Analysis of Baseline Methodologies

NAESB Phase II DR M&V Issues Working Group

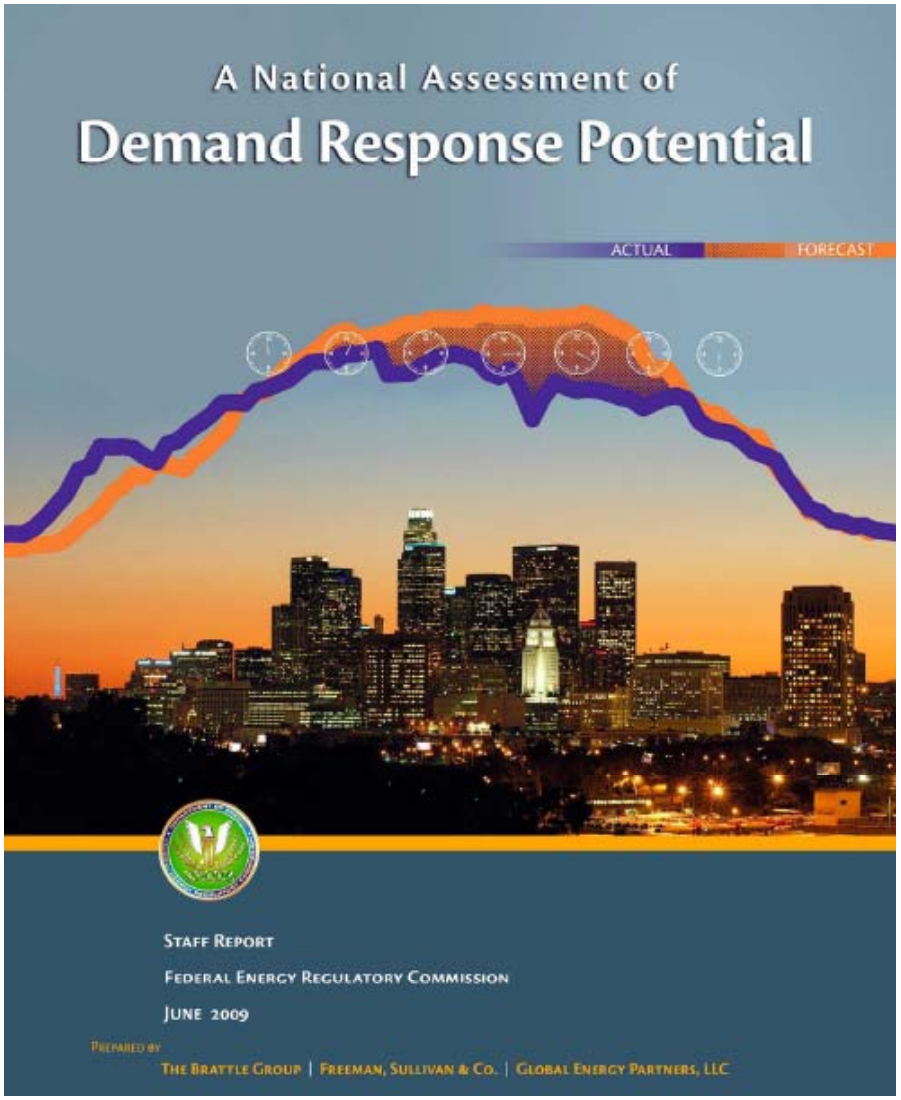
September 20, 2010

Agenda

- Background
- What Makes A “Good” Baseline
- Maximum Base Load Analysis
- High X of Y Analysis

ENERNOC

FERC has requested development of standards around DR measurement



…development of standardized practices for quantifying demand reductions would greatly improve the ability of system operators to rely on demand response programs

…Central to the issue of measurement is a determination of the customer baseline

ENERNOC

What is a DR "Baseline" – FERC (NAESB)

FERC

- **"43. The Commission is requiring, consistent with our regulation at 18 CFR 35.28(c)(vi), each ISO and RTO to revise its OATT to include the NAESB Phase I M&V Standards we are incorporating by reference herein. (FERC Order 676-F at 22)**

NAESB

- **"Baseline**
  - **A Baseline is an estimate of the electricity that would have been consumed by a Demand Resource in the absence of a Demand Response Event. The Baseline is compared to the actual metered electricity consumption during the Demand Response Event to determine the Demand Reduction Value. " (NAESB Phase I M&V Standards at 9, emphasis added)**

