# The Myth of the "Ideal" DR Evaluation Methodology

Authors

Vijay Gopalakrishnan, DNV, Rocky Hill, CT
Aaron Schrader, DNV, Madison, WI
Ken Agnew, DNV, Madison, WI
Antonio Larson, National Grid, Waltham, MA
Jon Maxwell, DNV, College Station, TX

27 September 2022

# Acknowledgements

## This work would not have been possible without:

- Ralph Prahl, Prahl Associates
- Miles Ingram, formerly Eversource
- Tracy Dyke-Redmond, Eversource
- Alexandra Abbruscato, formerly Eversource

# Contents

# Introduction to Baselines
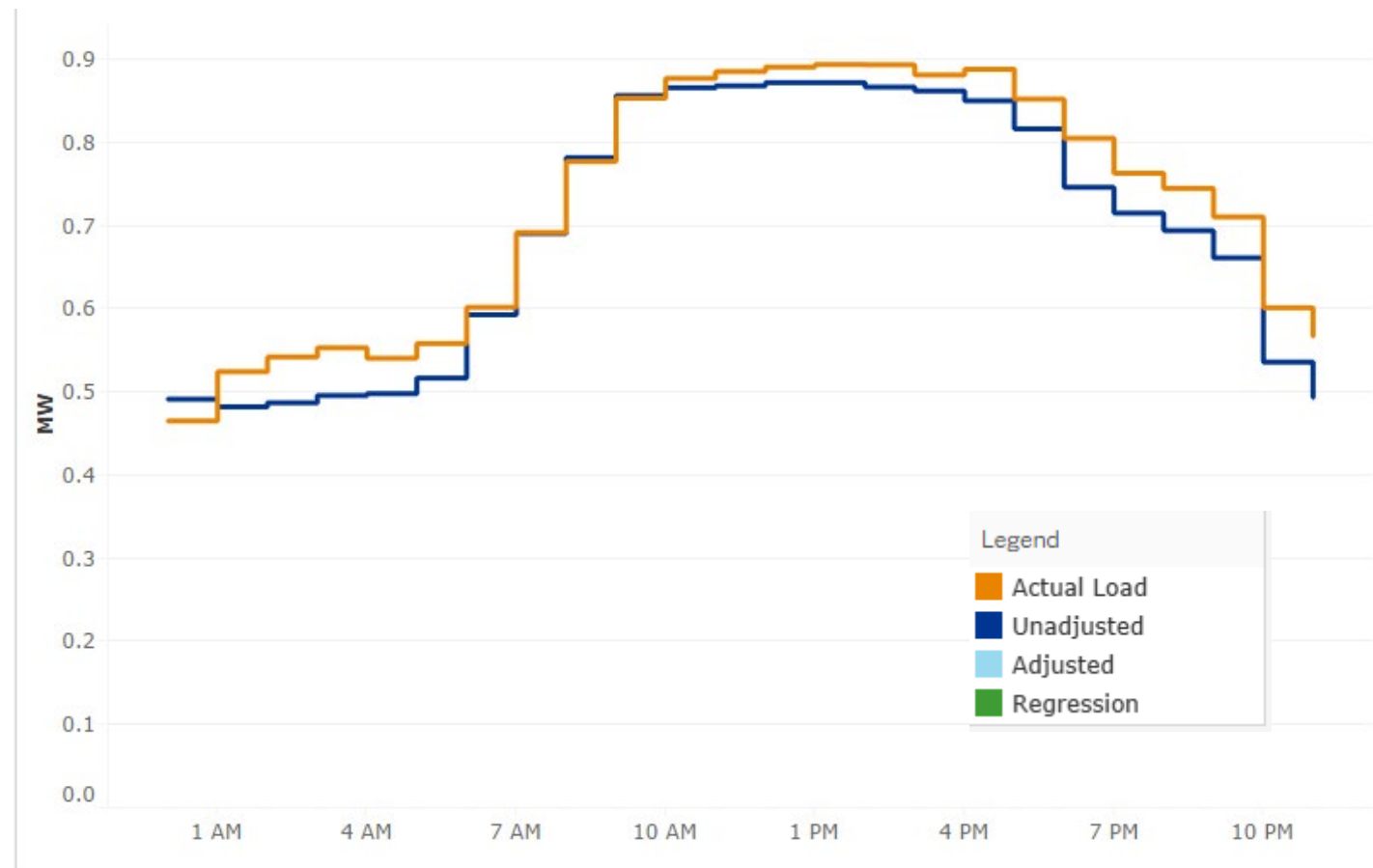
# Basic DR Example – Baseline Days

- Settlement baselines are usually built with some subset of prior days.

