How Accurate is Automated, M&V 2.0?

Funded by US Department of Energy Building Technologies Office,
Cody Taylor

Research Team:
Jessica Granderson, Samir Touzani, Claudine Custodio, Samuel Fernandes, Michael Sohn, LBNL

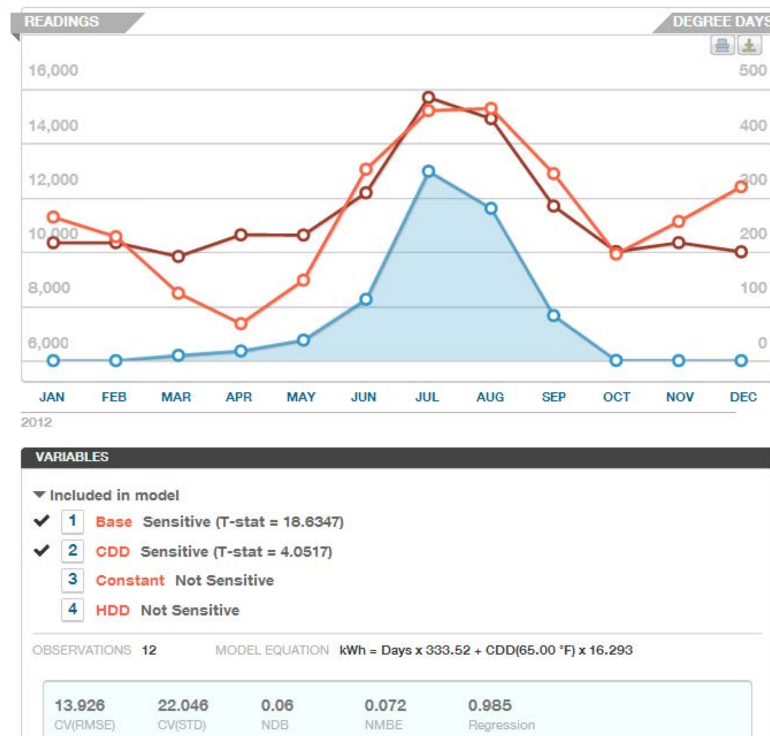
David Jump, QuEST

Outline

- Motivation, background, scope
- Baseline model testing procedure
- Models tested
- Model analysis methodology and accuracy results
- Conclusions, implications
- Ongoing and future work

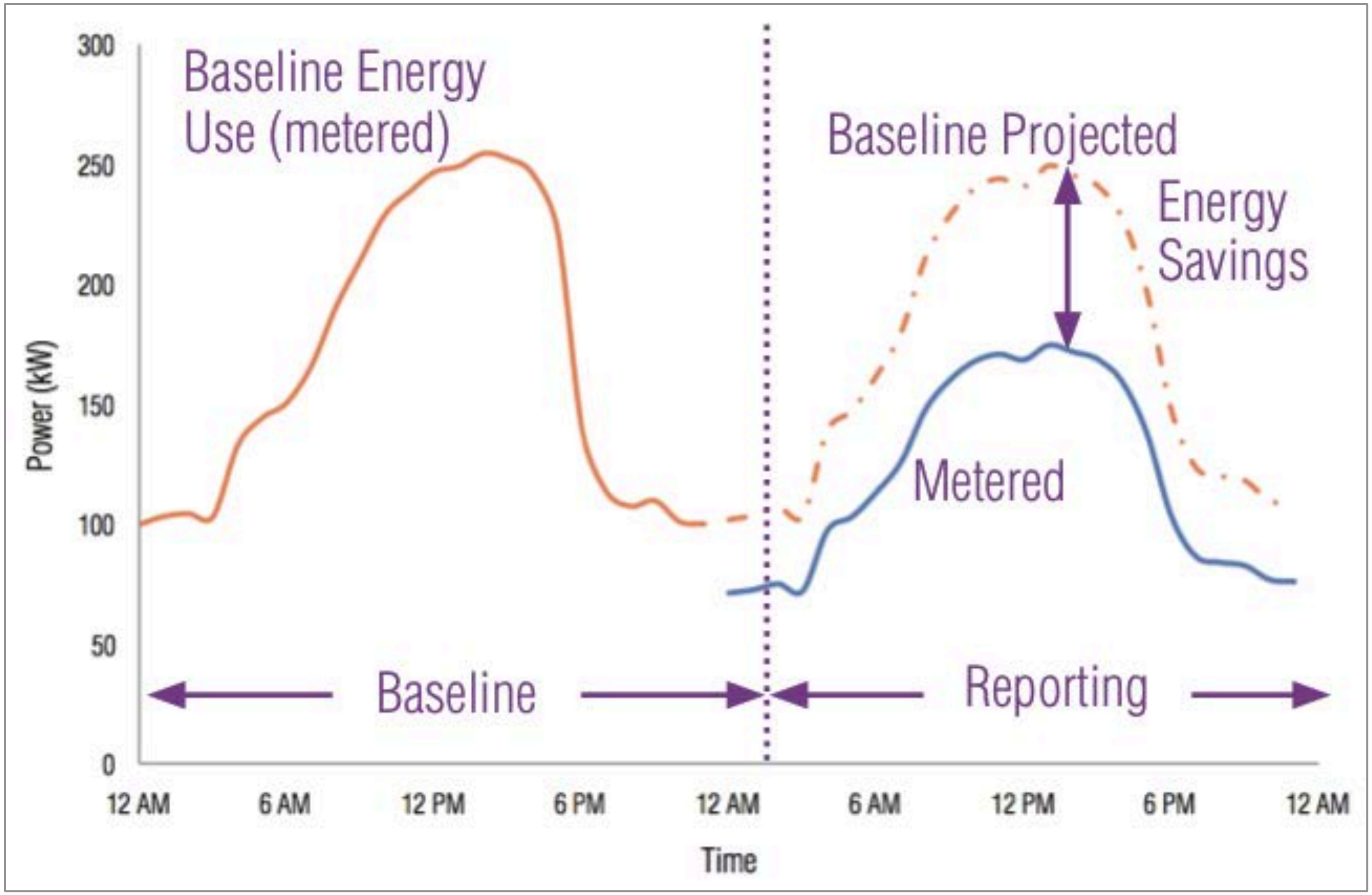
Motivation

- **High level goal**: Enable the industry to harness emerging tools and devices to conduct M&V at dramatically lower cost, with comparable or improved accuracy – M&V 2.0



