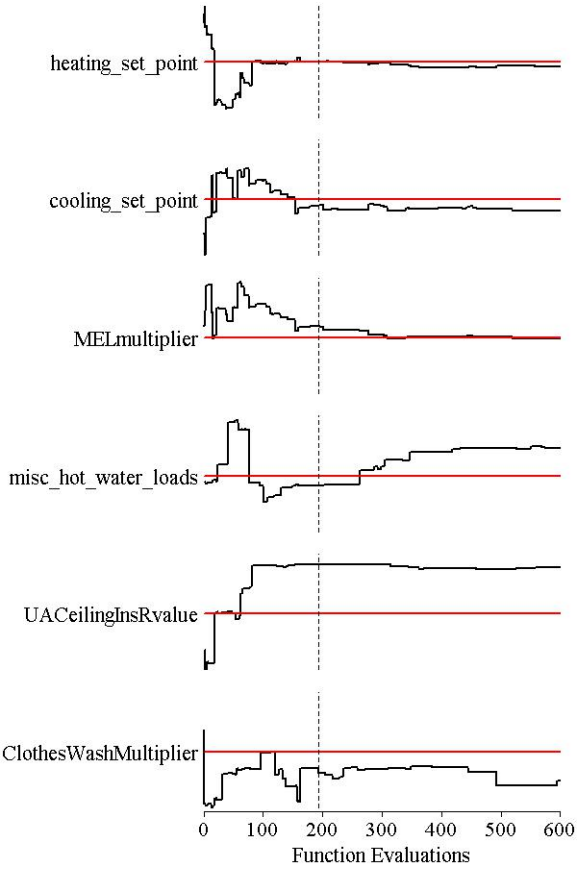


**Figure 34. Procedure 3.3 parameter convergence for calibrations to monthly data, Scenario 1**

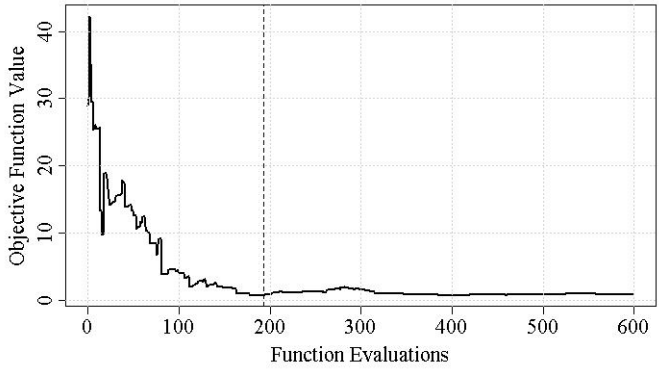


**Figure 35. Procedure 3.3 residual convergence for calibrations to monthly data, Scenario 1**

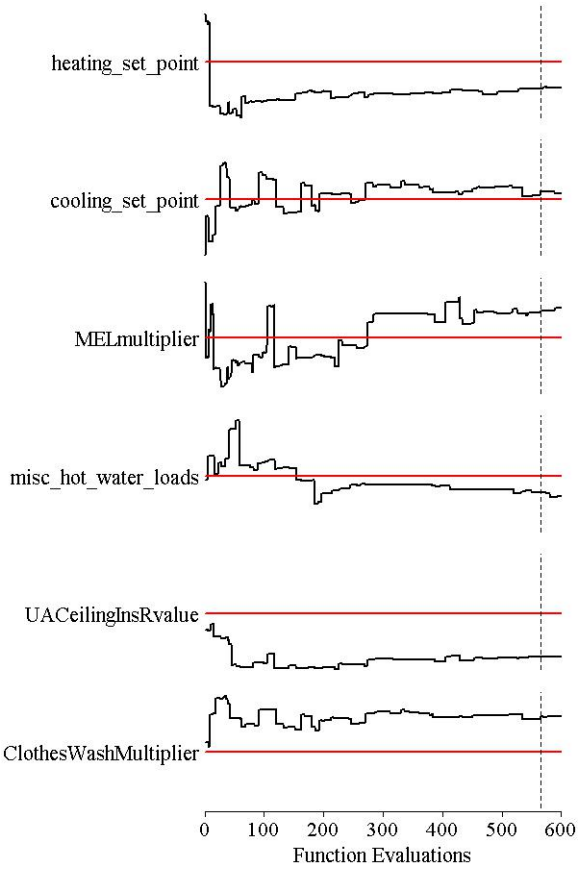


**Figure 36. Procedure 3.3 parameter convergence for calibrations to daily data, Scenario 1**

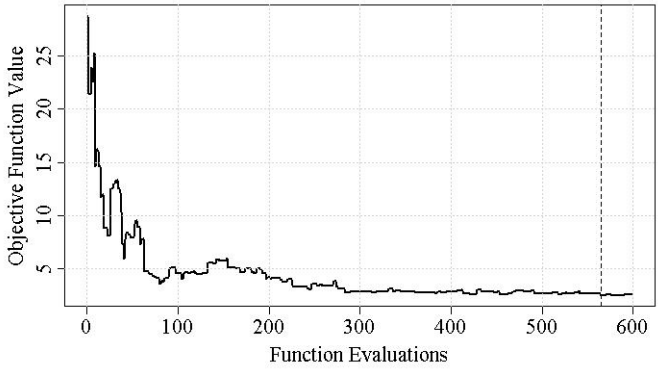


**Figure 37. Procedure 3.3 residual convergence for calibrations to daily data, Scenario 1**

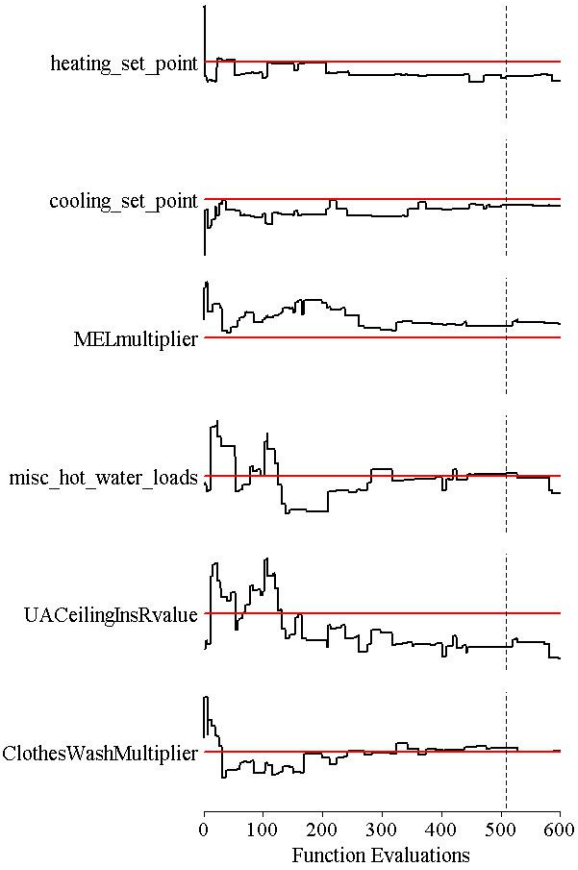


**Figure 38. Procedure 3.3 parameter convergence for calibrations to hourly data, Scenario 1**

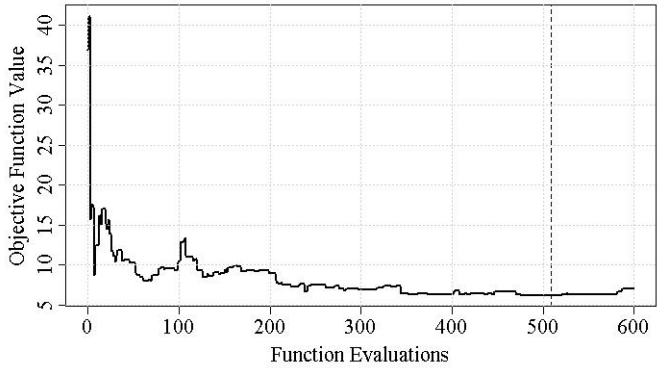


**Figure 39. Procedure 3.3 residual convergence for calibrations to hourly data, Scenario 1**

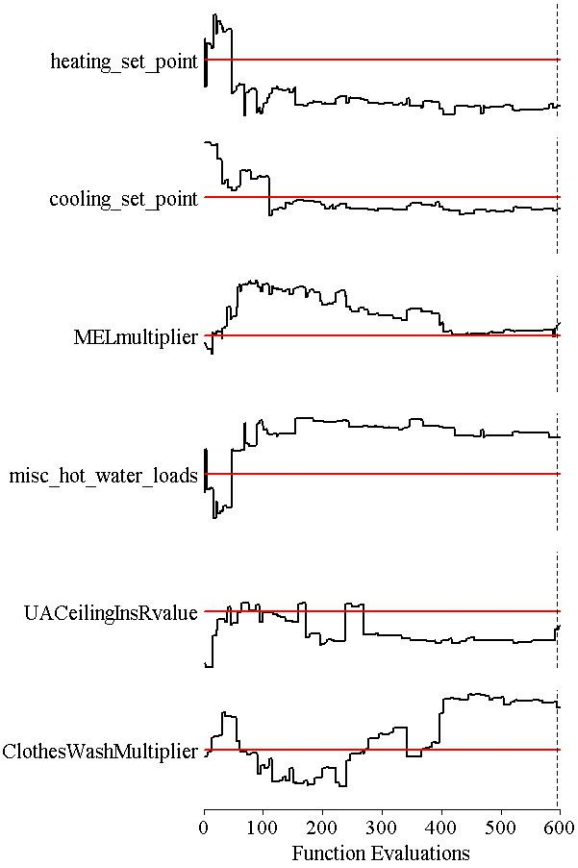


**Figure 40. Procedure 3.3 parameter convergence for calibrations to monthly data, Scenario 2**

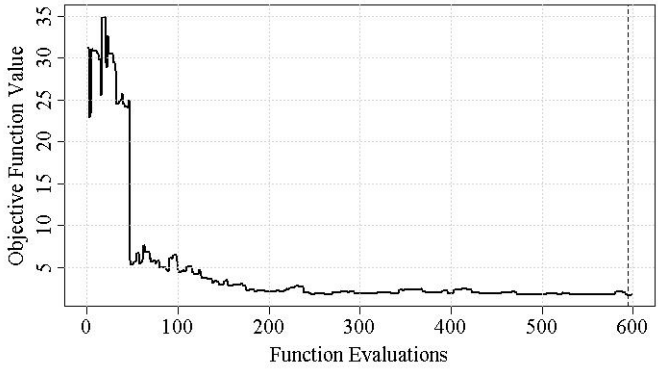


**Figure 41. Procedure 3.3 residual convergence for calibrations to monthly data, Scenario 2**

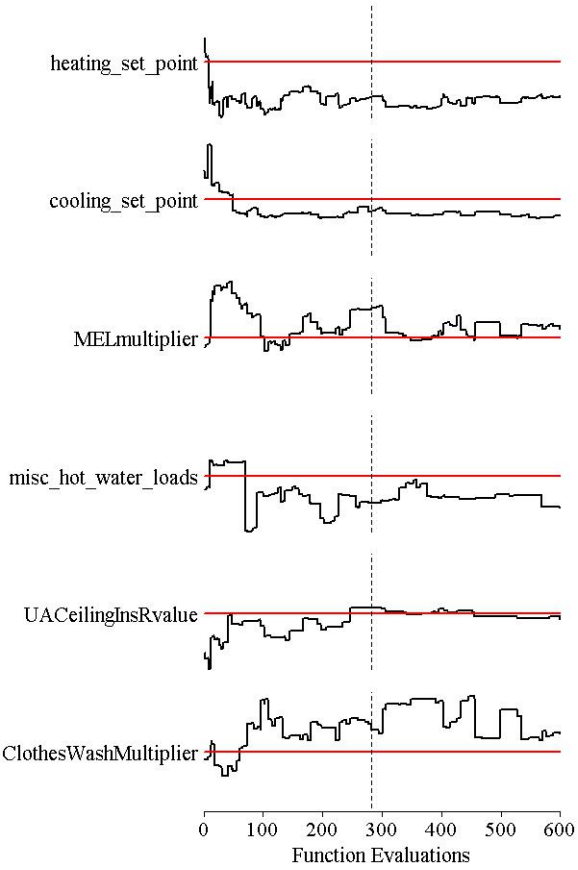


**Figure 42. Procedure 3.3 parameter convergence for calibrations to daily data, Scenario 2**

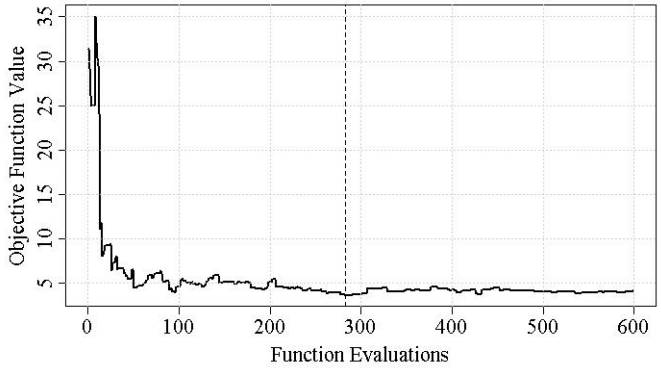


**Figure 43. Procedure 3.3 residual convergence for calibrations to daily data, Scenario 2**

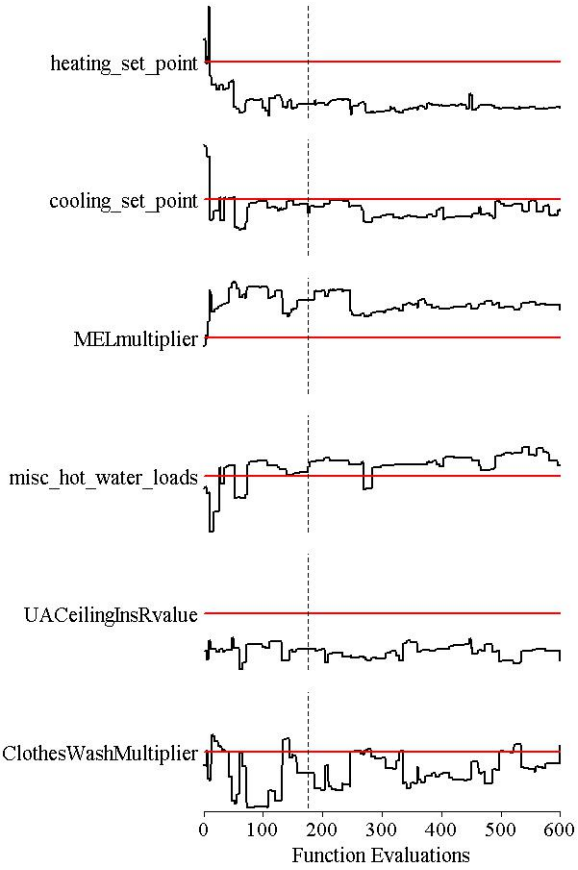


**Figure 44. Procedure 3.3 parameter convergence for calibrations to hourly data, Scenario 2**

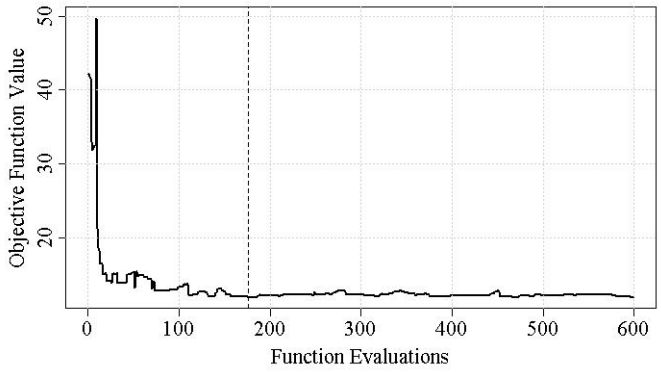


**Figure 45. Procedure 3.3 residual convergence for calibrations to hourly data, Scenario 2**

Graphical representations of output agreements are shown in Figure 79 through Figure 84 of Appendix H. Again, the daily and hourly output was summed into monthly output so that a meaningful comparison between errors for the monthly, daily, and hourly calibrated models can be made.

Each retrofit measure given in Table 3 was applied to the calibrated model, and calibrated model energy savings predictions were calculated. Resulting BoC values were then calculated using Eq. (4) to assess the BoC. These results are summarized in Figure 46 and Figure 47.

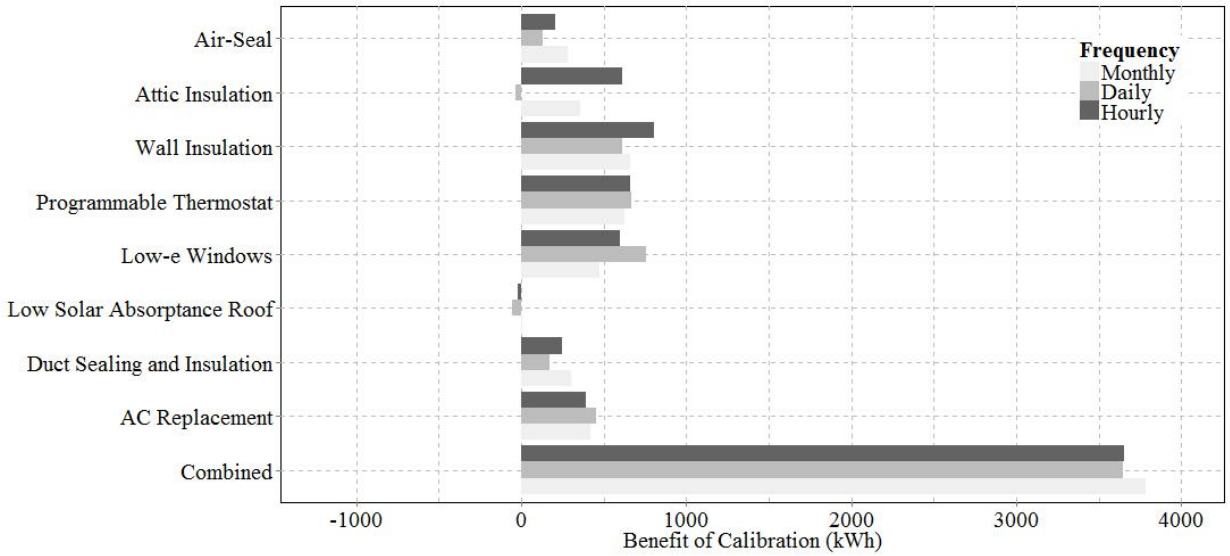


**Figure 46. Procedure 3.3 BoC, Scenario 1**

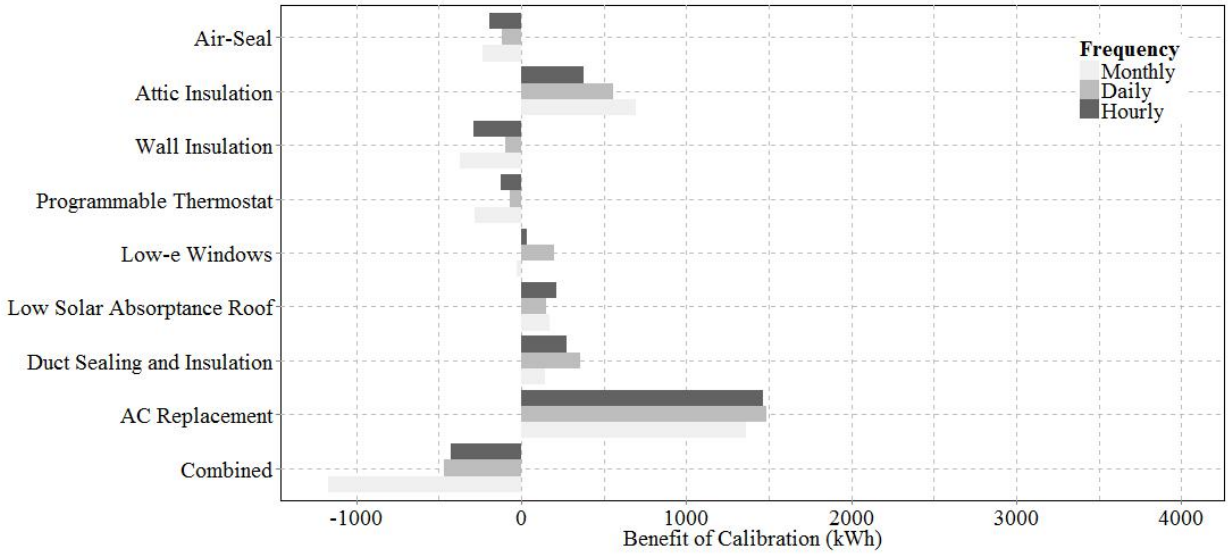


**Figure 47. Procedure 3.3 BoC, Scenario 2**

## 3.4 Simple Output Ratio Calibration Approach

Whereas Procedures 3.1–3.3 involve the adjustment of input parameters, this approach applies a simple output correction factor to uncalibrated energy savings predictions. Its results serve as a baseline of comparison for the more complex and computationally intensive input calibration methods. We denote the predicted pre-retrofit annual energy use from the uncalibrated model

$\Gamma_{uncal}$. This output calibration method simply assumes that the calibrated model energy savings $\psi^c$ is given by:

$$\psi^c = \frac{\Gamma_{ref}}{\Gamma_{uncal}} \psi^u, \tag{12}$$

where $\psi^u$ is defined as in Section 2.6 and $\Gamma_{ref}$ is defined as in Section 2.7. The benefit of calibration results are summarized in Figure 48 and Figure 49.