- ISO-NE and MA DR programs all use 10 of 10, meaning load at the same times of day as the event, but for the prior 10 non-holiday weekdays.

- For this example, these ten days are remarkably similar

- A mean for each interval across these most recent days produces an average shape.



**Baseline Pool**
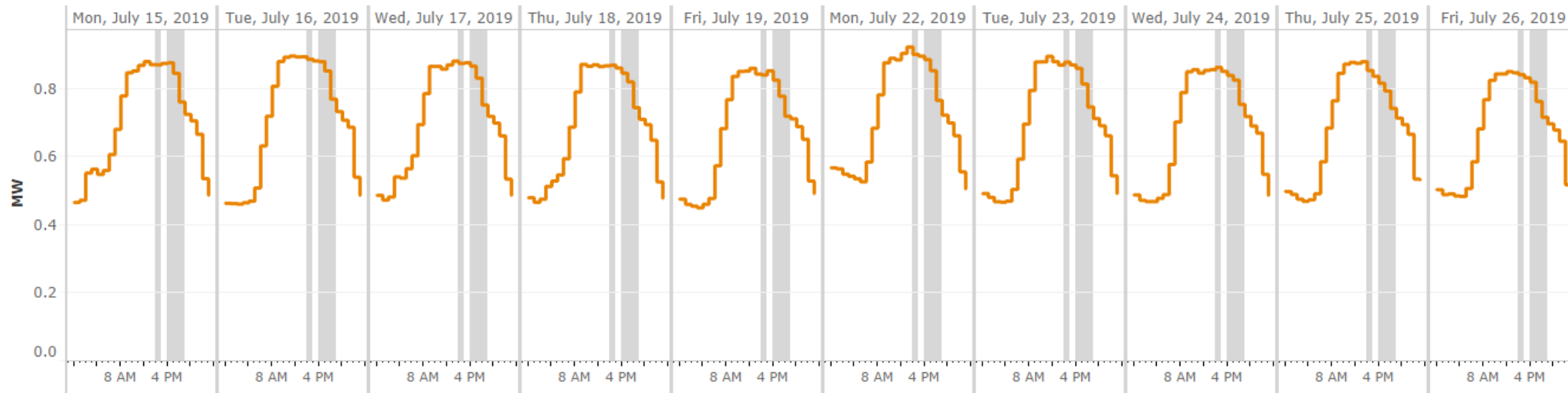
# Basic DR Example – Unadjusted Baseline

- For this customer, the unadjusted 10 of 10 provides a reasonably accurate baseline (counterfactual load) but it diverges by 7pm.

# Basic DR Example – Adjusted Baseline

- The 10 of 10 baseline is adjusted up to meet the load. The load and adjusted baseline are identical in the 2-3pm hour.