Software images: Noesis

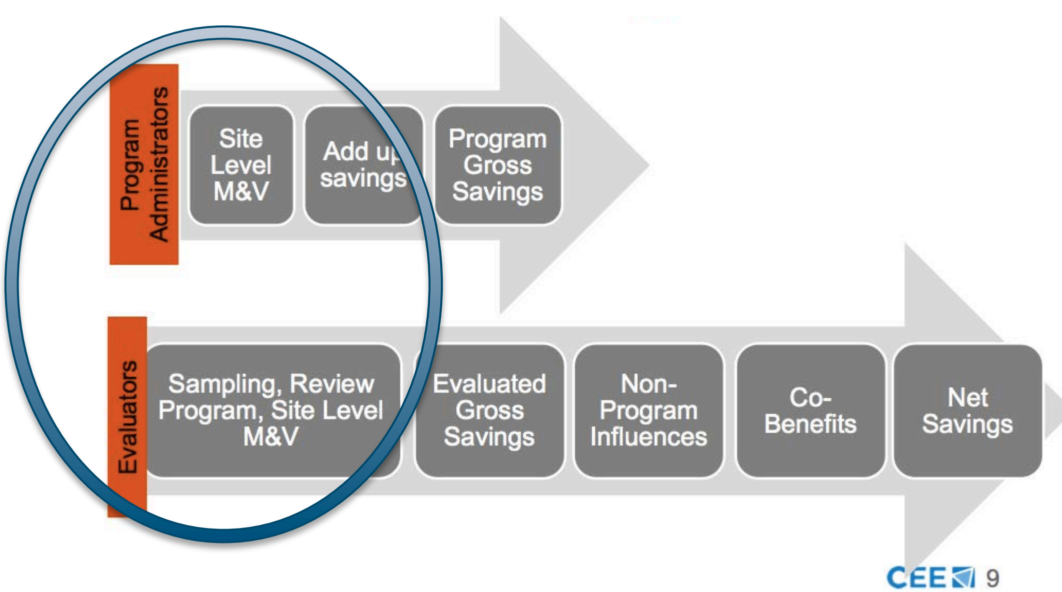
M&V Use Case



Baseline Energy Use (metered)
Baseline Projected
Energy Savings
Metered
Baseline
Reporting
Power (kW)
300
250
200
150
100
50
0
12 AM
6 AM
12 PM
6 PM
12 AM
6 AM
12 PM
6 PM
12 AM
Time

The Relationship Between M&V and EM&V

- M&V quantifies site, project-specific and program level *gross* energy savings
- EM&V evaluates according to policy goals, considering things like attribution, free ridership, code and standard baselines, and incremental costs, to quantify *net* energy savings
- M&V is only one part of EM&V



Industry Trends, Overarching Context

- Increased access to data via smart meters, devices, and analytics tools
- Performance based outcomes, incentives, codes
- Interest in multi-measure whole-building programs that can generate deeper savings
  - traditional M&V approaches can become too complicated or expensive
- Desire to reduce time, cost, complexity of M&V, and EM&V

Automated "M&V 2.0" Is Here

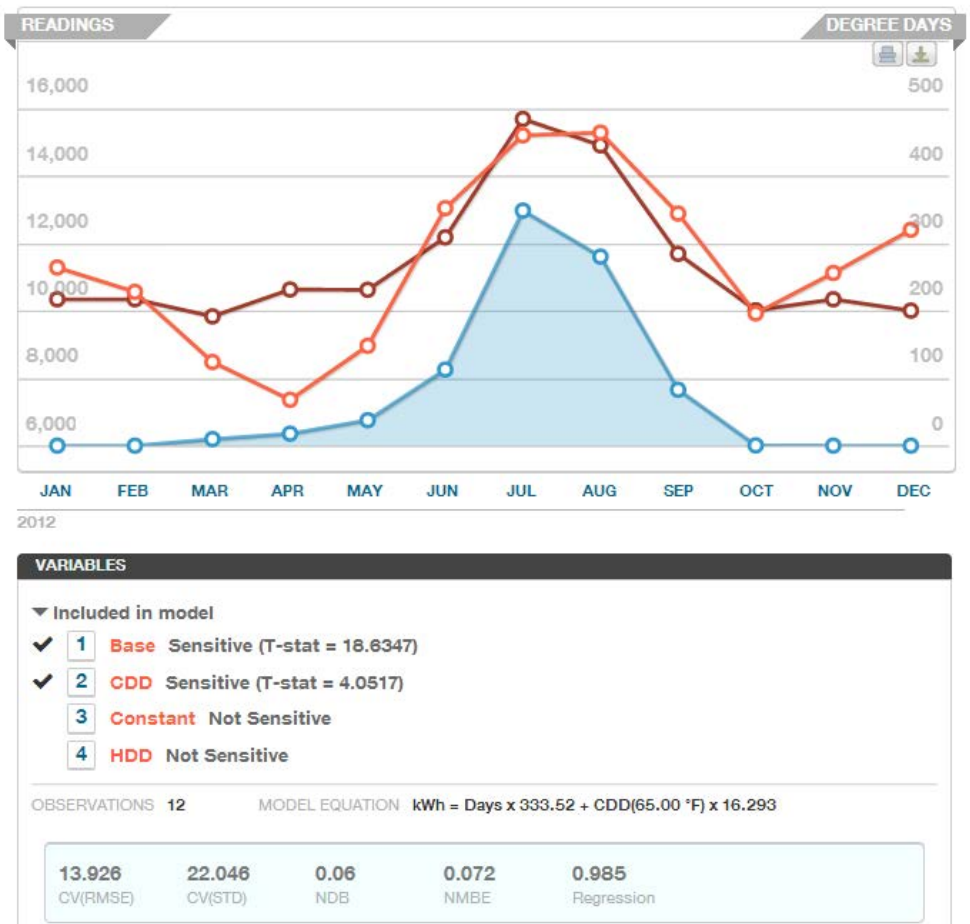
Automated M&V is beginning to be offered in energy management and information systems

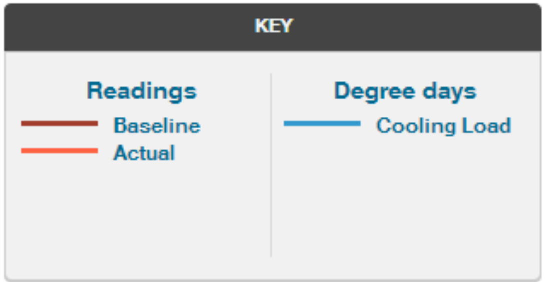
Baselines are automatically created using historic interval meter data (system level or whole-building) and weather data feeds

User enters the date of ECM implementation, savings automatically calculated

Image: Noesis

Automated M&V May Use Interval, Daily, Monthly Data





Example at left from Noesis Energy

While this example uses monthly data; interval data offers the most promise for reducing time, and maintaining accuracy

What Questions Are Being Asked*?

