Lawrence Berkeley National Laboratory

Behavior Analytics

Providing insights that enable evidence-based, data-driven decisions

Taking advantage of smart meter data:

combining behavioral economics with data science analytics

Peter Cappers & Annika Todd, PhD

IRP Contemporary Issues Technical Conference

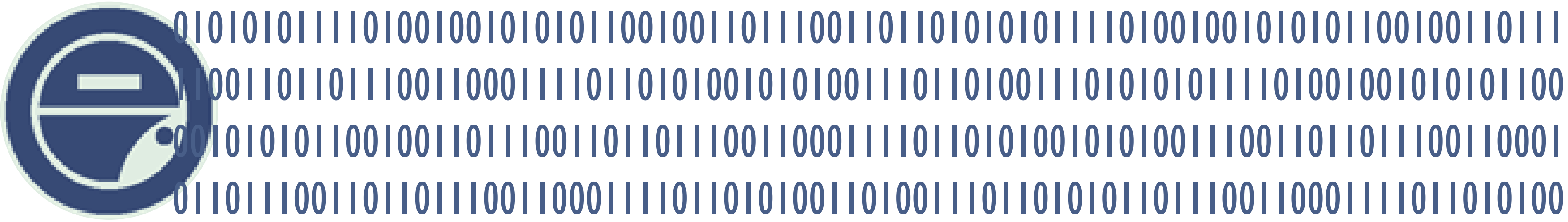
April 24, 2018

Data explosion in energy

- AMI, thermostats, appliances, cars
- Linked to other time and location-specific information (temperature, census, satellite)
- Provide vast, constantly growing streams of rich data

Smart meter data enables many possibilities for cutting edge analyses

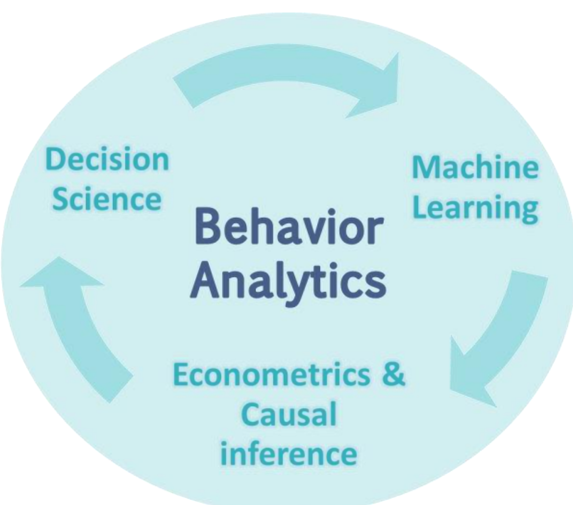
- What can we do with all of this data?



- Many possibilities!
- Insights from the data → tremendous potential value for a wide range of energy programs, policies, and overall grid integration.

Our solution: *Combine behavioral economics with data science*

Using only *easily accessible* data from smart meters and other sources



Better understand:

- Customers' **energy characteristics**
- Customers' energy **usage behaviors**

Implications and uses for:

1. Load forecasting
2. Utility planning
3. Increasing cost effectiveness of rates and DSM programs (existing or new)

Main Takeaway:

1. **Lots of things you can do with smart meter data** (5 examples)

2. **Some can be really useful, and some aren't** (insist on seeing results)

3. **Let's just do a lot of quick A/B testing and analysis – what actually works? What should we try next?**
    - Test big things (program validity), small things (best wording for marketing messages), test continuously

Examples that we have done

Dataset for these examples

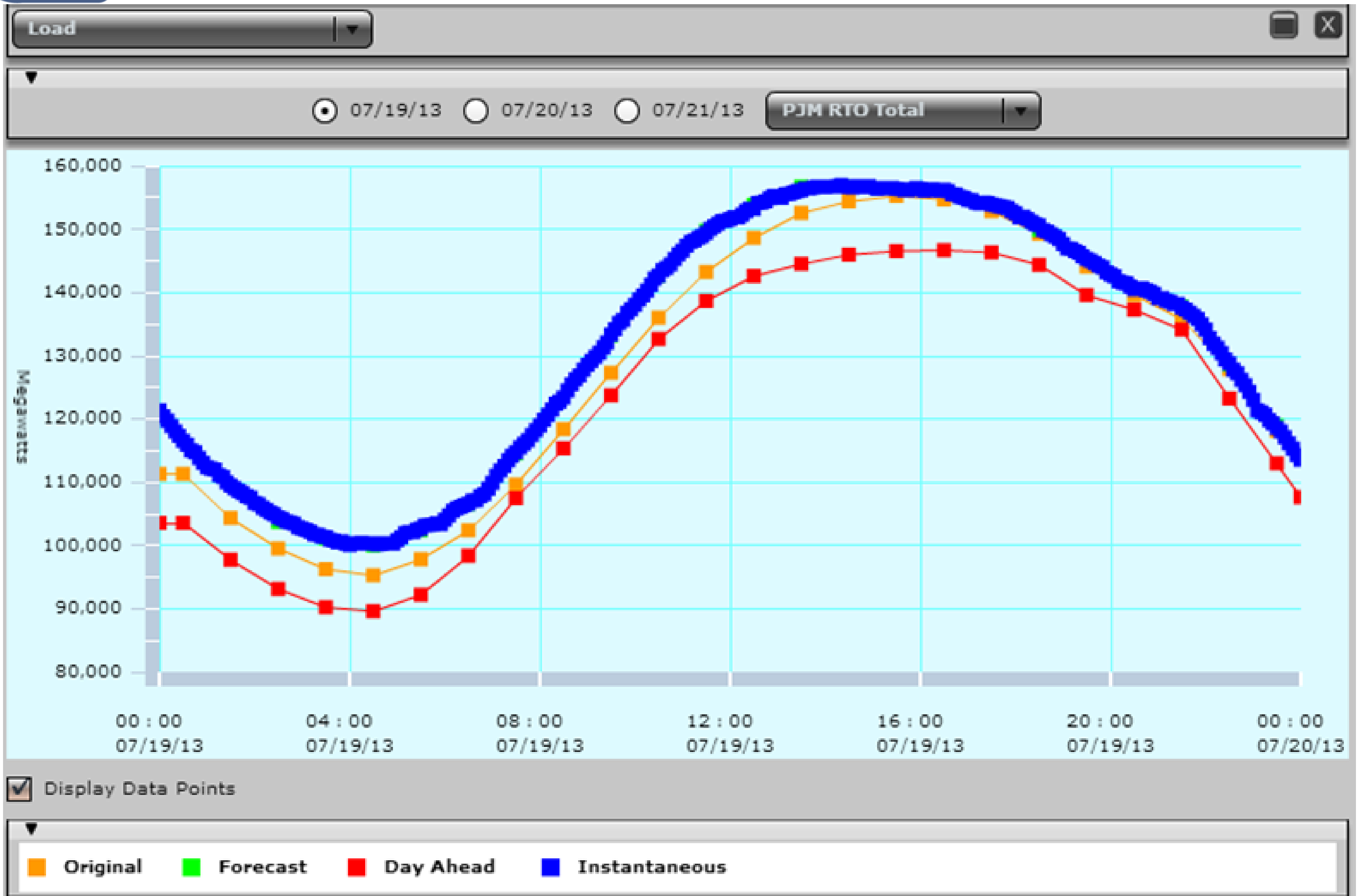
- Residential hourly electricity data
  - 100,000 households
- A region with usage peaks in the summer time
- Pilots for TOU and CPP rates
- Randomized controlled trial of these new rates
  - Households are randomly placed in different treatment groups
  - Randomized control group to compare to
- Over 3 years of data
  - One year prior to new rates
  - Two years once rates start

Example 1

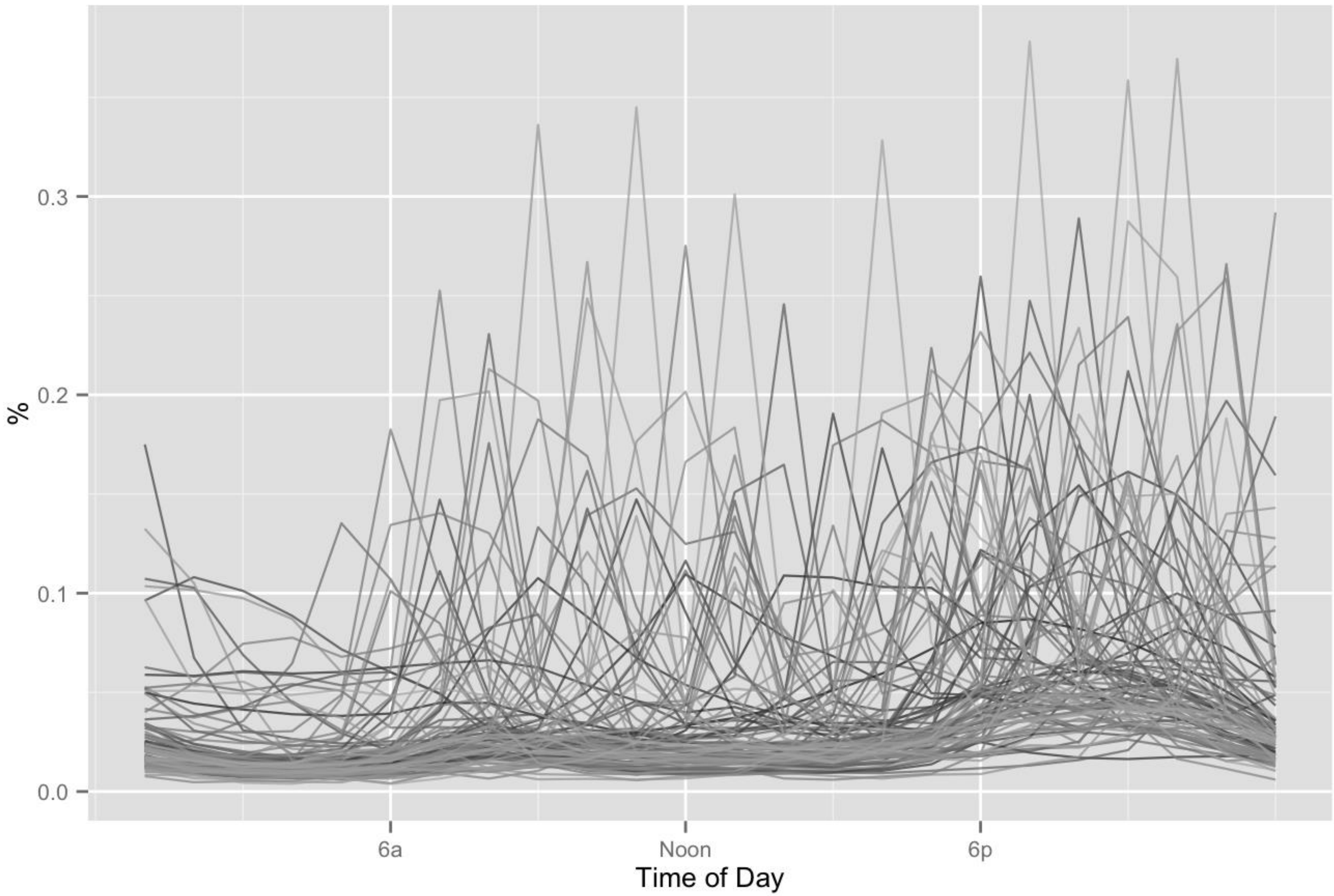
Cluster load shape patterns

Form groups of households with similar load shapes

What the grid sees – aggregate load shape from everyone on the system



Wide variety of load patterns across customers
(even customers who appeared to be similar)

