



Analysis of Baseline Methodologies

NYISO ICAP Working Group Meeting

September 15, 2010

Agenda

- ① Background
- ① What Makes A “Good” Baseline
- ① Maximum Base Load Analysis
- ① High X of Y Analysis

What is Demand Response to the NYISO?

Treatment of Long-Term SCR Shutdowns

David J. Lawrence
Manager, Auxiliary Market Products
New York Independent System Operator

NYISO Installed Capacity Working Group Meeting
June 5, 2009

Draft – for Discussion Purposes Only



Relevant Tariff Provisions

- ◆ MST Tariff - 2.52.1 Expected Load Reduction
 - For purposes of determining the Real-Time Locational Based Marginal Price, the reduction in Load expected to be realized in real-time from activation of the Emergency Demand Response Program and from Load reductions requested from Special Case Resources, as established pursuant to ISO Procedures.
- ◆ MST Tariff - 2.172c Special Case Resource
 - Demand Side Resources capable of being interrupted upon demand...

4

Draft – for Discussion Purposes Only



Operational Concerns and Principles

- ◆ Emergency Operations Manual expects SCRs to reduce Operating Reserves Deficiency:
 - Count NYISO Special Case Resources (ICAP) as NYISO Operating Reserve and determine if it is of sufficient quantity to eliminate the deficiency.¹
- ◆ SCR modeling in the IRM study is consistent with the use of these resources to eliminate a capacity deficiency.
 - SCRs are ICAP resources that only provide energy/load curtailment when activated in accordance with the NYISO Emergency Operating Manual.²

Recognizing SCR long-term shutdowns is consistent with the way mothballed/retired generators are treated.

EOP Manual, Sec. 4.4.1, Step 5
2009 IRM Report, p. 47

3

It is “Demand” that
“Responds” during capacity
deficiencies to alleviate or
eliminate reserve shortages.

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

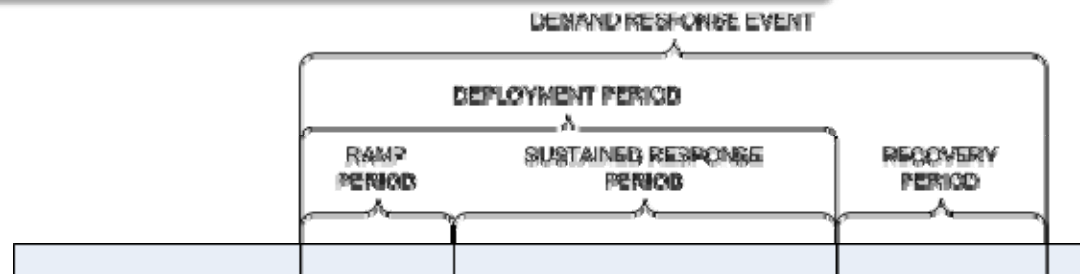
- “Baseline
 - A Baseline is an estimate of the electricity that would have been consumed by a Demand Resource in the absence of a Demand Response Event. The Baseline is compared to the actual metered electricity consumption during the Demand Response Event to determine the Demand Reduction Value. “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