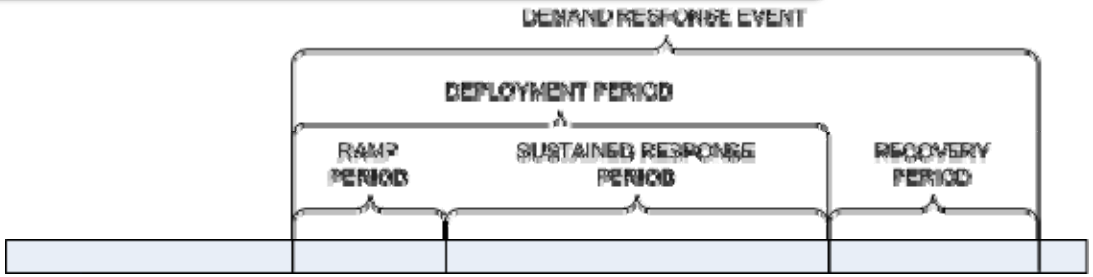
ENERNOC

NAESB's DR M&V efforts resulted in a useful semantic framework…

Definitions and Glossary

- 4 Service Types (Capacity, Energy, Reserves, Regulation)
- 5 Performance Evaluation Methods
- Example of terms: Baseline, Performance Window, Telemetry

Illustrations of Events



…but fell short of actual standards

Business Practices

- There are ~100 instances of the phrase, "the System Operator shall specify"

Narrowing the focus…

ANALYSES

NAESB Standards			
Performance Evaluation Type	Valid For Service Type		
	Energy	Capacity	Reserves
Maximum Base Load	✓	✓	✓
Meter Before / Meter After	✓	✓	✓
Baseline Type-I	✓	✓	✓
Baseline Type-II	✓	✓	✓
Metering Generator Output	✓	✓	✓

This does not tell us anything about which performance measurements are best

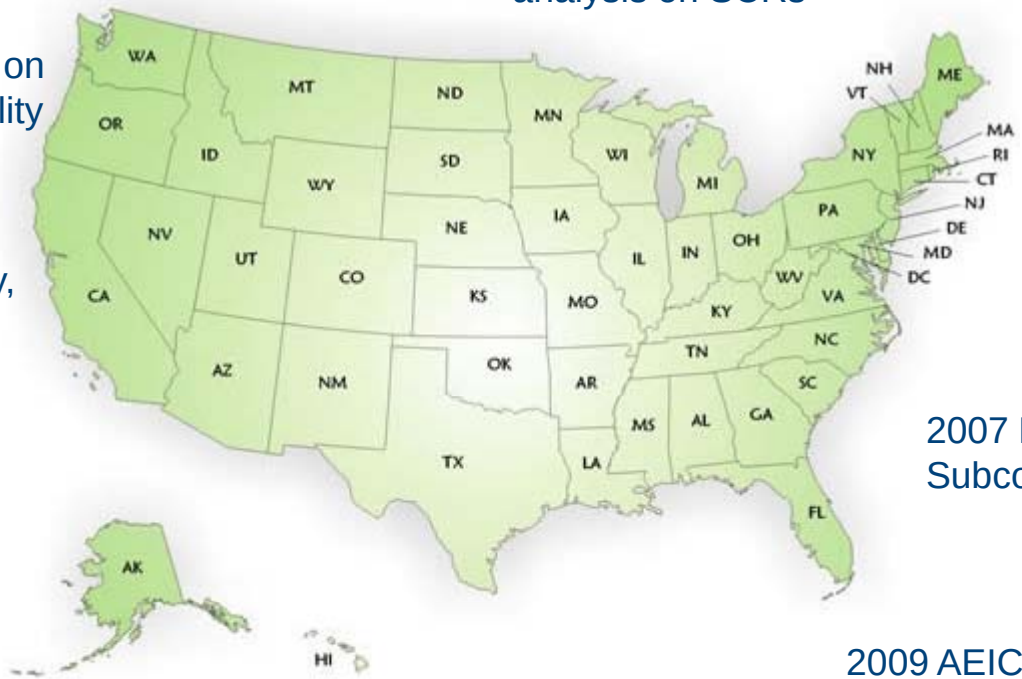
ENERNOC

Previous baseline studies by consultants, utilities, research centers, and market participants