- The adjustment improves the estimated baseline load to match the actual load more closely.
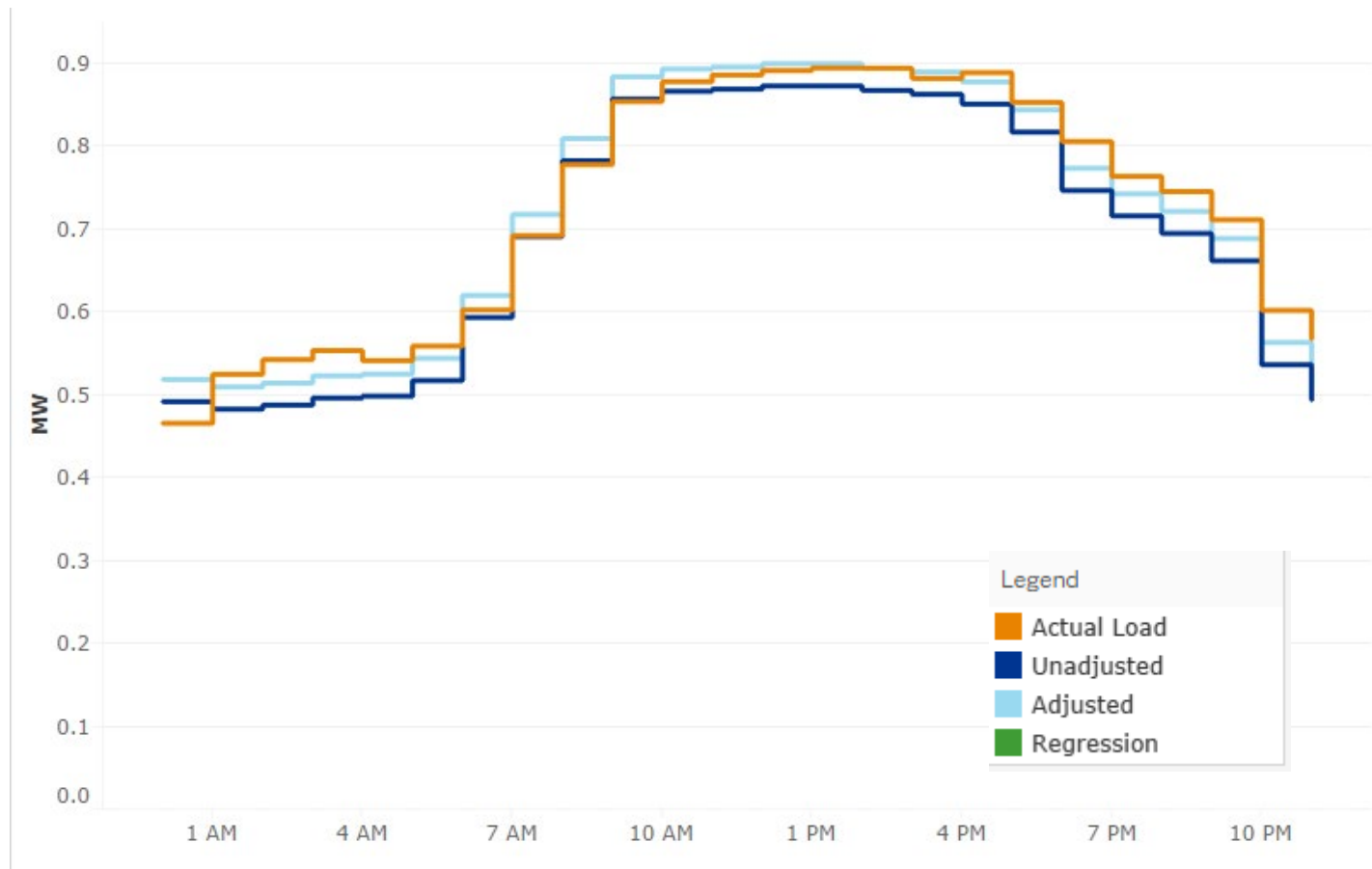
# Basic DR Example – Full Season Load

- Zoom out and look at this site load over the whole summer.

- Well behaved, yes, but at two different levels. The regression gets the daily shape approximately right, but it splits the difference between modes within months.

- In August it appears to do better, but that is because almost the whole month is at that level.