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

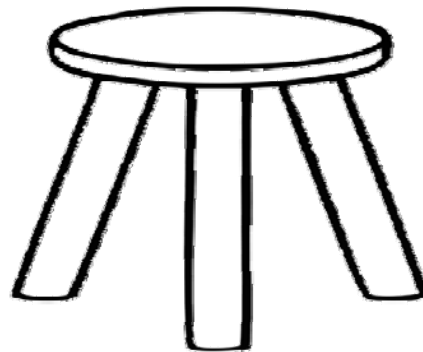


This does not tell us anything about which performance measurements are best

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

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



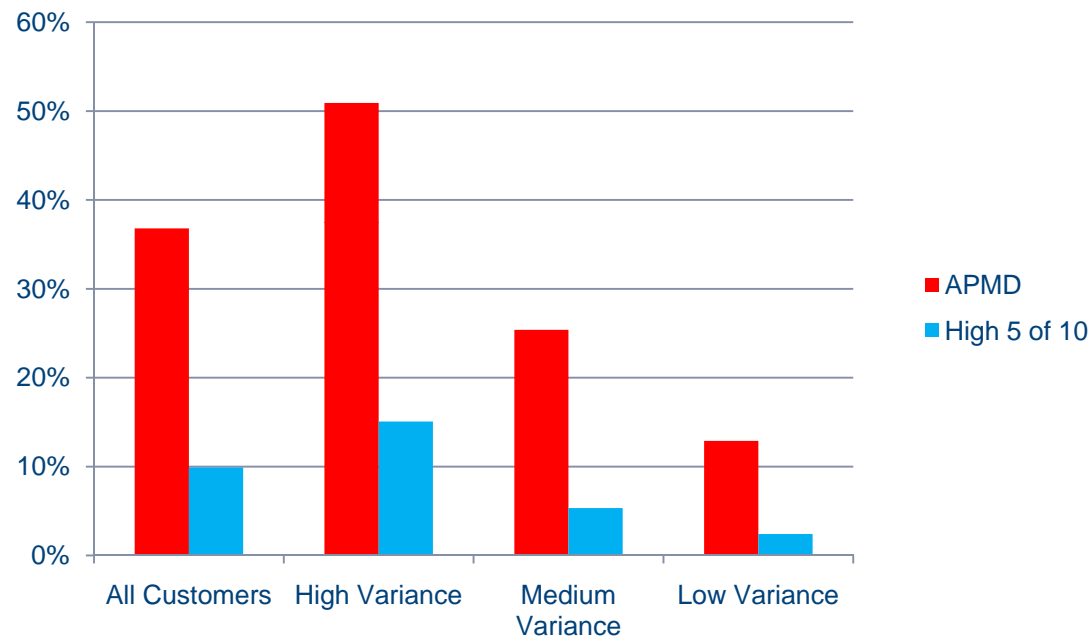
Generally suspected that both MBL methods do not correctly estimate what load would have been if the site did not curtail given the load data used for the baseline is a year old.