2003 KEMA-XENERGY Report

2009 & 2010 NYISO analysis on SCRs

2006 Quantum report on day-ahead and reliability DR programs



2006 EnerNOC whitepaper

2008 LBNL DR Study, 33 commercial sites

2008 EnerNOC whitepaper

2008 EnerNOC & CDRC testimony before CA PUC

2007 PJM CBL Subcommittee Analysis

2009 AEIC report on Estimation Errors

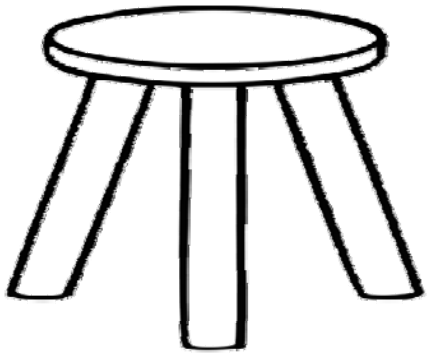
2009 CAISO Proxy DR Program Proposal

ENERNOC

What Makes A “Good” Baseline?

Accuracy

- Customers should receive credit for no more and no less than the curtailment they actually provide



Integrity

- Baseline method should protect against attempts to “game the system” and should not encourage irregular consumption

Simplicity

- The baseline and resulting curtailment calculations should be simple enough for all stakeholders to calculate, including end-user customers, during events

ENERNOC

Baseline Candidate #1: Maximum Base Load

NAESB Definition

A performance evaluation methodology based solely upon a Demand Resource's ability to reduce to a specified level of electricity demand

The "Drop To" Method

Examples

- Average Peak Monthly Demand (APMD) – NYISO SCR/ICAP
- Peak Load Contribution (PLC) – PJM FSL

ACCURACY ?

SIMPLICITY ✓

INTEGRITY ✓

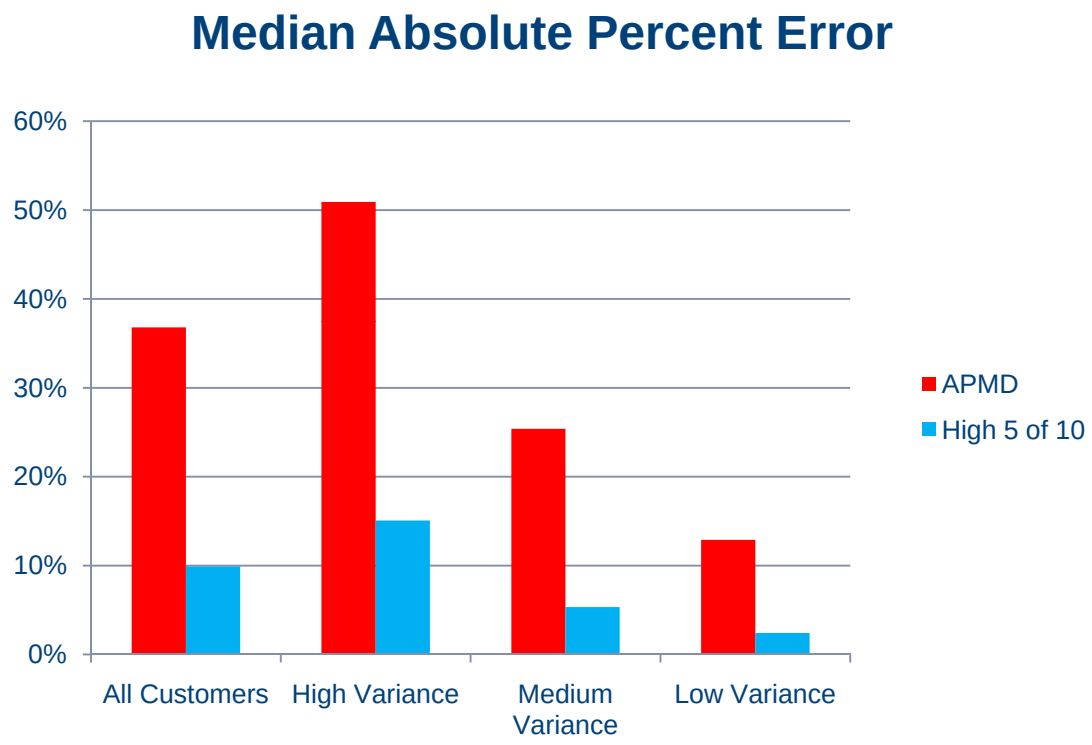
Generally suspected that both MBL methods do not correctly estimate what load would have been if the site did not curtail given the load data used for the baseline is a year old.