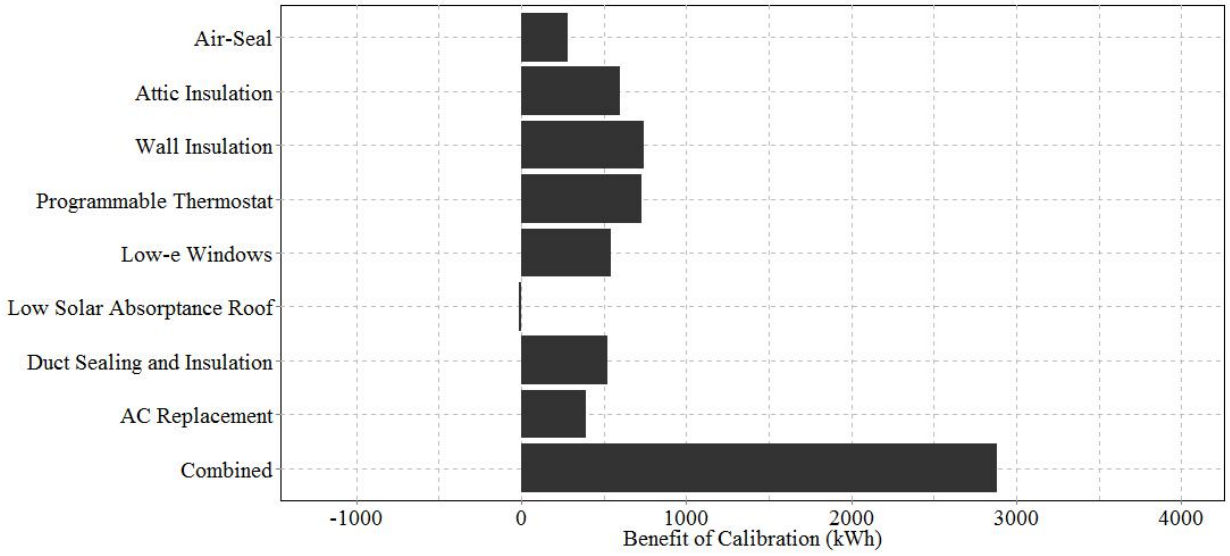


**Figure 48. Procedure 3.4 BoC, Scenario 1**

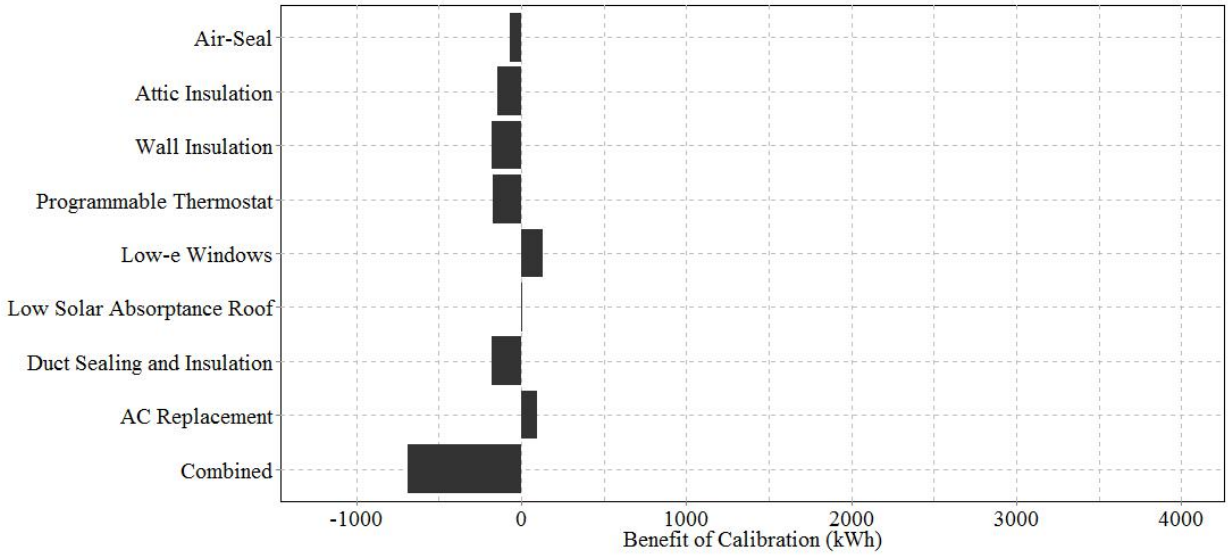


**Figure 49. Procedure 3.4 BoC, Scenario 2**

# 4 Discussion

Procedures 3.1–3.4 were evaluated based on the accuracy of predicted energy savings, computational cost, repeatability, automation, and ease of implementation. These evaluations are summarized in Sections 4.1–4.5.

## 4.1 Accuracy of Predicted Energy Savings

The calibration techniques were evaluated based on their ability to improve the accuracy of energy savings predictions for retrofit measures by tabulating and comparing the total benefit of calibration TBoC. Table 6 and Table 7 report TBoC, which is calculated as the sum of the BoC (kWh) across all retrofit measures listed in Table 3 (including the combined retrofit measure).

**Table 6. TBoC, Scenario 1**

| Technique | $TBoC$ | | |
|---|---|---|---|
| | $n = 12$ | $n = 365$ | $n = 8760$ |
| ASHRAE 1051-RP-based | 7,182.5 | 7,117.7 | 7,431.7 |
| Simplified simulated annealing | 6,580.7 | 6,457.4 | 7,184.4 |
| Regression metamodeling | 6,915.3 | 6,338.2 | 7,148.1 |
| Simple output ratio | 6,671.5 | 6,671.5 | 6,671.5 |

**Table 7. TBoC, Scenario 2**

| Technique | $TBoC$ | | |
|---|---|---|---|
| | $n = 12$ | $n = 365$ | $n = 8760$ |
| ASHRAE 1051-RP-based | –257.5 | 2,322.9 | 2,426.7 |
| Simplified simulated annealing | 467.0 | 2,198.1 | 2,186.0 |
| Regression metamodeling | 277.4 | 1,990.6 | 1,329.6 |
| Simple output ratio | –1,206.1 | –1,206.1 | –1,206.1 |

Comparisons of each calibrated savings prediction to the uncalibrated savings prediction and to the reference savings are also given; Eq. (5) and (6) were used to produce graphical representations of energy savings prediction accuracy (see Figure 50 and Figure 51) for monthly, daily, and hourly data calibration cases. For "ASHRAE 1051-RP-based," the value represents the mean savings prediction for the 10 calibration solutions.

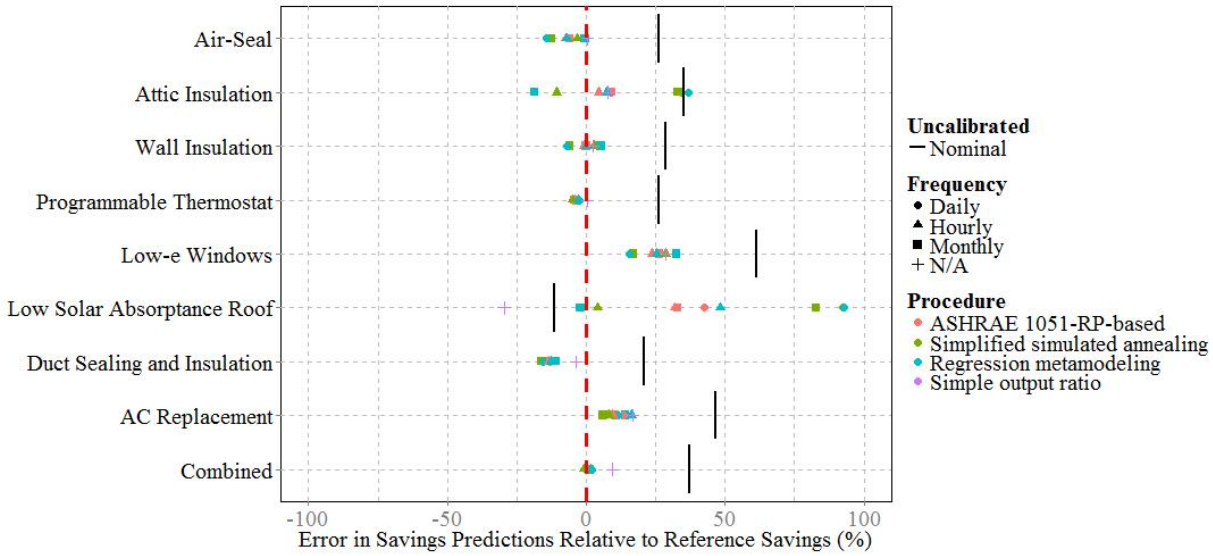


**Figure 50. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 1**

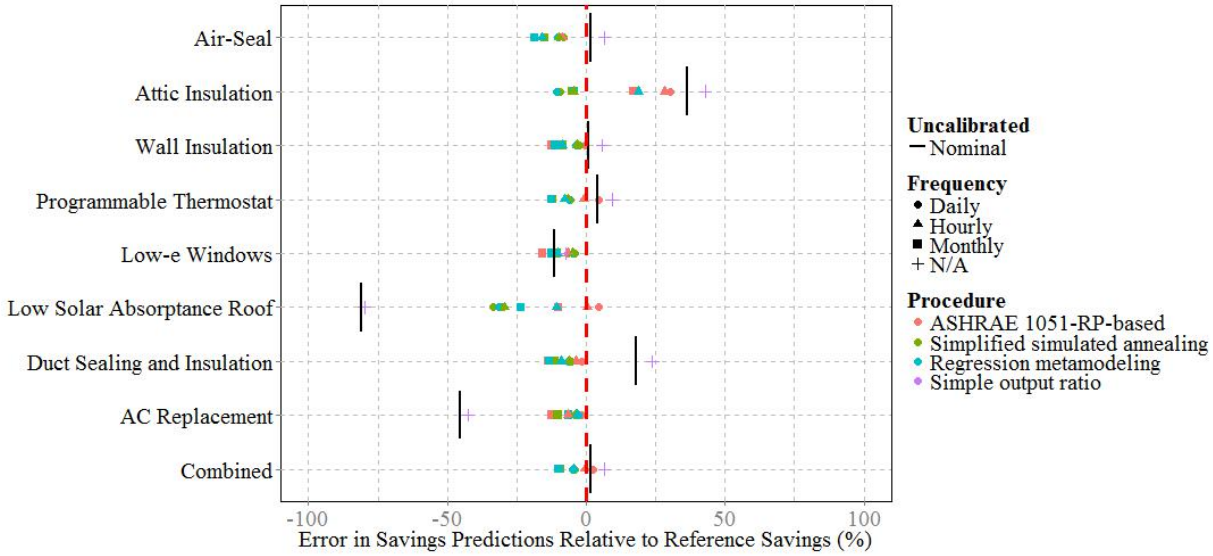


**Figure 51. Predicted annual percent energy savings error relative to reference savings for each retrofit measure, Scenario 2**

Based on the results summarized in Table 6 and Table 7 and in Figure 50 and Figure 51, the following observations can be made:

- With the exception of Procedure 3.4 for Scenario 2, calibration generally improved the accuracy of energy savings predictions relative to reference energy savings.
- Procedures 3.1–3.3 (input calibration approaches) performed considerably better than Procedure 3.4 (output calibration approach) for Scenario 2.

- For Scenario 1, Procedure 3.1 provided the highest TBoC of any calibration procedure for all utility data frequency cases; for Scenario 2, this was true for the daily and hourly cases, but not for the monthly case.
- For Scenario 1, Procedures 3.1–3.3 produced higher TBoC for hourly calibrations than for daily calibrations; surprisingly for Scenario 2, this was *not* always the case.
- For Scenario 2, Procedures 3.1–3.3 produced higher TBoC for daily calibrations than for monthly calibrations; surprisingly for Scenario 1, this was *not* the case.
- Calibration provided more benefit for Scenario 1 than for Scenario 2; for Scenario 1, the uncalibrated model overpredicts reference utility data (e.g., see energy use data in Figure 73), whereas for Scenario 2, the uncalibrated model overpredicts heating energy but underpredicts cooling energy (e.g., see energy use data in Figure 76).
- For the Combined retrofit measure in Scenario 1, the approximately 40% overprediction with the uncalibrated model was eliminated by applying Procedures 3.1–3.4.
- The tolerance intervals for Procedure 3.1 did not always contain the reference savings values (see Figure 13 and Figure 14). For both scenarios, the intervals for monthly calibrations contained the reference savings less often than for daily and hourly calibrations.
- For Scenario 1, tolerance intervals for Procedure 3.1 for monthly calibrations were narrower than for daily and hourly calibrations; for Scenario 2, tolerance intervals for monthly calibrations were narrower than for daily calibrations, and tolerance intervals for hourly calibrations were narrower than for monthly calibrations (see Figure 13 and Figure 14).
- For Scenario 2 the total benefit of calibration TBoC is small compared to Scenario 1, but calibration removed large discrepancies for individual measures (e.g., attic insulation, low solar absorptance roof, AC replacement).

## 4.2 Computational Cost

Table 8 gives a breakdown of computational costs by calibration procedure component; the table also shows calibration procedure equivalent computational costs when parallel computing resources are maximized.

**Table 8. Computational Cost**

| Proc. | Component | Total Simulations[1] | Minimum Series[2] | Notes |
|---|---|---|---|---|
| 3.1 | Sensitivity analysis | 2,500[a] | 1 | |
| | Central composite design | N/A | N/A | |
| | Gradient calculations | 350 | 1 | |
| | Simulated annealing | 6,010[b] | 601 | Each optimization may be parallelized. |
| | Retrofit measures | 90 | 1 | Measures applied to each model. |
| | **Total** | **8,950** | **604** | |
| 3.2 | Sensitivity analysis | 2,400 | 1 | 100 samples for each of the 24 inputs. |
| | Central composite design | N/A | N/A | |
| | Gradient calculations | 350 | 1 | |
| | Simulated annealing | 601[b] | 601 | |
| | Retrofit measures | 9 | 1 | |
| | **Total** | **3,360** | **604** | |
| 3.3 | Sensitivity analysis | 2,400 | 1 | 100 samples for each of the 24 inputs. |
| | Central composite design | 77 | 1 | $2^6 + (2)(6) + 1 = 77$. |
| | Gradient calculations | 0 | 0 | |
| | Simulated annealing | 0 | 0 | |
| | Retrofit measures | 9 | 1 | |
| | **Total** | **2,486** | **3** | |
| 3.4 | Sensitivity analysis | N/A | N/A | |
| | Central composite design | N/A | N/A | |
| | Gradient calculations | N/A | N/A | |
| | Simulated annealing | N/A | N/A | |
| | Retrofit measures | 9 | 1 | |
| | **Total** | **9** | **1** | |

[1] Total number of BEopt/DOE-2.2 simulations.
[2] The equivalent computational time in terms of BEopt/DOE-2.2 simulations when parallel computing is maximized (note: this does not include the computational time that may be needed to initiate batch runs and process results).
[a] As described in ASHRAE 1051-RP, more realizations may be needed when calibrating to empirical billing data (see Section 5.2 on future work for further discussion).
[b] For this study, the total simulations performed during simulated annealing was predetermined; the number of simulations may be reduced (perhaps using a goodness-of-fit threshold value), but this may reduce method accuracy in some cases.

Based on the information provided in Table 8, and considering the results summarized in Section 4.1, the following observations can be made:

- Generally, the more computationally expensive calibration procedures produced more accurate energy savings predictions for retrofit measures (see TBoC values in Table 6 and Table 7, and savings prediction errors in Figure 50 and Figure 51).
- Simulations for the sensitivity analysis approaches for Procedures 3.1–3.3 may all be run in parallel.

- If parallelized, Procedures 3.1–3.2 have equivalent overall procedure runtime.
- Simulations within each component of Procedures 3.3 and 3.4 may be parallelized; simulations for Procedure 3.3 can be parallelized such that the calibration procedure includes the equivalent of three simulations in series.

## 4.3 Automation

Procedures 3.1–3.3 require that the analyst begin the calibration process by selecting the set of 24 influential model inputs. Sensitivity analysis and optimization portions of these calibration procedures are then fully automated from start to finish, requiring no further manual adjustments or intervention. Procedure 3.4 is completely automated once the uncalibrated model and corresponding savings predictions are developed.

## 4.4 Repeatability

Procedures 3.1–3.4 are much more repeatable than manual calibrations since they are automated and require limited assessor decision-making and input. Specifically, Procedure 3.4 is deterministic once an uncalibrated model and corresponding savings predictions are developed, so results are fully repeatable. On the other hand, Procedures 3.1–3.3 each contain some subjective elements in their methodologies:

- The analyst's initial selection of the 24 most influential model inputs may vary.
  - Repeatability across calibrations may not be guaranteed if this initial set differs.
- Randomly selected values are determined by pseudo-random number generators.[15]
  - The LHMC sensitivity analysis in Procedure 3.1 simulates a random subset of $3^{24}$ total factor combinations; therefore, $\chi^2$ values may fluctuate slightly if the overall process is repeated.
  - The sensitivity analysis method used in Procedures 3.2 and 3.3 include random selections from triangular distributions; sensitivity rankings from analysis to analysis become more stable as the number of samples increases.
  - The simulated annealing algorithm uses random selections to calculate gradients and also to determine perturbation vector scaling (i.e., there are random elements to the search path).
- The number of adjustable input parameters considered for optimization may vary.
  - We fixed the number of adjustable inputs during calibration for this study to allow easier comparison across calibration methods. This allows the analyst to decide how many adjustable inputs to include; however, it may not guarantee repeatability of results, because agreement between predicted and utility billing data may vary.