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

Median Absolute Percent Error

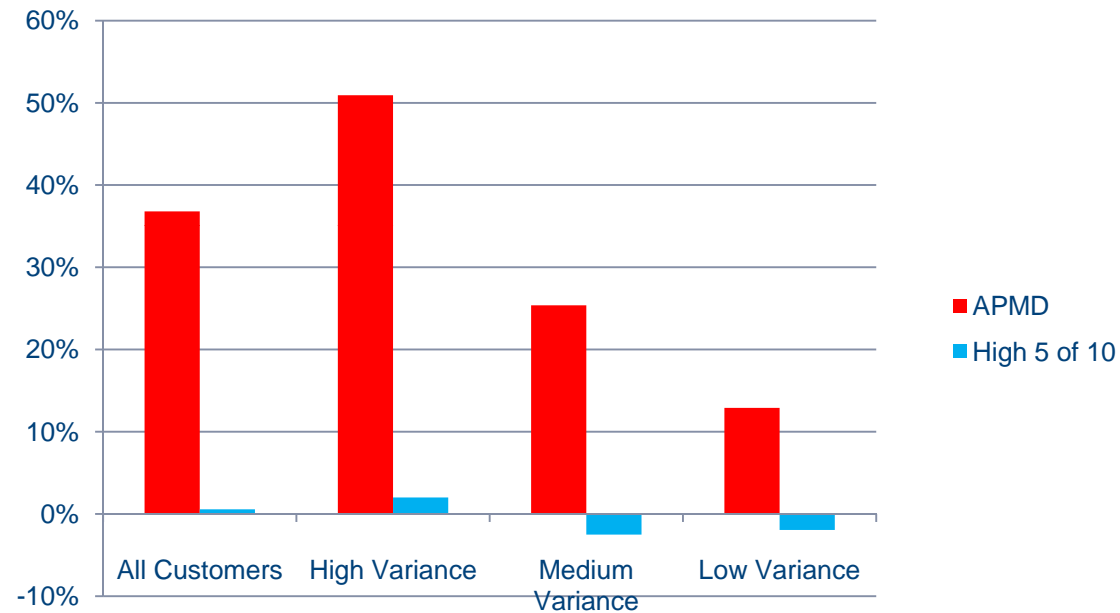


METRICS

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

APMD consistently overstates actual load

Median Percent Difference



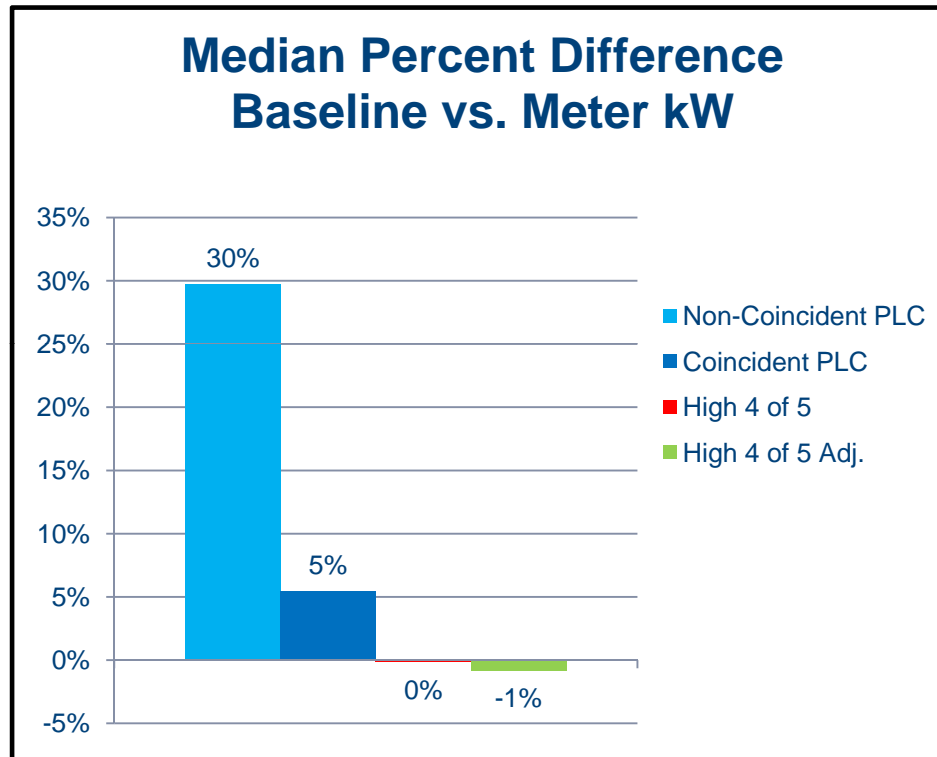
METRICS

Median Percent Difference is a measure of *bias*

Analysis of PJM sites and baselines

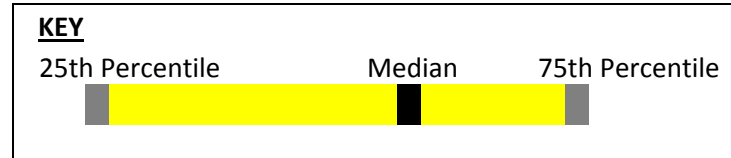
<u>Parameter</u>	<u>Detail</u>
➤ Five event-like days	<ul style="list-style-type: none">• Identified based on load, pricing and weather conditions
➤ All PJM sites selected	<ul style="list-style-type: none">• Curtailment only sites, must have complete data for summer 2008 and 2009
➤ Coincident Peak Load Contribution	<ul style="list-style-type: none">• Average customer load coincident with 5 system peak hours from last summer
➤ Non-Coincident Peak Load Contribution	<ul style="list-style-type: none">• Average of 5 peak hours of each individual customer, not coincident with system peak hours
➤ High 4 of 5 Methods	<ul style="list-style-type: none">• Summer 2009 usage data with and without symmetric additive adj.
➤ Timeframe	<ul style="list-style-type: none">• Summer 2009, program hours 12-8pm