ENERNOC

Analysis of NYISO sites and baselines

- Criteria for Customers
  - Customers: All
  - Load Type: Curtailment
  - Status: Project Complete or Asset Ready to Respond
  - Project Complete Date: Prior to and including June 1, 2007
- Selection of Event-like Days
  - Used proprietary internal application that rates for each hour the likelihood of an event (low, medium, high)
  - Choose 4 days per month, starting with High ratings, than Medium ratings, than Low ratings of days closest to July 31
  - Considered the summer months June, July, August, September of 2009
- High 5 of 10
  - Excluded event days
  - Excluded weekends
  - Reflected average of 8 event-like days in June-July 2009

Large errors between the APMD and actual load

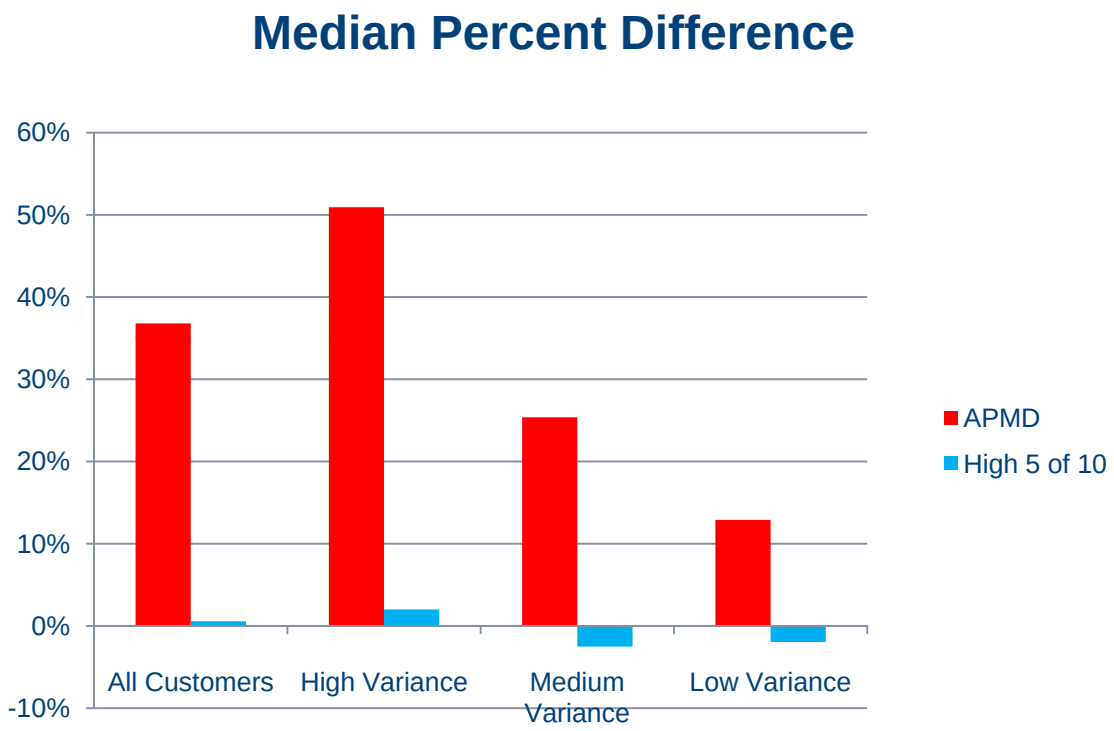


METRICS

Median Absolute Percent Difference is a measure of ***magnitude of error***

ENERNOC

APMD consistently overstates actual load



METRICS