- How can we reduce the time and costs necessary to quantify savings?
- How can I determine whether a given model or commercial tool is robust and accurate?
- How can I compare and contrast proprietary tools and 'open' modeling methods for M&V?
- What repeatable test procedures can be used to evaluate model and tool performance, and which metrics provide critical performance insights?
- Can I use a whole-building approach for my programs and projects?

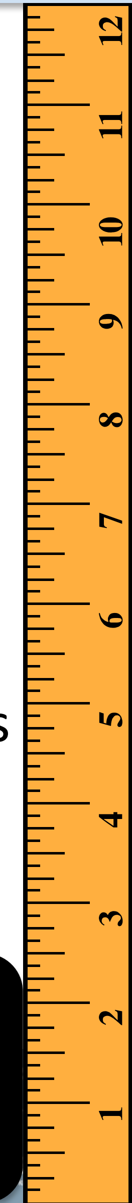
***These are all questions asked before a project is conducted; after a project, we want to know how much was saved, what was the uncertainty, how confident are we in those savings?**

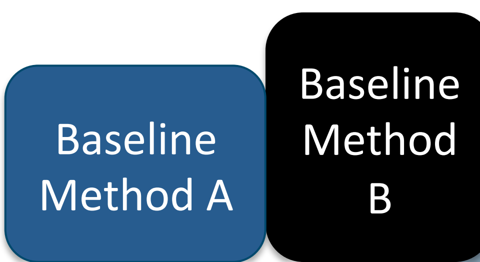
What Have We Done to Address These Questions?

- Developed a testing procedure to quantify baseline model accuracy
- Solicited *new interval* baseline models from industry, tools, and academic communities
- Applied the test procedure to evaluate performance of 10 baseline models
- Worked with advisory group, utility program managers to identify most critical performance metrics for M&V
- Developed conclusions regarding potential for wider adoption of AMI data + analytics for M&V

Value Proposition

- Transparent statistical tests and metrics can be used to evaluate automated baseline methods, tools
  - To determine and compare accuracy of both **proprietary** and '**open**' methods
- Objective performance assessment methodology can provide a win/win/win
  - Allow vendors to retain proprietary IP underlying the algorithms
  - Allow users to gauge performance of the tool/approach
  - Provide evidence, confidence needed for scaled deployment, widespread adoption



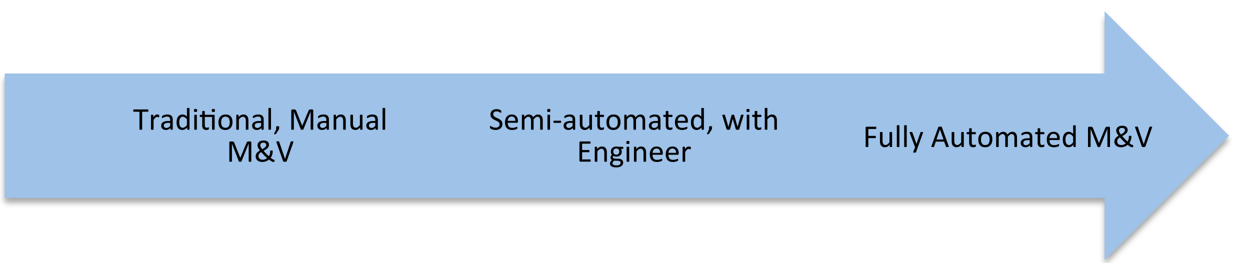


Scope of Work, Results To-Date

- Whole-building avoided energy use calculations, IPMVP Option C, interval data
  - prior work addressed models used for monthly data
  - test procedure can also be applied to models used for DR savings, Option B retrofit isolation, energy anomaly/fault detection

- M&V focus, not yet EM&V, attribution, code baseline, net to gross …

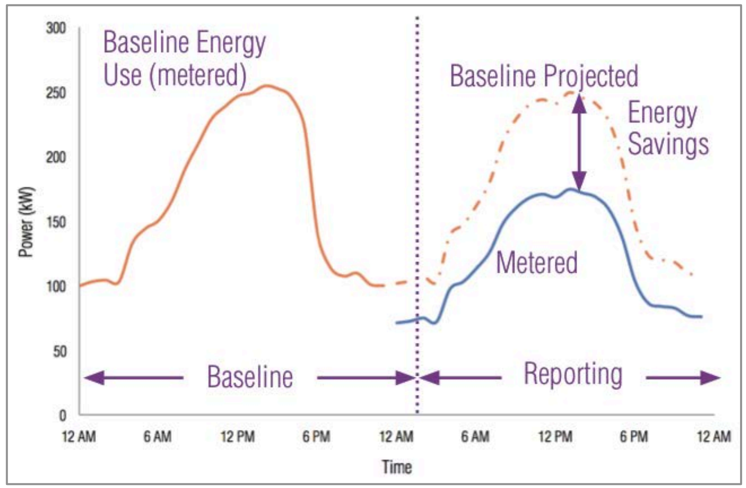
- Streamlining and scaling M&V in practice:
  - Analysis of *fully automated* baseline model capabilities
  - Establishes a *floor of performance* that can be improved by the oversight of engineer, used to reduce costs and time



Baseline Model Testing Procedure

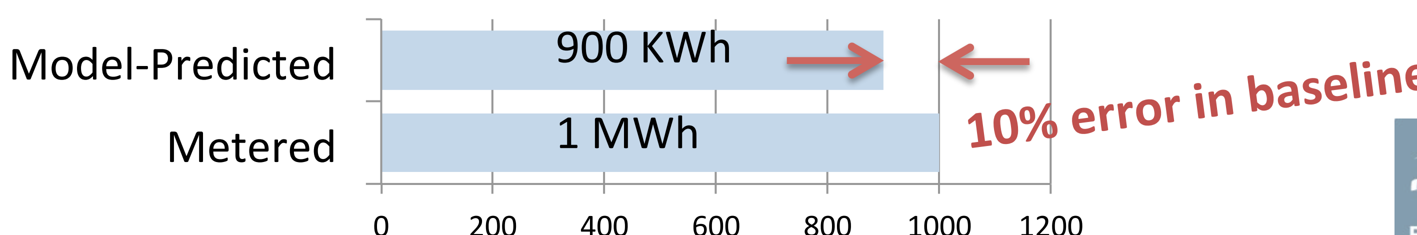
How Accurate is the Baseline Model?