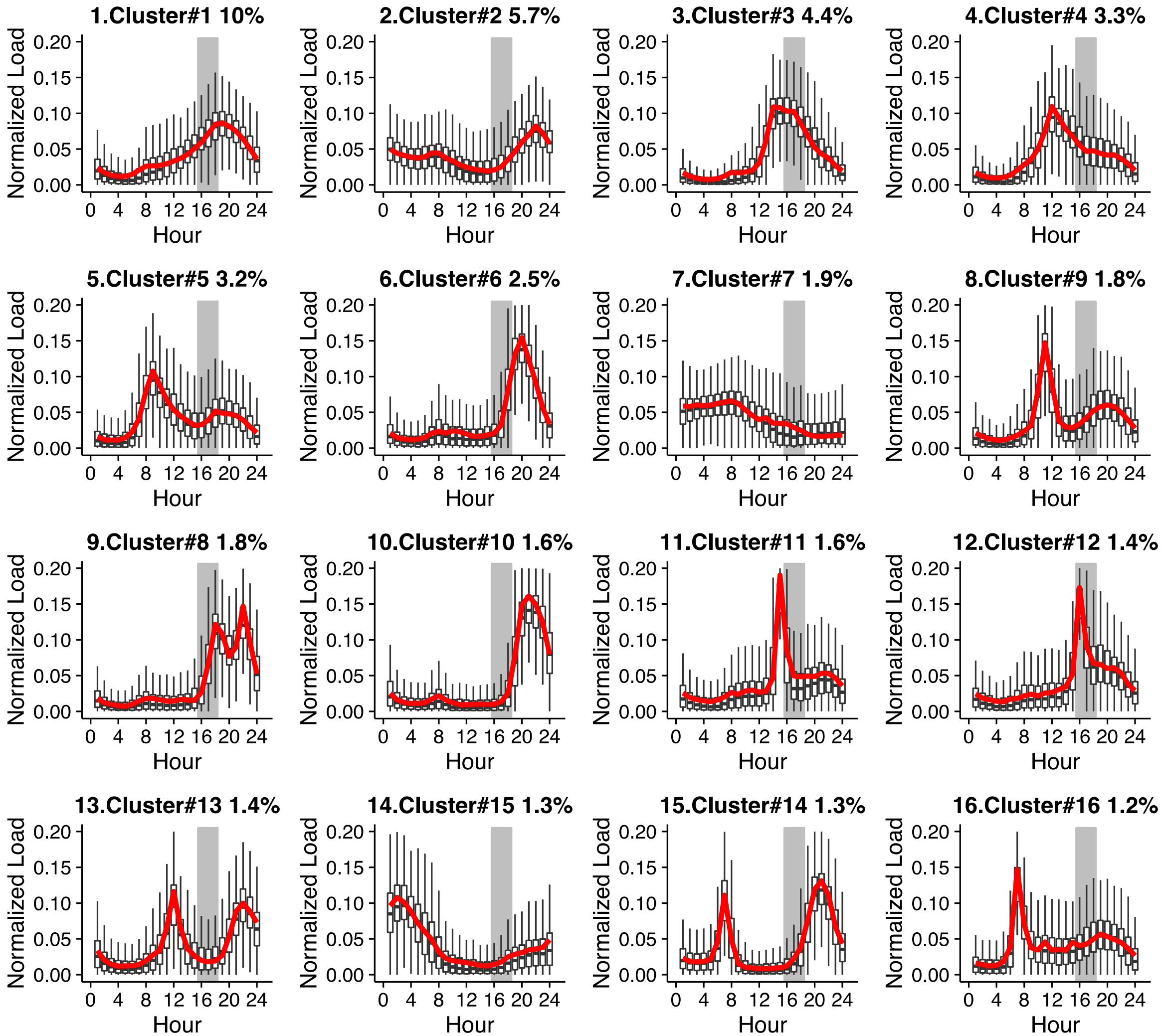


Cluster load shape patterns
Form groups of similar load shapes

- Let machine learning show you patterns of energy usage characteristics
- Cluster all of the various types of households' daily load shape patterns, to form groups that are similar to each other

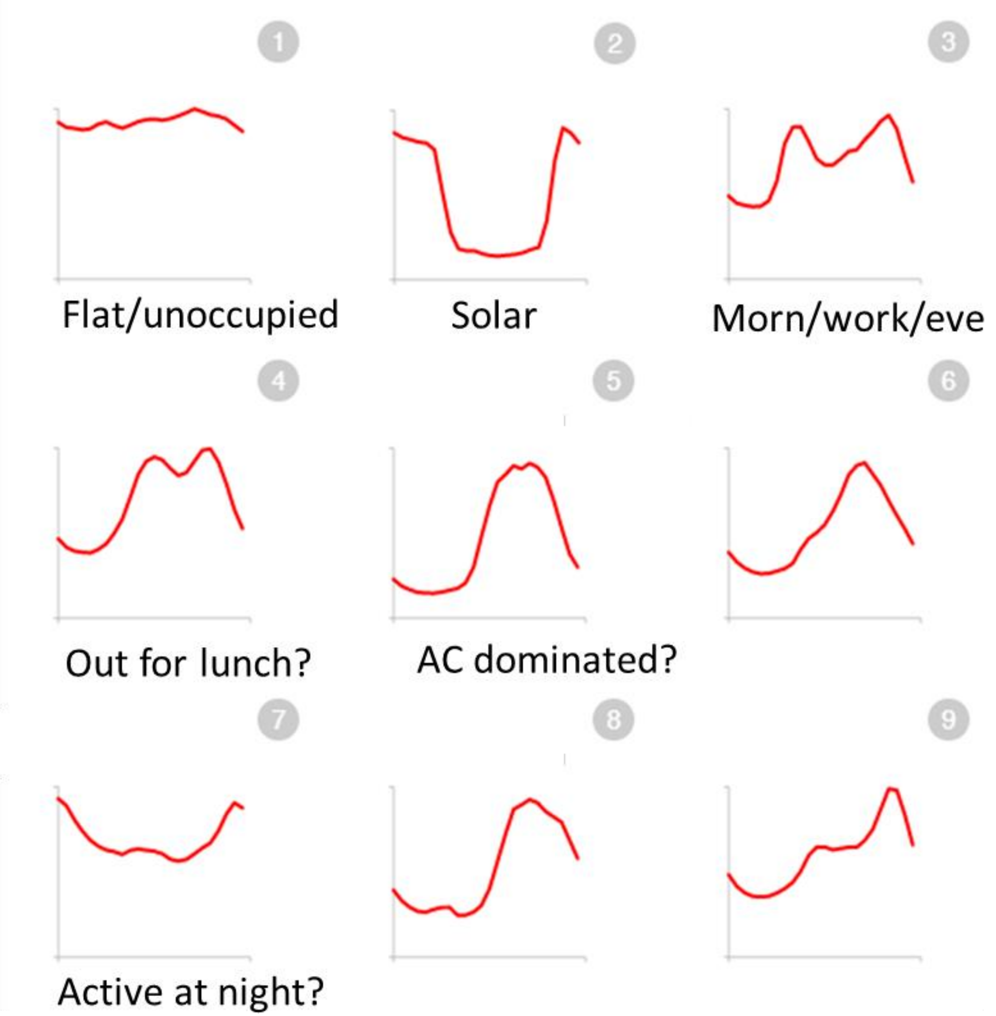
Use algorithms to cluster load shapes.

99 cluster groups: these 16 are the biggest



Look at “Representative Load shapes”

Better predictions of current/future energy use



Example 2

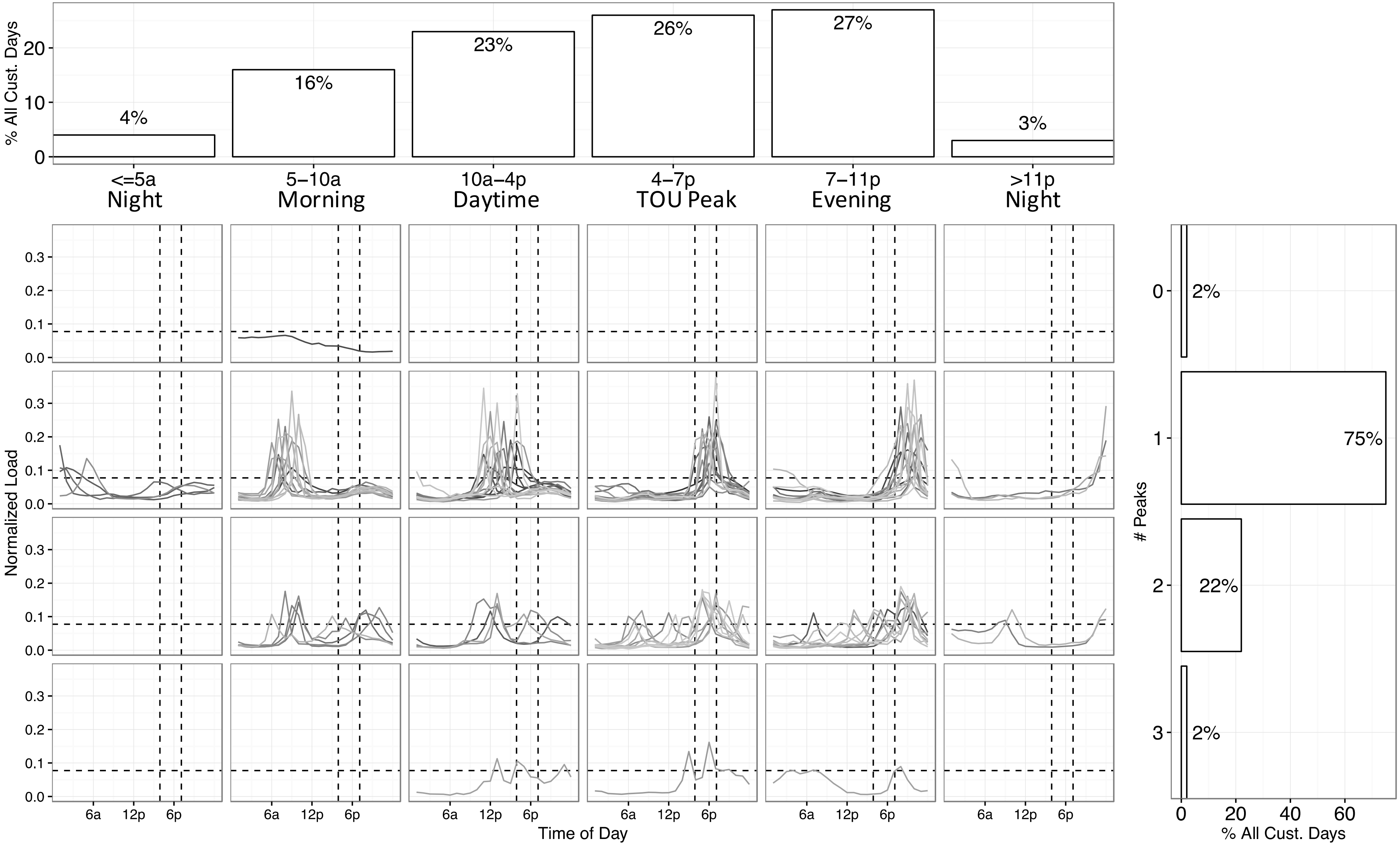
Look at distribution of load shape clusters across…

Number of peaks

When the peaks occur

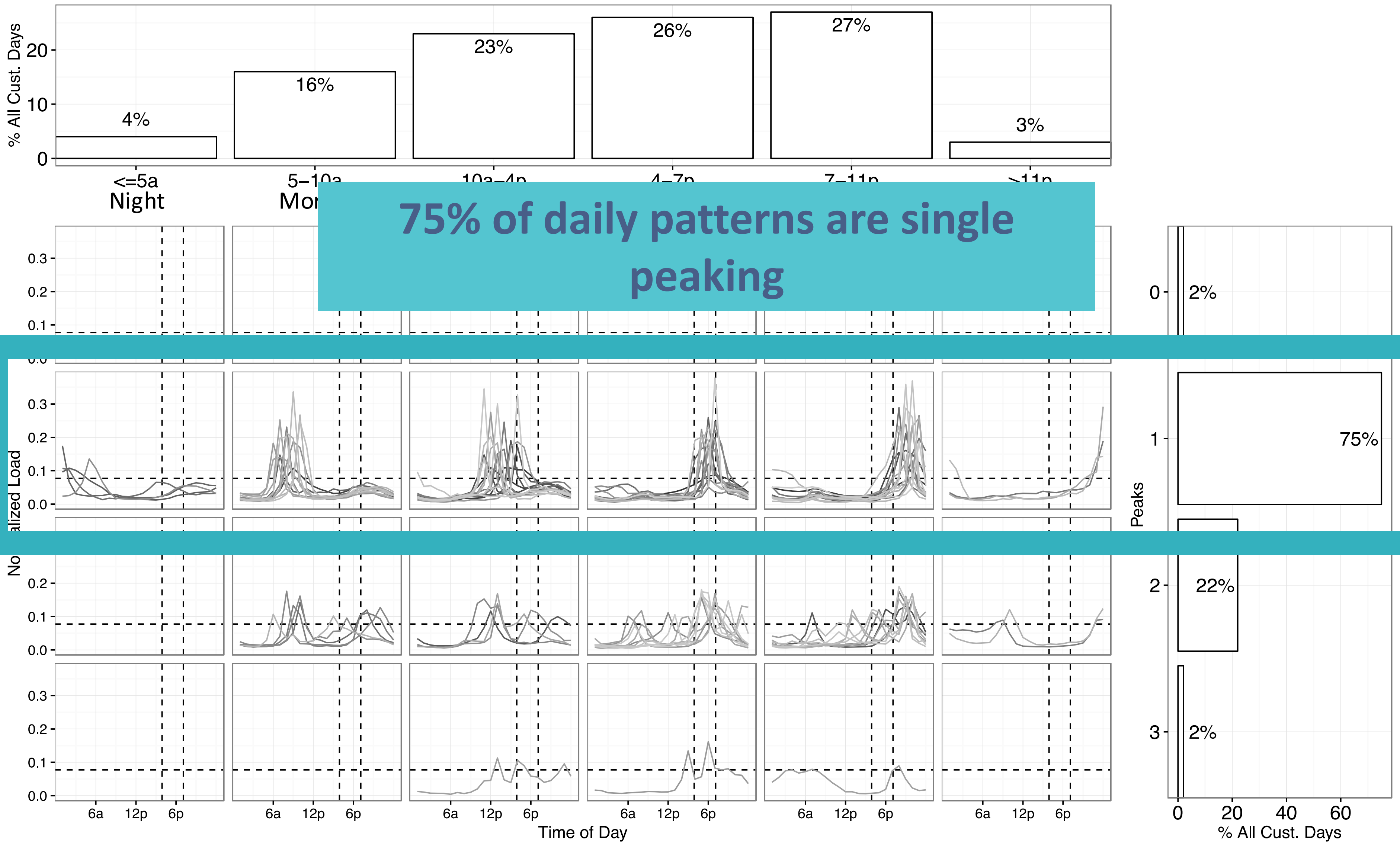
Group clusters based on when peaks occur