## 4.5 Ease of Implementation

This section provides a brief summary of the scripting processes for implementing Procedures 3.1–3.4. Procedures 3.1–3.3 require adjustments to nominal input values; Procedure 3.4 requires calculating a correction factor, minimal text file input/output, one nominal building simulation,

[15] Randomly selected values may be stored for later use so that procedure results may be reproduced and verified.

one simulation per retrofit measure considered, and only few additional arithmetic calculations only. Procedures 3.1–3.3 are more involved as they incorporate more text file input/output, algorithms for discretizations and probability distribution selections, statistical procedures, inversion algorithms, etc. Specifically, the following computer software programs and scripting languages were used to implement these procedures:

- BEopt/DOE-2.2:
    - Developing nominal pre- and post-retrofit input files (XML format).
    - Simulating input files in batch mode using the BEopt Run Manager to produce simulation output files.
- Custom Python 2.7 (Rossum and Drake 2012) scripts for:
    - Conducting Monte Carlo-based sensitivity analyses.
    - Implementing the simulated annealing algorithm.
    - Replacing XML input file element values.
    - Parsing and composing simulation output data.
    - Managing files of various types.
    - Performing arithmetic calculations.
    - Executing external commands (such as calling the BEopt Run Manager).
- Custom R 2.12.2 (R 2012) scripts for:
    - Plotting and graphing.
    - Statistical procedures and intervals.
    - Arithmetic calculations.
    - Experimental designs.
    - Multiple linear regression (i.e., solving normal equations using QR factorization).

# 5 Conclusions and Future Work

Section 5.1 provides conclusions drawn from results reported in Section 3 and observations made in Section 4. Section 5.2 provides ideas for further developing automated residential calibration techniques.

## 5.1 Conclusions

For this study, four mathematical calibration methods were implemented using BEopt/DOE-2.2 and tested using monthly, daily, and hourly synthetic utility data, for scenarios in which the uncalibrated model overpredicts (Scenario 1) and underpredicts (Scenario 2) the reference utility data. The four methods implemented were an ASHRAE 1051-RP-based approach (Procedure 3.1), a simplified simulated annealing optimization approach (Procedure 3.2), a regression metamodeling optimization approach (Procedure 3.3), and a simple output ratio approach (Procedure 3.4). Various retrofit measures were applied to the models obtained using the calibration methods and the methods were evaluated based on the accuracy of predicted savings, computational cost, repeatability, automation, and ease of implementation. This study led to the following primary conclusions:

- Generally, the more computationally expensive calibration procedures led to more accurate energy savings predictions.
- Procedure 3.1 used multiple calibration solutions to estimate ranges for predicted savings. These ranges did not always contain the reference savings value, especially for calibrations to monthly data. Reducing the number of calibration solutions investigated may streamline the approach without much loss of accuracy.
- Scenario 1 had consistent overprediction (i.e., uncalibrated model overpredicted reference billing data the entire season), whereas Scenario 2 had compensating errors (i.e., uncalibrated model overpredicted heating energy but underpredicted cooling energy); compensating errors created a more difficult calibration problem, as the procedures generally performed better for Scenario 1 than for Scenario 2.
- Generally, calibrations to higher frequency data produced more accurate energy savings predictions, and the improvement for daily and hourly calibrations over monthly calibrations was greatest in the presence of compensating errors (for example, the average $TBoC$ values across Procedures 3.1–3.3 for Scenario 2 were 1,981 kWh [hourly], 2,171 kWh [daily], and 162.3 kWh [monthly]).
- Procedure 3.3 performed similarly to the more computationally expensive Procedures 3.1–3.2 for Scenario 1, but did not for Scenario 2; Procedure 3.3 could be drastically streamlined using parallel computing, but more scenarios are needed to evaluate and understand its performance across a range of conditions.
- Across sensitivity analysis methods used in this study, a subset of strong inputs (i.e., "heating_set_point," "cooling_set_point") had considerably higher sensitivity analysis coefficients than the other strong inputs, and calibrated values for this subset generally agreed much better with reference values than the other adjusted values.

Overall, the results suggest that:

1. The optimization problem is still significantly underdetermined when calibrating to monthly, daily, and hourly data using the approaches investigated in this study, which are largely based on previous studies performed mostly in the context of commercial buildings.
2. Additional research is needed to develop improved or alternate approaches that take full advantage of the additional information content contained in the high-frequency residential billing data.

In the nearer term, calibration methods similar to those described in this study could be implemented in residential simulation tools and tested in the field for automated calibrations to monthly billing data. They could be implemented in the context of emerging industry standards for residential model calibration (such as Building Performance Institute Standard 2400 [BPI 2011]). Software developers who have the capability to run batch simulations in parallel (e.g., through cloud-computing) could employ methods similar to the regression metamodeling optimization approach to reduce the time required for automated calibration.

## 5.2 Future Work

This study was an initial step in investigating four automated residential calibration techniques over monthly, daily, and hourly synthetic utility data. Further research is needed to evaluate residential calibration techniques in the following areas:

- **More scenarios and house types:** Two calibration scenarios were investigated in this study: scenarios in which the uncalibrated model overpredicts and underpredicts the utility billing data for a 1960s-era ranch-style home. More scenarios (e.g., the uncalibrated model underpredicts heating energy but overpredicts cooling energy) and house types (e.g., different vintages and energy efficiency features) should be investigated in future studies as they may provide more information about the strengths and limitations of the calibration techniques (i.e., whether positive BoC can be expected for other scenarios or house types).
- **Calibration to submetered/disaggregated smart-meter utility data:** Researchers are developing techniques for disaggregating smart-meter data by device (Kolter and Jaakkola 2012). Future work is needed to develop and evaluate residential calibration methods in the context of disaggregated utility data.
- **Calibration to mixed-fuel utility data:** This study considered an all-electric home, which presented more difficulties than mixed-fuel calibrations in that there was no disaggregation by fuel type, but fewer difficulties in that all energy consumption was measured at higher frequency in the daily and hourly cases (for mixed fuel, gas data are typically available at a monthly frequency). Future studies should develop and evaluate calibration methods in the context of mixed-fuel scenarios.
- **Alternative sensitivity analyses:** The sensitivity analysis used in this study to identify the adjustable inputs considers only annual energy totals and not seasonal energy totals. Future studies should consider designating more adjustable inputs related to heating and cooling as opposed to baseload consumption. Adjusting inputs dominated by baseload

inputs may encourage matching utility billing data for the wrong reasons and has negligible impact on retrofits that are not baseload related.

- **Alternative objective functions:** For this study, we used Eq. (18) as our optimization objective function. This objective function characterizes the variability of the errors between model-predicted and measured utility data (ASHRAE 2002). Alternative objective functions should be investigated (e.g., a *penalty approach* in which statistically unlikely calibration solutions are penalized [Carroll and Hitchcock 1993]), as they impact how well optimizations minimize residuals, avoid local minima, and recover better estimates of reference values.
- **Validation using high-quality empirical data:** After calibration methods are further evaluated and refined using synthetic utility data, they should be tested using high-quality, empirical audit and utility data. For example, it will be important to investigate whether 2,500 LHMC realizations are sufficient for Procedure 3.1.
- **Schedules:** Similar to BESTEST-EX, this study used multipliers to vary some schedules in magnitude. Future studies could use probabilistic techniques to generate unique hourly schedules for lighting, occupancy, etc.

# Glossary

| | |
|---|---|
| 2-level full factorial design | An experimental design taking on all possible 2-level combinations of factors included. |
| Adjustable parameters | Model inputs whose values are allowed to change during the calibration process. |
| Approximate inputs | Building model inputs with assigned triangular probability distributions. |
| Benefit of calibration | Difference in absolute error of savings predictions between calibrated and uncalibrated models. |
| Calibrated model | The model whose inputs have been adjusted through calibration. |
| Explicit input value | Value randomly selected from the triangular probability distribution of approximate inputs; used in the reference model. |
| Feasible calibration solutions | A term used in ASHRAE 1051-RP denoting calibration solutions satisfying ASHRAE Guideline 14 criteria. |
| Heuristic template | Database consisting of 20–24 influential input parameters with assigned best-guess, minimum, and maximum values. |
| Influential inputs | Those 24 inputs identified heuristically by an expert in a walk-through audit. |
| Reference model | Model consisting of randomly selected explicit input values. |
| Reference utility data | Utility data generated from BEopt/DOE-2.2 output from the reference model. |
| Response surface methodology | The study of relationships between several explanatory variables and one or more response variables. |
| SLA (specific leakage area) | Ratio of the effective leakage area of the living space to the floor area of the space. |
| Strong parameters | Top influential parameters identified from each procedure's sensitivity analysis. |
| Triangular probability distribution | Probability distribution type assigned to approximate inputs; characterized by minimum, nominal, and maximum values. |
| Uncalibrated model | The model consisting of "best-guess" or nominal input values. |
| Weak parameters | Influential parameters not identified as strong parameters. |

# References


ASHRAE. (2002). *ASHRAE Guideline 14-2002: Measurement of Energy and Demand Savings*. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Barton, R.R. (2009). "Simulation Optimization: Using Metamodels." *Proceedings of the 2009 Winter Simulation Conference*; pp. 230–238.

Batmaz, I.; Tunali, S. (2002). "Second-Order Experimental Designs for Simulation Metamodeling." *Simulation* (78:12); pp. 699–715.

Bertagnolio, S.; Lemort, V.; Andre, P. (2010). *Simulation Assisted Audit & Evidence Based Calibration Methodology*. International Energy Agency Energy Conservation in Buildings and Community Systems Annex 53: Total Energy Use in Buildings – Analysis and evaluation methods.

BPI. (2011). *BPI-2400-S-2011: Standardization Qualification of Whole House Energy Savings Estimates*. Building Performance Institute, Inc.

Carroll, W.L.; Hitchcock, R.J. (1993). "Tuning Simulated Building Description To Match Actual Utility Data: Methods and Implementation." *ASHRAE Transactions* (99:2); pp. 928–934.

Christensen, C.; Anderson, R.; Horowitz, S.; Courtney, A.; Spencer, J. (2006). *BEopt Software for Building Energy Optimization: Features and Capabilities.* Golden, CO: National Renewable Energy Laboratory. NREL/TP-550-39929.

Collins, M.D.; Kuperman, W.A.; Schmidt, H. (1992). "Nonlinear Inversion for Ocean-Bottom Properties." *Journal of Acoustical Society of America* (92:5); pp. 2770–2783.

Hamby, D.M. (1994). "A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models." *Environmental Monitoring and Assessment* 32; pp. 135–154.

Judkoff, R.; Polly, B.; Bianchi, M.; Neymark, J. (2011). *Building Energy Simulation Test for Existing Homes (BESTEST-EX): Instructions for Implementing the Test Procedure, Calibration Test Reference Results, and Example Acceptance-Range Criteria*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5500-52414.

Judkoff, R.; Polly, B.; Bianchi, M.; Neymark, J. (2010). *Building Energy Simulation Test for Existing Homes (BESTEST-EX); Phase 1 Test Procedures: Building Thermal Fabric Cases.* Golden, CO: National Renewable Energy Laboratory. NREL/TP-550-47427.

Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. (1983). "Optimization by Simulated Annealing." *Science, New Series* (220:4598); pp. 671–680.

Kissock, J.K.; Haberl, J.S.; Claridge, D.E. (2003). "Inverse Modeling Toolkit (1050RP): Numerical Algorithms for Best-Fit Variable-Base Degree-Day and Change-Point Models." *ASHRAE Transactions-Research* (109:2); pp. 425–434.

Kolter, J.Z.; Jaakkola, T. (2012). "Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation." *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*.

Kotz, S.; van Dorp, J.R. (2004). *Beyond Beta, Other Continuous Families of Distributions with Bounded Support and Applications.* Chapter 1, "The Triangular Distribution." Singapore: World Scientific Publishing Co.

Kuperman, W.A.; Collins, M.D.; Perkins, J.S.; Davis, N.R. (1990). "Optimal Time-Domain Beamforming With Simulated Annealing Including Application of *A Priori* Information. *Journal of Acoustical Society of America* (88:4); pp. 1802–1810.

Lee, S.U.; Claridge, D.E. (2002). "Automatic Calibration of a Building Energy Simulation Model Using a Global Optimization Program." *Proceedings of the Second International Conference for Enhanced Building Operations*.

"National Solar Radiation Data Base." (2008). National Renewable Energy Laboratory. Accessed February 27, 2012: http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/tmy3/.

Manfren, M.; Aste, N.; Moshksar, R. (2013). "Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation." *Applied Energy* 103; pp. 627–641.

New, J.; Sanyal, J.; Bhandari, M.; Shrestha, S. (2012). *Autotune E+ Building Energy Models.* International Building Performance Simulation Association.

Norford, L.K.; Socolow, R.H.; Hsieh, E.S.; Spadaro, G.V. (1994). "Two-to-One Discrepancy Between Measured and Predicted Performance of a 'Low-Energy' Office Building: Insights From a Reconciliation Based on the DOE-2 Model." *Energy and Buildings* 21; pp. 121–131.

Pappas, A.; Reilly, S. (2011). "Streamlining Energy Simulation To Identify Building Retrofits." *ASHRAE Journal* 53; pp. 32–42.

Pedrini, A.; Westphal, F.S.; Lamberts, R. (2002). "A methodology for building energy modeling and calibration in warm climates." *Building and Environment* 37; pp. 903–912.

"Treat: high-performance software for building energy analysis." (2010). Performance Systems Development. Accessed February 27, 2012: www.psdconsulting.com/software/treat.

R. (2012). *R: A Language and Environment for Statistical Computing.* Foundation for Statistical Computing.

Raftery, P.; Keane, M.; Costa, A. (2009). "Calibration of a Detailed Simulation Model to Energy Monitoring System Data: A Methodology and Case Study." *Eleventh International IBPSA Conference*.

Raftery, P.; Keane, M.; O'Donnell, J. (2011). "Calibrating Whole Building Energy Models: An Evidence-Based Methodology." *Energy and Buildings* 43; pp. 2356–2364.

Reddy, T.A.; Maor, I. (2006). *Procedures for Reconciling Computer-Calculated Results With Measured Energy Data.* Philadelphia, PA: ASHRAE Research Project 1051-RP.

Rossum, G.V.; Drake, Jr., F.L. (2012). *The Python Language Reference.* Python Software Foundation.

Subbarao, K.; Burch J.D.; Hancock, C.E.; Lekov, A.; Balcomb, J.D. (1988). *Short-Term Energy Monitoring (STEM): Application of the PSTAR Method to a Residence in Fredericksburg, Virginia.* Golden, CO: Solar Energy Research Institute. SERI/TR-254-3356.

Sun, J.; Reddy T.A. (2006). "Calibration of Building Energy Simulation Programs Using the Analytic Optimization Approach (RP-1051)." *HVAC&R Research* (12:1); pp. 177–196.

Szu, H.; Hartley, R. (1987). "Fast Simulated Annealing." *Physics Letters* 122; pp. 157–162.

Trefethen, L.N.; Bau, III, D. (1997). *Numerical Linear Algebra.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Walpole, R.E.; Myers, R.H. (1989). *Probability and Statistics for Engineers and Scientists.* New York, NY: Macmillan Publishing Company.

Westphal, F.S.; Lamberts, R. (2005). "Building Simulation Calibration Using Sensitivity Analysis." *Ninth International IBPSA Conference*.

Yoon, J.; Lee, E.J.; Claridge, D.E. (2003). "Calibration Procedure for Energy Performance Simulation of a Commercial Building." *Journal of Solar Energy Engineering* 125; pp. 251–257.

# Appendix A Approximate Inputs, Uncertainty Ranges, and Explicit Input Values

**Table 9. Operational Inputs and Uncertainty Ranges**

| Approximate Input[1] | Minimum | Nominal | Maximum | Units |
|---|---|---|---|---|
| **AnnualExteriorLightingEnergy** | 45.0 | 179.0 | 446.0 | kWh/yr |
| **AnnualInteriorLightingEnergy** | 234.0 | 935.0 | 2336.0 | kWh/yr |
| **cooling_set_point** | 71.0 | 78.0 | 86.0 | °F |
| **ClothesWashMultiplier**[2] | 0.2 | 0.8 | 2.0 | - |
| FractionWindowAreaOpen | 0.000 | 0.133 | 0.467 | frac |
| FurnitureAreaFraction | 0.3 | 0.4 | 0.5 | frac |
| FurnitureConductivity | 0.5603 | 0.8004 | 1.0405 | Btu·in./h·ft$^2$·°F |
| FurnitureSolarAbsorptance | 0.4 | 0.6 | 0.8 | frac |
| FurnitureSpecHeat | 0.261 | 0.290 | 0.319 | Btu/lb·°F |
| **FurnitureWeight** | 2.0 | 8.0 | 14.0 | lb/ft$^2$ |
| **heating_set_point** | 60.0 | 68.0 | 75.0 | °F |
| **interior_shading** | 0.5 | 0.6 | 1.0 | frac |
| **KitchenAppliancesMultiplier**[3] | 0.2 | 0.8 | 2.0 | - |
| **misc_hot_water_loads** | 0.2 | 0.8 | 2.0 | - |
| **MELmultiplier** | 0.2 | 0.8 | 2.0 | - |
| MGLmultiplier | 0.2 | 0.8 | 2.0 | - |
| RangeHoodExhaust | 80.0 | 100.0 | 120.0 | cfm |
| RefrigeratorAnnualEnergy | 303.8 | 434.0 | 564.2 | kWh/yr |
| **WaterHeaterSetpoint** | 110.0 | 125.0 | 140.0 | °F |

[1] Bolded inputs are influential inputs.
[2] Values selected from this range are used for the following BEopt inputs: CWBABMultiplierElec, CWBABMultiplierHotWater, DryerBABMultiplierElec.
[3] Values selected from this range are used for the following BEopt inputs: DWBABMultiplierElec, DWBABMultiplierHotWater, CookingRangeBABMultiplierElec.