M&V Use Case



Error in reported savings is proportional to error in baseline projection

Error = **% difference** between total **metered** energy use, total **model-predicted use**



Testing Procedure

Baseline Model

Test Data*: Many buildings, metered data

*No efficiency interventions

Model

- **Split data set** into training & prediction period
- **Train the model** by showing data, hiding prediction-period data
- **Generate post-period predictions**

Compare

Compare predicted data to actual data that was 'hidden' from model to quantify error

Repeat for many buildings

Assess

Calculate Performance Metrics, e.g. %Error, R^2, CV(RMSE) …

Illustration of Test Procedure

- Metered building data, no known efficiency measures, interventions (green)
- Model predictions (orange) are compared to actual meter data (green)
- Repeat for hundreds of buildings to understand overall model accuracy, fraction of buildings for which errors are small vs. large
- Repeat for many models to understand relative model performance
- Shorten training/prediction periods to understand impact on accuracy when length of measure pre-, post- period is changed

BERKELEY LAB
Lawrence Berkeley National Laboratory

Questions On Motivation, Scope, or Testing Procedure?

(Please type into the webex conference interface)

Interval Data Baseline Models Tested

Description of Interval Data Models Using Time, Outside Air Temperature

- M1. *Combination principle component analysis and bin modeling,* developed by Buildings Alive Pty. Ltd. of Sydney Australia.
- M2. *Combination Random Forest, Extra-Trees (extremely randomized trees) and Mean Week,* developed by Paul Raftery and Tyler Hoyt at the Center for the Built Environment, University of California, Berkeley.
- M3. *Advanced regression including a term for drift,* developed by Gridium Inc.
- M4. *Mean Week*[1] - predictions depend on day and time only. For example, the prediction for Tuesday at 3 PM is the average of all of the data for Tuesdays at 3 PM.
- M5. *Time-of-Week-and-Temperature*[2] - the predicted load is a sum of (1) a "time of week effect, and (2) a piecewise-continuous effect of temperature. The temperature effect is estimated separately for periods of the day with high and low load, to capture different temperature slopes for occupied and unoccupied building modes.

1. Granderson, et al. Assessment of automated measurement and verification (M&V) methods. Lawrence Berkeley National Laboratory, July 2015, LBNL#-187225.
2. Mathieu, J.L., P.N. Price, S. Kiliccote, & M.A. Piette. Quantifying Changes in Building Electricity Use, With Application to Demand Response. SmartGrid, IEEE Transactions, vol. 2, Issue 3, pp.507-518. August 2011.

Description of Interval Data Models

- M6. *Weighted Time-of-Week-and-Temperature*[3] - the *Time-of-Week and-Temperature* model with a weighting factor to give more statistical weight to days that are nearby to the day being predicted.

- M7. *Ensemble approach combining nearest neighbors and a generalized linear model*, developed by Lucid Design Group.

- M8. *Combination Multivariate Adaptive Regression Splines (MARS) and other advanced regression*

- M9. *Combination bin modeling and other advanced regression,* developed by Performance Systems Development of New York, LLC.

- M10. *Nearest neighbor advanced regression*

3. Piette, M.A., Brown R.E., Price P.N., Page, J., Granderson, J., Riess, D., et al. (2013). Automated measurement and signaling systems for the transactional network. Lawrence Berkeley National Laboratory, December 2013. LBNL-6611E.

Model Run Times*

*** Based on 12 month training period, run on iMac 3.5GHz Quad Core i7**

Model	Less than 5 min	More than 30 min	More than 1 hour
M1	X		
M2	X		
M3	X		
M4	X		
M5	X		
M6	X		
M7		X	
M8			X
M9	X		
M10			X

Analysis Methodology

Summary of Analyses

- Predictions generated for 12-month prediction/post period
  - Using 12-, 9-, 6-, and 3-month training/pre periods
- Standard practice and guidance for whole-building M&V is 12 months pre/post

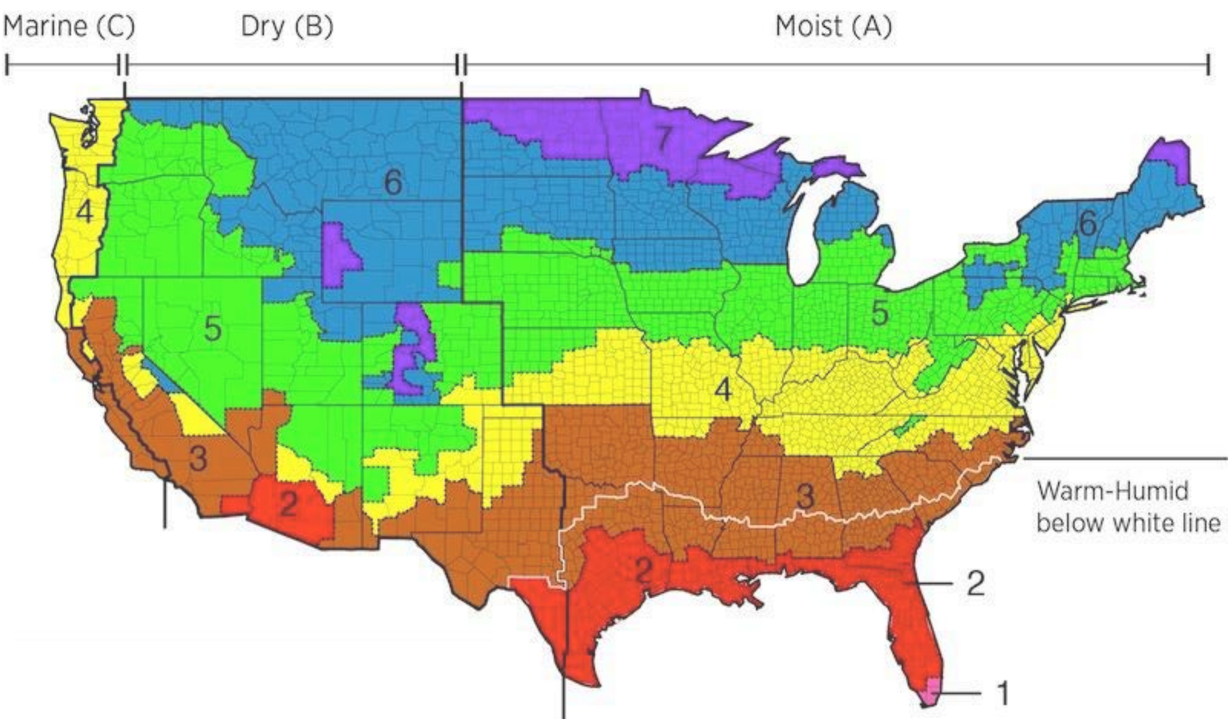
Characterization of Test Dataset

- 537 commercial buildings
  - 15-minute electric load data
  - Outside air temperature based on zip code

- No *known* efficiency interventions, significant changes in operations, occupancy

Characterization of Test Dataset

ASHRAE Climate Zone	1 (Very Hot)	2 (Hot)	3 (Warm)	4 (Mixed)	5 (Cool)	6 (Cold)	7 (Very Cold)
Number of Buildings analyzed to-date	1	15	277	237	5	1	1



US Map with ASHRAE-IECC Climate Zones

BERKELEY LAB
Lawrence Berkeley National Laboratory

Performance Metrics of Focus

- Many possible goodness of fit metrics to choose from
  - different insights into performance accuracy
  - but also high degree of overlap

- Analyzing too many metrics makes it hard to draw conclusions about model performance

- ~20 representatives from efficiency program management evaluation, implementation voted on top two metrics of choice for M&V use case

- There actually was strong consensus!