- The regression baseline does not deal well with non-weather correlated variability.
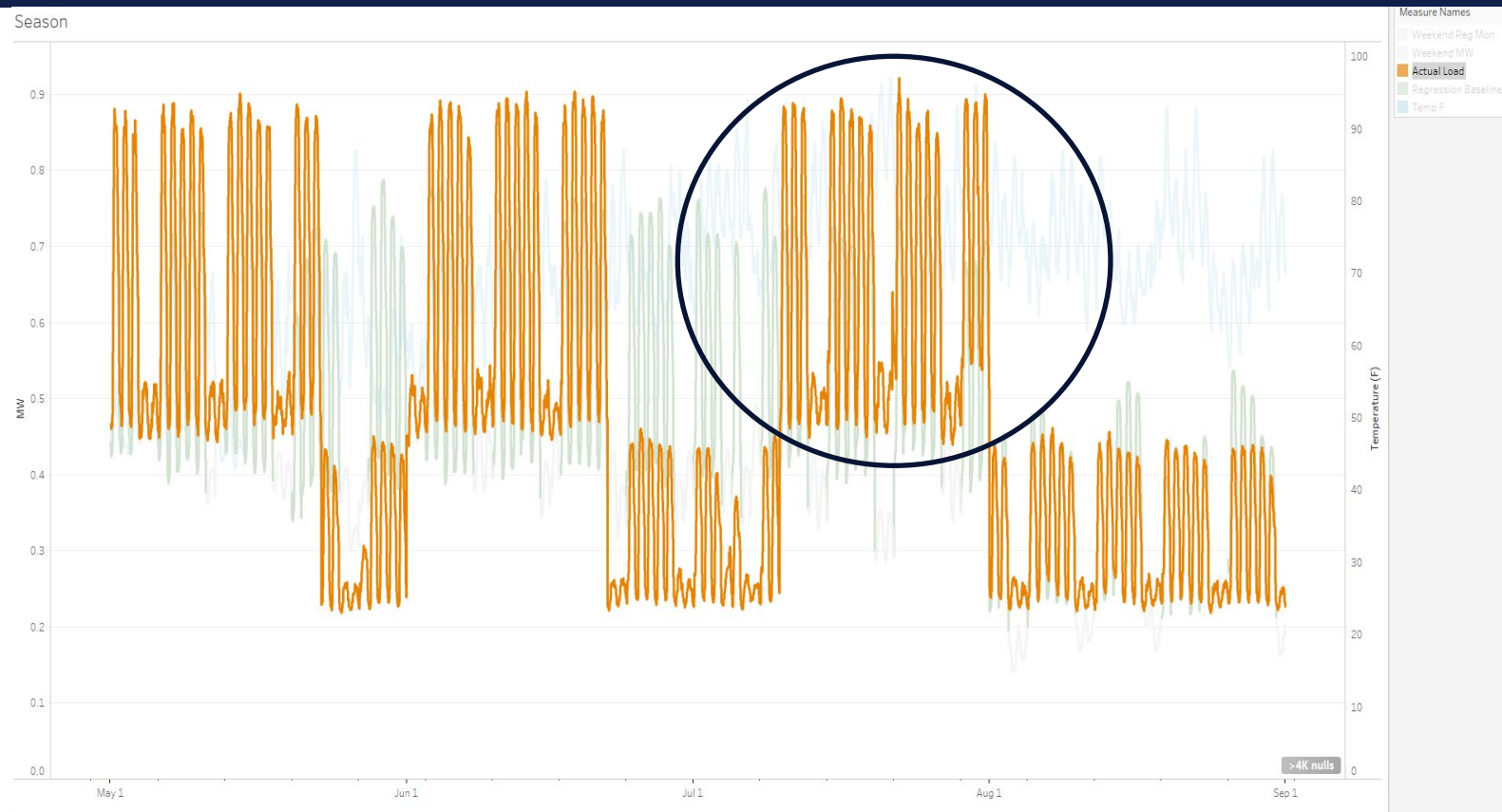
# Basic DR Example – Regression Baseline

- Add in the regression baseline. It has approximately the right shape but misses the magnitude by 25%.