PJM Analysis Results

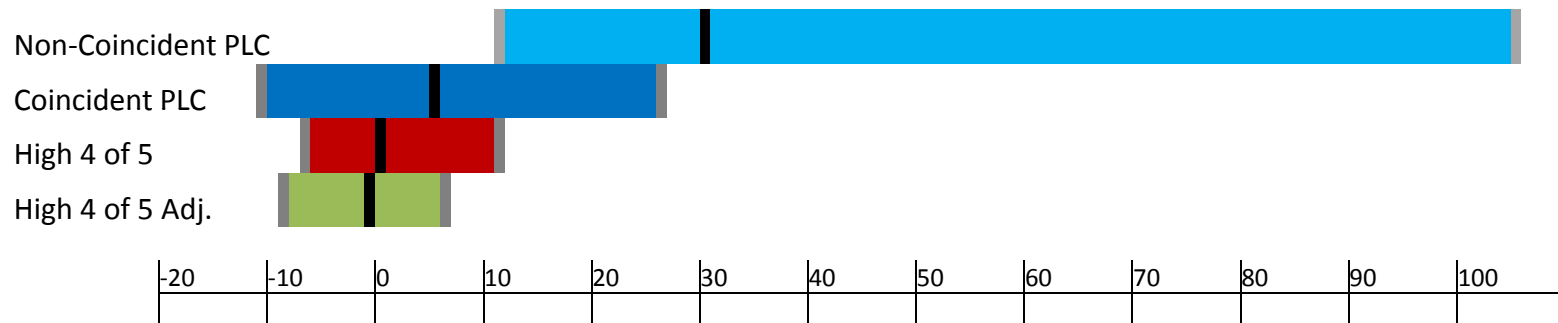


- Non-coincident PLC greatly overestimates load
- Coincident PLC only slightly overestimates load, and median is close to High 4 of 5 median