Two Primary Performance Metrics

Normalized mean bias error, $NMBE = \frac{\frac{1}{N}\sum_i^N (y_i - \hat{y}_i)}{\bar{y}} \times 100$

CV of the root mean squared error, $CV(RMSE) = \frac{\sqrt{\frac{1}{N}\sum_i^N (y_i - \hat{y}_i)^2}}{\bar{y}} \times 100$

- Provide a complement in understanding model performance
- NMBE is total percent difference between predicted and actual energy use
- CV(RMSE) indicates model's ability to predict the overall load shape
  - familiar to practitioners
  - prominent in resources such as ASHRAE Guideline 14

Questions On Models, Data, or Metrics?

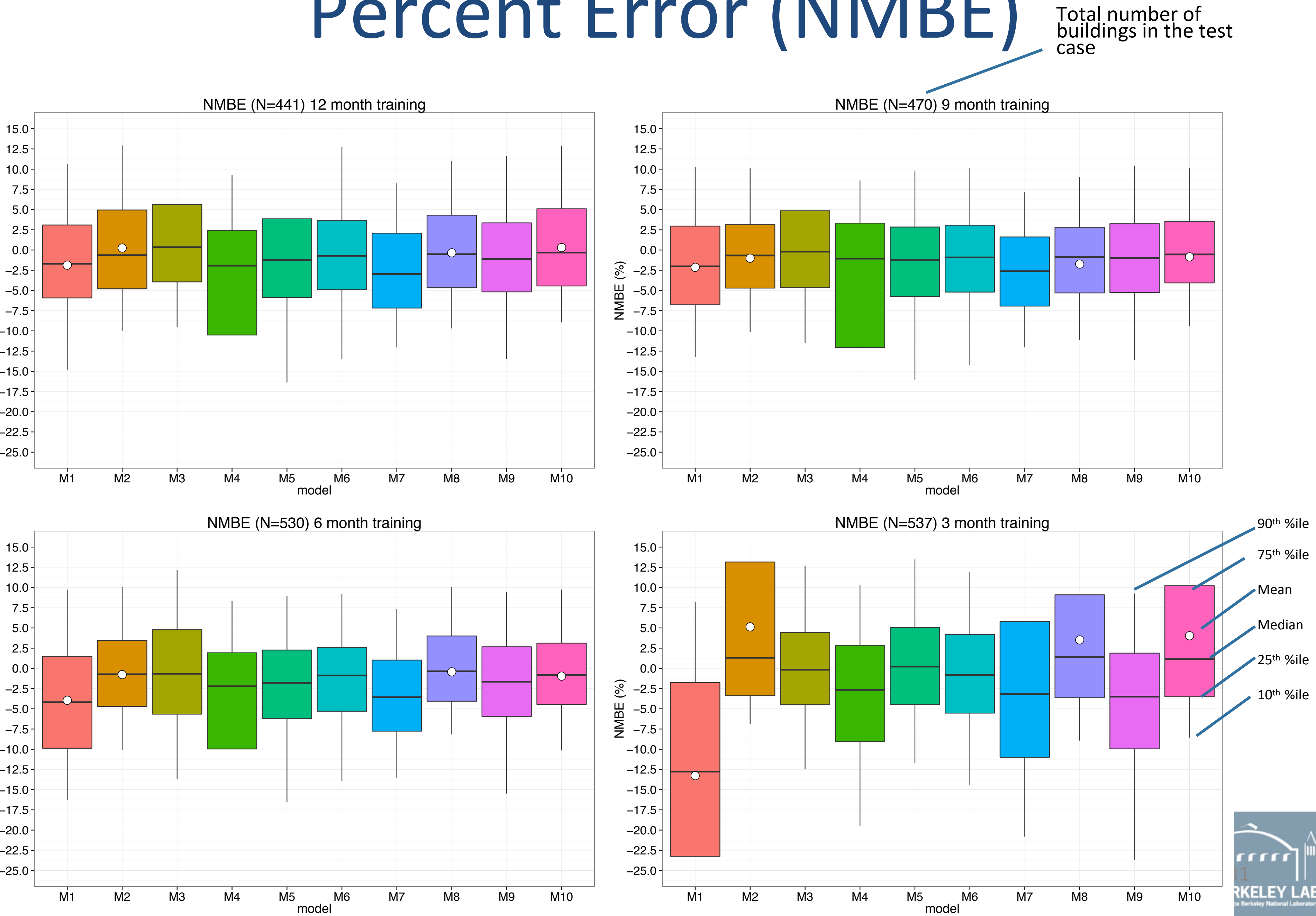
(Please type into the webex conference interface)

Model Accuracy Results

Form of the Results

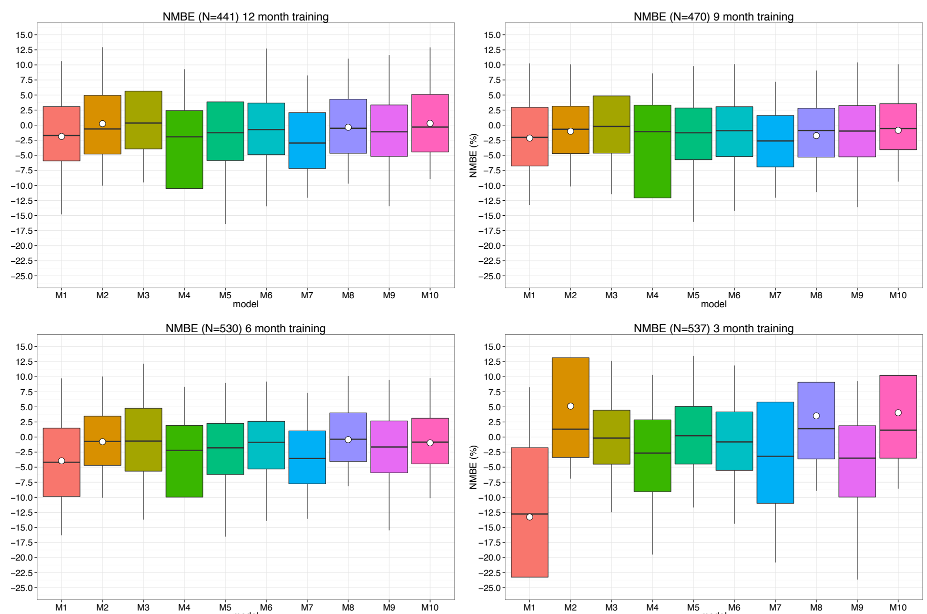
- To get a sense of *general, overall model accuracy*, we look at prediction errors across *many* buildings
- Some buildings are predicted with very little error, some buildings with higher error
- So we consider distributions/percentiles of errors, as in standardized test scores
  - Median is the midpoint, or "average": errors for 50% of the buildings are higher, and for 50% of the buildings are lower
  - Half of the population falls between the 25th and 75th percentile

