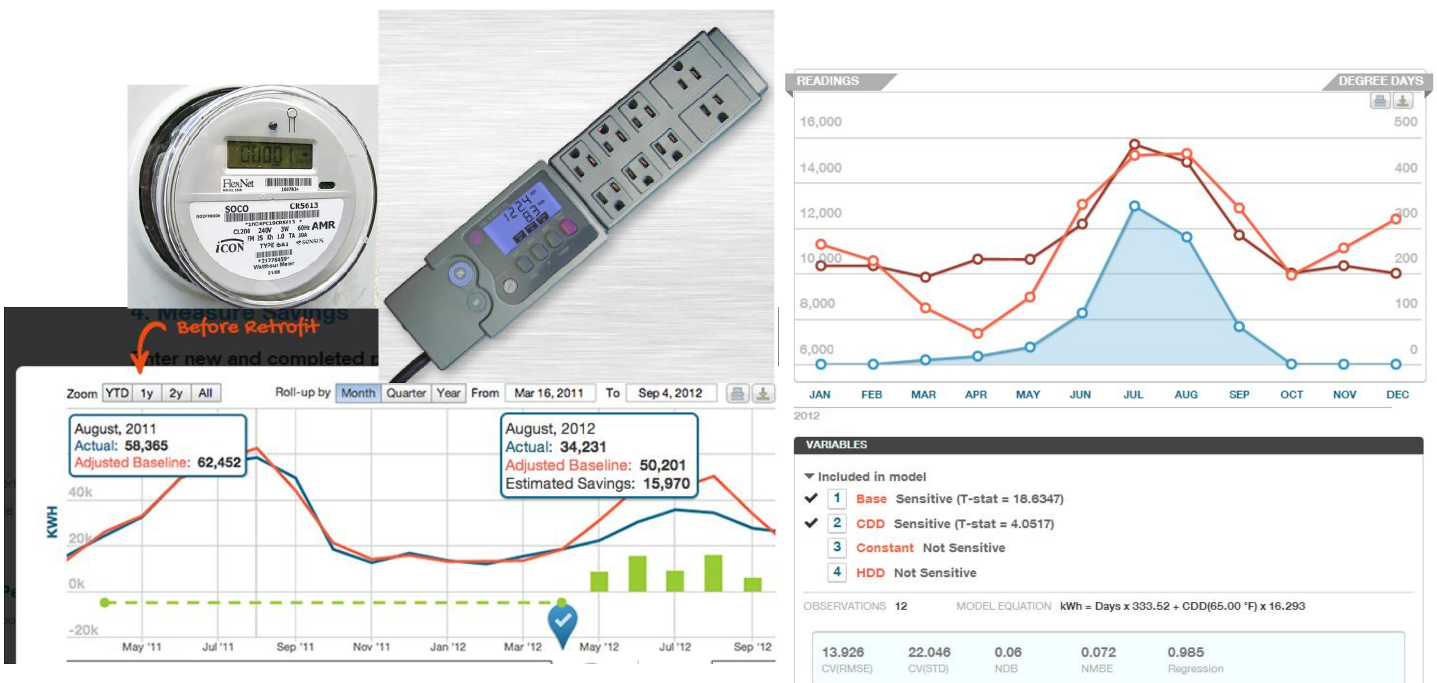
M&V Shootout: Testing the Performance of New Energy Baseline Models