Choice of Peak Days Significantly Affects Accuracy of MBL Methods



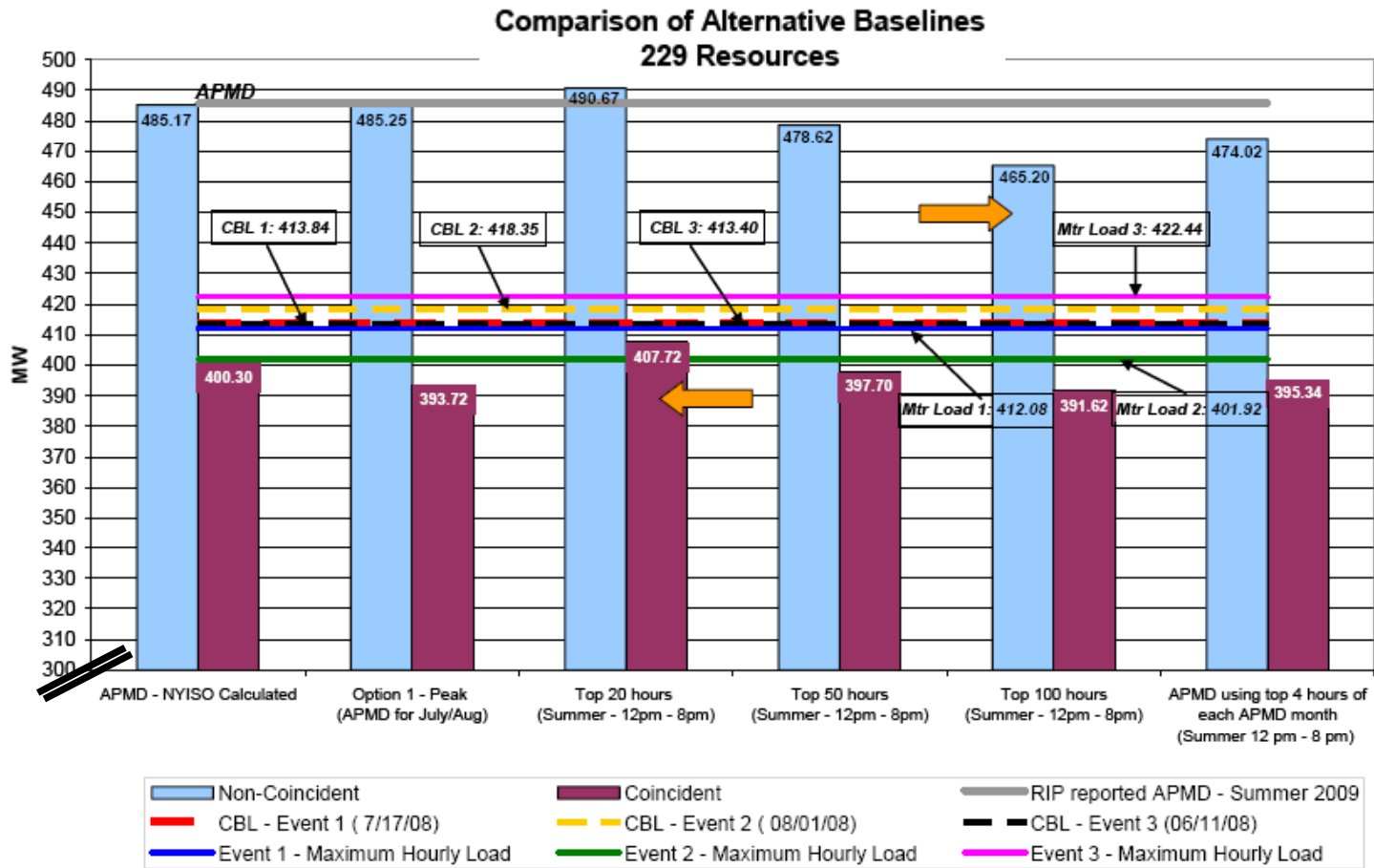
Median Percent Difference



- 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

NYISO study confirms this analysis



The coincident APMD is close to the CBL methods, while the non-coincident APMD predicts much higher load values