**Table 10. Asset Inputs and Uncertainty Ranges**

| Approximate Input[1] | Minimum | Nominal | Maximum | Units |
|---|---|---|---|---|
| **AC_CoolingSEER** | 9.0 | 10.0 | 10.5 | kBtu/kWh |
| CarpetPadRValue | 1.456 | 2.080 | 2.704 | h·ft$^2$·°F/Btu |
| CavityDepth1 | 3.325 | 3.500 | 3.675 | In. |
| **CavityInsRvalue1** | 0.80 | 1.01 | 1.80 | h·ft$^2$·°F/Btu |
| **CrawlACH** | 1.0 | 2.0 | 4.0 | air changes/h |
| CrawlCeilingFramingFactor | 0.08 | 0.10 | 0.12 | frac |
| CrawlCeilingRcavity | 13.3 | 19.0 | 24.7 | h·ft$^2$·°F/Btu |
| **DuctLeakage**[2] | 0.24 | 0.30 | 0.36 | frac of AH fan flow |
| **FinishAbsorptivity** | 0.5 | 0.6 | 0.8 | frac |
| **FinishConductivity** | 0.4570 | 0.6528 | 0.8486 | Btu·in./h·ft$^2$·°F |
| FinishDensity | 36.0 | 40.0 | 44.0 | lb/ft$^3$ |
| FinishEmissivity | 0.87 | 0.90 | 0.93 | frac |
| FinishSpecificHeat | 0.252 | 0.280 | 0.308 | Btu/lb·°F |
| FinishThickness | 0.418 | 0.440 | 0.462 | in |
| **FramingFactor1** | 0.20 | 0.25 | 0.30 | frac |
| GypsumThicknessCeiling | 0.4754 | 0.5004 | 0.5254 | in |
| GypsumThicknessExtWall | 0.4754 | 0.5004 | 0.5254 | in |
| **gypsum_conductivity** | 0.7778 | 1.1112 | 1.446 | Btu·in./h·ft$^2$·°F |
| gypsum_density | 45.0 | 50.0 | 55.0 | lb/ft$^3$ |
| gypsum_specific_heat | 0.234 | 0.260 | 0.286 | Btu/lb·°F |
| **LivingSpaceSLA** | 0.000619 | 0.000886 | 0.000974 | in$^2$/ft$^2$ |
| PartitionWallMassConductivity | 0.7778 | 1.1112 | 1.4446 | Btu·in./h·ft$^2$·°F |
| PartitionWallMassDensity | 45.0 | 50.0 | 55.0 | lb/ft$^3$ |
| PartitionWallMassFractionOfFloorArea | 1.064 | 1.330 | 1.596 | frac |
| PartitionWallMassSpecificHeat | 0.234 | 0.260 | 0.286 | Btu/lb·°F |
| PartitionWallMassThickness | 0.4754 | 0.5004 | 0.5254 | in |
| RoofingMaterialAbsorptivity | 0.5 | 0.6 | 0.8 | frac |
| RoofingMaterialEmissivity | 0.87 | 0.90 | 0.93 | frac |
| **ShelterCoefficient** | 0.4 | 0.5 | 0.7 | frac |
| UACeilingFramingFactor | 0.08 | 0.10 | 0.12 | frac |
| **UACeilingInsRvalue**[3] | 5.0 | 11.0 | 16.0 | h·ft$^2$·°F/Btu |
| UARoofFramingFactor | 0.08 | 0.10 | 0.12 | frac |
| UARoofFramingThickness | 5.225 | 5.500 | 5.775 | in |
| UnfinishedAtticSLA | 0.00167 | 0.00333 | 0.00667 | in.$^2$/ft$^2$ |

| Approximate Input[1] | Minimum | Nominal | Maximum | Units |
|---|---|---|---|---|
| WaterHeaterEnergyFactor | 0.86 | 0.92 | 0.93 | frac |
| WindowSHGC | 0.645 | 0.679 | 0.713 | coefficient |
| **WindowUvalue** | 0.619 | 0.774 | 0.833 | Btu/h·ft$^2$·°F |
| **wood_conductivity** | 0.5603 | 0.8004 | 1.0405 | Btu·in./h·ft$^2$·°F |
| wood_density | 28.8 | 32.0 | 35.2 | lb/ft$^3$ |
| wood_specific_heat | 0.261 | 0.290 | 0.319 | Btu/lb·°F |

[1] Bolded inputs are influential inputs.
[2] Values selected for this range are implemented by proportionally scaling the following BEopt inputs: AHLeakSA, AHLeakRA, ReturnLeak, SupplyLeak.
[3] After the value was selected from this range, the nominal insulation conductivity value was used to calculate the value of UACeilingJoistThickness, which was then also used for the value of UACeilingInsThickness.

Random numbers $\alpha_1, \ldots, \alpha_\lambda$ were generated from a uniform distribution on the interval $(0,1)$. Then, the set of explicit values $v_i$ were calculated as:

$$v_i = \begin{cases} x_i^{min} + \sqrt{\alpha_i\left(x_i^{nom} - x_i^{min}\right)\left(x_i^{max} - x_i^{min}\right)}, \text{ if } \alpha_i \leq \dfrac{x_i^{nom} - x_i^{min}}{x_i^{max} - x_i^{min}}, \\ x_i^{max} - \sqrt{(1-\alpha_i)(x_i^{max} - x_i^{nom})\left(x_i^{max} - x_i^{min}\right)}, \text{ if } \alpha_i > \dfrac{x_i^{nom} - x_i^{min}}{x_i^{max} - x_i^{min}}, \end{cases} \quad (13)$$

for $i = 1, \ldots, \lambda$, the number of approximate inputs perturbed to generate the synthetic utility data: $x_i^{min}, x_i^{nom}, x_i^{max}$ are the probability distribution's minimum, nominal, and maximum values, respectively (Kotz and van Dorp 2004).

**Table 11. Reference Model Input Values for Scenario 1[16]**

| Approximate Input | $v_i$ | Approximate Input | $v_i$ |
|---|---|---|---|
| AC_CoolingSEER | 10.12 | heating_set_point | 65.47 |
| AnnualInteriorLightingEnergy | 574.46 | interior_shading | 0.74 |
| AnnualExteriorLightingEnergy | 270.57 | KitchenAppliancesMultiplier | 1.20 |
| CarpetPadRValue | 2.01 | LivingSpaceSLA | 0.00085 |
| CavityDepth1 | 3.51 | MELmultiplier | 0.55 |
| CavityInsRvalue1 | 1.37 | MGLmultiplier | 0.95 |
| ClothesWashMultiplier | 0.64 | misc_hot_water_loads | 0.87 |
| cooling_set_point | 82.01 | PartitionWallMassConductivity | 0.94 |
| CrawlACH | 3.05 | PartitionWallMassDensity | 49.18 |
| CrawlCeilingFramingFactor | 0.11 | PartitionWallMassFractionOfFloorArea | 1.36 |
| CrawlCeilingRcavity | 21.81 | PartitionWallMassSpecificHeat | 0.26 |
| DuctLeakage | 0.31 | PartitionWallMassThickness | 0.49 |
| FinishAbsorptivity | 0.57 | RangeHoodExhaust | 110.49 |
| FinishConductivity | 0.63 | RefrigeratorAnnualEnergy | 480.27 |
| FinishDensity | 40.85 | RoofingMaterialAbsorptivity | 0.56 |
| FinishEmissivity | 0.90 | RoofingMaterialEmissivity | 0.91 |
| FinishSpecificHeat | 0.27 | ShelterCoefficient | 0.47 |
| FInishThickness | 0.43 | UACeilingFramingFactor | 0.10 |
| FractionWindowAreaOpen | 0.16 | UACeilingInsRvalue | 12.90 |
| FramingFactor1 | 0.24 | UARoofFramingFactor | 0.11 |
| FurnitureAreaFraction | 0.38 | UARoofFramingThickness | 5.56 |
| FurnitureConductivity | 0.93 | UnfinishedAtticSLA | 0.0031 |
| FurnitureSolarAbsorptance | 0.61 | WaterHeaterEnergyFactor | 0.87 |
| FurnitureSpecHeat | 0.27 | WaterHeaterSetpoint | 138.37 |
| FurnitureWeight | 8.20 | WindowSHGC | 0.66 |
| GypsumThicknessCeiling | 0.49 | WindowUvalue | 0.70 |
| GypsumThicknessExtWall | 0.49 | wood_conductivity | 0.79 |
| gypsum_conductivity | 1.29 | wood_specific_heat | 0.29 |
| gypsum_specific_heat | 0.26 | wood_density | 29.58 |
| gypsum_density | 45.73 | | |

[16] Units for these approximate inputs are found in Tables 9 and 10.

**Table 12. Reference Model Input Values for Scenario 2[17]**

| Approximate Input | $v_i$ | Approximate Input | $v_i$ |
|---|---|---|---|
| AC_CoolingSEER | 9.70 | heating_set_point | 66.65 |
| AnnualInteriorLightingEnergy | 1703.27 | interior_shading | 0.88 |
| AnnualExteriorLightingEnergy | 272.04 | KitchenAppliancesMultiplier | 0.73 |
| CarpetPadRValue | 2.19 | LivingSpaceSLA | 0.00079 |
| CavityDepth1 | 3.42 | MELmultiplier | 0.93 |
| CavityInsRvalue1 | 0.99 | MGLmultiplier | 1.00 |
| ClothesWashMultiplier | 0.92 | misc_hot_water_loads | 0.96 |
| cooling_set_point | 73.69 | PartitionWallMassConductivity | 1.19 |
| CrawlACH | 2.74 | PartitionWallMassDensity | 53.03 |
| CrawlCeilingFramingFactor | 0.10 | PartitionWallMassFractionOfFloorArea | 1.30 |
| CrawlCeilingRcavity | 19.69 | PartitionWallMassSpecificHeat | 0.25 |
| DuctLeakage | 0.29 | PartitionWallMassThickness | 0.49 |
| FinishAbsorptivity | 0.74 | RangeHoodExhaust | 103.64 |
| FinishConductivity | 0.63 | RefrigeratorAnnualEnergy | 494.23 |
| FinishDensity | 39.64 | RoofMaterialAbsorptivity | 0.71 |
| FinishEmissivity | 0.93 | RoofingMaterialEmissivity | 0.92 |
| FinishSpecificHeat | 0.27 | ShelterCoefficient | 0.52 |
| FInishThickness | 0.44 | UACeilingFramingFactor | 0.098 |
| FractionWindowAreaOpen | 0.19 | UACeilingInsRvalue | 15.25 |
| FramingFactor1 | 0.26 | UARoofFramingFactor | 0.09 |
| FurnitureAreaFraction | 0.43 | UARoofFramingThickness | 5.69 |
| FurnitureConductivity | 0.65 | UnfinishedAtticSLA | 0.0031 |
| FurnitureSolarAbsorptance | 0.71 | WaterHeaterEnergyFactor | 0.89 |
| FurnitureSpecHeat | 0.51 | WaterHeaterSetpoint | 122.91 |
| FurnitureWeight | 7.11 | WindowSHGC | 0.69 |
| GypsumThicknessCeiling | 0.51 | WindowUvalue | 0.76 |
| GypsumThicknessExtWall | 0.51 | wood_conductivity | 0.68 |
| gypsum_conductivity | 1.24 | wood_specific_heat | 0.28 |
| gypsum_specific_heat | 0.28 | wood_density | 31.01 |
| gypsum_density | 53.15 | | |

[17] Units for these approximate inputs are found in Tables 9 and 10.

# Appendix B Sensitivity Analysis Results

**Table 13. The 24 Influential Inputs**

| Influential Input | $\xi_j$ |
|---|---|
| heating_set_point | 0.135 |
| cooling_set_point | 0.063 |
| MELmultiplier | 0.031 |
| misc_hot_water_loads | 0.028 |
| UACeilingInsRvalue | 0.026 |
| CrawlACH | 0.019 |
| ClothesWashMultiplier | 0.018 |
| DuctLeakage | 0.017 |
| wood_conductivity | 0.013 |
| WindowUvalue | 0.012 |
| KitchenAppliancesMultiplier | 0.011 |
| AC_CoolingSEER | 0.009 |
| AnnualInteriorLightingEnergy | 0.009 |
| LivingSpaceSLA | 0.009 |
| CavityInsRvalue1 | 0.008 |
| interior_shading | 0.008 |
| FurnitureWeight | 0.005 |
| ShelterCoefficient | 0.004 |
| FinishConductivity | 0.004 |
| gypsum_conductivity | 0.003 |
| AnnualExteriorLightingEnergy | 0.003 |
| WaterHeaterSetpoint | 0.003 |
| FramingFactor1 | 0.003 |
| FinishAbsorptivity | 0.002 |

**Table 14. Procedure 3.2 and 3.3 Results of $\xi$ Sensitivity Analysis**

| Strong Parameter | $\xi_j$ |
|---|---|
| heating_set_point | 0.126 |
| cooling_set_point | 0.064 |
| MELmultiplier | 0.032 |
| misc_hot_water_loads | 0.027 |
| UACeilingInsRvalue | 0.021 |
| ClothesWashMultiplier | 0.020 |

# Appendix C Process Flow

Figure 52 shows the process flow for developing approximate input ranges, identifying influential inputs, selecting adjustable inputs, and finally, adjusting these select inputs to recover the calibrated inputs. The steps in the "Initial Assumptions" process are performed for the purpose of comparing calibration procedures; the steps in the "Calibration" are performed as a part of each calibration procedure.

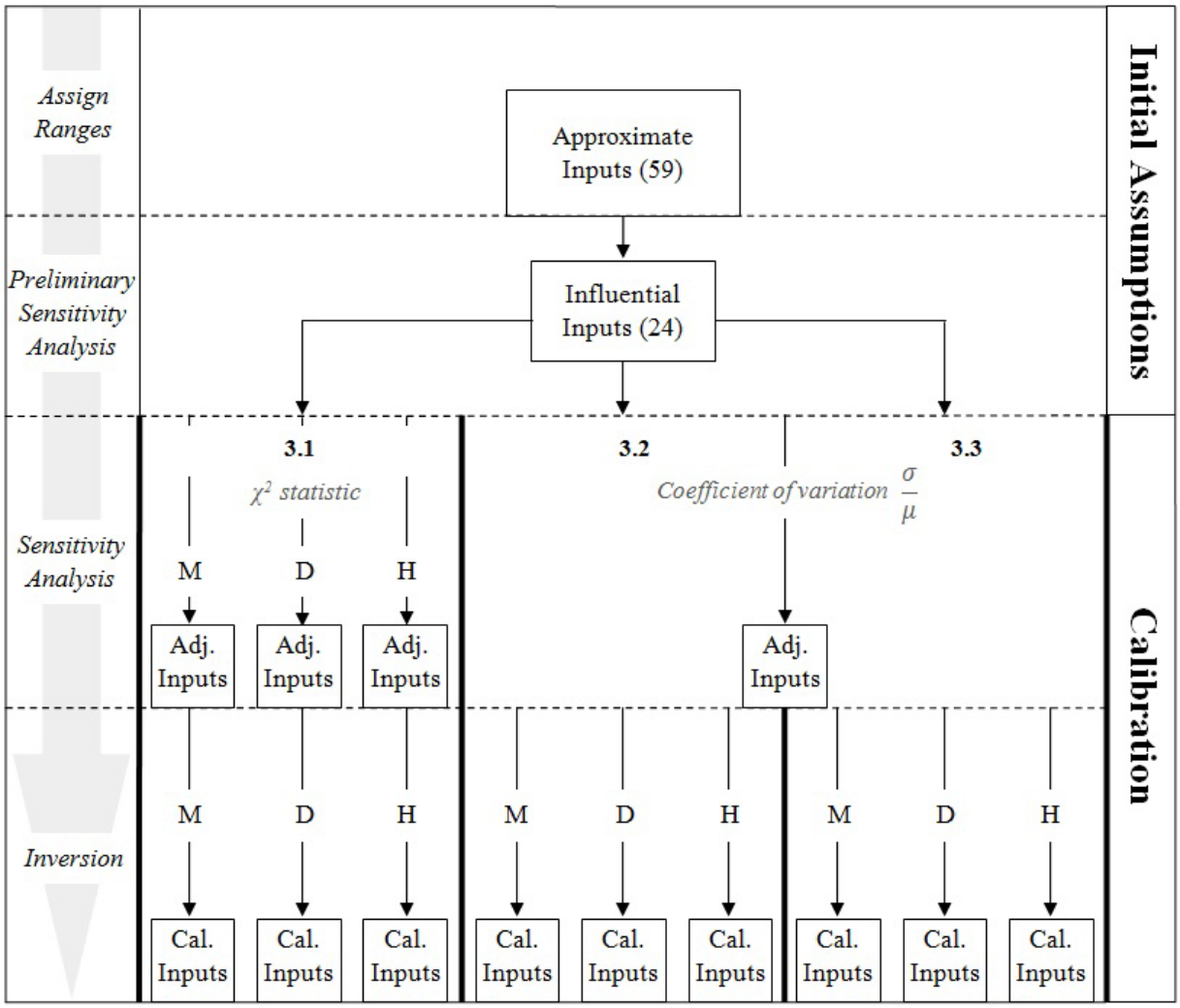


**Figure 52. Process flow for input calibration methods**

# Appendix D Latin Hypercube Monte Carlo

This appendix presents the steps for performing a three-level Latin Hypercube Monte Carlo (LHMC) (Reddy and Maor 2006). These steps were followed for the implementation of Procedure 3.1 (see Section 3.1.2).

1. Discretize each of the 24 influential parameters' triangular probability distributions into three ranges of equal probability (i.e., areas of one-third). Determine the median values (levels), $x_i^{low}, x_i^{high}, x_i^{mid}$, of each of these three ranges as follows:

$$x_i^{low} = x_i^{min} + \sqrt{\frac{\left(x_i^{nom} - x_i^{min}\right)\left(x_i^{max} - x_i^{min}\right)}{6}}, \tag{14}$$

$$x_i^{high} = x_i^{max} - \sqrt{\frac{\left(x_i^{max} - x_i^{nom}\right)\left(x_i^{max} - x_i^{min}\right)}{6}}, \tag{15}$$

$$x_i^{mid} = \begin{cases} x_i^{min} + \sqrt{\frac{\left(x_i^{nom} - x_i^{min}\right)\left(x_i^{max} - x_i^{min}\right)}{2}}, \text{ if } x_i^{nom} \geq \frac{x_i^{max} + x_i^{min}}{2} \\ x_i^{max} - \sqrt{\frac{\left(x_i^{max} - x_i^{nom}\right)\left(x_i^{max} - x_i^{min}\right)}{2}}, \text{ if } x_i^{nom} < \frac{x_i^{max} + x_i^{min}}{2} \end{cases} \tag{16}$$

2. Generate LHMC combinations of all parameters. That is, use uniform sampling without replacement to randomly select a subset of the $3^{24}$ total LHMC parameter combinations.
3. Create realizations by replacing the influential input values of the base-case building description file with each LHMC parameter combination.
4. Simulate in BEopt/DOE-2.2 all building description files in automated batch mode.

# Appendix E Central Composite Design

Central composite design matrix $\mathbf{X}$ relies on a *2-level full factorial*, axial points, and center run vertically concatenated design composite, and is given by:

$$\mathbf{X} = \left[\begin{array}{c|cccc|ccc|ccc} \text{Intercept} & & \text{Design} & & & & \text{Interaction} & & & \text{Squared} & & \\ 1 & -\alpha & -\alpha & \cdots & -\alpha & (-\alpha)(-\alpha) & \cdots & (-\alpha)(-\alpha) & (-\alpha)(-\alpha) & \cdots & (-\alpha)(-\alpha) \\ 1 & \alpha & -\alpha & \cdots & -\alpha & (\alpha)(-\alpha) & \cdots & (-\alpha)(-\alpha) & (\alpha)(\alpha) & \cdots & (-\alpha)(-\alpha) \\ & & & \ddots & & & \ddots & & & \ddots & \\ 1 & -\alpha & \alpha & \cdots & \alpha & (-\alpha)(\alpha) & \cdots & (\alpha)(\alpha) & (-\alpha)(-\alpha) & \cdots & (\alpha)(\alpha) \\ 1 & \alpha & \alpha & \cdots & \alpha & (\alpha)(\alpha) & \cdots & (\alpha)(\alpha) & (\alpha)(\alpha) & \cdots & (\alpha)(\alpha) \\ 1 & -1 & 0 & \cdots & 0 & 0 & \cdots & 0 & (-1)(-1) & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & (1)(1) & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ & & & \ddots & & & \ddots & & & \ddots & \\ 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right],$$

where $\alpha = 2^{-\frac{\tau}{4}}$ and $\tau$ is the number of parameters considered in the design (i.e., number of 'Design' columns). Each row of the submatrix whose columns are the 2nd through the $(\tau + 1)^{\text{th}}$ columns of $\mathbf{X}$ are coded realizations (corresponding to the 1st through $\tau^{\text{th}}$ parameter) which, when replaced in the base-case building description file and simulated in BEopt/DOE-2.2, produce response matrix $\mathbf{Y}$ whose dimensions are $(2^{\tau} + 2\tau + 1)$ x $n$, where $n$ is the number of data points to which we are calibrating. The coded values $\mathrm{X}_{i,j}$ linearly map to input values $\mathrm{X}'_{i,j}$ by:

$$\mathrm{X}'_{i,j} = \mathrm{X}_{i,j}\left(\frac{1}{2}\left(x_j^{max} - x_j^{min}\right)\right) + \left(\frac{1}{2}\left(x_j^{max} + x_j^{min}\right)\right), \tag{17}$$

for $j = 1, \ldots, \tau$, where the $\mathrm{X}_{i,j}$ are elements from the 'Design' submatrix of $\mathbf{X}$, and $x_j^{min}$ and $x_j^{max}$ are the probability distribution's minimum and maximum values, respectively, from Table 9 and Table 10. There exist several alternative designs for constructing quadratic metamodels for use in response surface methodology. For this analysis, the central composite design was chosen and implemented due to its low cost and widespread use (Barton 2009; Batmaz and Tunali 2002).