Different # of peaks at different times of the day



Group clusters based on when peaks occur

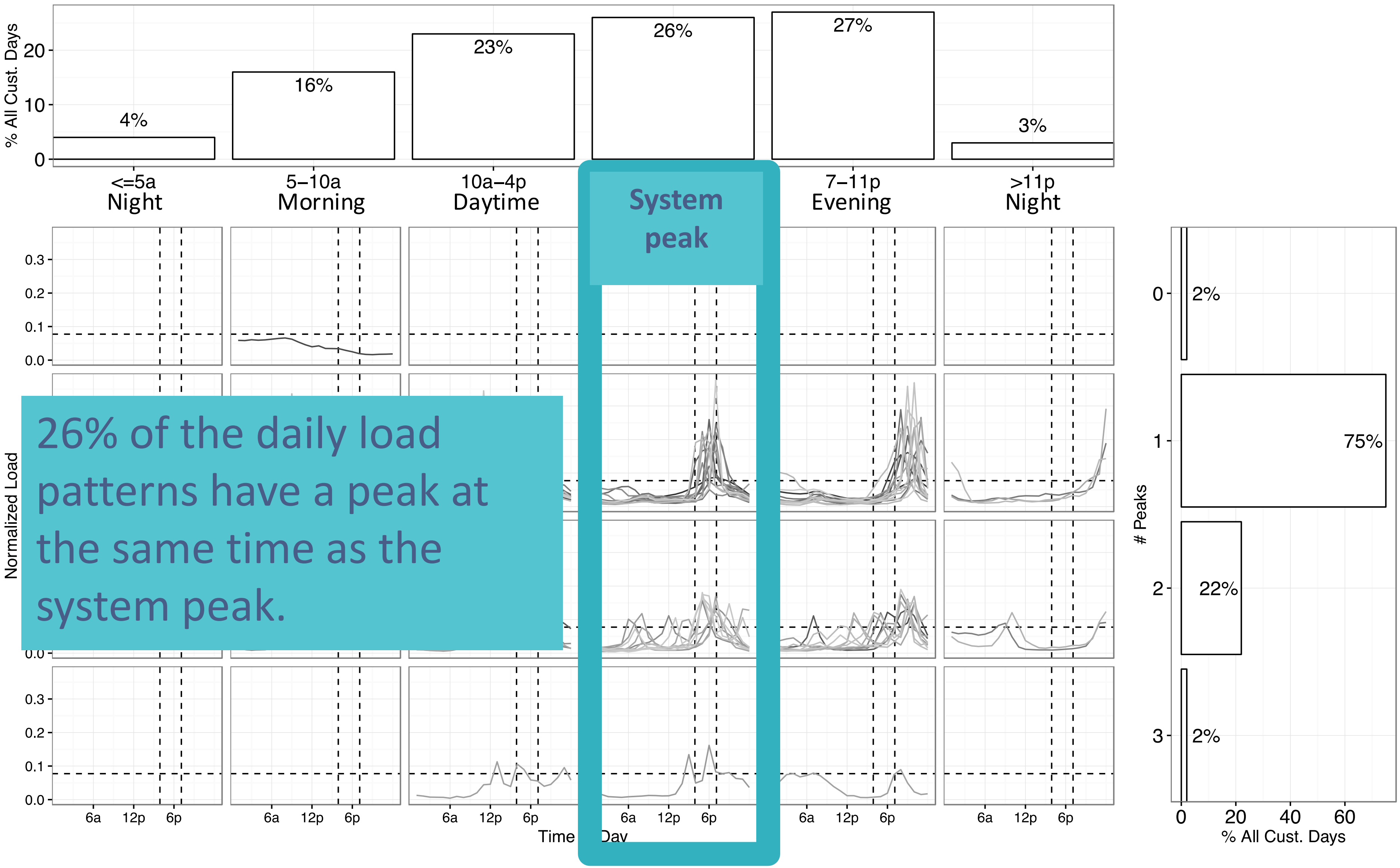
Different # of peaks at different times of the day



75% of daily patterns are single peaking

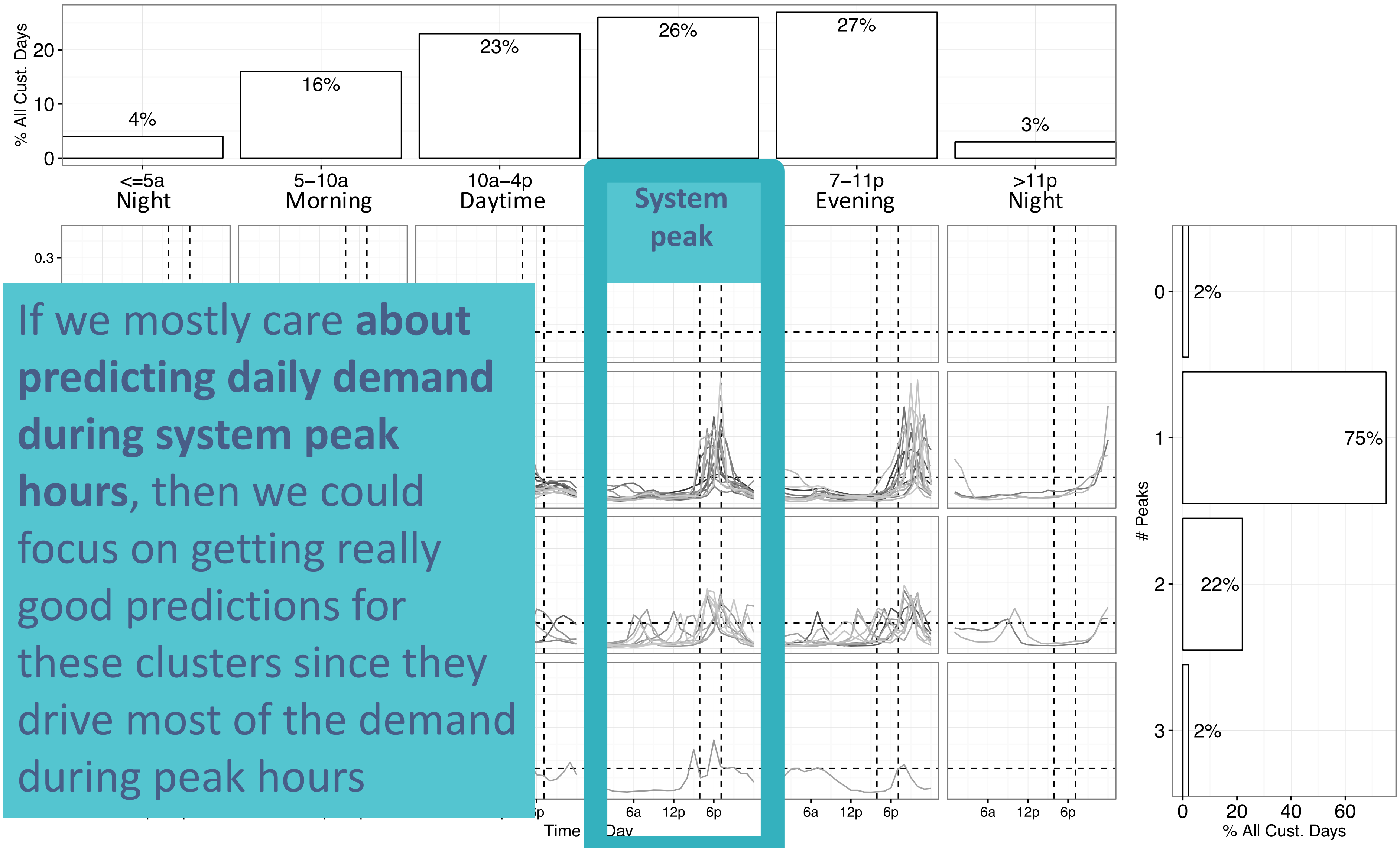
Group clusters based on when peaks occur

Different # of peaks at different times of the day



Group clusters based on when peaks occur

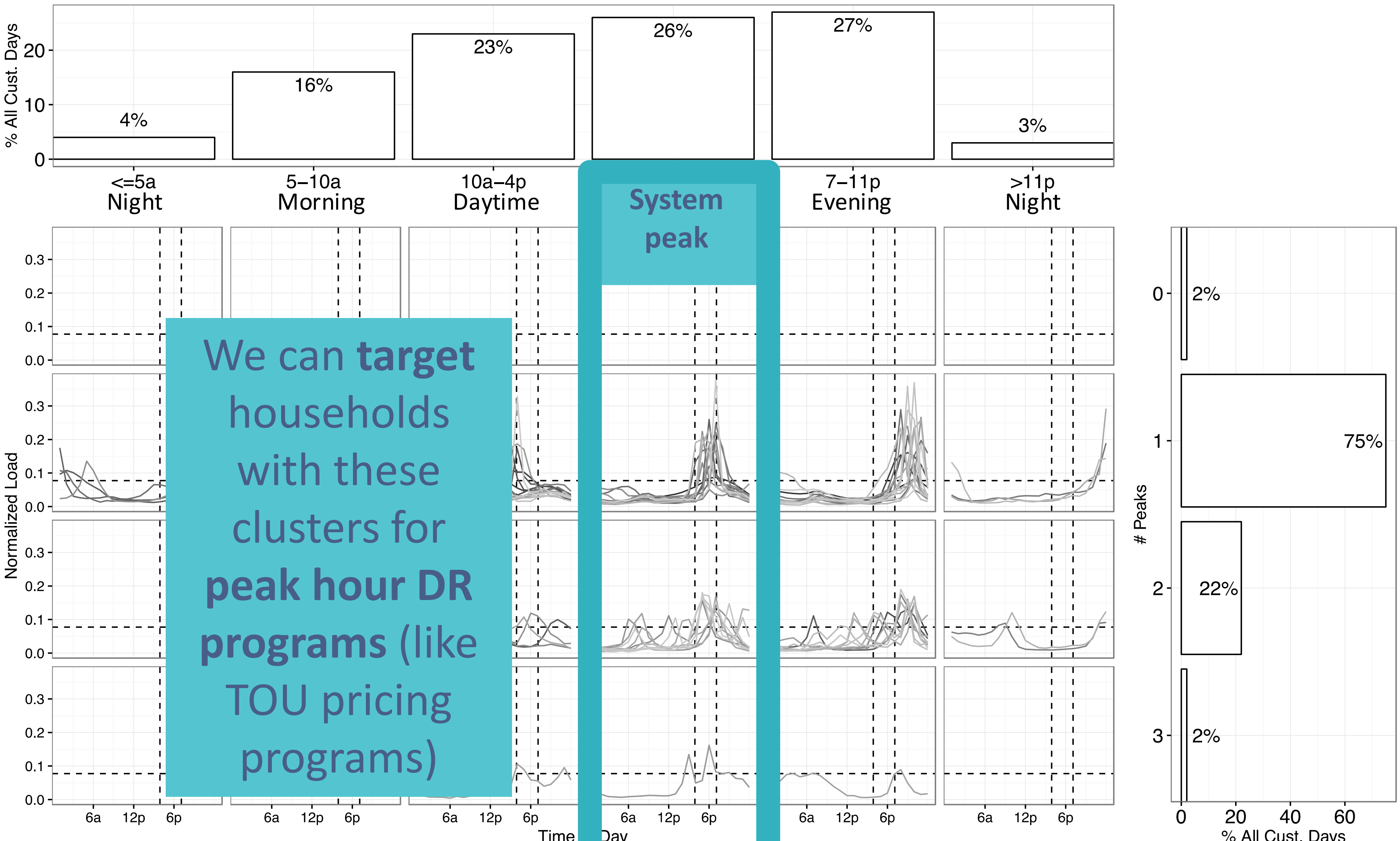
Different # of peaks at different times of the day