- If this customer reduced load by up to 25%, this baseline would not give credit. In fact, it would appear as negative load reduction and, when aggregated, would negate load reduction identified with other customers.
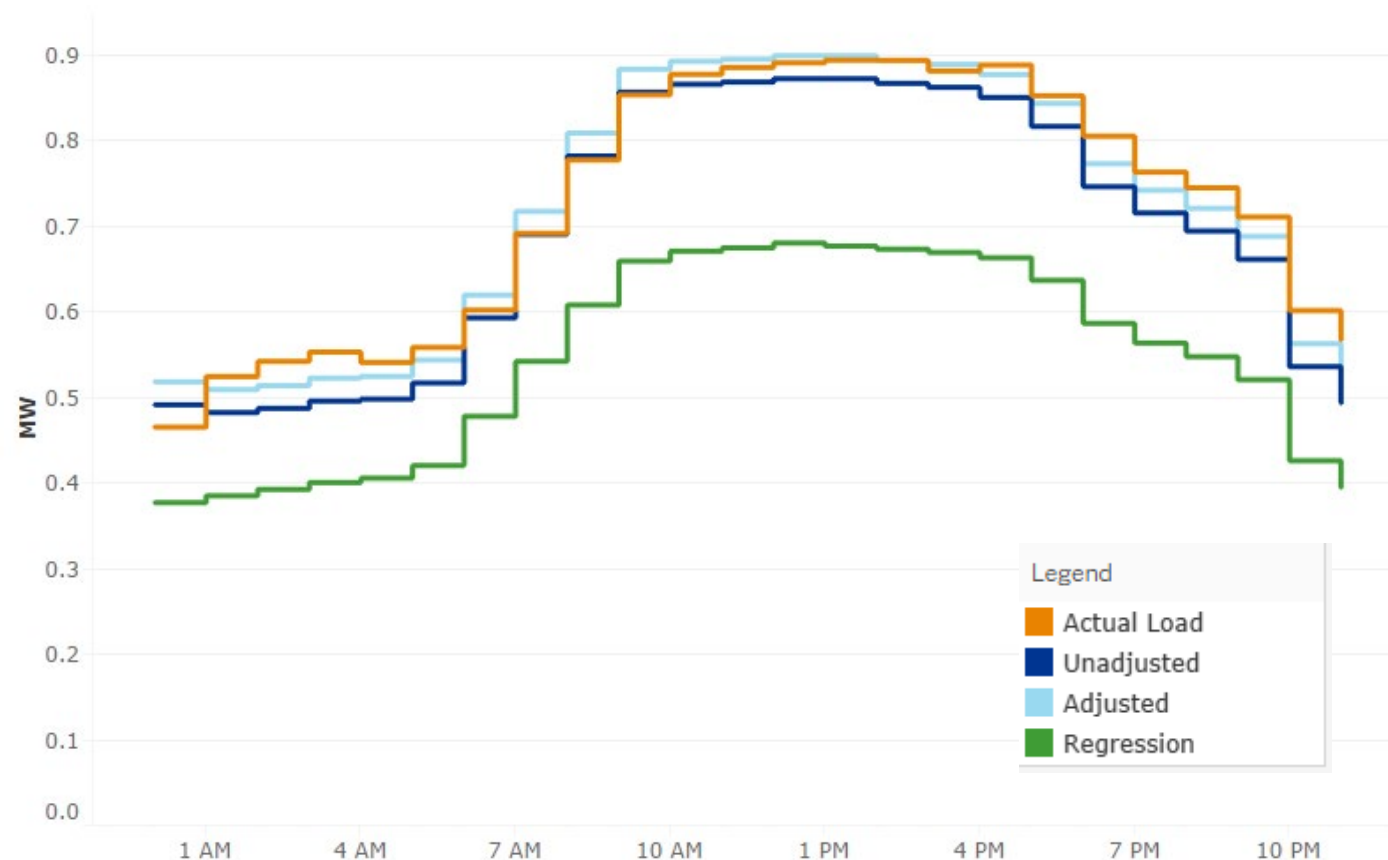
# Weather correlated load- Shut down

- July 5th, this customer stayed shut down for that Friday.

- Both regression and unadjusted could give too much load reduction.

- In the interest of not penalizing the customer who shuts down early, the asymmetric baseline rewards an unusually low load relative to baseline