Jessica Granderson

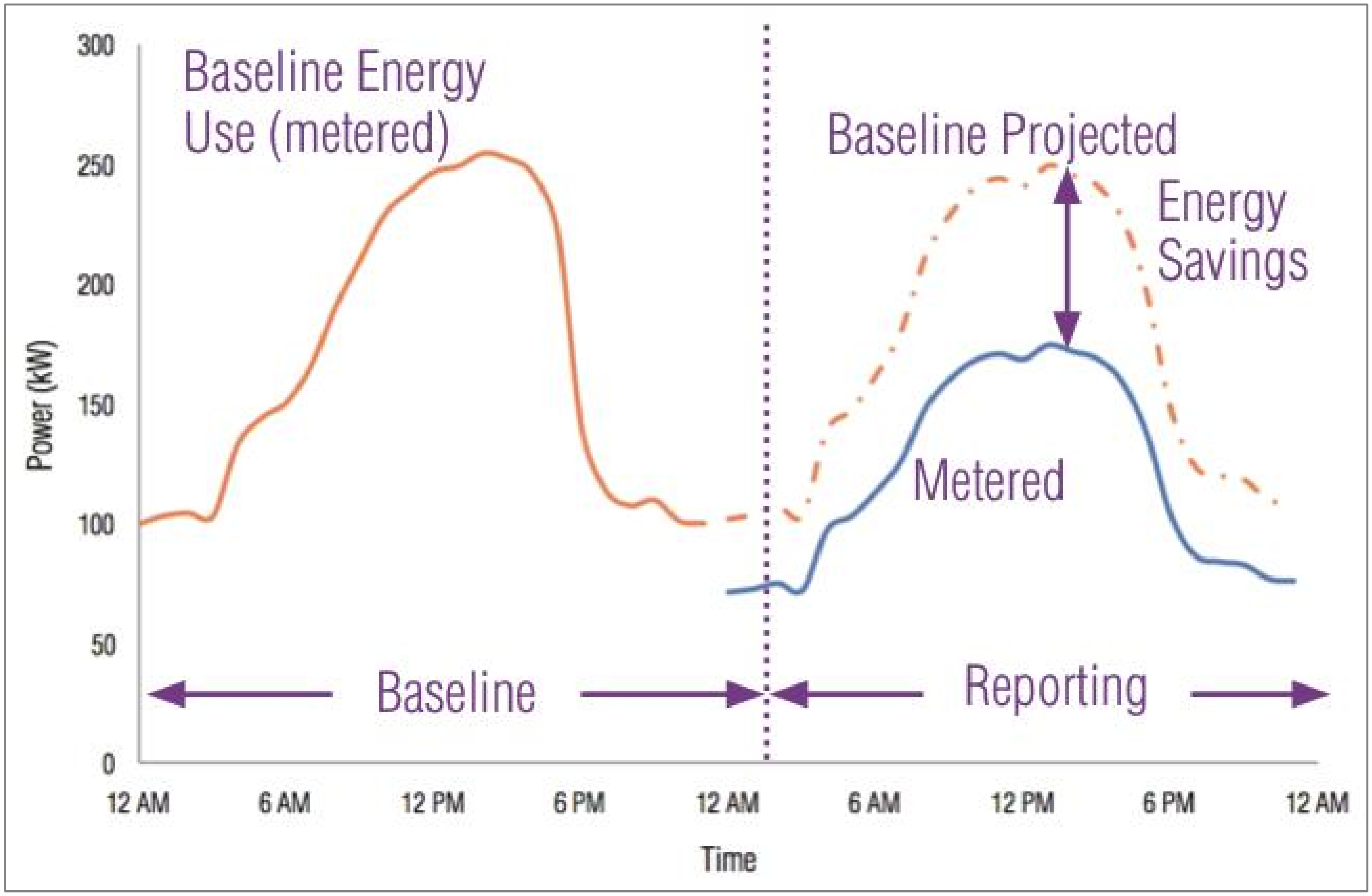
2015 IEPEC Conference — Long Beach, California

Motivation

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



Avoided Energy Use, Existing Use Baseline



Automated M&V 2.0 Is Here

- Offered in energy management and information systems
- Baselines 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

What Questions Are Being Asked*?

- How can we reduce the time and costs necessary to quantify savings?
- How can we know if a model or commercial tool is robust and accurate?
- How can we compare and contrast proprietary tools and ‘open’ methods?
- What test procedures can be used to evaluate model and tool performance, and which metrics are most important?
- Can I use a whole-building approach for my programs and projects?