# Appendix F Goodness-of-Fit Indices

For assessing agreement between model-predicted and measured utility data, goodness-of-fit indices are defined. The "coefficient of variation of the root mean square error" CV(RMSE) characterizes the variability of the errors between model-predicted and measured utility data (ASHRAE 2002) and is given by:

$$\text{CV(RMSE)} = \frac{1}{\bar{y}}\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}} \text{ x } 100\%. \tag{18}$$

The "normalized mean bias error" NMBE quantifies the percentage error between model-predicted and measured utility data summed over the number of data points (ASHRAE 2002) and is given by:

$$\text{NMBE} = \frac{1}{\bar{y}}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n-p} \text{ x } 100\%. \tag{19}$$

For Eqns. (18) and (19) the $y_i$ are measured utility data, the $\hat{y}_i$ are simulation-predicted data, $\bar{y}$ is the measured utility data mean, $n$ is the number of measured utility data points used in calibration, and $p$ is the number of adjustable model parameters.[18]

The goodness-of-fit indices given in Eqs. (18) and (19) can be consolidated into a single statistical index representing overall goodness-of-fit (Reddy and Maor 2006). For simultaneously considering both the CV(RMSE) and the NMBE, the "total goodness-of-fit" $\delta_{TOTAL}$ is defined as:

$$\delta_{TOTAL} = \sqrt{\frac{w_{\text{CV(RMSE)}}^2 \text{CV(RMSE)}^2 + w_{\text{NMBE}}^2 \text{NMBE}^2}{w_{\text{CV(RMSE)}}^2 + w_{\text{NMBE}}^2}}, \tag{20}$$

where $w_{\text{CV(RMSE)}}, w_{\text{NMBE}}$ are the weights of goodness-of-fit indices CV(RMSE), NMBE, respectively. For this analysis, we assume equal weight for CV(RMSE) and NMBE (i.e., $w_{\text{CV(RMSE)}} = w_{\text{NMBE}} = 0.5$), and so Eq. (20) reduces to:

$$\delta_{TOTAL} = \frac{\sqrt{2}}{2}\sqrt{\text{CV(RMSE)}^2 + \text{NMBE}^2}. \tag{21}$$

[18] Note that in accordance with suggestions from Reddy and Maor, $p = 1$ for Eqn. (18) and $p = 0$ for Eqn. (19).

# Appendix G Comparing Wall and Ceiling R-Values to BESTEST-EX

For this study, BEopt inputs corresponding to wall and ceiling assemblies were perturbed at a detailed level (e.g., conductivity of framing members). For the study in BESTEST-EX, overall wall and ceiling R-values were perturbed. Therefore, a small study was conducted to verify that the perturbations at a detailed level resulted in wall and ceiling assembly R-value ranges similar to those in BESTEST-EX. Specifically, random selections for detailed inputs related to wall and ceiling assemblies were made and used to calculate overall R-values of the assemblies according to the steady-state, parallel-path method. This process was repeated multiple times to develop distributions of overall wall and ceiling R-values. Then, these distributions were compared to distributions resulting from selections from ranges specified in BESTEST-EX (see Figure 53 through Figure 56). The comparisons revealed that the default, hard-wired OSB sheathing R-value in BEopt/DOE-2.2 was R-0.69 less than the nominal wallboard R-value defined in BESTEST-EX. To correct for this, R-0.69 of insulated foam sheathing was modeled in BEopt and scaled appropriately for all simulations in which the wood conductivity values were randomly selected. After this correction, random selections from input ranges (specified in Table 9 and Table 10) related to wall and ceiling assemblies resulted in very similar wall and ceiling assembly R-value ranges found in BESTEST-EX.

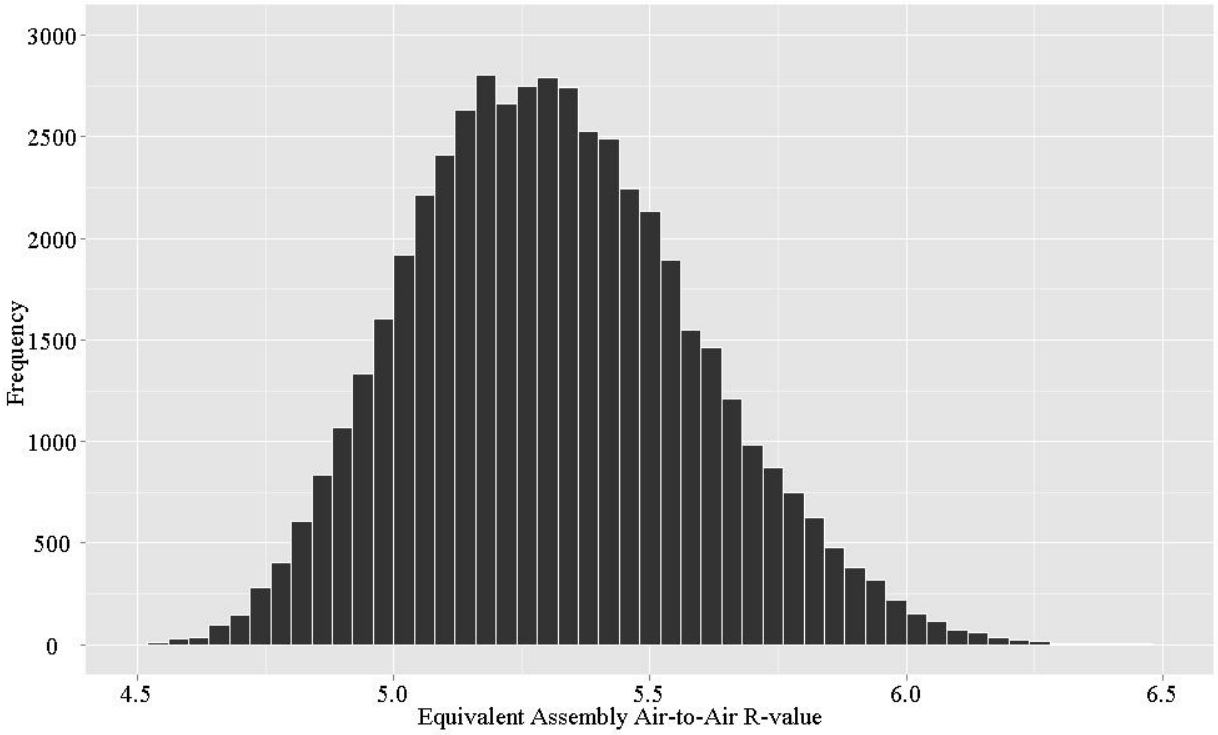


**Figure 53. Histogram of equivalent wall assembly R-values resulting from 50,000 sets of random input selections from ranges specified in this study**

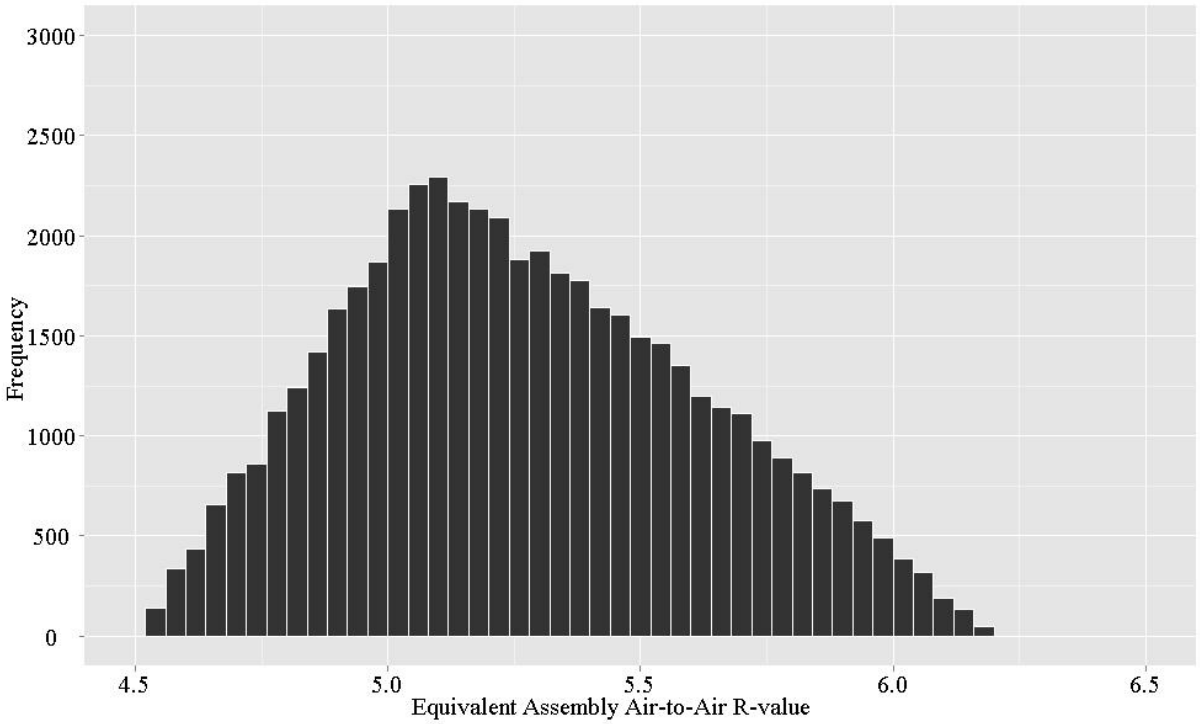


**Figure 54. Histogram of equivalent wall assembly R-values resulting from 50,000 sets of random input selections from ranges specified in BESTEST-EX**

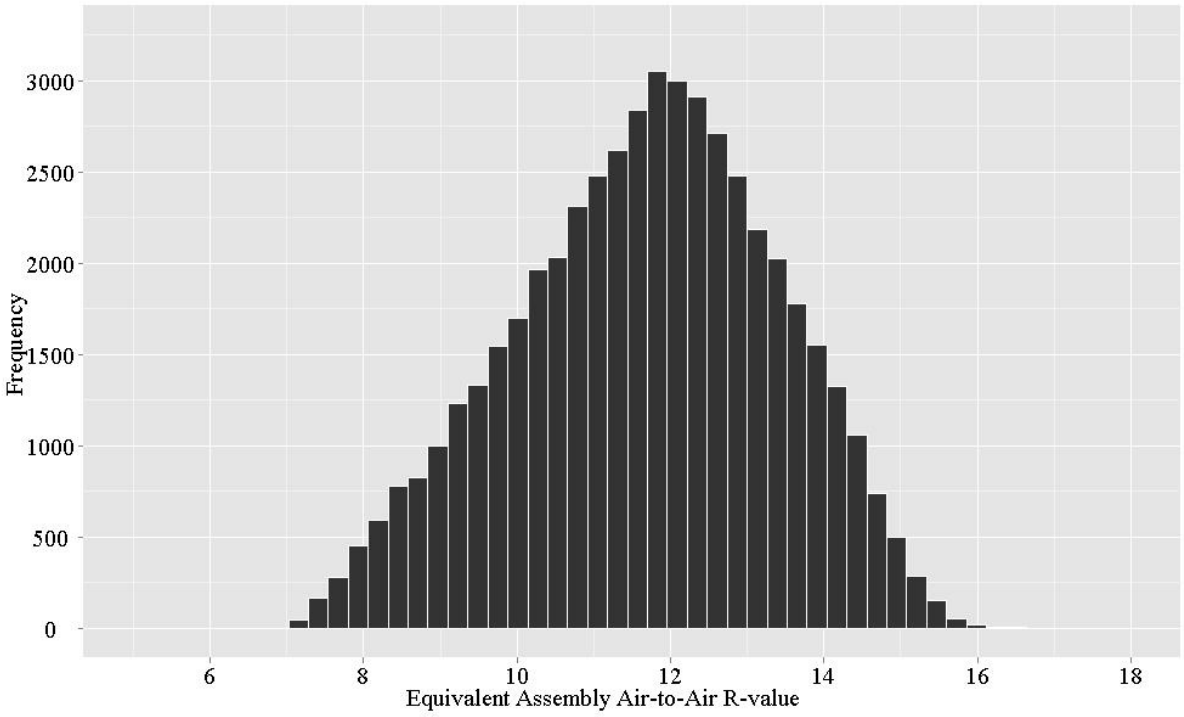


**Figure 55. Histogram of equivalent ceiling assembly R-values resulting from 50,000 sets of random input selections from ranges specified in this study**

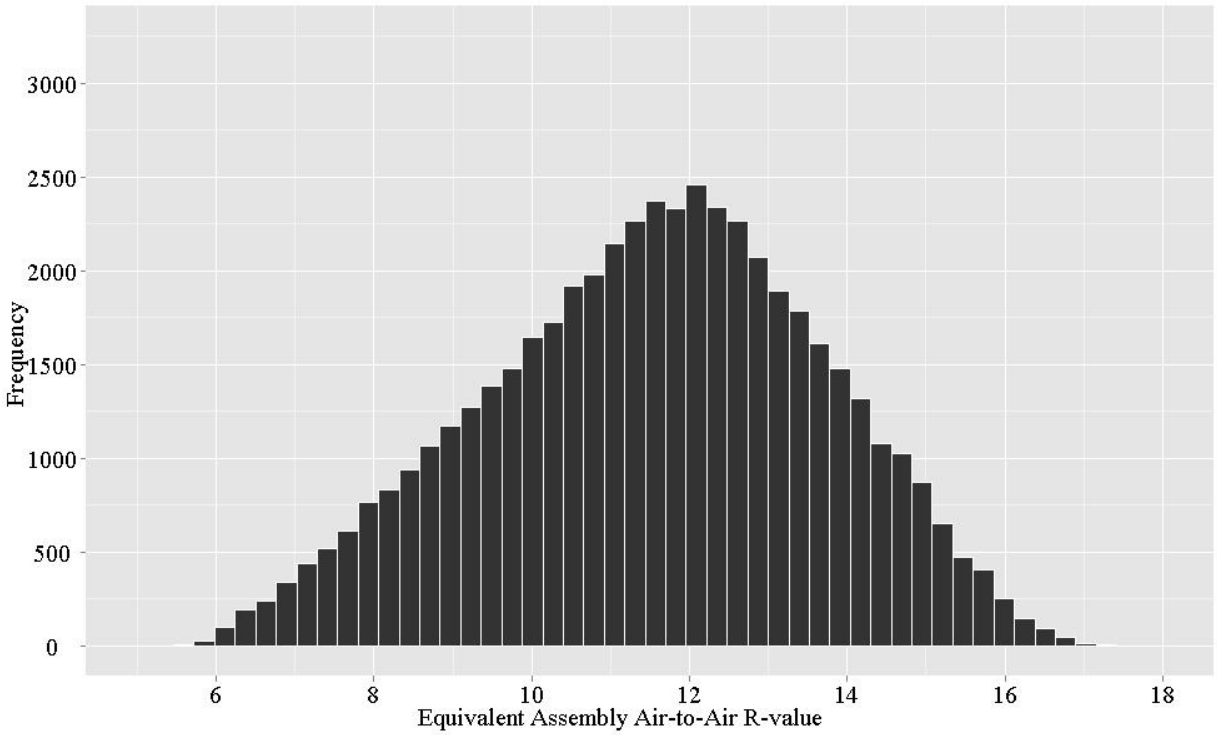


**Figure 56. Histogram of equivalent ceiling assembly R-values resulting from 50,000 sets of random input selections from ranges specified in BESTEST-EX**