Percent Error (NMBE)



What Do These Distributions Of Percent Errors Tell Us?

- Differences between models are mostly small
- Across all models, for 12-month training
  - *Average* median percent error ~-1.2%
  - *Range* of median errors is ~-3% to 0.4%
- No clear "winner" across models and training periods – all are good, especially for the case of 12-months training

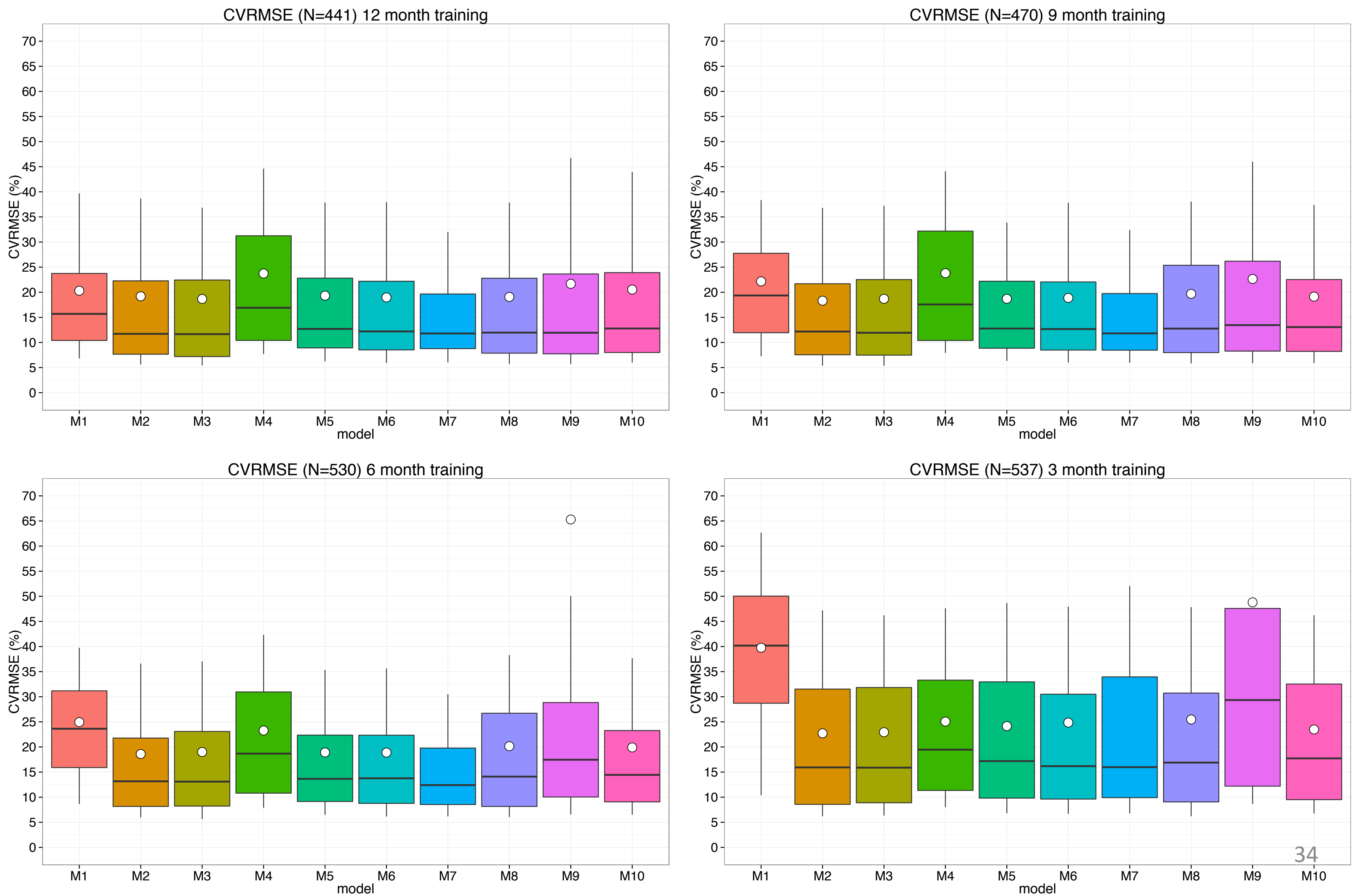


What Happens As We Shorten the Training Period?

Model	Model Training Period			
	12 months	9 months	6 months	3 months
M1	-1.7	-2.02	-4.19	-12.77
M2	-0.63	-0.68	-0.73	1.3
M3	0.35	-0.2	-0.67	-0.17
M4	-1.93	-1.07	-2.22	-2.66
M5	-1.25	-1.26	-1.79	0.21
M6	-0.73	-0.92	-0.88	-0.81
M7	-2.97	-2.62	-3.57	-3.19
M8	-0.51	-0.88	-0.36	1.38
M9	-1.1	-0.98	-1.65	-3.5
M10	-0.32	-0.55	-0.84	1.14
Avg. of Absolute Median Values	1.15	1.12	1.69	2.71

- Difference in errors between 12- and 9-months training is small
- For (some) models, accuracy begins to degrade when training period shortened to 6 months, more when shortened to 3 months
- Some models are more robust to shorter training periods