Median Percent Difference is a measure of ***bias***

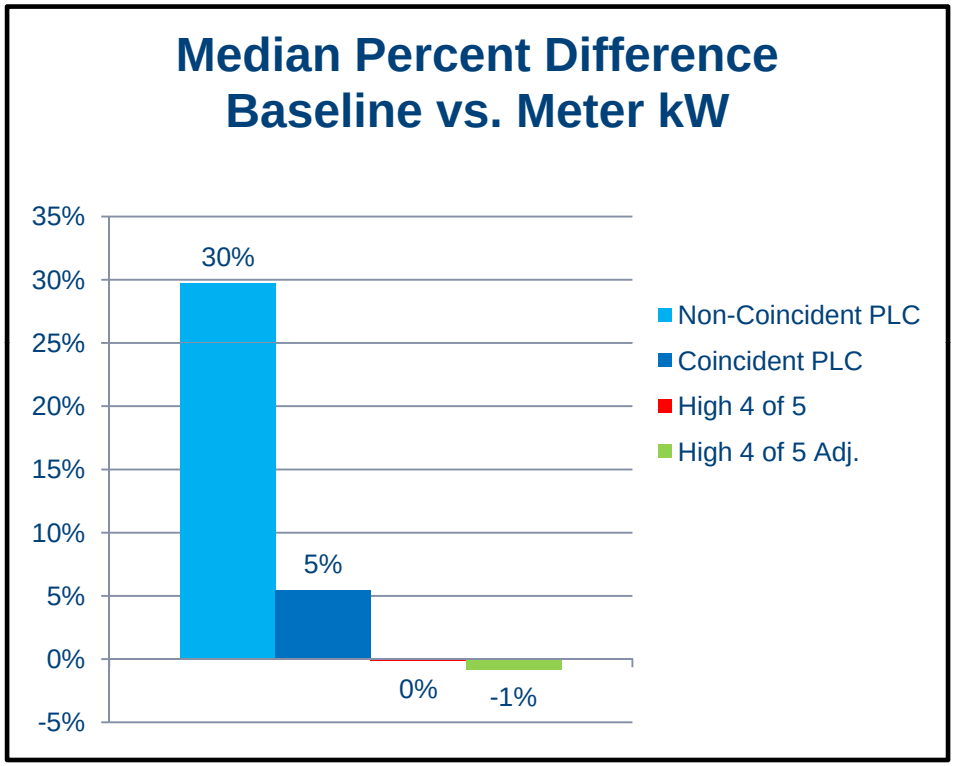
ENERNOC

Analysis of PJM sites and baselines

Parameter	Detail
➢ Five event-like days	• Identified based on load, pricing and weather conditions
➢ All PJM sites selected	• Curtailment only sites, must have complete data for summer 2008 and 2009
➢ Coincident Peak Load Contribution	• Average customer load coincident with 5 system peak hours from last summer
➢ Non-Coincident Peak Load Contribution	• Average of 5 peak hours of each individual customer, not coincident with system peak hours
➢ High 4 of 5 Methods	• Summer 2009 usage data with and without symmetric additive adj.
➢ Timeframe	• Summer 2009, program hours 12-8pm

ENERNOC

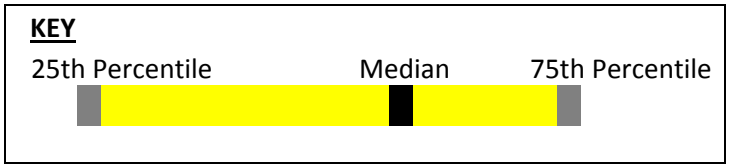
PJM Analysis Results

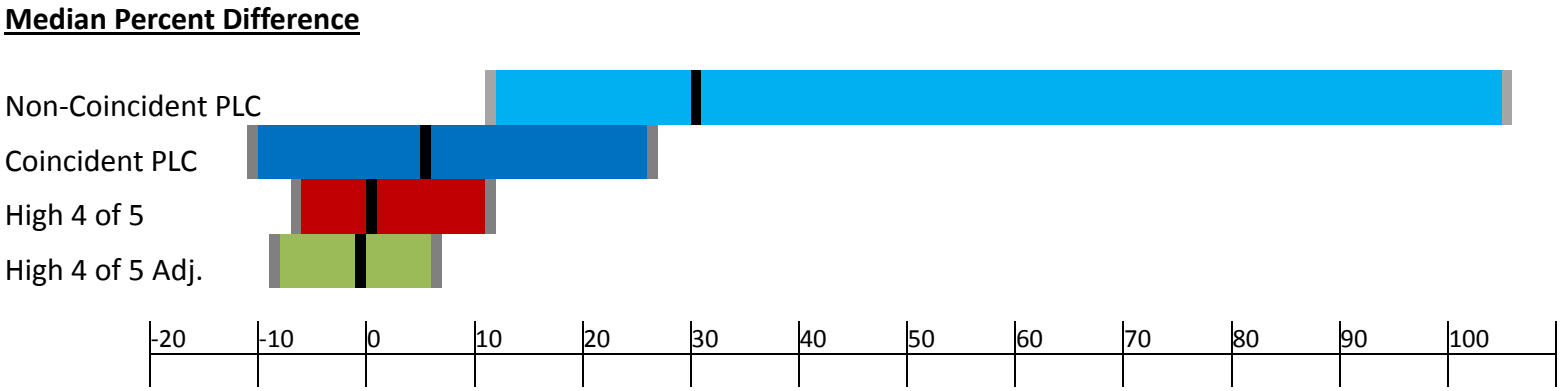


- Non-coincident PLC greatly overestimates load
- Coincident PLC only slightly overestimates load, and median is close to High 4 of 5 median