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

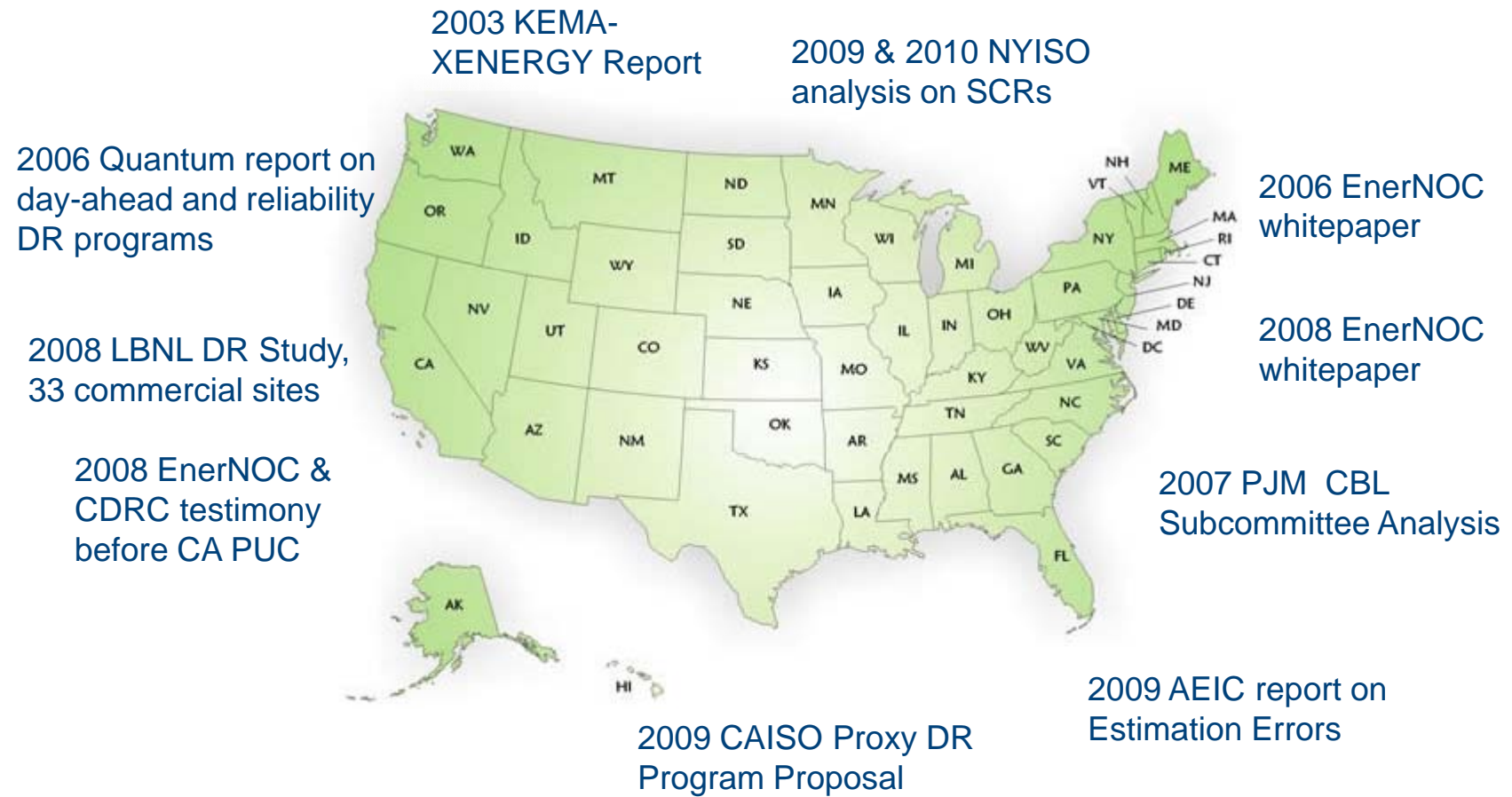
Examples

- 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

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



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

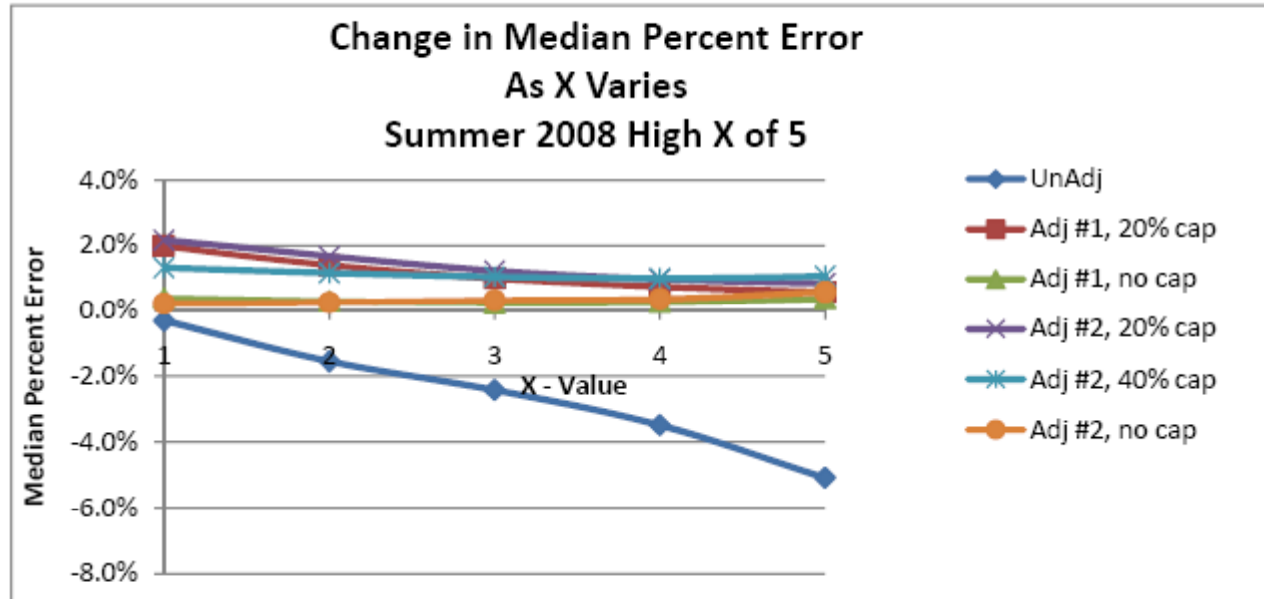
- Best X of Y method
 - All 10 of 10 days, Mid 6 of 10 days, High 5 of 10 days?
- Adjustment Type
 - How many hours? What hours? Additive or scalar?