***All are 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 model performance
- With advisory group **identified most critical performance metrics for M&V**
- **Developed conclusions regarding potential** for wider adoption of AMI data + analytics for M&V

Baseline Model Testing Procedure

Baseline Model

Test Data*: Many buildings, metered data

*No efficiency interventions

Model

- **Split data set** into hypothetical training & prediction period
- **Train the model** training data, hide the rest
- **Generate post-period predictions**

Compare

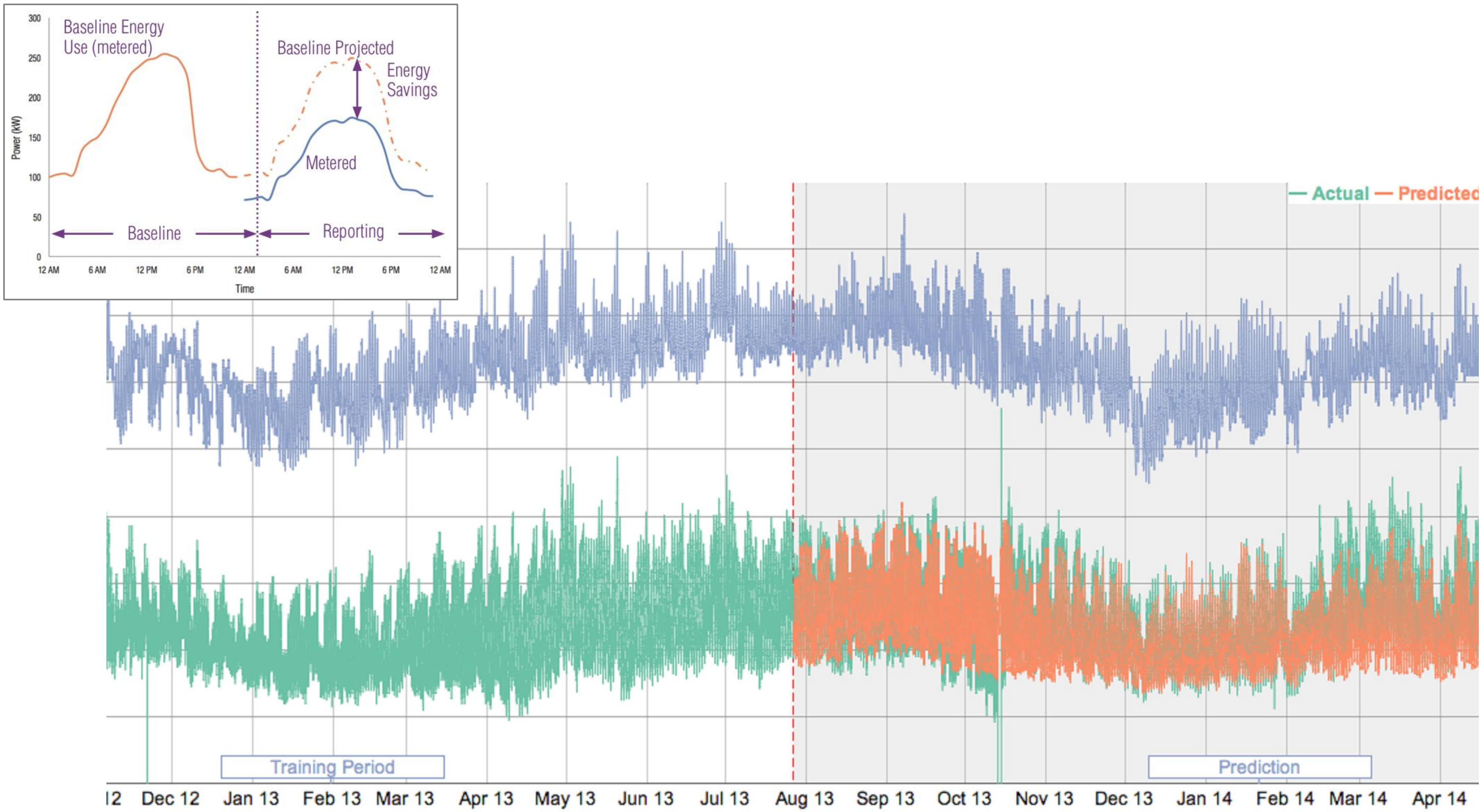
Compare predicted data to actual data that was ‘hidden’ from model to quantify error