ENERNOC

Choice of Peak Days Significantly Affects Accuracy of MBL Methods





- Both PLC methods have wider ranges of percent differences, so for any given site, the baseline could be far from a fair predictor of load and enhance opportunity for selection bias

25th Percentile to 75th Percentile shows the range of the middle 50% of data points and shows the range of errors; this range is smaller and more descriptive than a range from the minimum to the maximum

Baseline Candidate #2: Baseline Type I

NAESB Definition

A baseline performance evaluation methodology based on a Demand Resource's historical interval meter data which may also include other variables such as weather and calendar data

The "Drop By" Method

Examples

- Regression – ERCOT
- Comparable Day – PJM, ERCOT
- Rolling Average – ISO-NE
- High X of Y – PJM, NYISO, SCE, BED, TVA, OPA

Numerous baseline methods, likely some are better than others

ENERNOC

Studies show. . .

There are some concepts that are well accepted “best practices”

- Exclude holidays, weekends, event days
- Use of an adjustment to a High X of Y method reduces bias and improves accuracy

There are more controversial practices/methods

- Best X of Y method
  - All 10 of 10 days, Mid 6 of 10 days, High 5 of 10 days?
- Adjustment Type
  - How many hours? What hours? Additive or scalar?

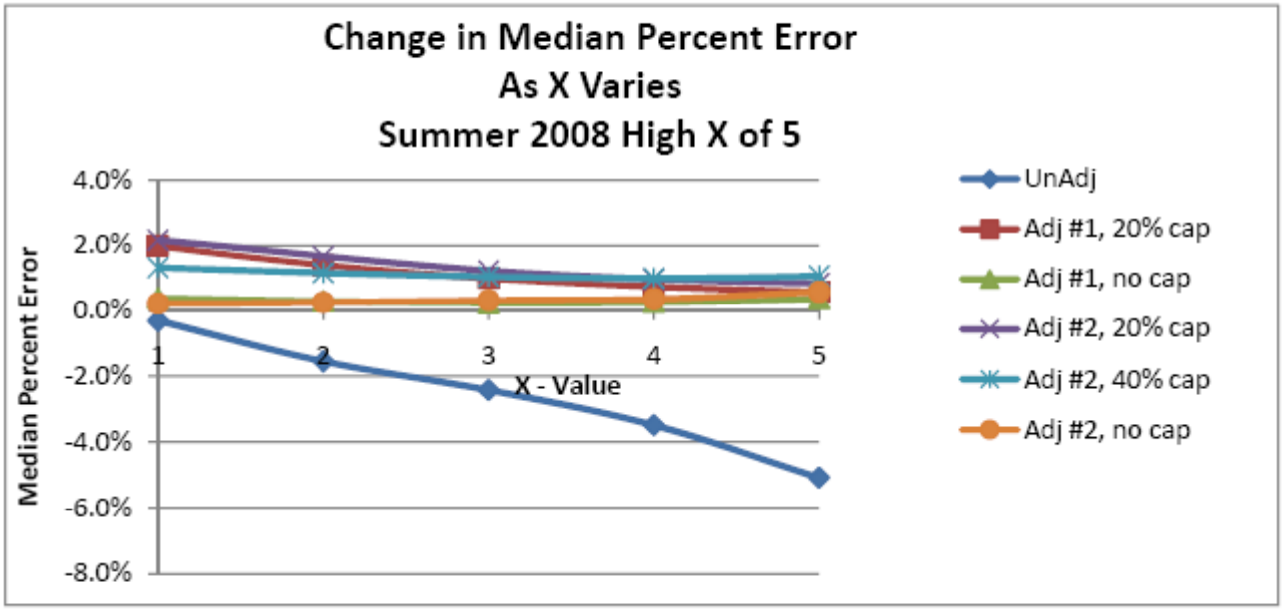
Overall, it is clear there is no perfect baseline method. . .