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

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

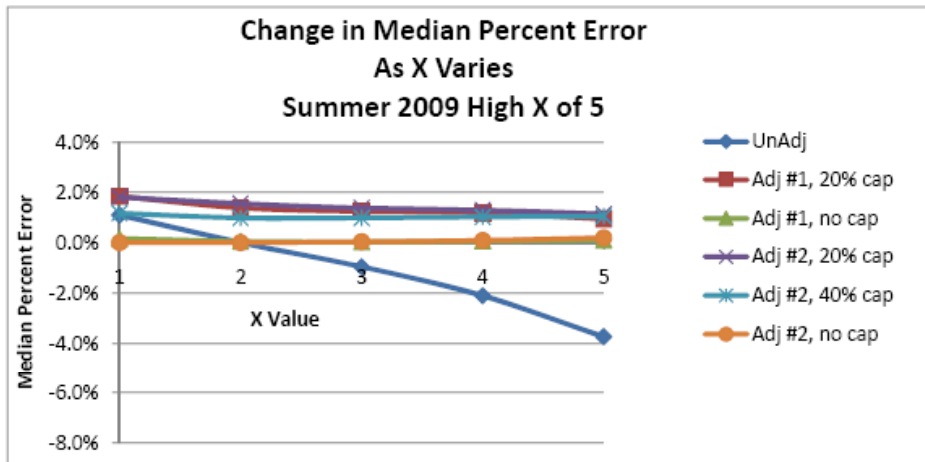
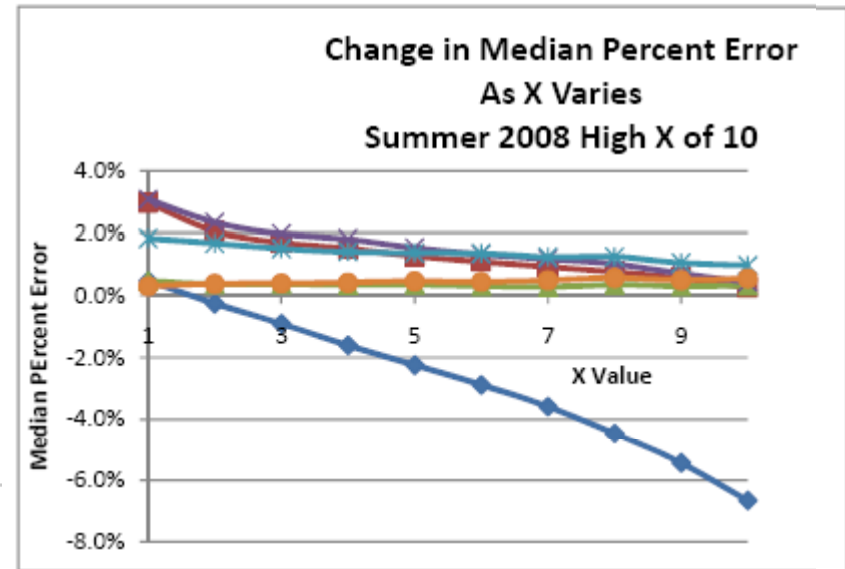
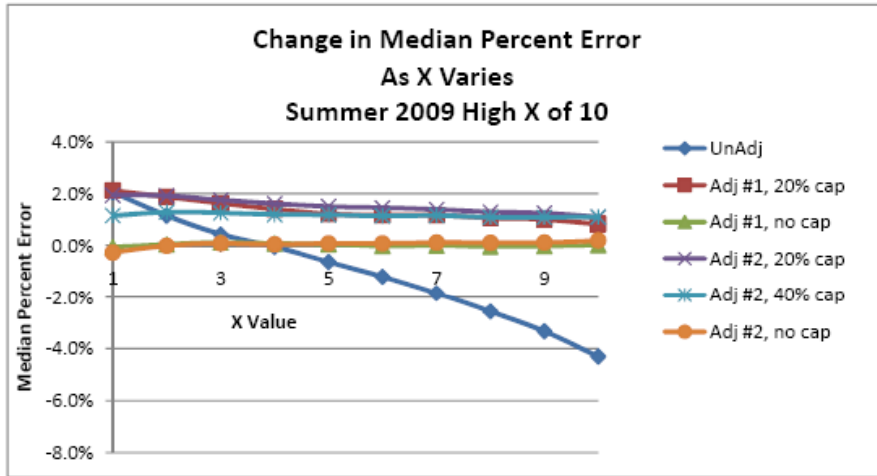
<u>Parameter</u>	<u>Detail</u>
➤ Three event-like days	<ul style="list-style-type: none">• Identified based on maximum daily load data for each region
➤ 306 sites randomly selected from pool	<ul style="list-style-type: none">• Pool of >1,000 pure curtailment customers from TX, CA, NY, ISO-NE, and PJM with complete data
➤ High X of 5	<ul style="list-style-type: none">• Each baseline, High 1 of 5, High 2 of 5, etc
➤ High X of 10	<ul style="list-style-type: none">• Each baseline, High 1 of 10, High 2 of 10, etc
➤ High X of 5 and High X of 10 Adj. #1	<ul style="list-style-type: none">• Adjustment using 3 hrs beginning 4 hrs before event; 20% cap & no cap
➤ High X of 5 and High X of 10 Adj.#2	<ul style="list-style-type: none">• 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