Repeat for many buildings

Assess

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

Illustration of Test Procedure



Scope of Analyses

- Whole-building avoided energy use calculations, IPMVP Option C, interval electricity data
  - 12 mo. prediction/'post' period
  - 12 mo. And shorter training/'pre' period

- M&V, not other elements of EM&V

- Streamlining and scaling M&V
  - Analysis of *fully automated* baseline model capabilities
  - Establishes *floor of performance* that can be improved by the oversight of an engineer

Models Tested

- 4 "open", 6 proprietary

Buildings Alive
Gridium
Lucid
Performance Systems Development
UCB Center for Built Environment

- Mix of mathematical approaches
  - Nearest neighbor
  - Advanced regressions
  - Principle component analysis
  - Hybrid, combined methods
  - Others

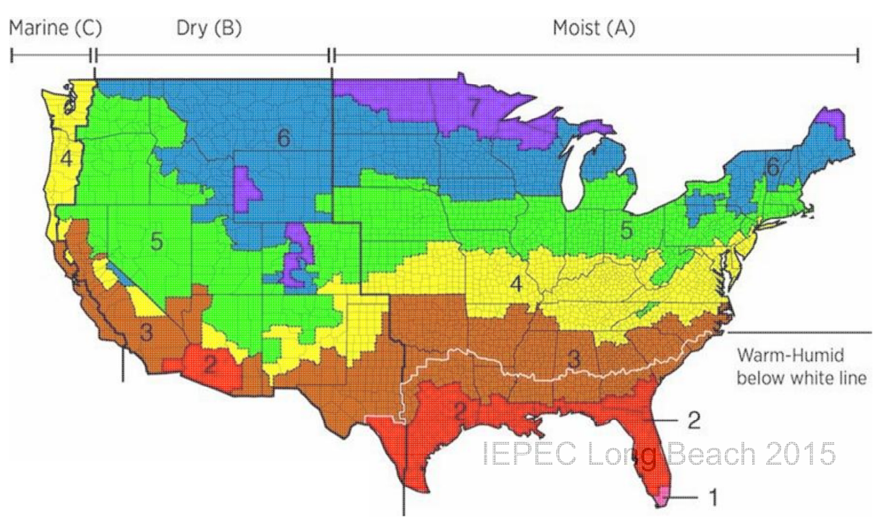
- Independent variables: time of day, day of week, outside air temperature