If we mostly care **about predicting daily demand during system peak hours**, then we could focus on getting really good predictions for these clusters since they drive most of the demand during peak hours

Group clusters based on when peaks occur

Different # of peaks at different times of the day



We can **target** households with these clusters for **peak hour DR programs** (like TOU pricing programs)

Example 3

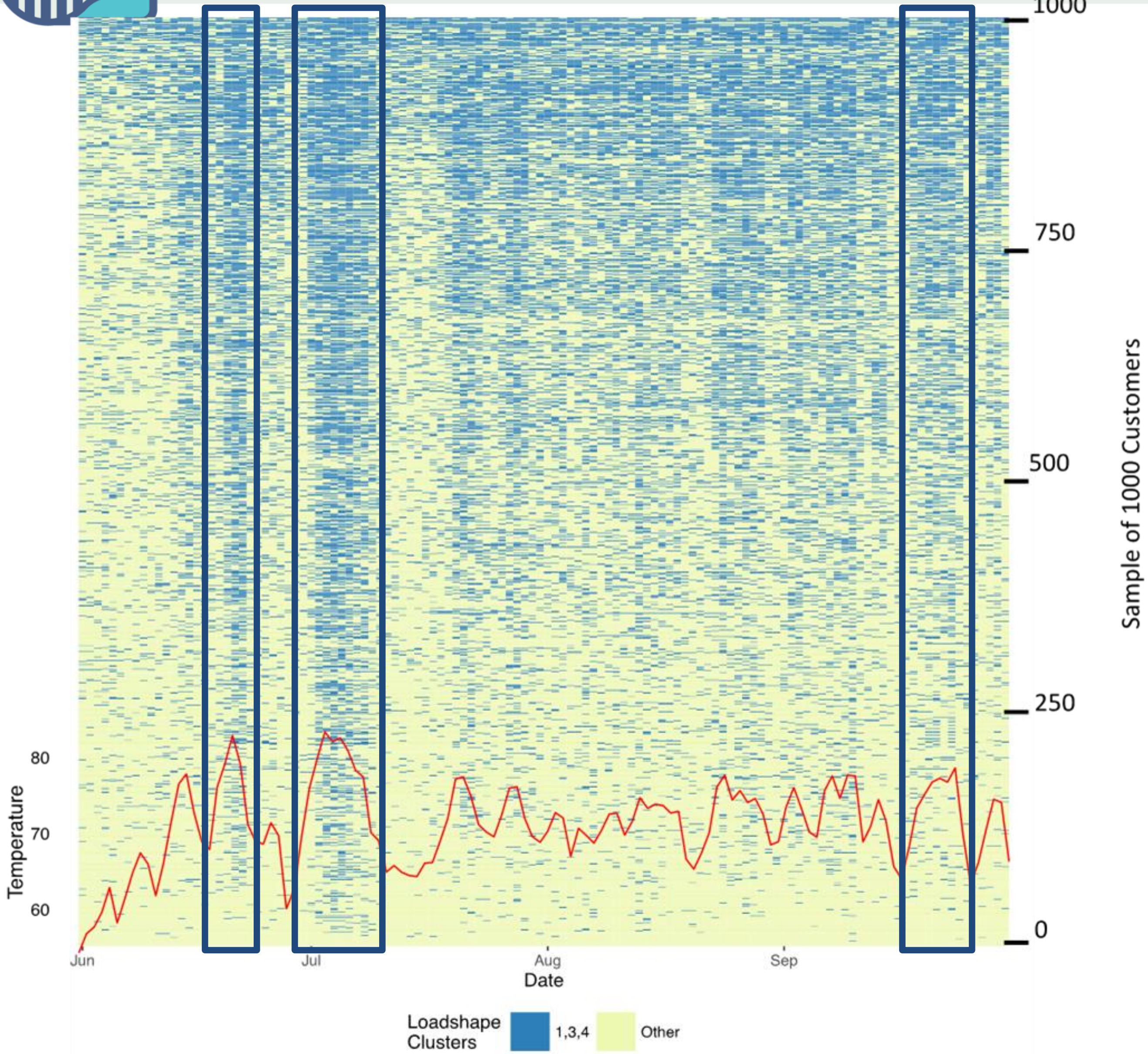
Look at distribution of load shape clusters across…..

Outdoor temperatures

Day of week

Season of the year

Group clusters by temperature and contribution to usage

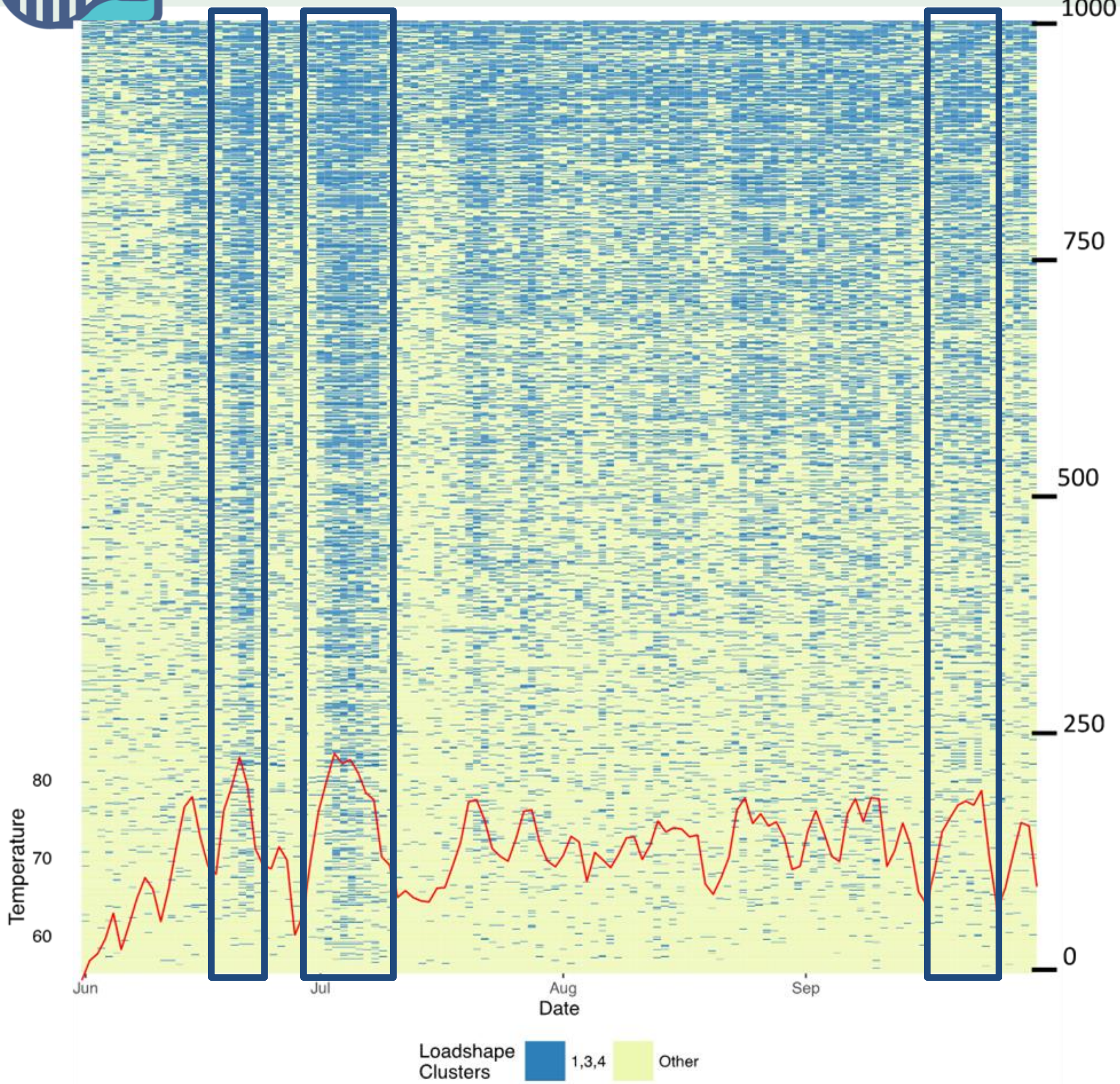


What we are seeing: The blue dots are only 3 of the clusters, and yellow is all of the others (96 other clusters)

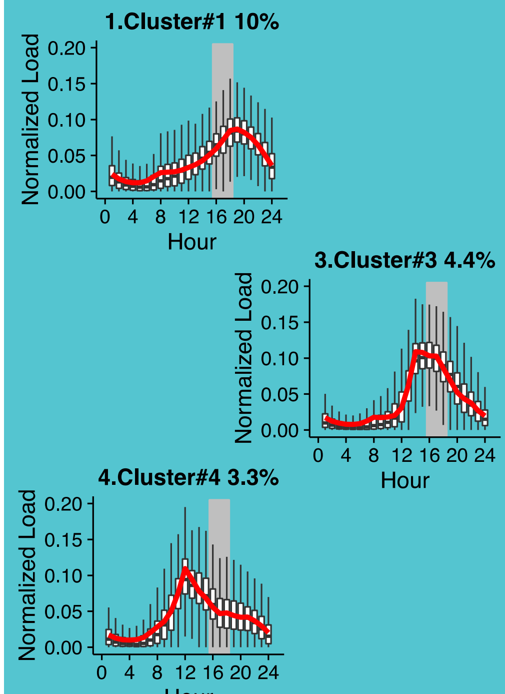
On hot days (where the red line is high), there are more blue dots than on other days.

This means….

Group clusters by temperature and contribution to usage



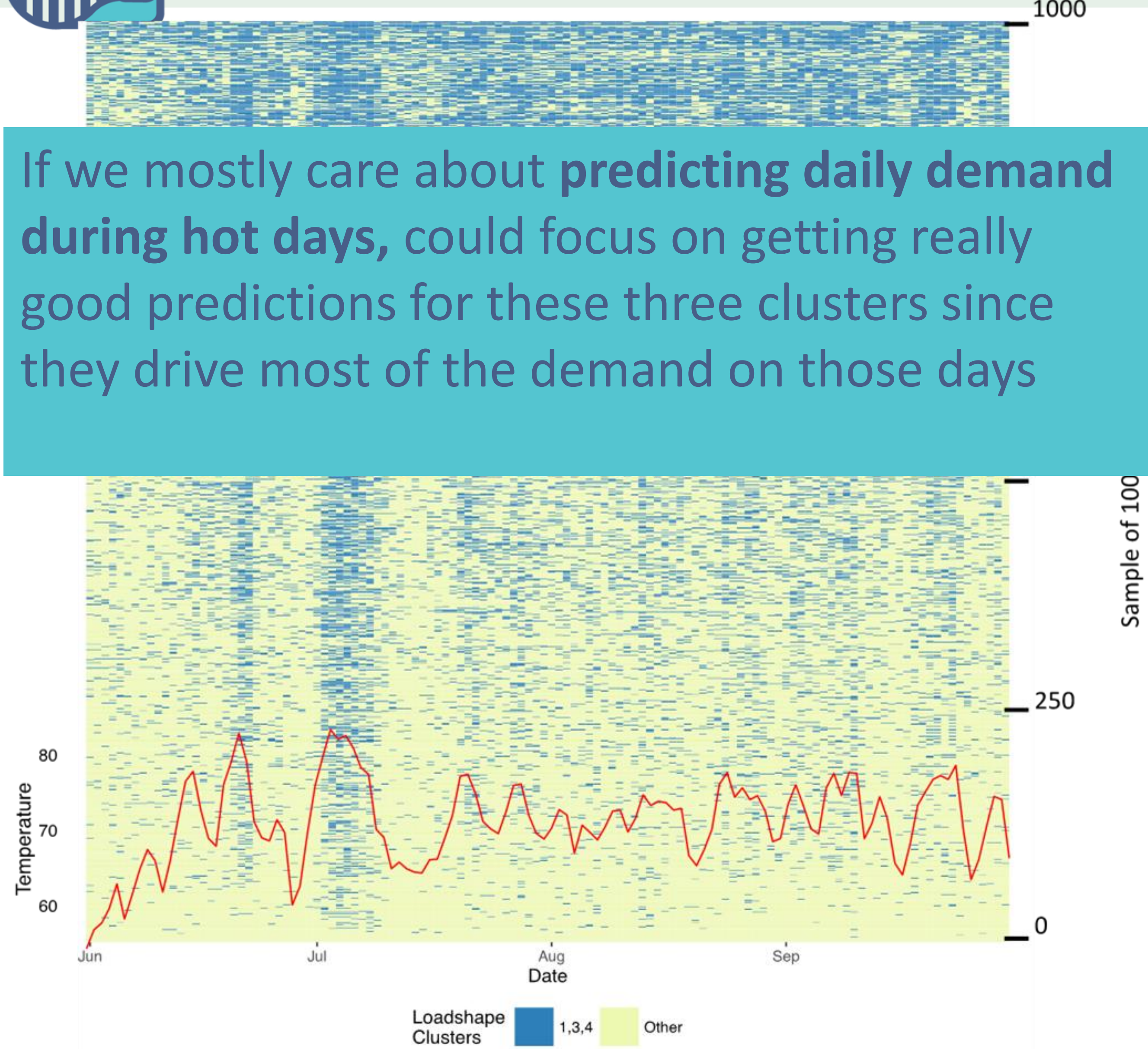
These 3 clusters:



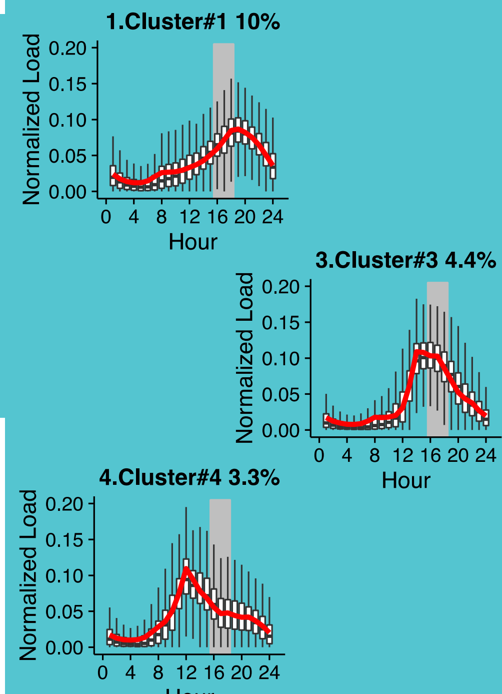
Cover 50% of electricity usage on hottest days

Group clusters by temperature and contribution to usage

If we mostly care about **predicting daily demand during hot days,** could focus on getting really good predictions for these three clusters since they drive most of the demand on those days



These 3 clusters:

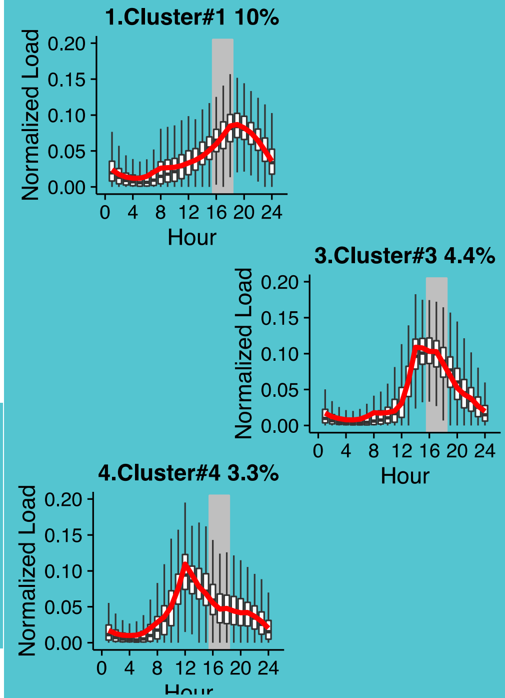


Cover 50% of electricity usage on hottest days

Group clusters by temperature and contribution to usage

We can **target** households with these three clusters for **event-driven DR programs** (like CPP pricing programs)

These 3 clusters:



Cover 50% of electricity usage on hottest days

Example 4

Identify energy characteristics and develop metrics to represent those characteristics

Segment household enrollment & response by energy characteristics

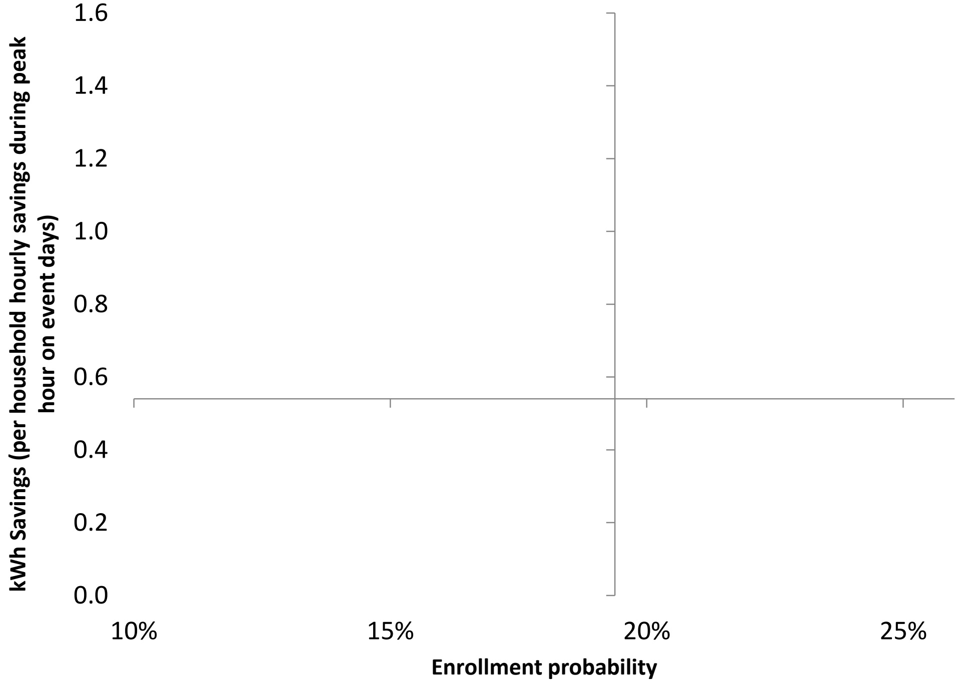
Apply segmentation for targeting, tailoring, and predicting to get better program outcomes

Decide what characteristics are useful, draw these characteristics out of the data

Identified a set of behavioral energy characteristics that we hypothesized should influence a household's willingness to enroll in and respond to time-varying pricing programs

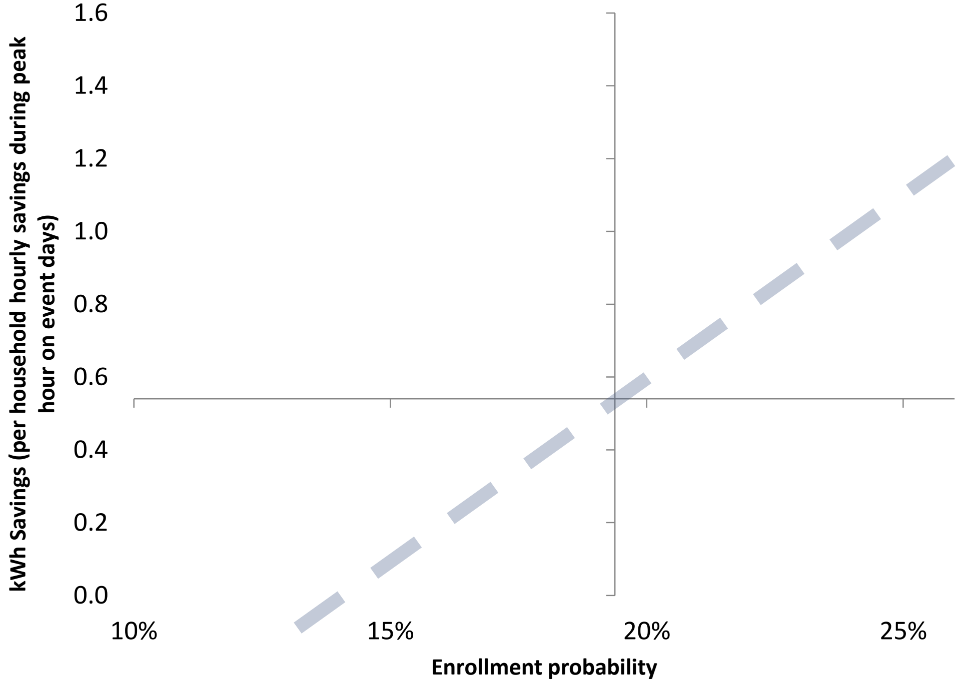
- **Baseload** usage
  - Metric: daily minimum usage
- **Flexibility** of a household's energy use schedule
  - More flexible households may be more able or willing to make changes
  - Metrics measuring variability in electricity usage patterns over time
- **Savings potential**
  - Metric of load magnitude on hot days;
- **Occupancy behavior** of a household
  - Presence of residents during times surrounding the peak periods may make them more able to respond, represented by
  - Metrics of usage during non-typical hours,
- **"Structural winningness"** for a particular type of program (e.g., new rate)
  - Structural winners are households that would receive lower bills on the new rate if they didn't make any changes in their energy usage relative to the prior year (while on the traditional time-invariant electricity rate)

Prototypical Load Shapes
Enrollment vs. Response



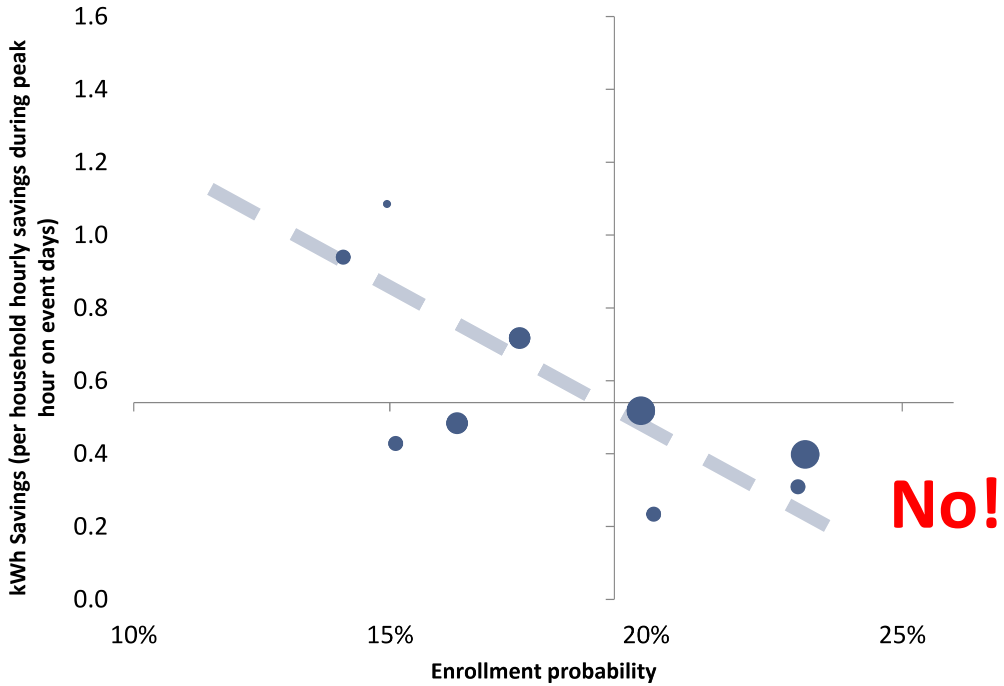
Source: Borgeson et al. (Forthcoming)

Do customers who are more likely to enroll also provide greater load response?