ENERNOC

But there are trends and some methods can be shown to be better or worse than others

Parameter	Detail
➢ Three event-like days	• Identified based on maximum daily load data for each region
➢ 306 sites randomly selected from pool	• Pool of >1,000 pure curtailment customers from TX, CA, NY, ISO-NE, and PJM with complete data
➢ High X of 5	• Each baseline, High 1 of 5, High 2 of 5, etc
➢ High X of 10	• Each baseline, High 1 of 10, High 2 of 10, etc
➢ High X of 5 and High X of 10 Adj. #1	• Adjustment using 3 hrs beginning 4 hrs before event; 20% cap & no cap
➢ High X of 5 and High X of 10 Adj.#2	• Adjustment using 2 hrs beginning 3 hrs before event; 20% cap & no cap

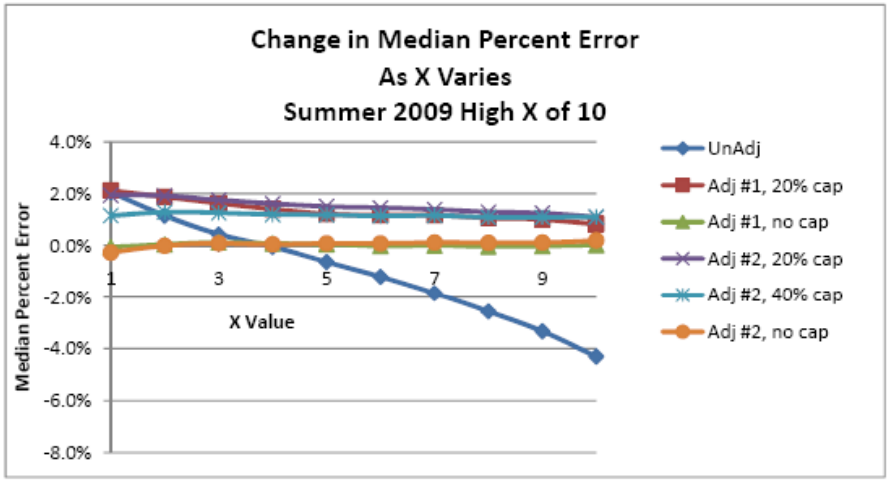
Comparison of X value variations on bias

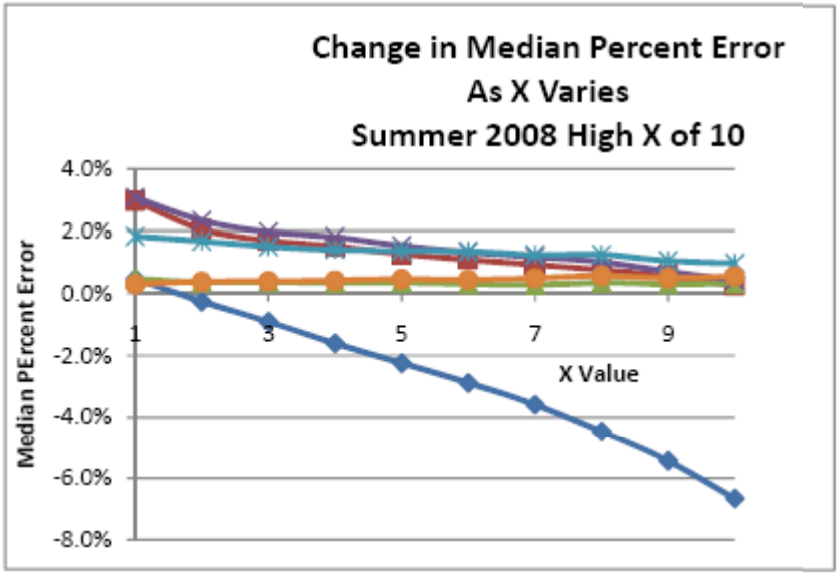


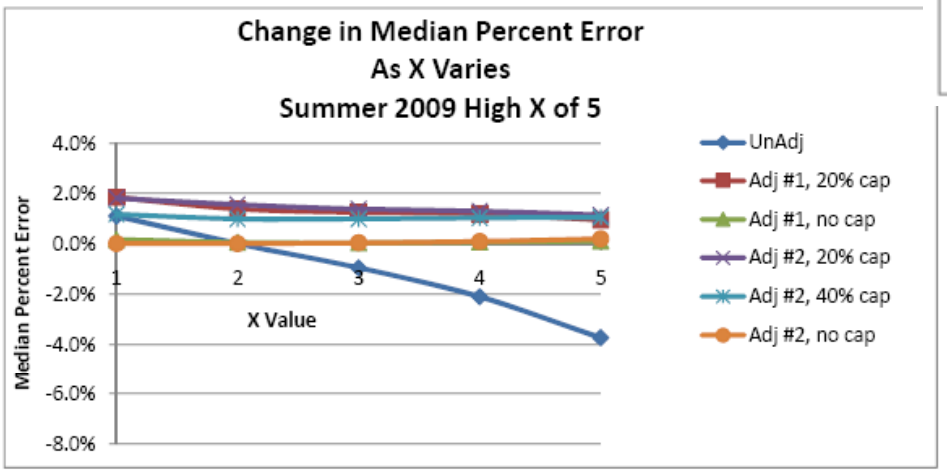
- Unadjusted baseline increasingly understates the load as X approaches Y
- Adjusted baselines have relatively level slopes, showing little change in bias as X varies
- Adjustments with no cap appear to be the least biased methods