Test Data Set

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

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

Most data from CA Zone 3, and Wash DC Zone 4; some from Seattle Zone 4



Performance Metrics of Focus

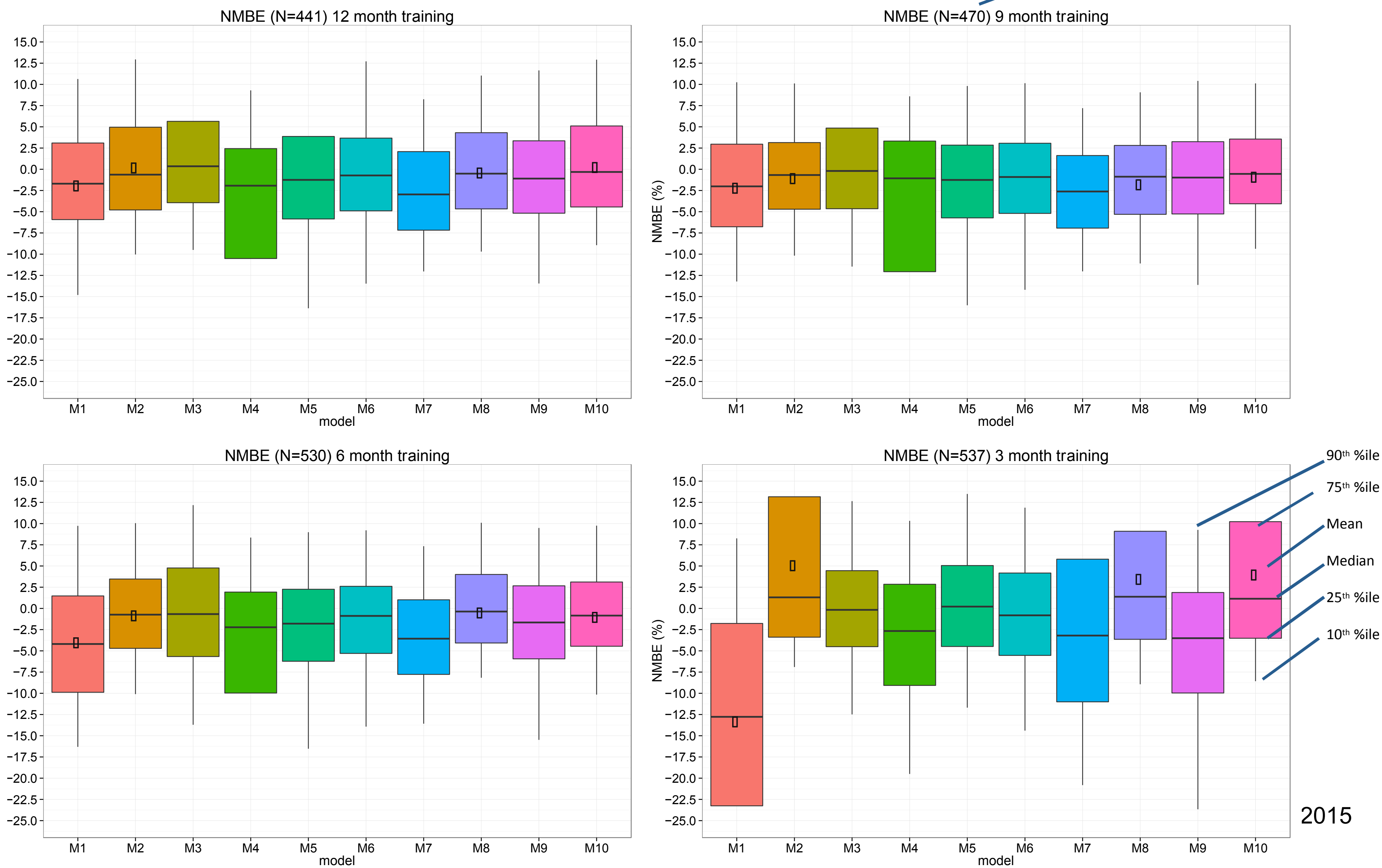
- Many possible goodness of fit metrics to choose from
- Analyzing too many metrics makes it hard to draw conclusions about model performance
- ~20 reps from efficiency program management evaluation, implementation voted on top two metrics of choice
- There actually was strong consensus!
  - CV(RMSE)
  - Normalized Mean Bias Error - total percent error in predicted vs. actual energy use

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)

Total number of buildings in the test case



What Do These Distributions Of Percent Errors Tell Us?

- Differences between models are mostly small

- Across the group of models, for 12-month training 12-mo prediction
  - *Average* median percent error ~-1.2%
  - *Range* of median errors is ~-3% to 0.4%

- All models perform well overall, especially for the case of 12-months training

What Happens As We Shorten the Training Period?