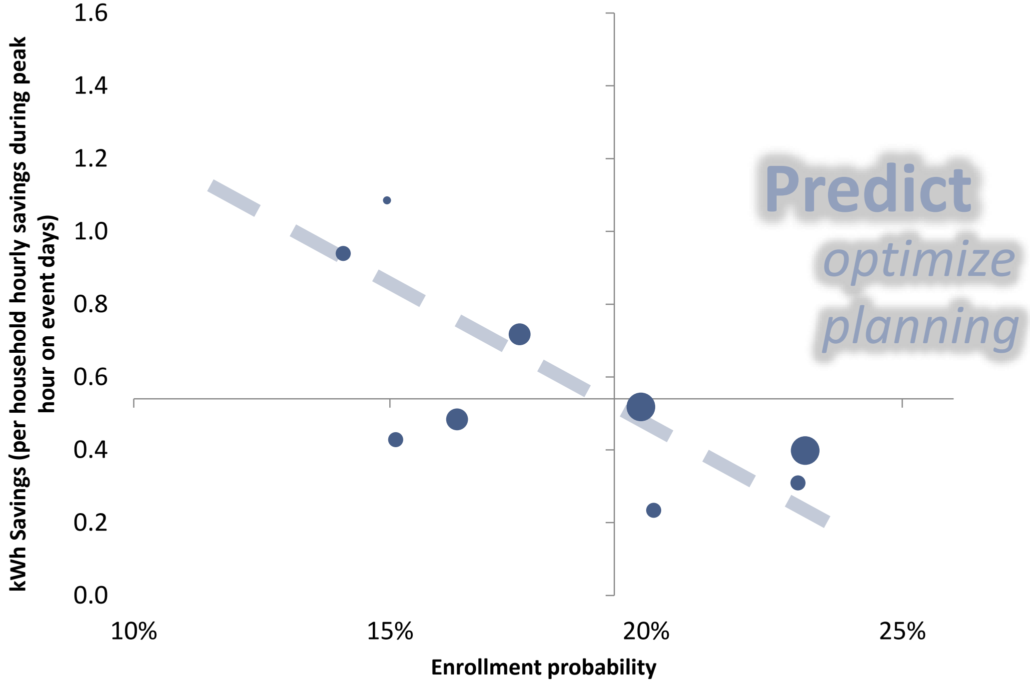
Source: Borgeson et al. (Forthcoming)

Do customers who are more likely to enroll also provide greater load response?



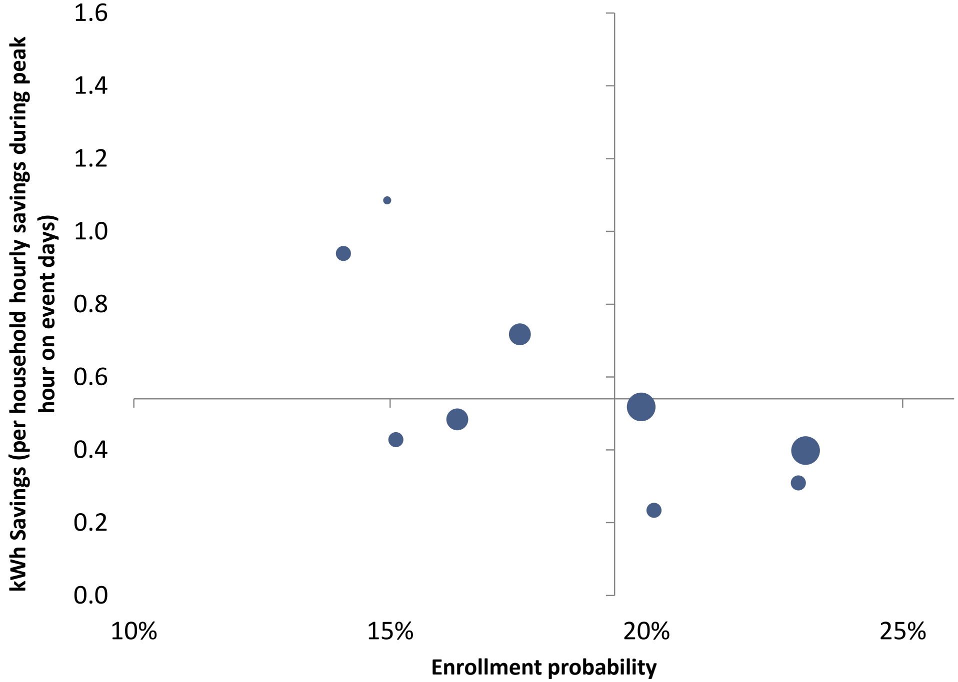
Source: Borgeson et al. (Forthcoming)

Planning Efforts Could Benefit from Knowing Types of Customers based on Enrollment and Responsiveness



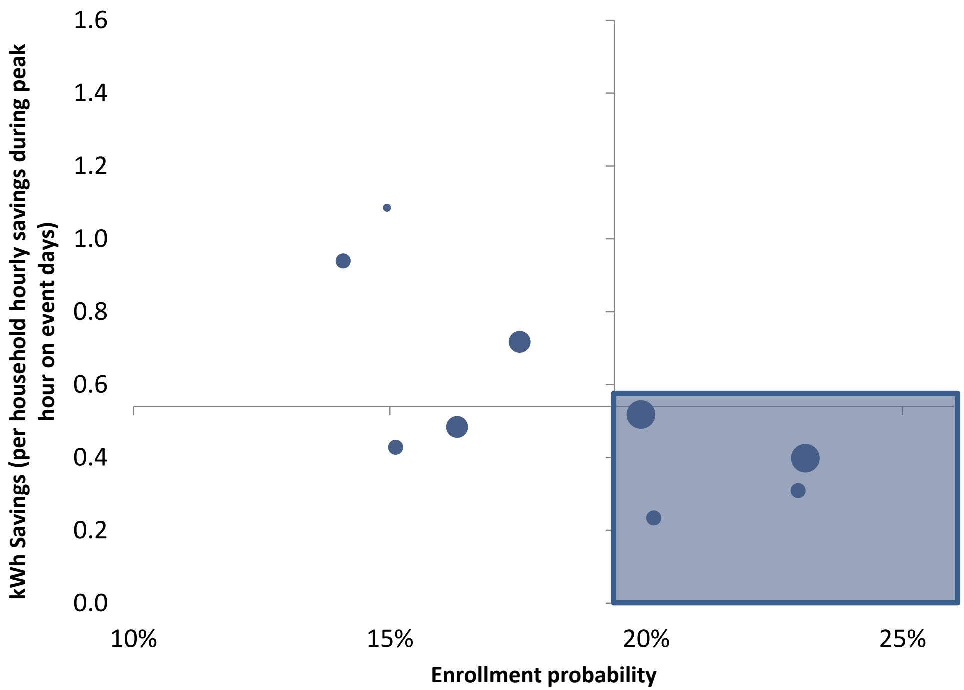
Source: Borgeson et al. (Forthcoming)

Do customers who see greater bill savings (i.e., structural winners) provide less load response?



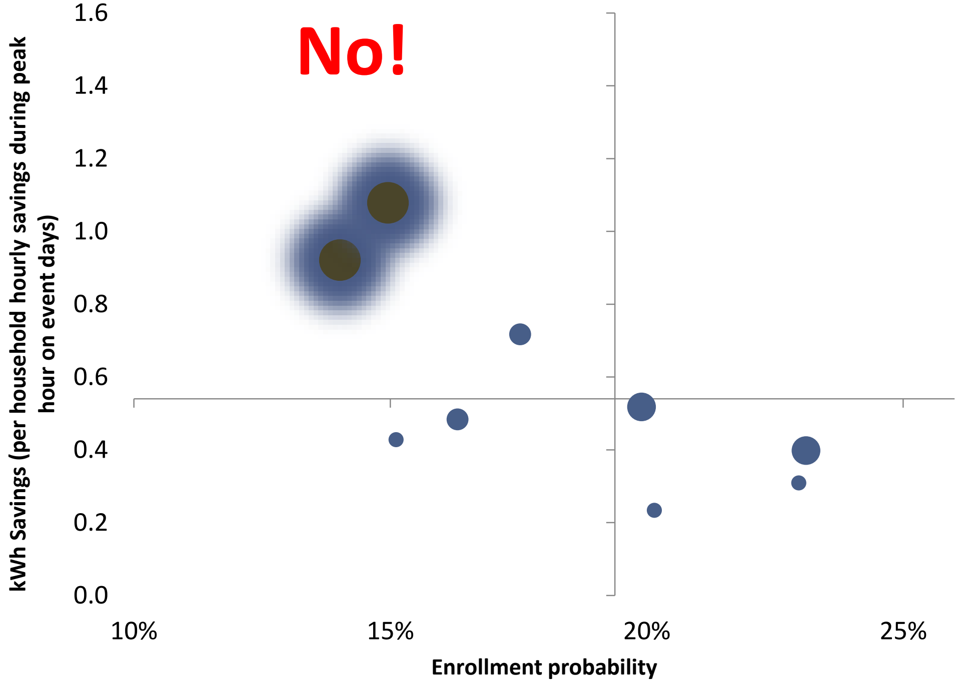
Source: Borgeson et al. (Forthcoming)

Do customers who see greater bill savings (i.e., structural winners) provide less load response?



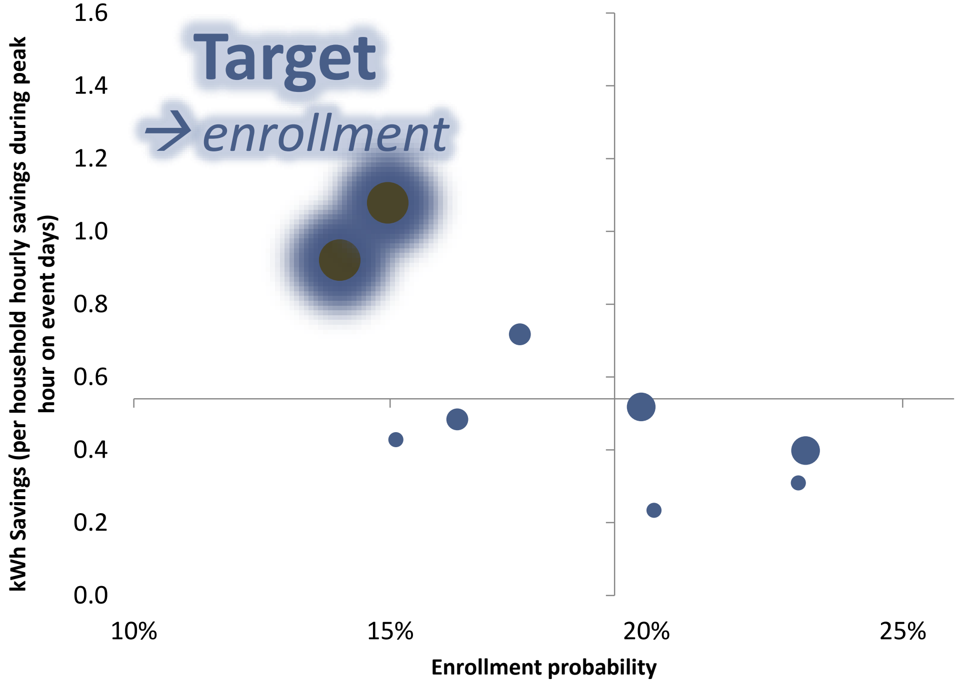
Source: Borgeson et al. (Forthcoming)

Do customers who see greater bill savings (i.e., structural winners) provide less load response?



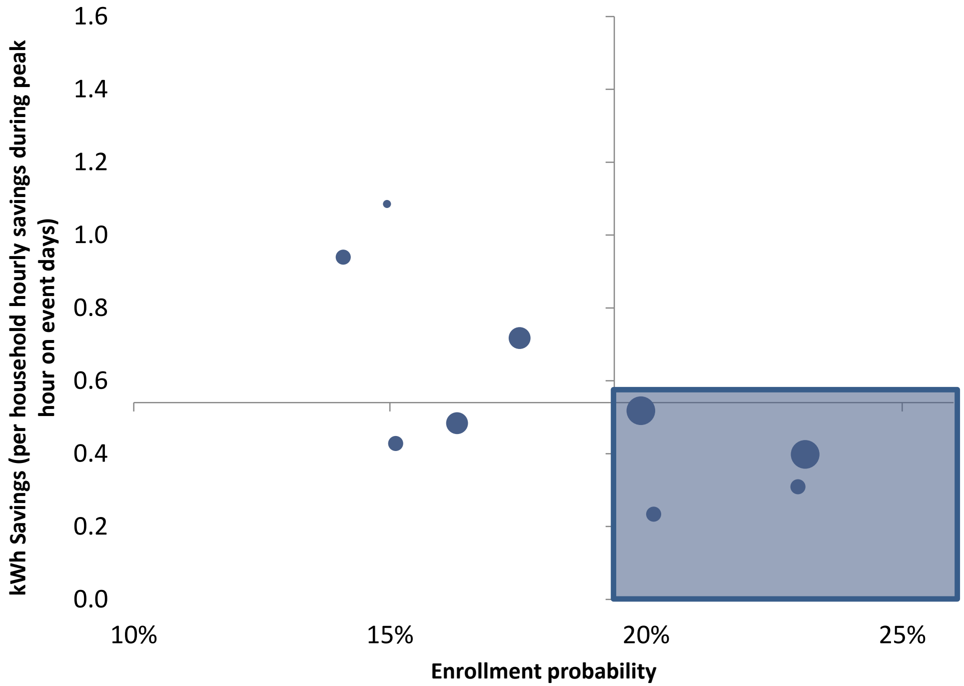
Source: Borgeson et al. (Forthcoming)