# Appendix H Additional Figures

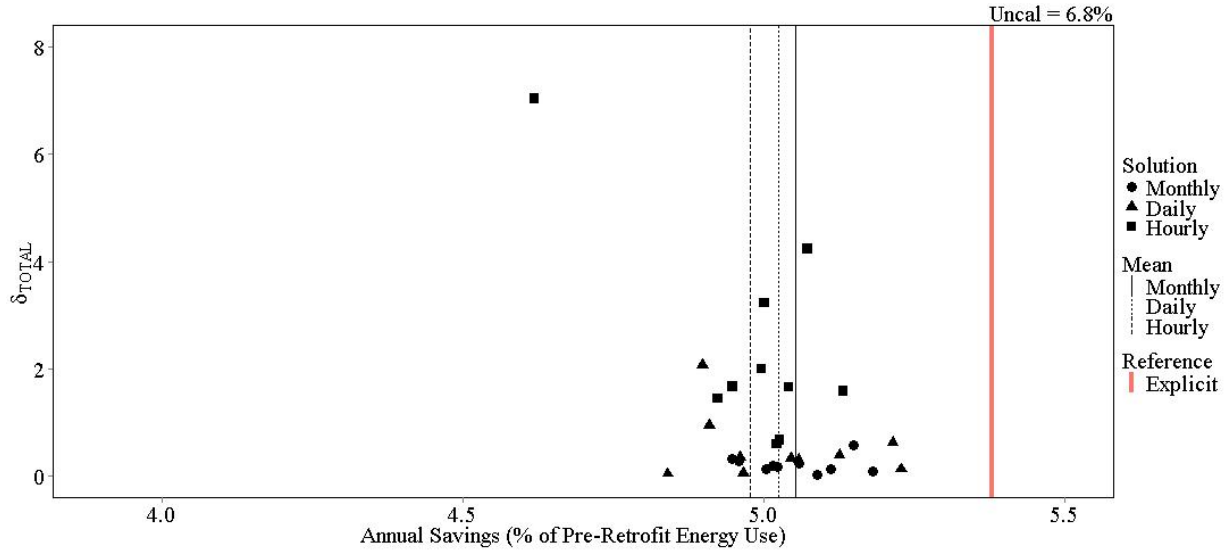


**Figure 57. Procedure 3.1 energy savings predictions for air-seal retrofit, Scenario 1**

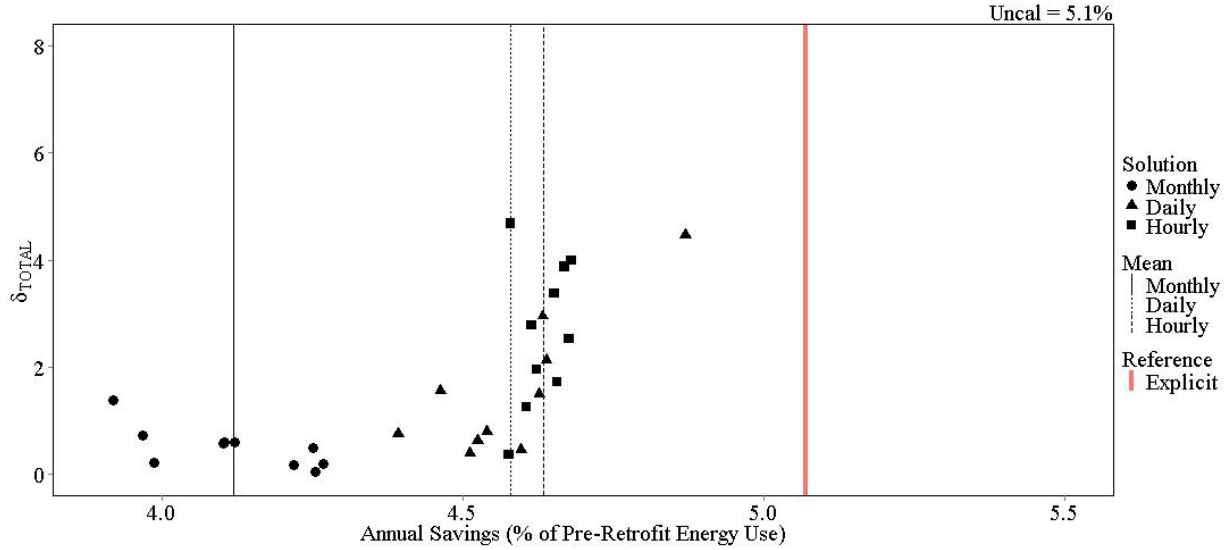


**Figure 58. Procedure 3.1 energy savings predictions for air-seal retrofit, Scenario 2**

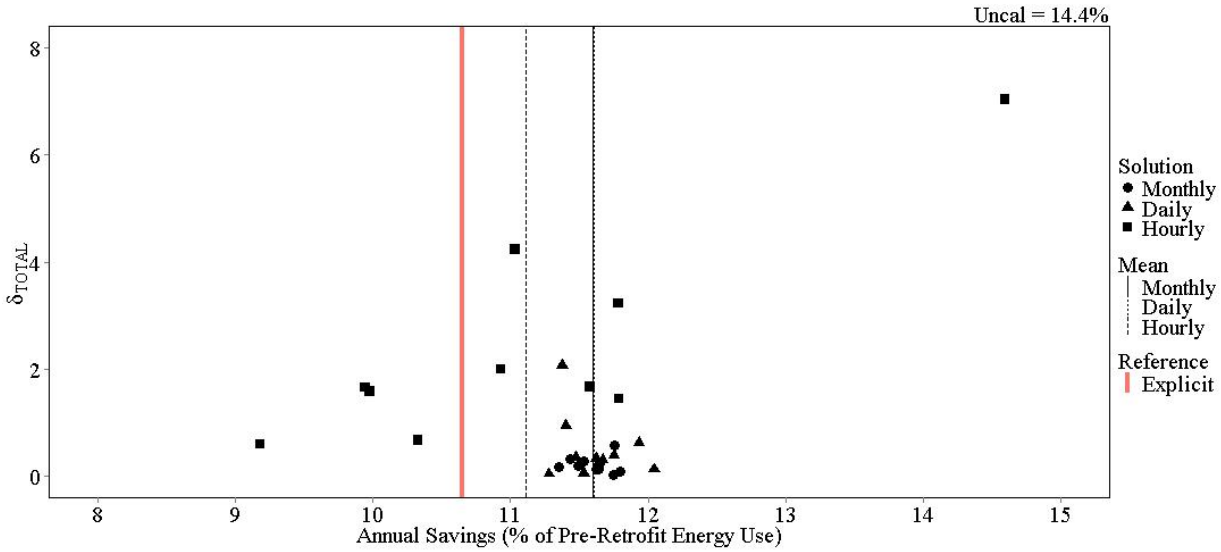


**Figure 59. Procedure 3.1 energy savings predictions for attic insulation retrofit, Scenario 1**

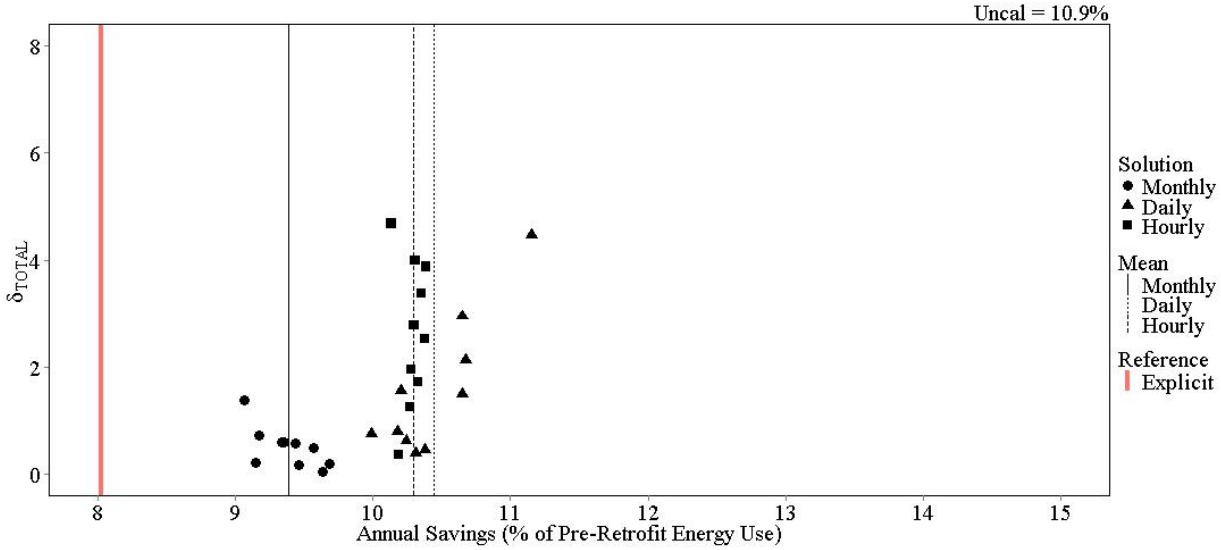


**Figure 60. Procedure 3.1 energy savings predictions for attic insulation retrofit, Scenario 2**

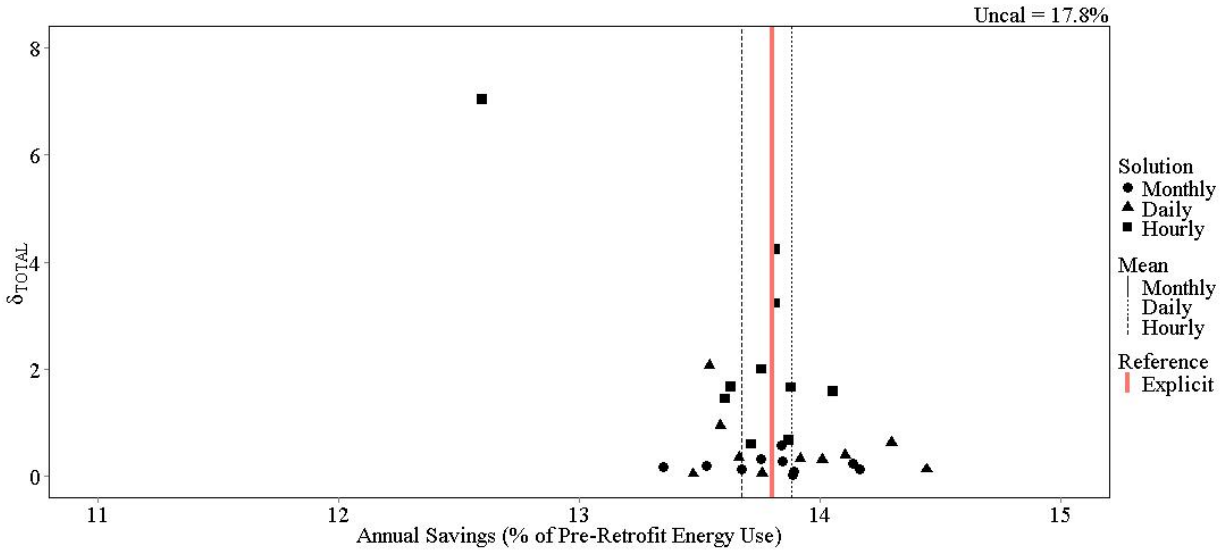


**Figure 61. Procedure 3.1 energy savings predictions for wall insulation retrofit, Scenario 1**

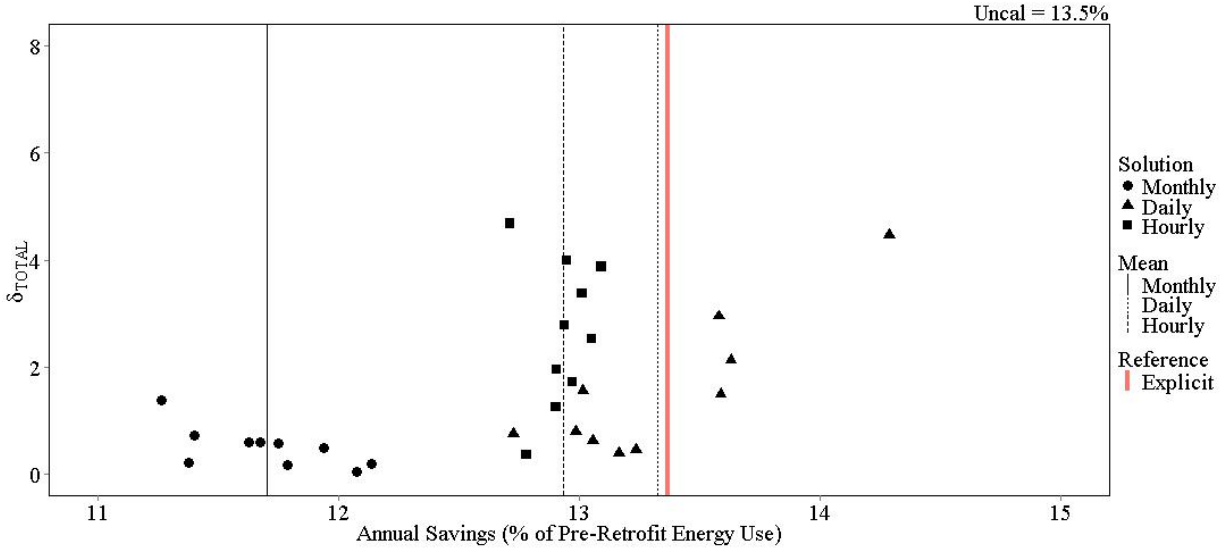


**Figure 62. Procedure 3.1 energy savings predictions for wall insulation retrofit, Scenario 2**

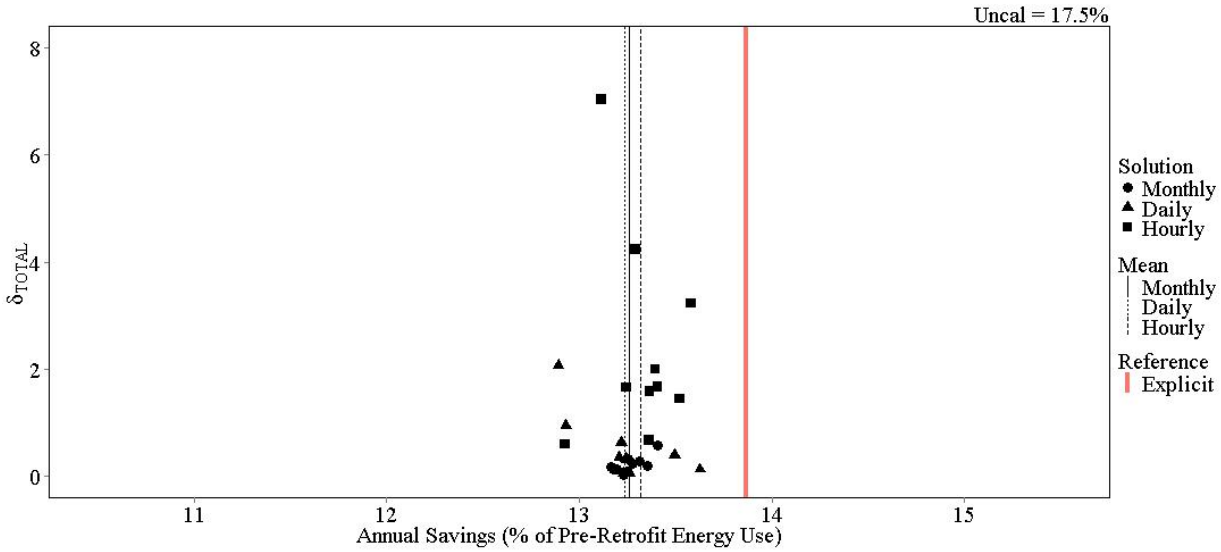


**Figure 63. Procedure 3.1 energy savings predictions for programmable thermostat retrofit, Scenario 1**

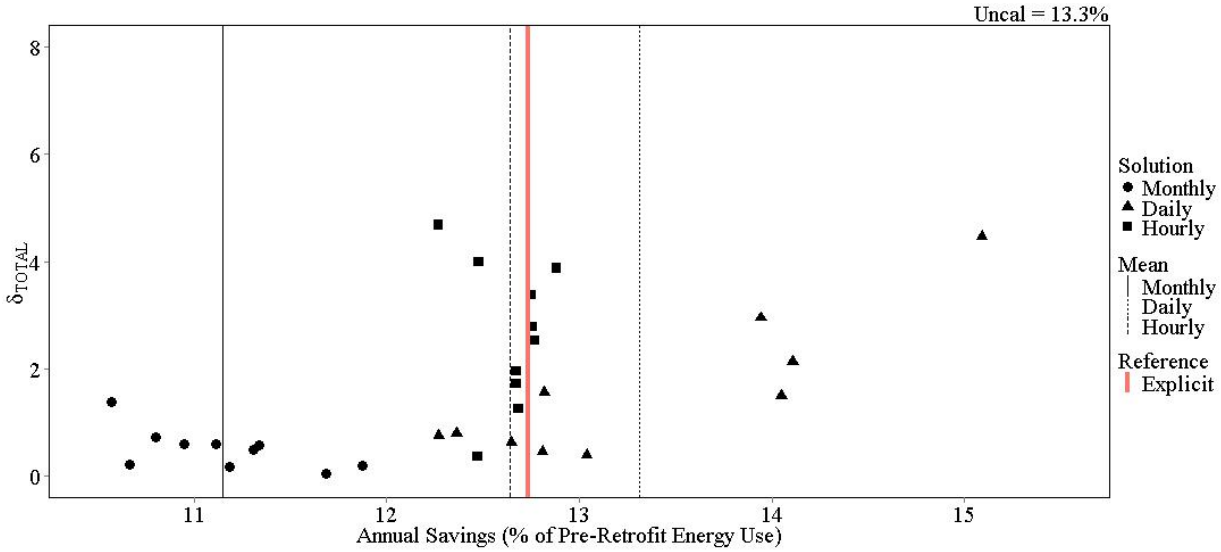


**Figure 64. Procedure 3.1 energy savings predictions for programmable thermostat retrofit, Scenario 2**

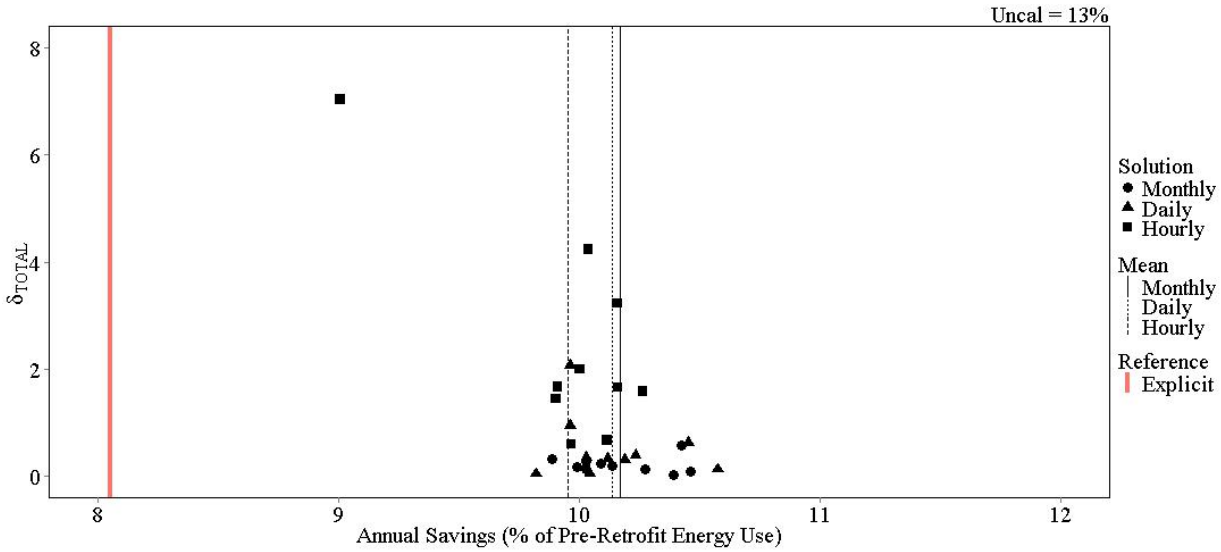


**Figure 65. Procedure 3.1 energy savings predictions for low-e windows retrofit, Scenario 1**

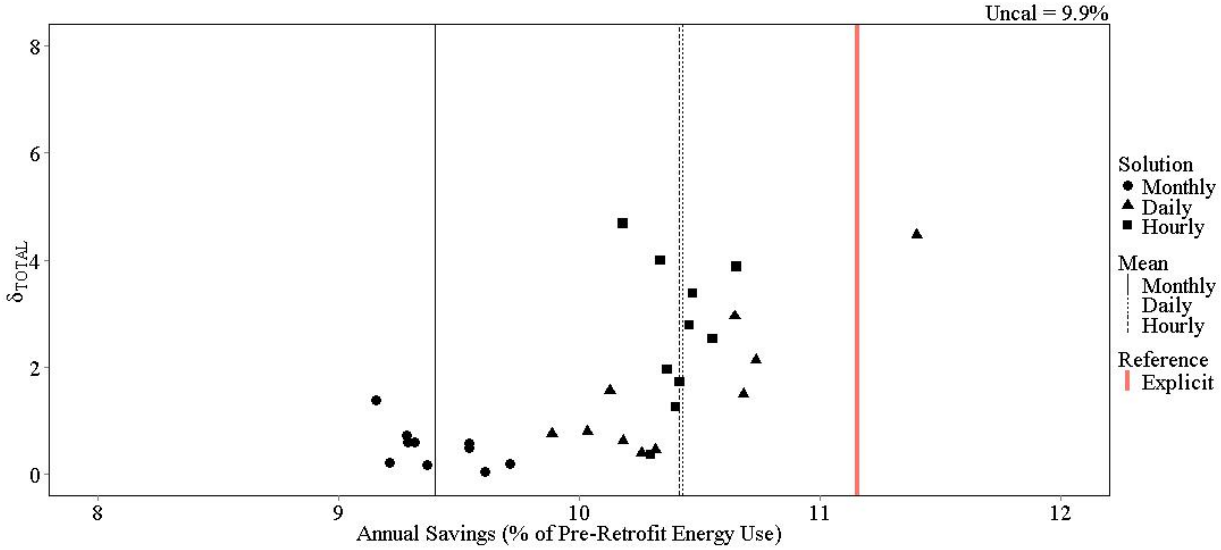