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



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

NYISO's EDRP CBL Baseline Meets Most Criteria

- Highest 5 of 10 structure is highly defensible
- 20% adjustment cap should be removed

So, where do we stand?

	ACCURACY	SIMPLICITY	INTEGRITY
• Regression	✓	X	✓
• Matching Day	✓ ¹	✓	✓
• High X of Y	✓	✓	✓
• cAPMD	✓ X	✓	✓ ²
• nAPMD	X	✓	✓ ²

1 Assumes day chosen is most similar and not most profitable

2 Assumes CSP does not exercise selection bias

However these assumptions are perilous...

	ACCURACY	SIMPLICITY	INTEGRITY
• Regression	✓	X	✓
• Matching Day	X	✓	✓
• High X of Y	✓	✓	✓
• cAPMD	✓	✓	X
• nAPMD	X	✓	X

EnerNOC's Recommendation

ENERNOC
get more

get more

Align capacity value with real-time performance

- Detailed performance data is not yet available, BUT...
- The APMD approach is inherently flawed:
 - It rewards providers for doing nothing or even INCREASING their load in real time
 - It can under-reward others for providing real reductions in actual load
 - It is inconsistent with standard industry practice
 - NO other ISO or utility uses this flawed approach
- Despite rigorously adhering to program rules, RIP are presented with strong incentive to enroll participants that do not benefit the system during emergencies
- EnerNOC urges the adoption of an alternative approach that better aligns the interests of RIPs and system operators:

Recommendation #3:

Use the EDRP CBL approach to determine ICAP/UCAP Translation Factors and eliminate APMD from the SCR program altogether

© 2006 EnerNOC, Inc. All Rights Reserved – Strictly Confidential, Permission Required for Distribution

ENERNOC

Demand Response in New York
Recommendations for The Next Phase

Price Responsive Load Working Group Meeting
November 2, 2007

 ENERNOC



Aaron Breidenbaugh
Director of Regulatory Affairs
101 Federal Street, Suite 1100
Boston, MA 02110
(617) 224-9918
abreidenbaugh@enernoc.com