CV(RMSE), Daily Energy Totals



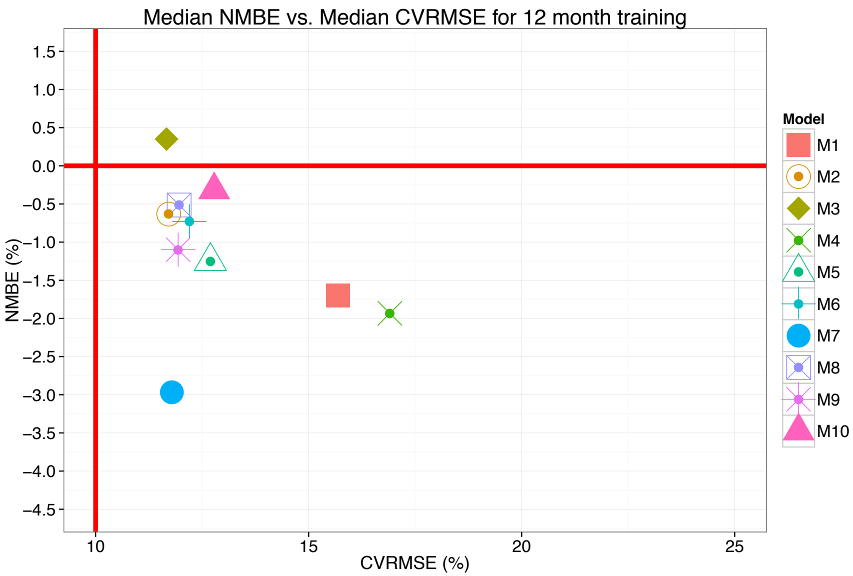
CV(RMSE), ASHRAE Guideline 14

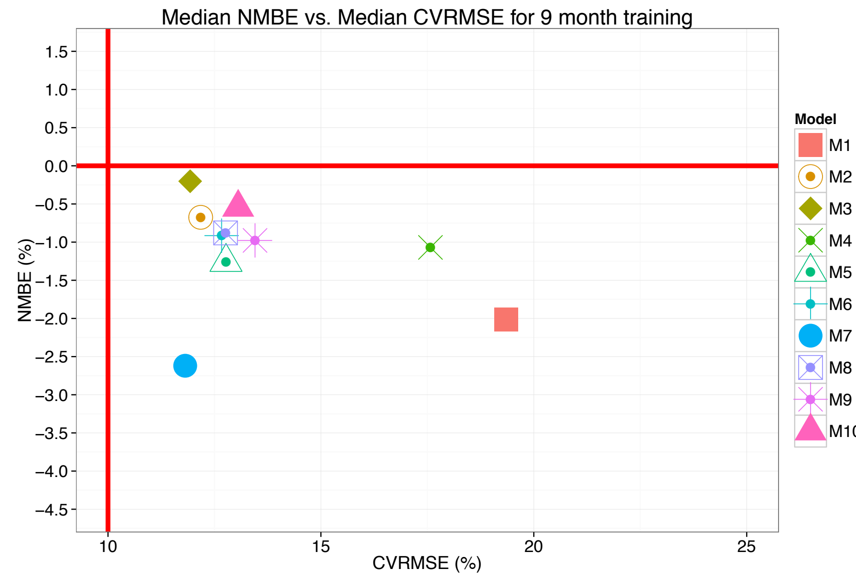
- ASHRAE Guideline 14 is the industry's reference on minimum acceptable levels of performance for measurement-based energy and demand savings in commercial transactions
- Models analyzed are likely to meet the Guideline 14 requirements
- Guideline 14 specifies CV(RMSE) during the *training*[4] period, should be <25%[5]
- In this study
  - Median CV(RMSE) for daily energy totals was <25% for every model, when twelve months of training data were used
  - This was true even when only 6 months of training data were used

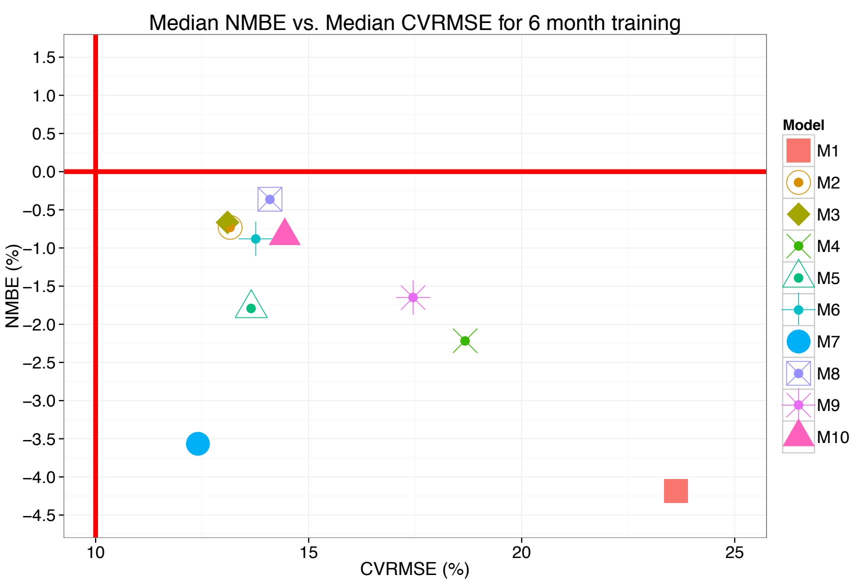
4. For a case of 12-month post/prediction data, where no uncertainty analysis is to be conducted
5. This study computed CV(RMSE) during the prediction period – which is expected to be even higher than that for the training period.

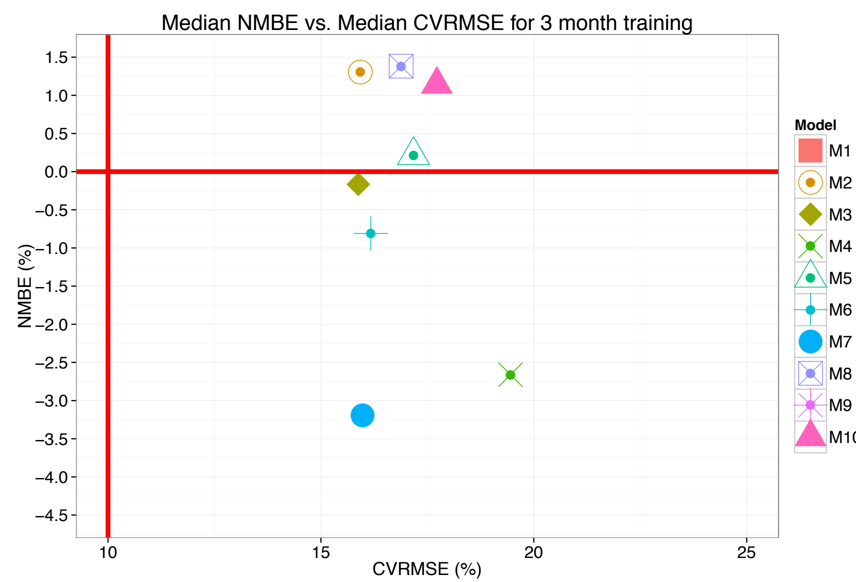
BERKELEY LAB
Lawrence Berkeley National Laboratory

Both Metrics Considered Simultaneously: Is There a Clear "Winner"?