- Effectively unbiased if you test across all days, but on any given day, there could be a significant bias.
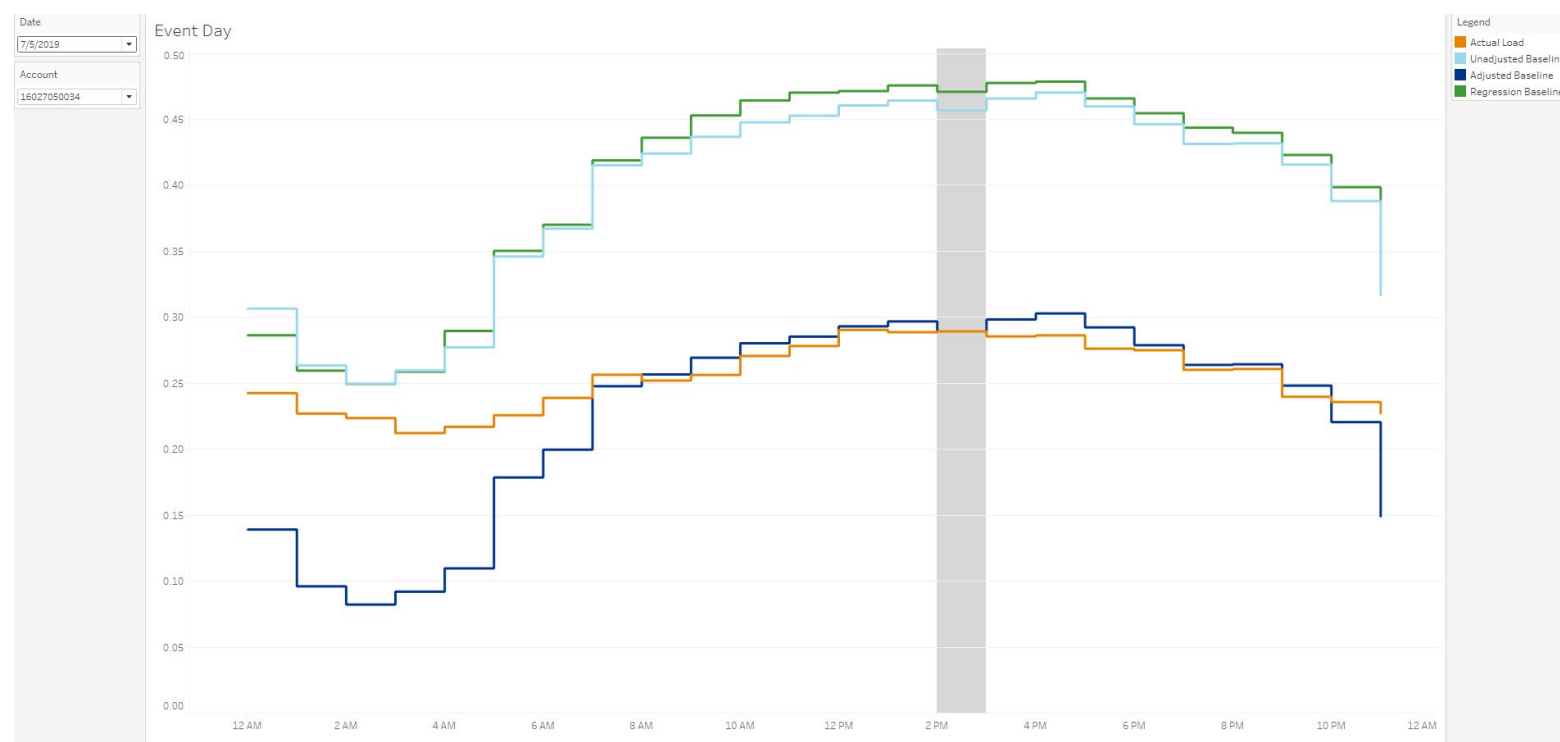
# Comparison of Two Program Years

# 2019 Results (Single Event)

- ICAP day event results for all ADR Initiative customers.

- 5-6pm is ICAP hour (right-most hour in shaded event period)