**Figure 66. Procedure 3.1 energy savings predictions for low-e windows retrofit, Scenario 2**

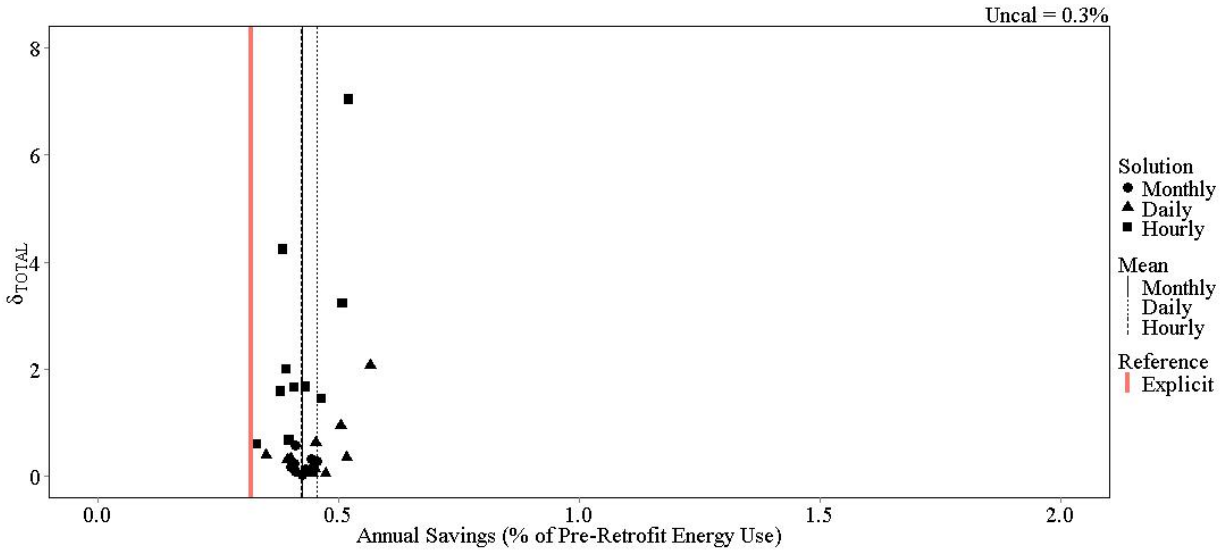


**Figure 67. Procedure 3.1 energy savings predictions for low solar absorptance roof retrofit, Scenario 1**

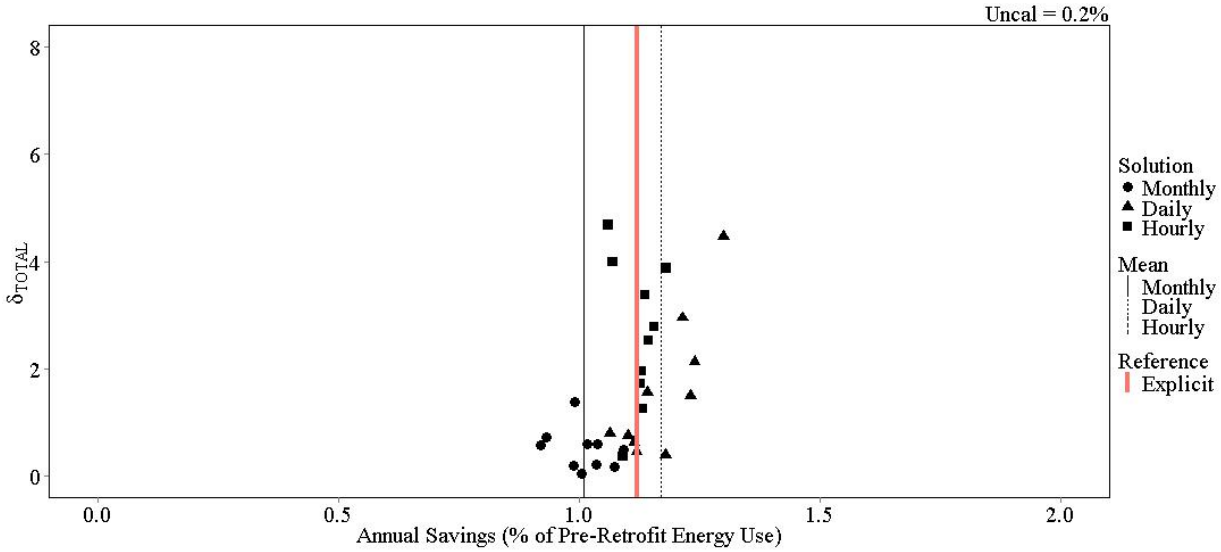


**Figure 68. Procedure 3.1 energy savings predictions for low solar absorptance roof retrofit, Scenario 2**

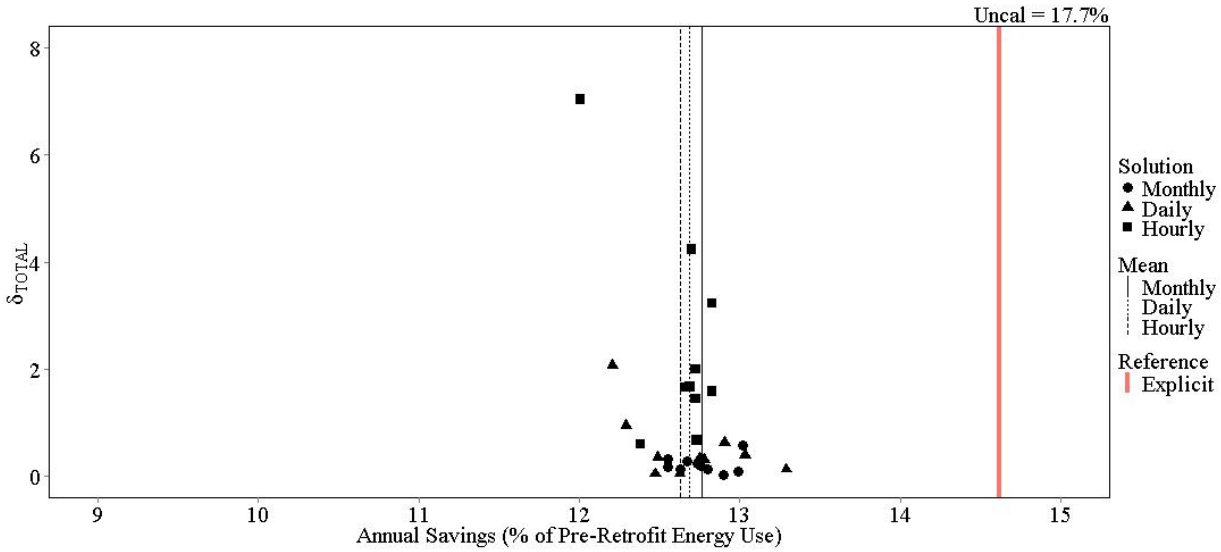


**Figure 69. Procedure 3.1 energy savings predictions for duct sealing and insulation retrofit, Scenario 1**

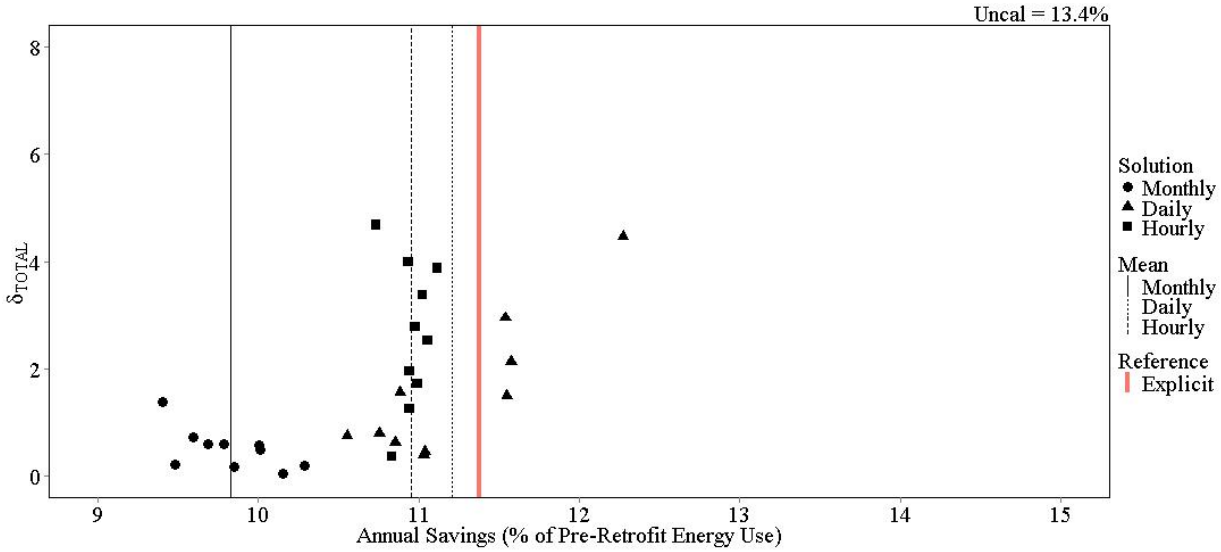


**Figure 70. Procedure 3.1 energy savings predictions for duct sealing and insulation retrofit, Scenario 2**

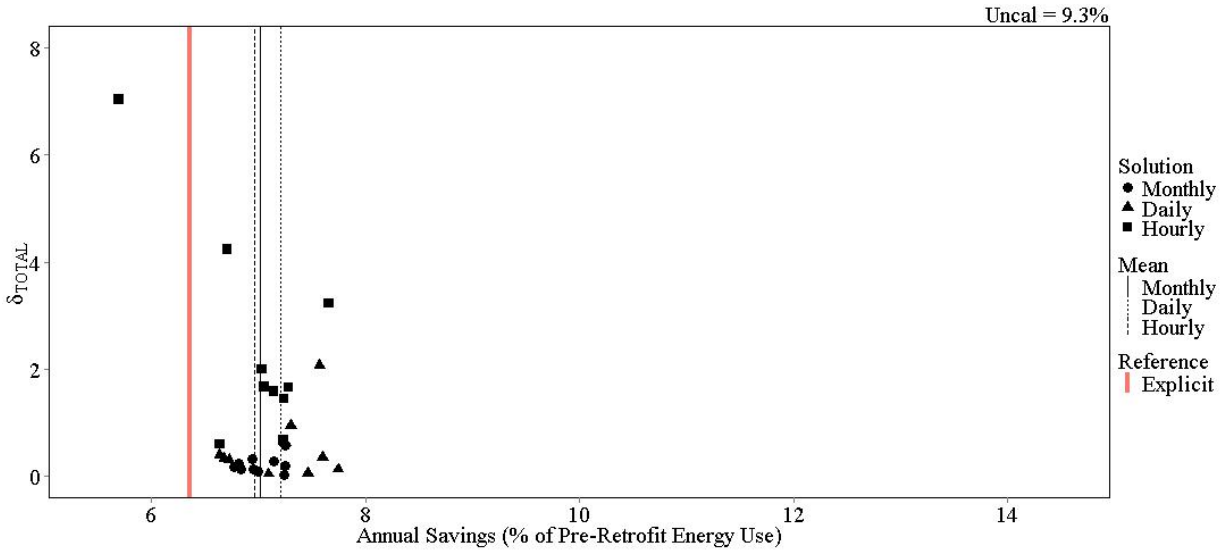


**Figure 71. Procedure 3.1 energy savings predictions for AC replacement retrofit, Scenario 1**

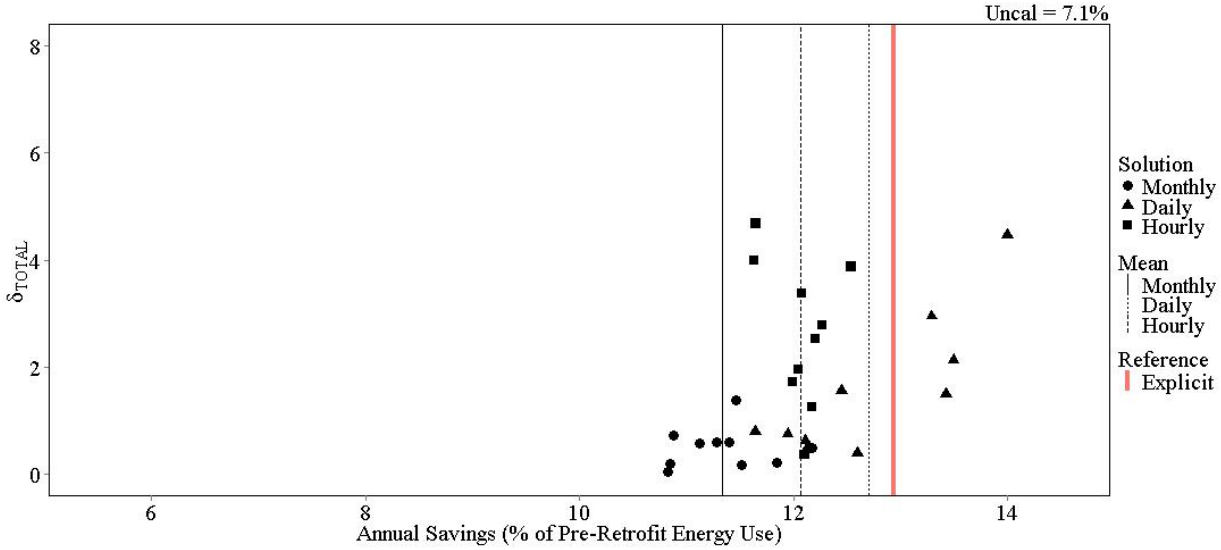


**Figure 72. Procedure 3.1 energy savings predictions for AC replacement retrofit, Scenario 2**

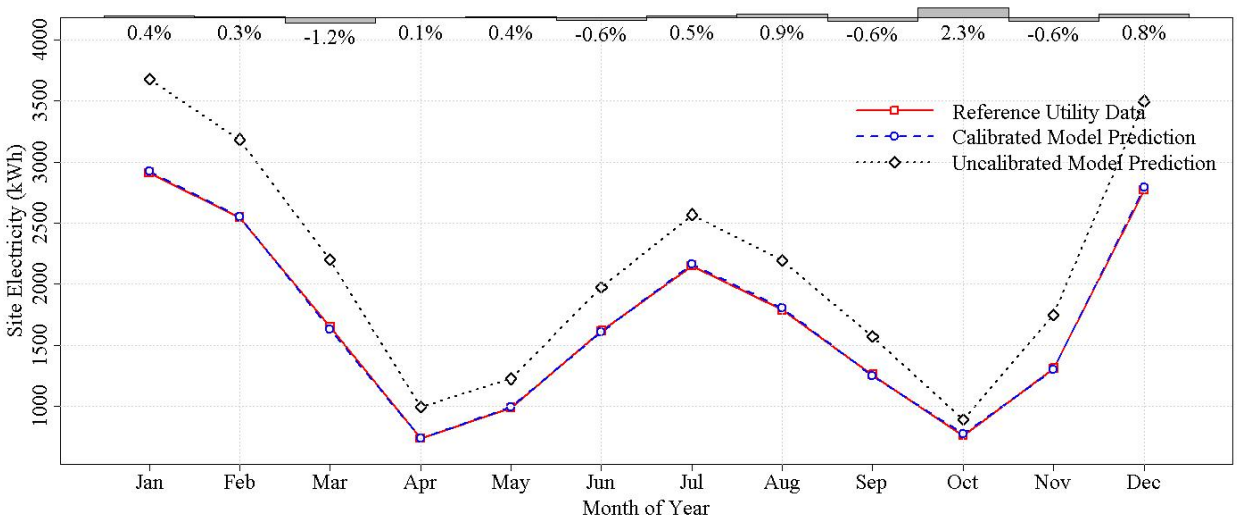


**Figure 73. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.2, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

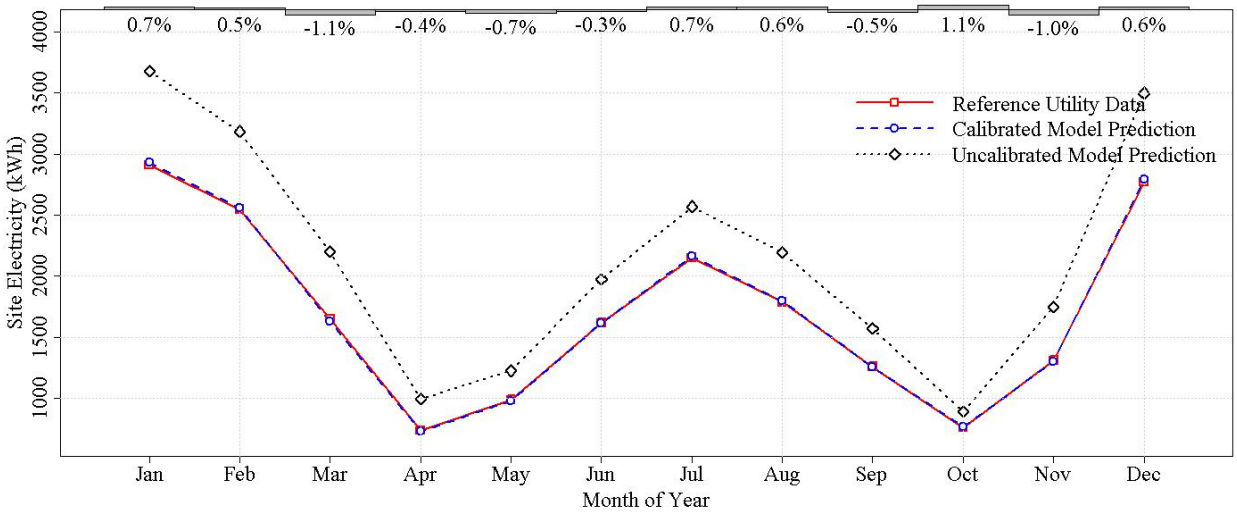


**Figure 74. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.2, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

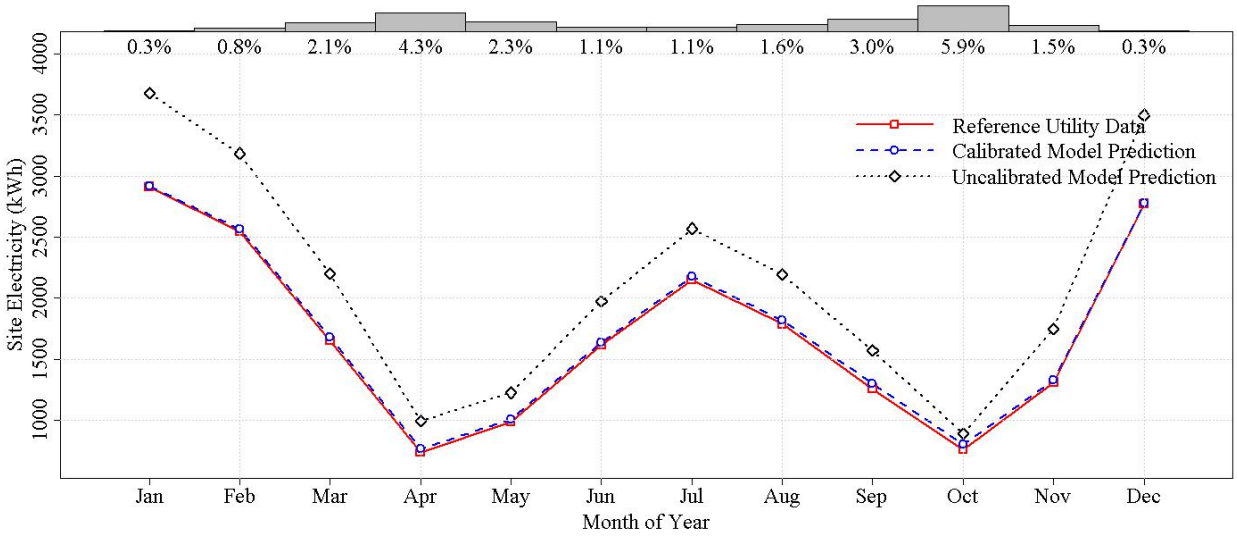


**Figure 75. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.2, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