ENERNOC

Conclusions hold true across High X of 5 and High X of 10 results for both 2008 and 2009







ENERNOC

Highlights of Findings

Maximum Base Load baselines

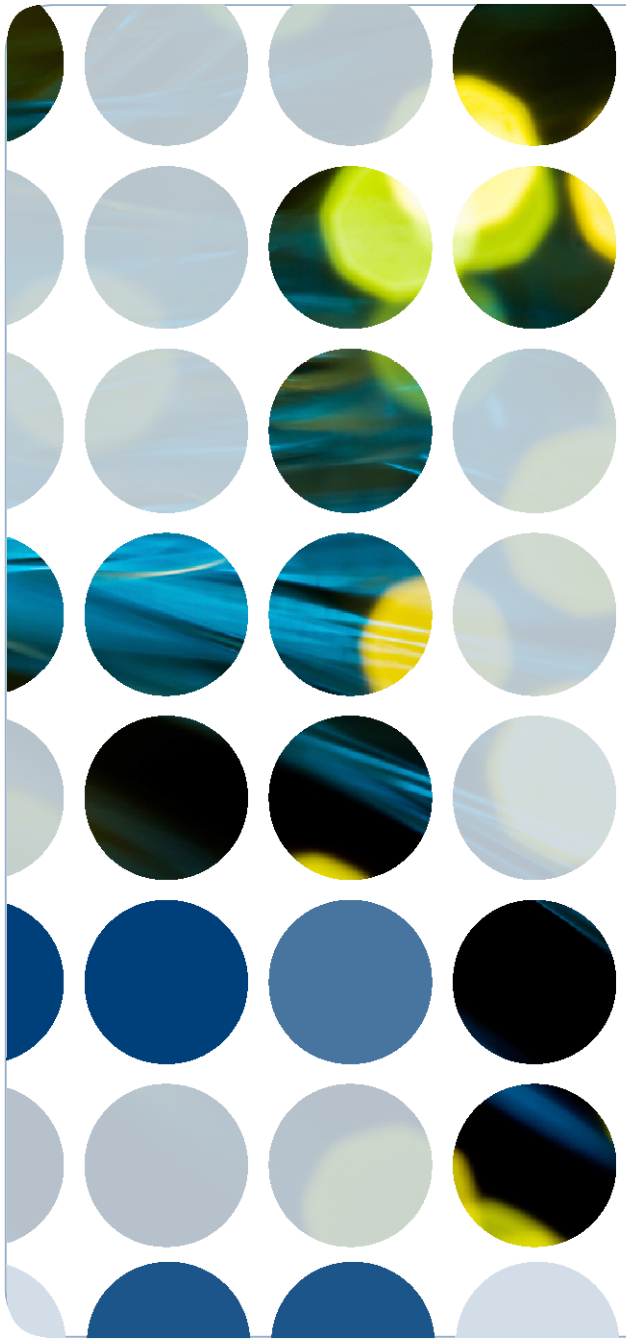
- Better to use coincident peak hours, rather than non-coincident peak days
- These methods still do not reliably measure what load would have been if the site did not curtail, as shown by the wider range of errors
- Given that there is a wide range of errors, programs that use these baselines leave more room for selection bias, i.e., selecting customers who will have inflated baselines and generate higher payments

High X of Y baselines

- Adjustment (using actual load prior to event) is necessary
- Adjustments to baselines should not be capped
- Adjusted baselines with X values closer to Y should be used as they are more accurate and take into consideration more data points

ENERNOC

Questions?



Aaron Breidenbaugh
Director of Regulatory Affairs
101 Federal Street, Suite 1100
Boston, MA 02110

(617) 224-9918
abreidenbaugh@enernoc.com