Event - 7/30/2019

Legend: Load, Unadjusted, Symmetric, Asymmetric, Regression
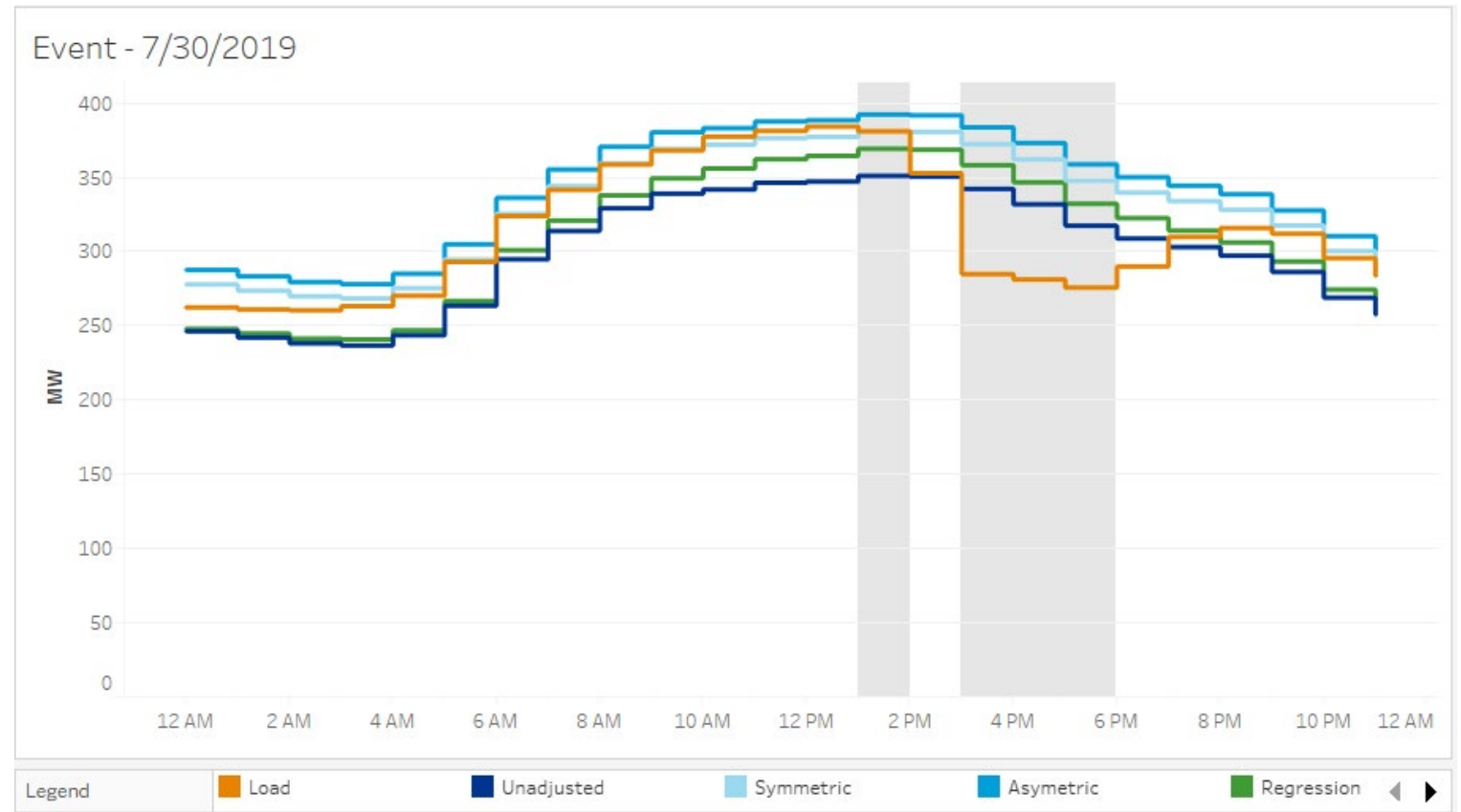
# 2018 Results

- Three days during the week of July 4<sup>th</sup> when the regression is as much as double the settlement baseline (symmetrically adjusted).