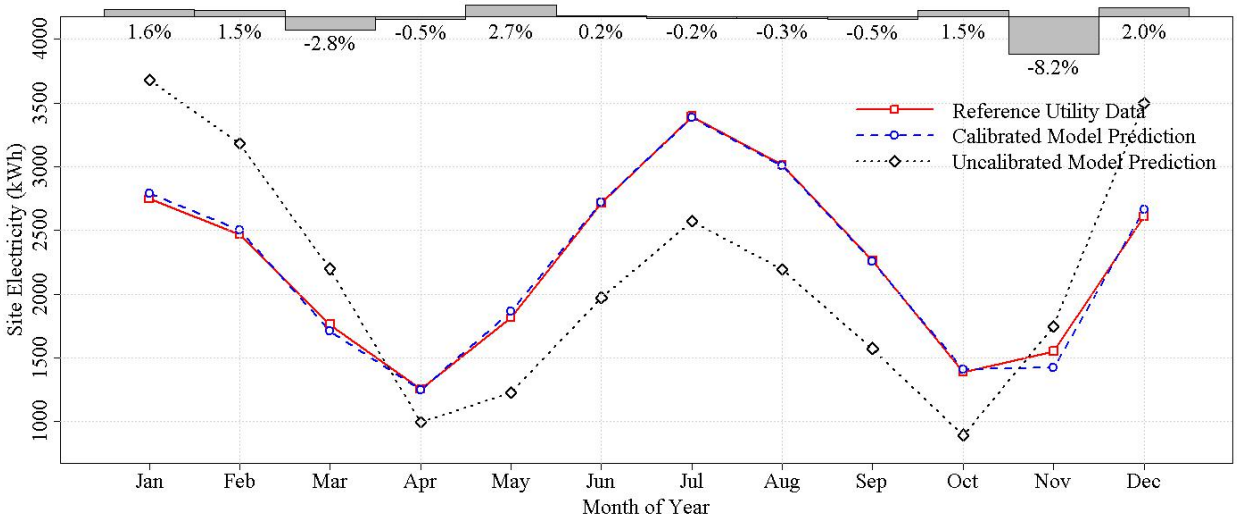


**Figure 76. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.2, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

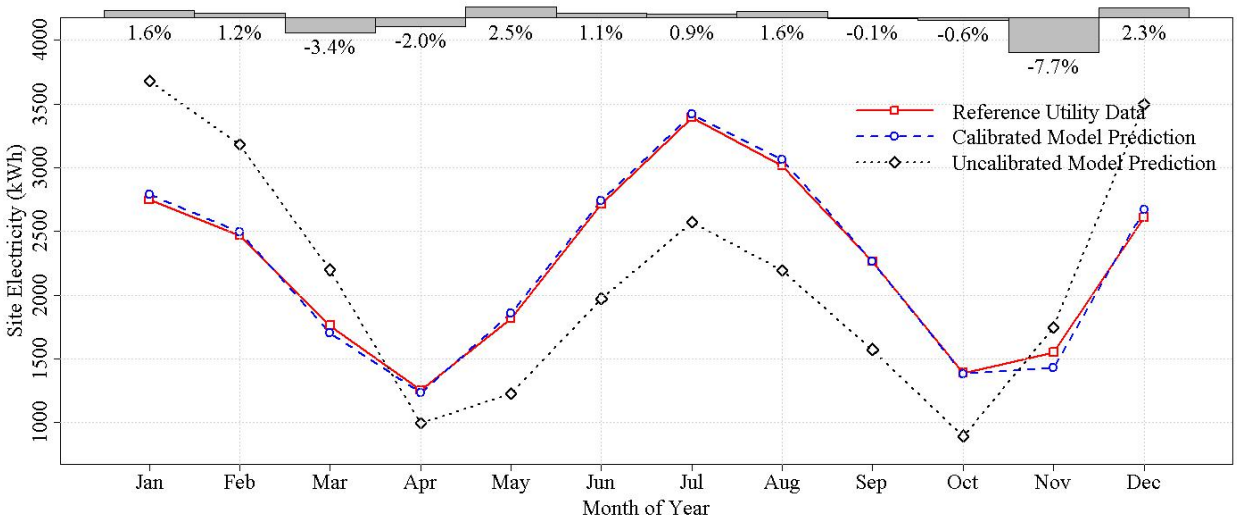


**Figure 77. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.2, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

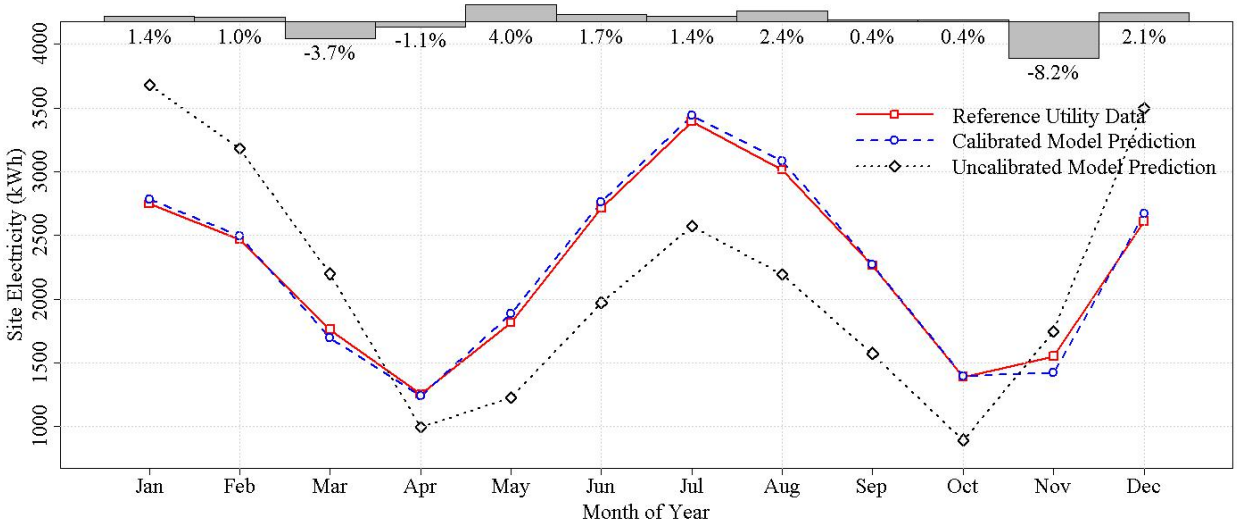


**Figure 78. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.2, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

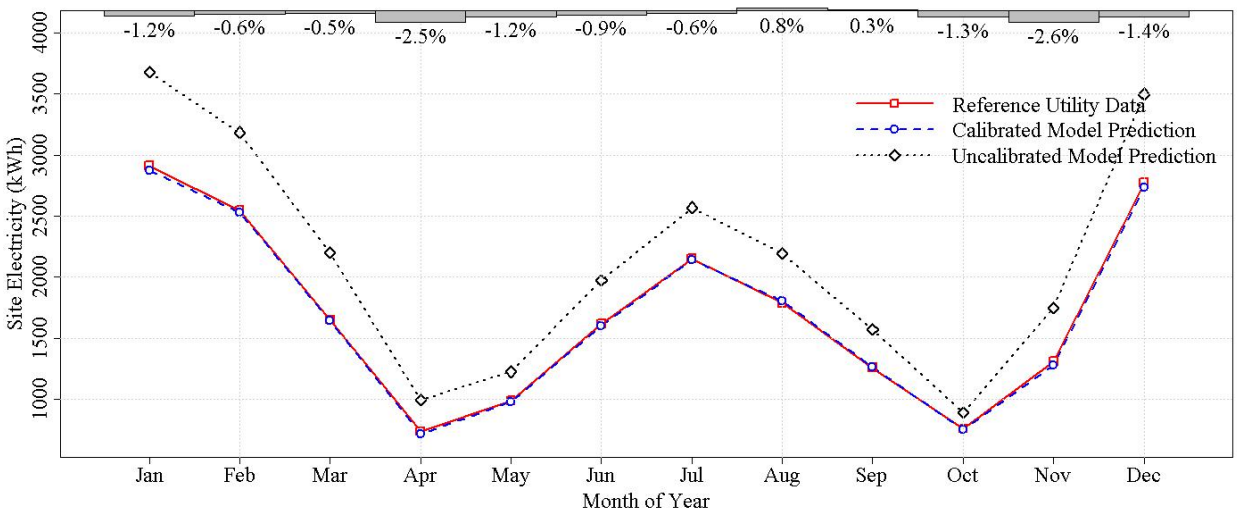


**Figure 79. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.3, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

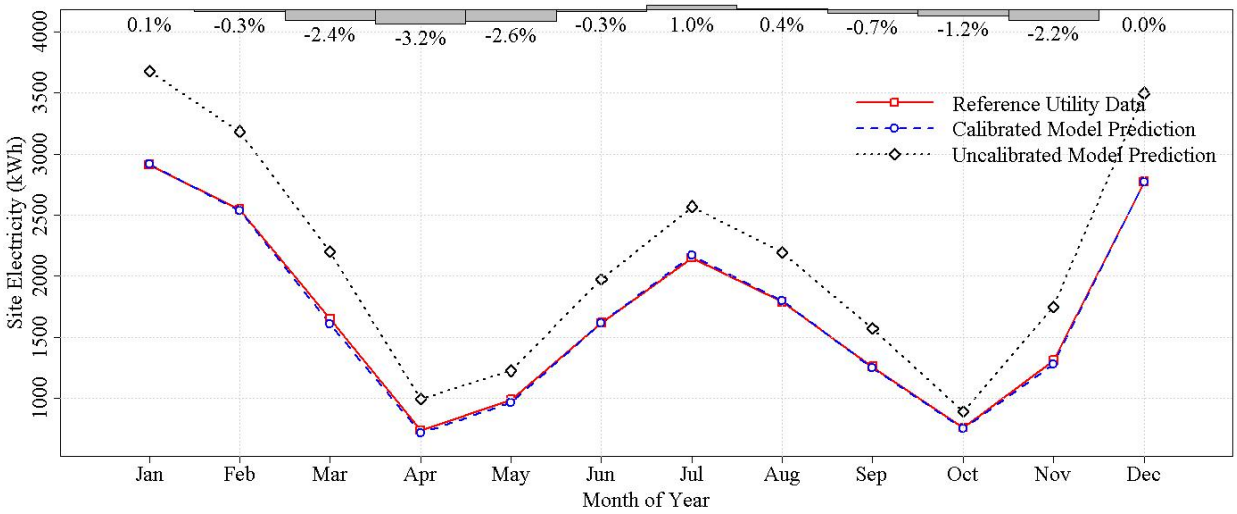


**Figure 80. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.3, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

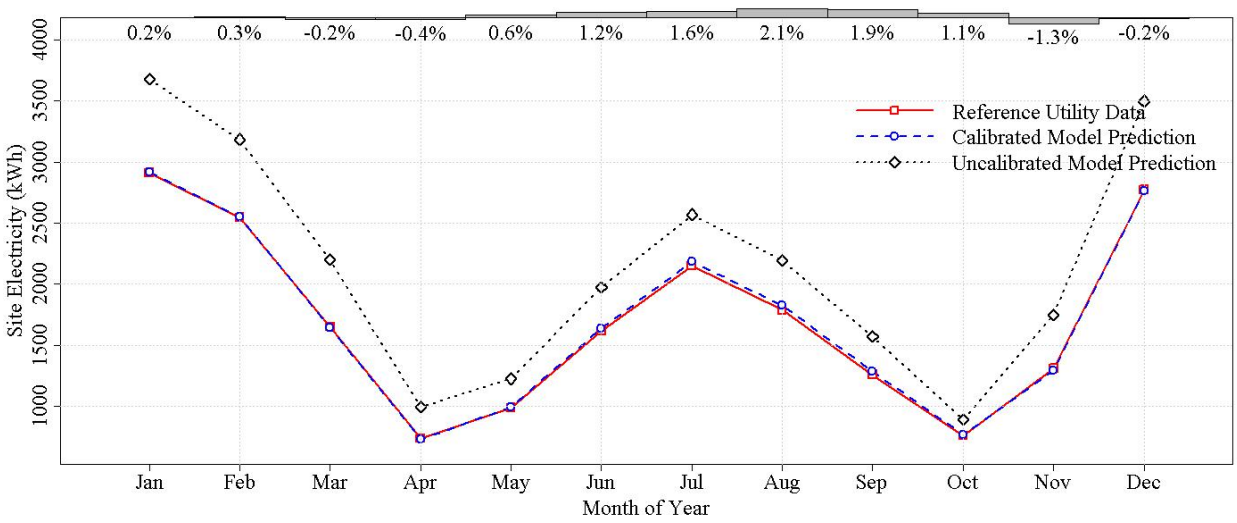


**Figure 81. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.3, Scenario 1. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

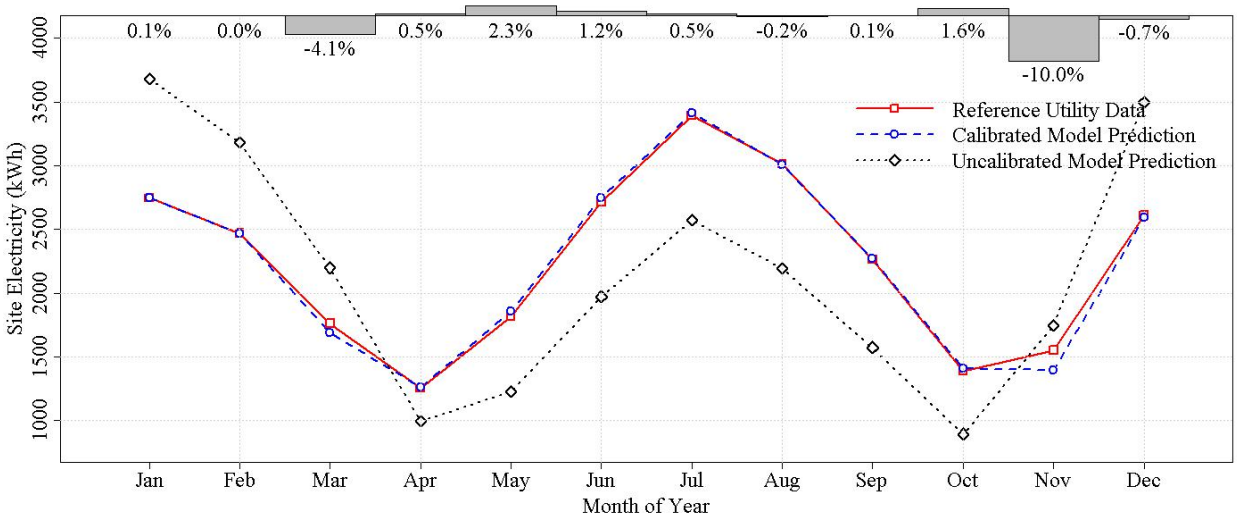


**Figure 82. Illustration of agreement between calibrated model predictions and reference utility data for calibration to monthly data using Procedure 3.3, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

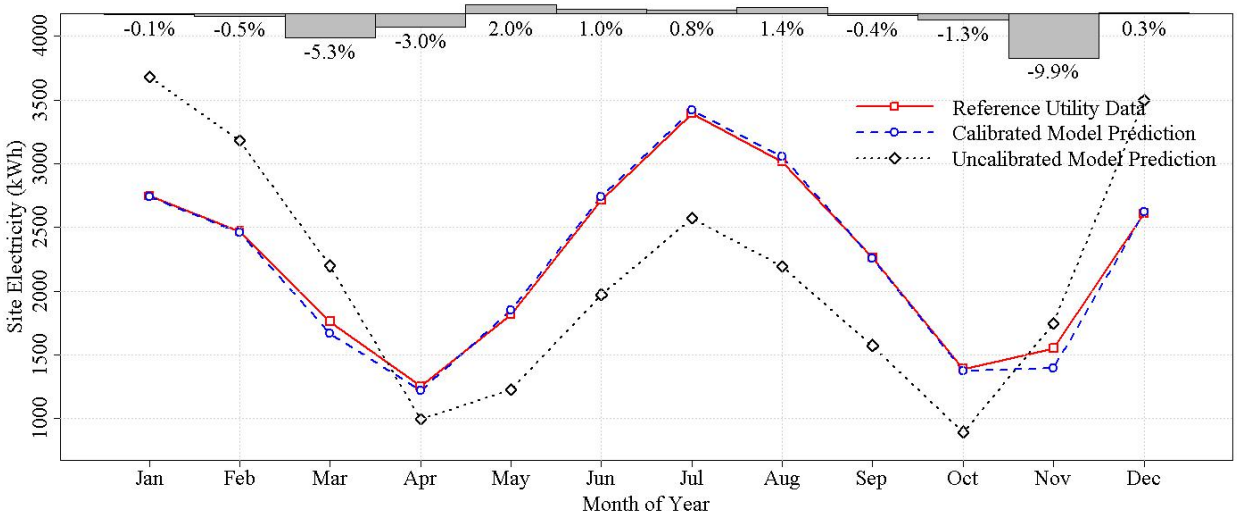


**Figure 83. Illustration of agreement between calibrated model predictions and reference utility data for calibration to daily data using Procedure 3.3, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**

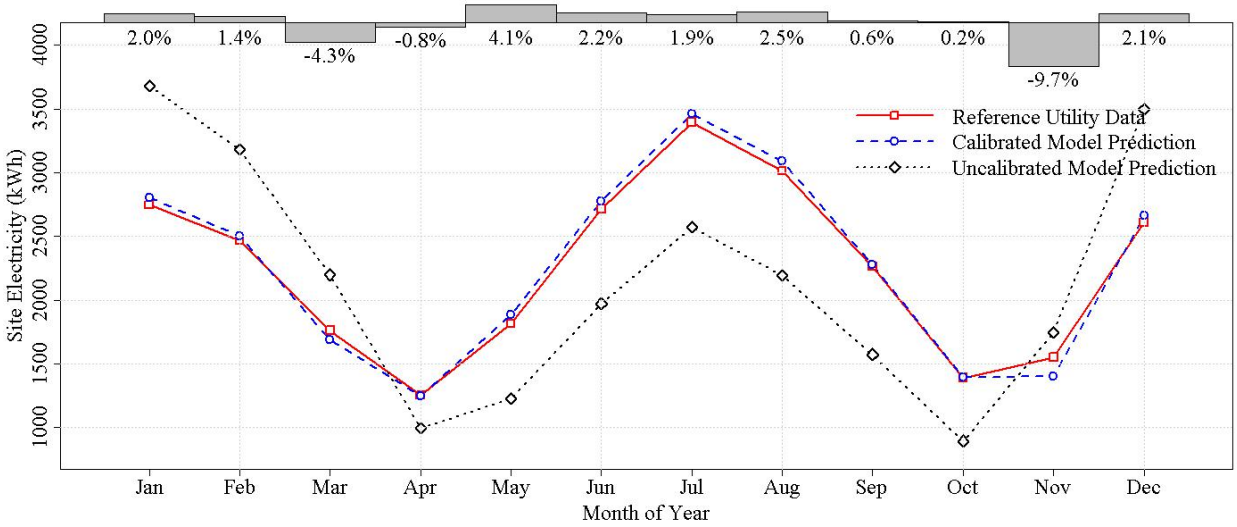


**Figure 84. Illustration of agreement between calibrated model predictions and reference utility data for calibration to hourly data using Procedure 3.3, Scenario 2. Percentages along top are percent errors of the calibrated model predictions relative to the reference utility data.**