Target market to the most responsive customers



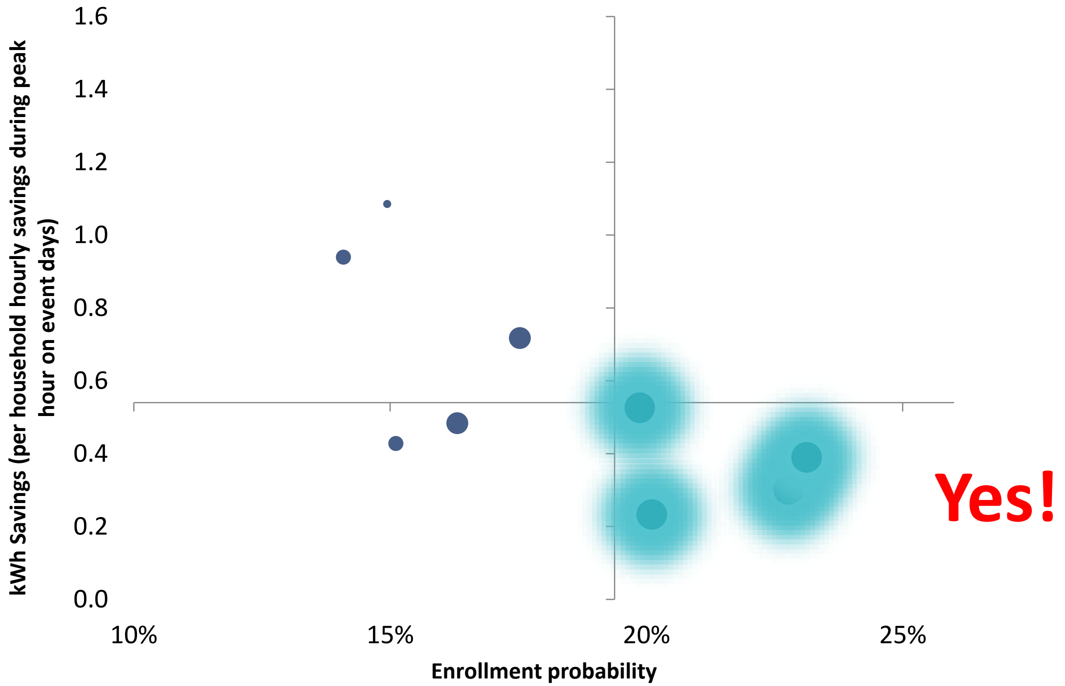
Source: Borgeson et al. (Forthcoming)

Can we identify customers who are highly likely to enroll and may be able to increase their responsiveness?



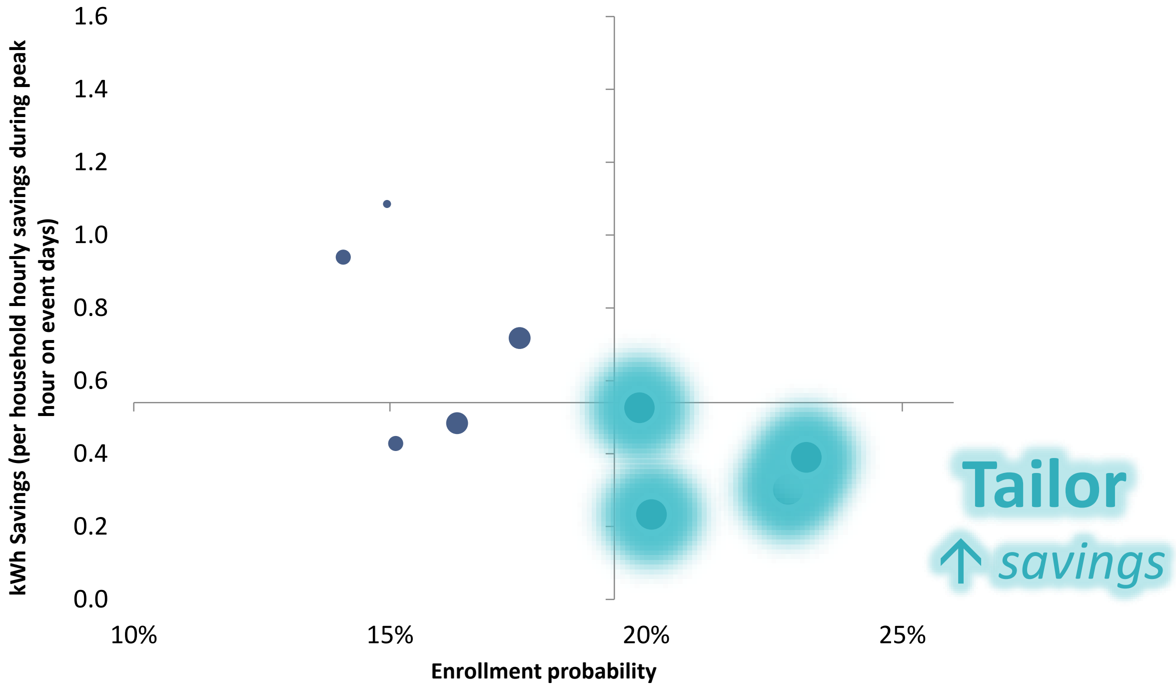
Source: Borgeson et al. (Forthcoming)

Can we identify customers who are highly likely to enroll and may be able to increase their responsiveness?



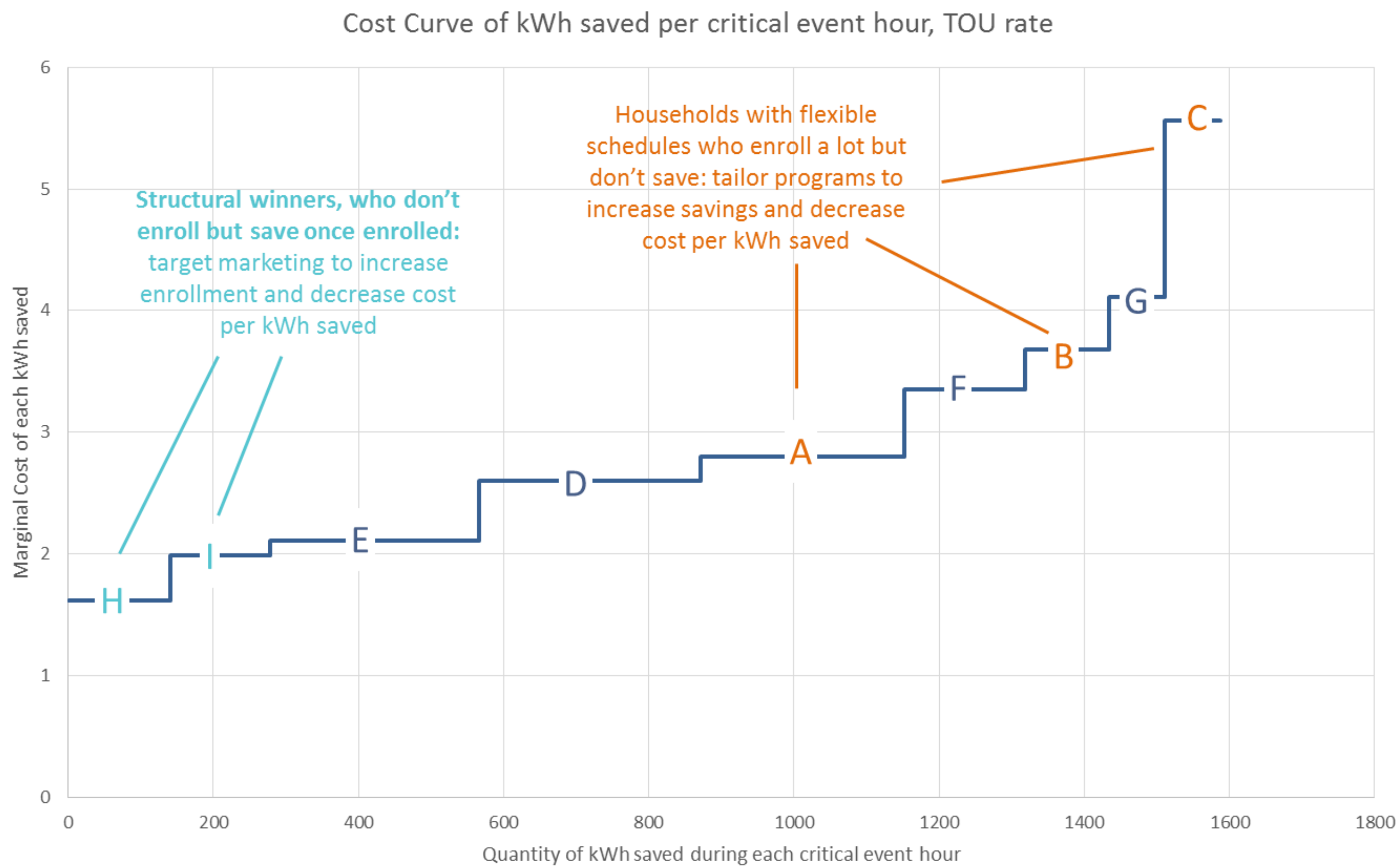
Source: Borgeson et al. (Forthcoming)

Tailor marketing and education material to better engage customers and increase their responsiveness



Source: Borgeson et al. (Forthcoming)

Which customers are more cost effective to pursue?



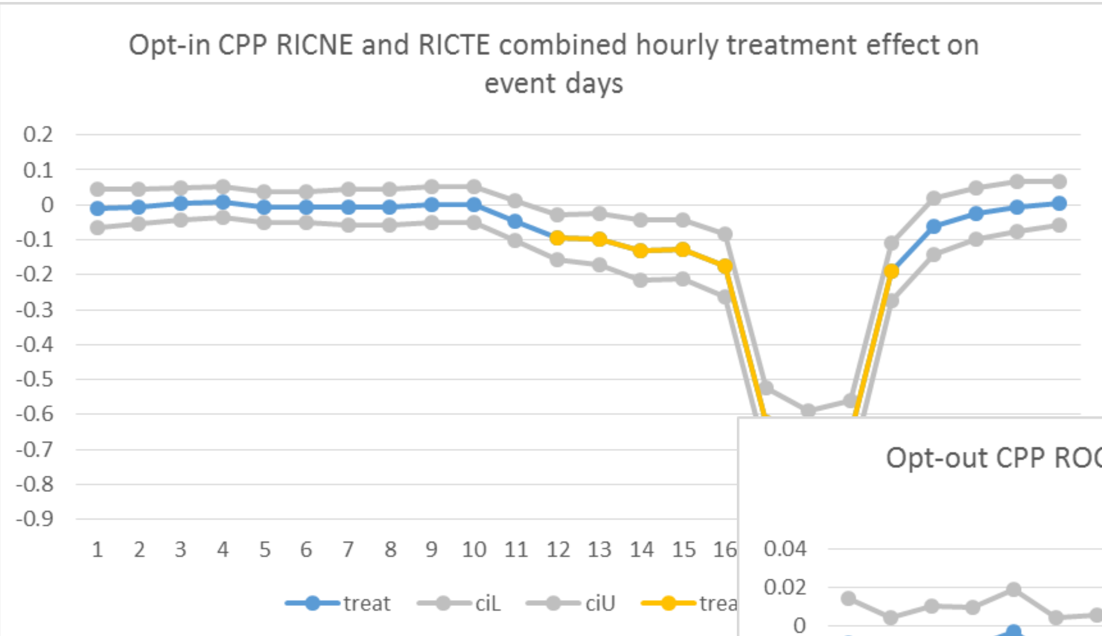
Source: Borgeson et al. (Forthcoming)