- 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

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

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

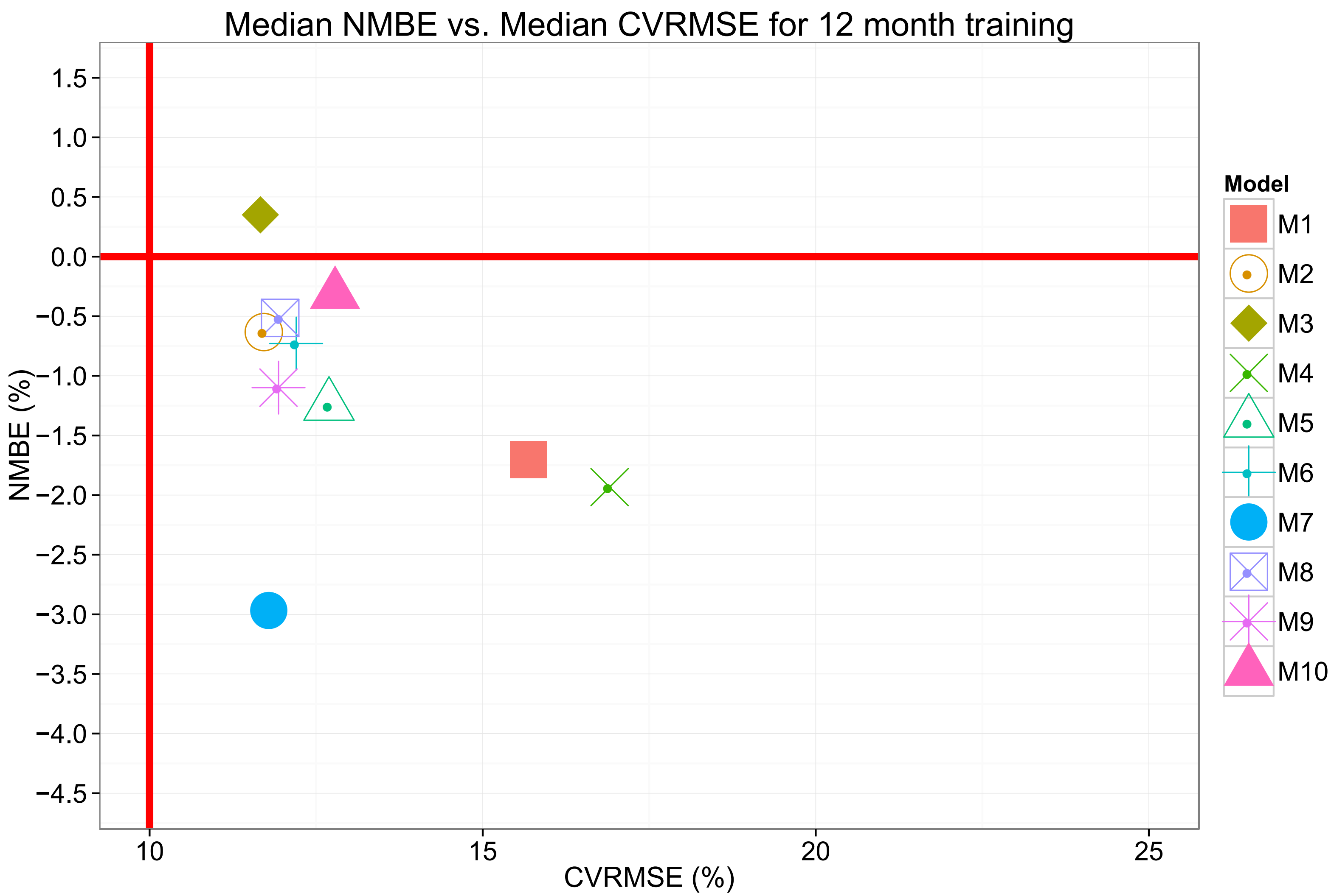
- Models analyzed are likely to meet the Guideline 14 requirements

- Guideline 14 specifies CV(RMSE) during the *training*[1] period, should be <25%[2]

- 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

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

Both Metrics At Once: No Clear “Winner”



Key Takeaways

- AMI data and interval data models/tools hold great promise to scale 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
- 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
  - Accuracy insights not generally available for stipulated savings
- Apply 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?
- Vet project-specific M&V plans
  - Use 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

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.

Thank You!

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