Conclusions, Implications

Key Takeaways

- AMI data and interval data models/tools hold great promise to scale adoption of whole-building measured savings calculations
  - Reducing time and costs, improving or maintaining accuracy
- Errors in predicting energy are on the order of a couple of percent for many buildings and many models
  - This is the floor of performance from the *fully automated case*, with no 'non-routine' adjustments from an engineer
  - Oversight of an engineer could improve accuracy even further
- 12 months pre/post data may not always be required for accurate whole-building M&V
- Models effectively meet ASHRAE guidelines in most cases

How Can You Use These Results: A Call to Action!

- Increase the use of these M&V methods
  - This study provides objective evidence that M&V models/tools are generally robust
  - This study provides accuracy insights that are not generally possible for deemed or stipulated savings
- Apply this test procedure and metrics to evaluate new tools/models
  - Use these results as a comparative benchmark
  - Consider accuracy and uncertainty requirements -- how good is good enough?
  - Pre-screen, or target the buildings in a population that are most predictable with the smallest errors
- Vet project-specific M&V plans
  - Use these findings to estimate expected ranges of uncertainty and confidence in reported savings
  - We can now be more precise than general guidelines that whole building M&V requires 12 months pre/post data, and 10% savings or greater

BERKELEY LAB
Lawrence Berkeley National Laboratory

Concluding Thoughts

- Growing availability of intelligent analytics tools, and metered building energy data present a tremendous opportunity for our industry
  - Leading-edge adopters already making powerful use of the technology

- The same technologies that drive significant savings also promise the ability to verify those savings
  - A win for the scaled adoption of cost-effective energy efficiency
  - Transparency and evidence that savings are achieved, value is delivered
  - Persistence of savings through continuous data-driven energy management

Questions On Results or Conclusions?

(Please type into the webex conference interface)

Ongoing and Future Work

Ongoing Work

- Demonstration of automated approaches with utilities/programs, and implementers or analytics vendors
  - Use data from buildings that have participated in whole-building (preferably) programs or pilots
  - Apply automated M&V alongside whatever M&V plan was/is already in place
  - Quantify savings with uncertainty and confidence
  - Publish and case studies on effectiveness

We are currently seeking utility/program and implementer or vendor partners who are interested in collaborating in this work. Please contact JGranderson@lbl.gov if you are interested in exploring this opportunity.

Future Work

- **Continued engagement of evaluator, program manager, and implementer communities**
  - Collectively define uncertainty and confidence requirements for reporting gross energy savings
  - Provide critical information necessary for adoption

- Transfer of test procedure to stakeholders, and industry bodies – interest from CA state, ASHRAE, utility community

Discussion?

(Please use the webex conference interface to raise your 'hand')

Acknowledgement

Technical Advisory Group

Mangesh Basarkar	Pacific Gas and Electric
Erik Boyer	Bonneville Power Administration
Kevin Cooney	Navigant Consulting, Inc.
Charles Culp	Texas A&M University
Collin Elliot	Itron
Daniel Harris	Independent consultant
Jennifer Huckett	Cadmus Group
Joshua Keeling	Portland Gas and Electric
Hannah Kramer	CLEAResult
David Kresta	Northwest Energy Efficiency Alliance
Ryan Lambert	Puget Sound Energy
Peter Miller	Natural Resources Defense Council
Jeffrey Perkins	Energy & Resource Solutions
David Rodenhizer	City of Seattle
Justin Spencer	Navigant Consulting, Inc.
Andrew Stryker	DNV GL

Model Developers

Buildings Alive, Pty. Ltd. Sydney Australia

Gridium, Inc.

Lucid Design Group

Paul Raftery and Tyler Hoyt, Center for the Built Environment, UC Berkeley

Performance Systems Development of New York, LLC

Contributors of Data

US Department of Energy, Building Technologies Office, Cody Taylor

Thank You!

For more information please contact Jessica Granderson
JGranderson@lbl.gov, 510.486.6792

Appendix: Scenario Uncertainty Analyses to Vet M&V Plans

Vetting Project-Specific M&V Plans

- Although not common today, uncertainty analysis supports evaluation and reduction of project and investment risk
  - E.g., savings = 12%, +/- 2%, with 80% confidence

- Before a project or program is launched, the M&V plan can be evaluated to understand expected ranges of uncertainties

Scenario Analysis

- Suppose your whole-building program is targeting 10-20% savings
- And you require 68% confidence in reporting those savings (ASHRAE levels)
- You want to know, before conducting the projects, whether your planned approach has any chance of meeting your requirements
  - You would like to use the time of week and temp. regression model
  - You want to use 12 months of 15-minute pre/post data

Scenario Analysis Using Results from Real Buildings

- For a set of CA buildings and the model of interest* CV(RMSE) are:
  - 0.151 in the median building
  - 0.087 in the best 20% of buildings
  - 0.068 in the best 10% of buildings

- Apply ASHRAE Guideline 14 (2002) equation for fractional savings uncertainty
  - Note that the 2014 version of Guideline 14 presents a modified formulation of this concept

* 15-min model and 15-min predictions; predictions rolled into daily energy totals from which CV(RMSE) was computed

BERKELEY LAB
Lawrence Berkeley National Laboratory

68% Confidence, Uncertainty Ranges

CV(RMSE)	Fractional Savings Uncertainties	
	10%	20%
0.151 for median building	0.100	0.050
0.087 for best 20% of buildings	0.058	0.029

- Fractional energy savings uncertainty in best 20% of buildings would range 0.058 to 0.029
  - 10% savings +/-3% and 20% savings +/- 1.4%
  - At 68% confidence

- Fractional energy savings uncertainty in a median building would range 0.1 to 0.05
  - 10% savings +/-5% and 20% savings +/- 2.5%
  - At 68% confidence

What If You Wanted to Shorten the Post Period to 6 Months?

CV(RMSE)	F	
	10%	20%
0.151 for median building	0.141	0.071
0.087 for best 20% of buildings	0.081	0.041

- Fractional energy savings uncertainty in best 20% of buildings would range 0.081 to 0.041
  - 10% savings +/-4% and 20% savings +/- 2%
  - At 68% confidence

- Fractional energy savings uncertainty in a median building would range 0.141 to 0.071
  - 10% savings +/-7% and 20% savings +/- 3.5%
  - At 68% confidence

BERKELEY LAB
Lawrence Berkeley National Laboratory