- Three days when almost half

- The more event days there are, the more likely it is that seasonal average performance will converge between metrics
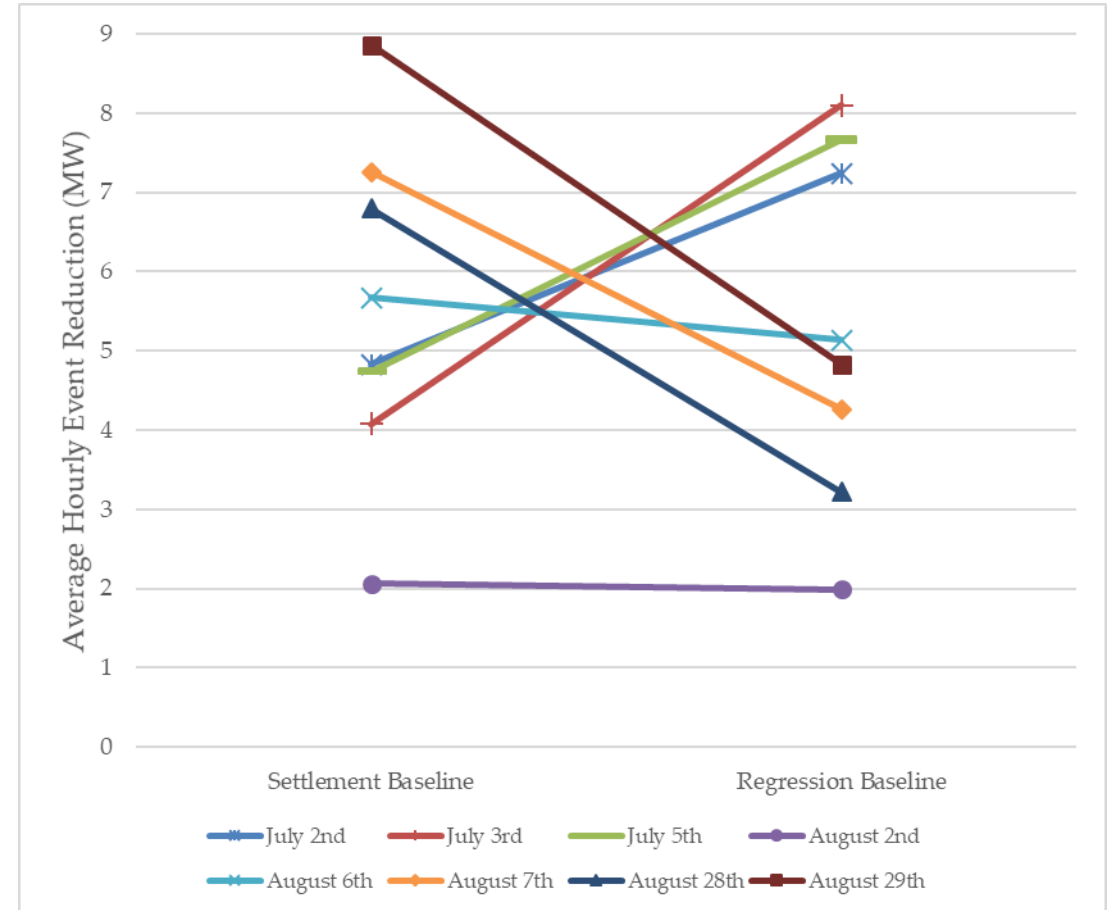
# Summary of Findings and Conclusions

# Summary of Findings and Conclusions

- A fundamental difference between load reduction and generation as a resource is that it is not possible to directly observe load reduction. Instead, measurement of load reduction necessarily means comparing observed load to a counterfactual load that would have occurred in the absence of dispatch.

- The evaluator knows only what we can infer from load data. Information such as nature of end uses under control, weather correlation of loads, shutdowns, etc. is challenging to collect and is typically not available.

- The unadjusted 10-of-10 baseline has minimal potential for baseline manipulation. However, the unadjusted 10-of-10 is inflexible for customers with weather-sensitive load or load variability unrelated to weather and can substantially understate load reduction for customers with weather sensitivity if events are called on extreme hot days.

DNV

IEPEC

# Summary of Findings and Conclusions (continued)

- The asymmetrically adjusted 10-of-10 baseline avoids penalizing customers that decrease load in preparation for an event (e.g., shift cancellation). However, it fails to make downward adjustments for customers where load is less than baseline due to event-day conditions unrelated to legitimate actions taken in preparation for the event. Its certain upward bias makes it unattractive for program evaluation.

- Symmetrically adjusted 10-of-10 baselines generally improve the accuracy of baseline estimates by reducing biases for customers with weather-sensitive load or load variability unrelated to weather. However, this baseline can substantially understate load reduction for customers that decrease load in anticipation or preparation for an event (e.g., shift cancellation), while pre-cooling can inflate the adjusted baseline for weather-sensitive customers.

# Summary of Findings and Conclusions (continued)

- The regression baseline has the advantage of considering more data (across the season, year, etc.), however, it still can only control for known independent variables such as weather, day of the week, and time of day. Unexplained variation such as periodic facility shutdowns or changes in production unrelated to the time of day and day of week can reduce the effectiveness of the regression baseline. Regression baselines work best when all independent variables that affect load are specified and known. The more events there are in a season, the more accurate a regression baseline is likely to be.

- Ultimately, no evaluation methodology is one size fits all. In many instances, multiple evaluation methodologies can provide context for each other to provide a better understanding of the load reduction that occurred.