Example 5

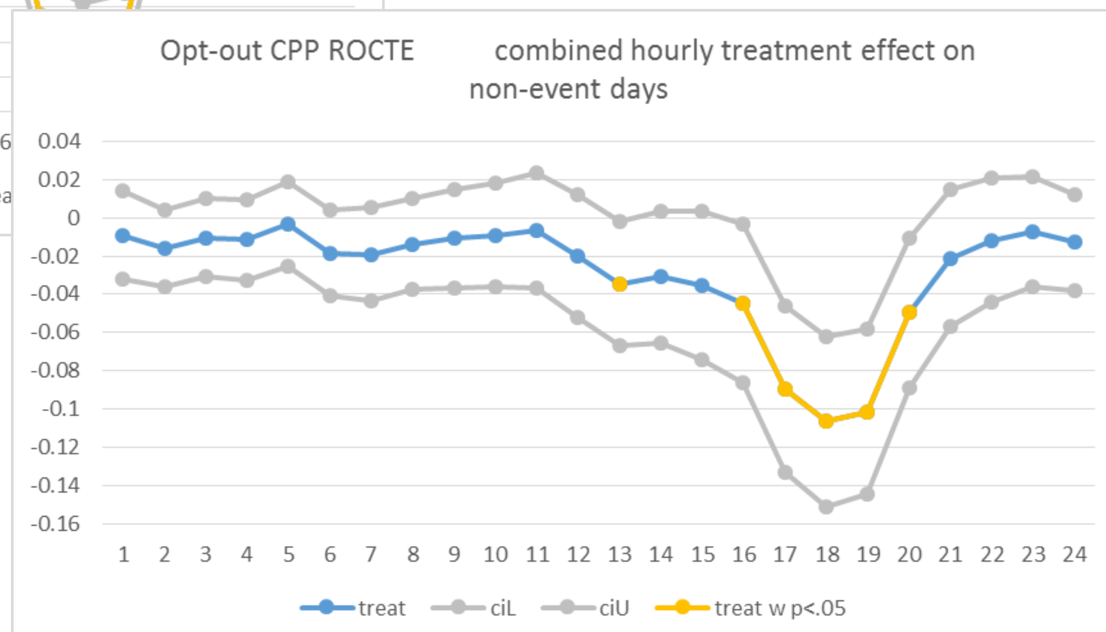
Identify timing of load response

Better understand its implications on other metrics of interest

Customer response to CPP: Consistent with Expectations on Event Days

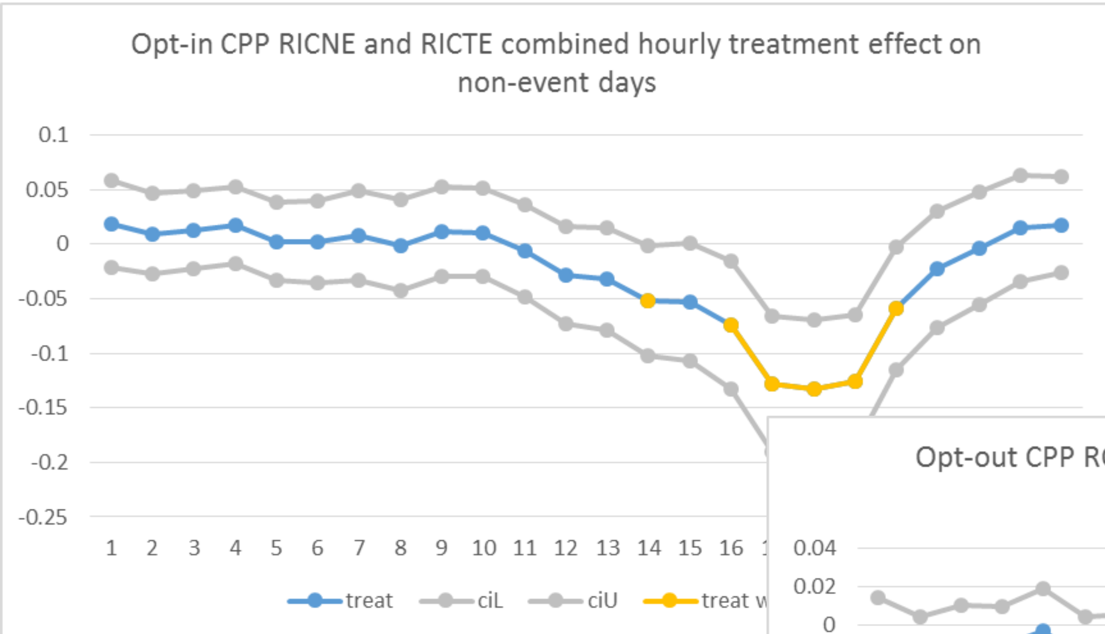


Participants reduce usage during CPP event hours on event days – as expected

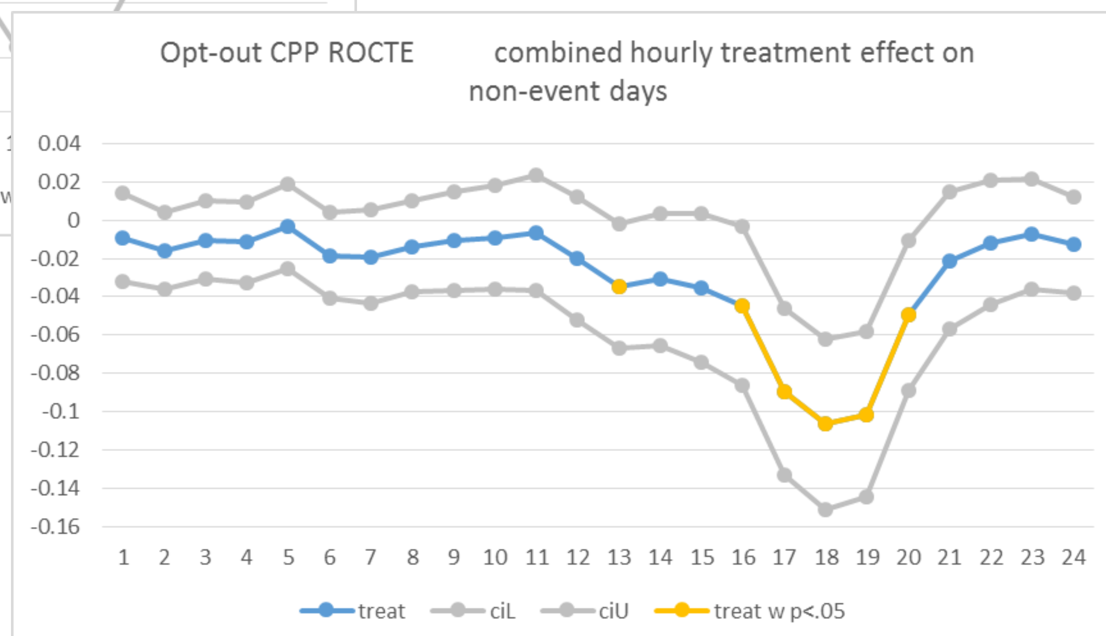


True for volunteers as well as those defaulted onto CPP

Do CPP Customers Alter Usage on Non-Event Days?



Participants reduce usage during CPP event hours on non-event days too!!!



True for volunteers as well as those defaulted onto CPP

Spillover can Undermine Lots of Other Metrics

- Unclear if this “spillover” effect would apply to PTR or other event-based DR programs…. But if it did:
  - Adversely affect baseline calculations that rely on previous non-event days usage
  - Adversely impact settlement calculations resulting in customers getting more/less than they actually deserve
  - Adversely impact load and peak demand forecasting, as well as allocation of coincident peak demand reductions for resource adequacy

Summary of Implications

- Improve prediction and forecasting
- Improve program cost-effectiveness
- Better EM&V methods

→ All of these help with utility planning, both short term (day-ahead DR planning), and long term portfolio planning

Main Takeaway:

1. **Lots of things you can do with smart meter data**

2. **Some can be really useful, and some aren't** (test with real insist on seeing results)

3. **Let's just do a lot of quick A/B testing and analysis – what actually works? What should we try next?**
   - Test big things (program validity), small things (best wording for marketing messages), test continuously

Berkeley Lab - *Behavior Analytics*

Providing insights that enable evidence-based, data-driven decisions

Contact:

Peter Cappers
pacappers@lbl.gov
315-637-0513

Annika Todd
atodd@lbl.gov
510-495-2165

Define relevant household energy behavior characteristics that you think are important

Flexibility Metrics (Variability of Usage)	
entropy	Entropy, meant to characterize overall variability in daily household consumption patterns, generated by clustering daily baseload usage patterns and calculating the entropy in load shape assignment for a given customer across days
pre-peak CV	Coefficient of variation (CV) of consumption in the two hours prior to the peak period across days
peak CV	CV of consumption during the peak period across days
post-peak CV	CV of consumption in the two hours following the peak period across days
Savings Potential Metrics (Load Magnitude during the Hottest Days)	
pre-peak mean (hot)	Average consumption during the two hours prior to the peak period
peak mean (hot)	Average consumption during the peak period
post-peak mean (hot)	Average consumption during the two hours following the peak period
Occupancy Metrics (Load Magnitude during the Non-Hottest Days)	
pre-peak mean	Average consumption during the two hours prior to the peak period
peak mean	Average consumption during the peak period
post-peak mean	Average consumption during the two hours following the peak period
Baseload Usage Metrics	
minimum	Average daily minimum consumption across all days (i.e., base load)
Structural Winningness	
Structural Winningness	The degree to which a household would get lower bills on the new rate if they didn't make any energy behavior changes (the amount of money a household would have saved in the pre-treatment year if they had been on the new rate instead of the old rate)

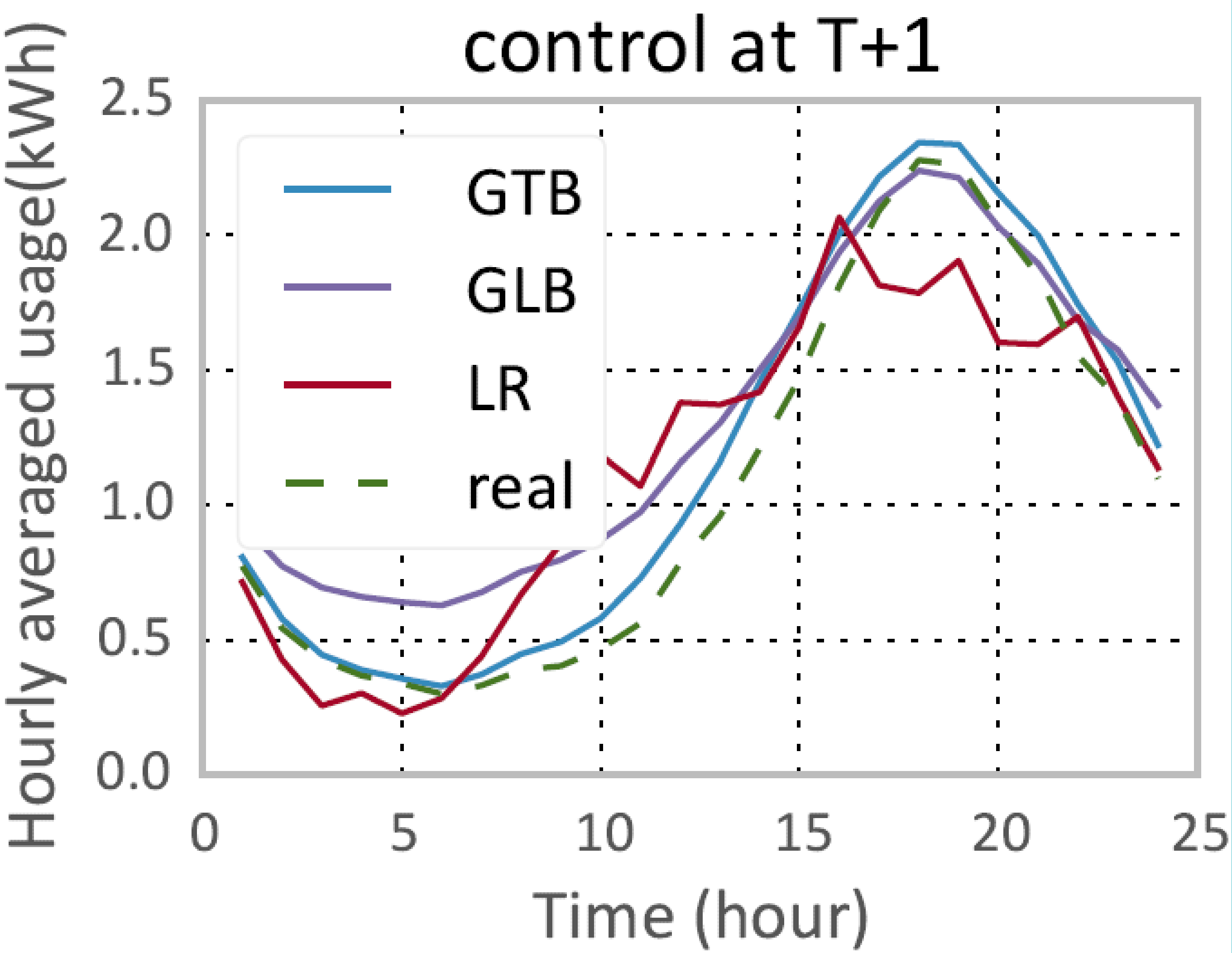
Appendix Example

Better baseline estimates

→ More accurate EM&V

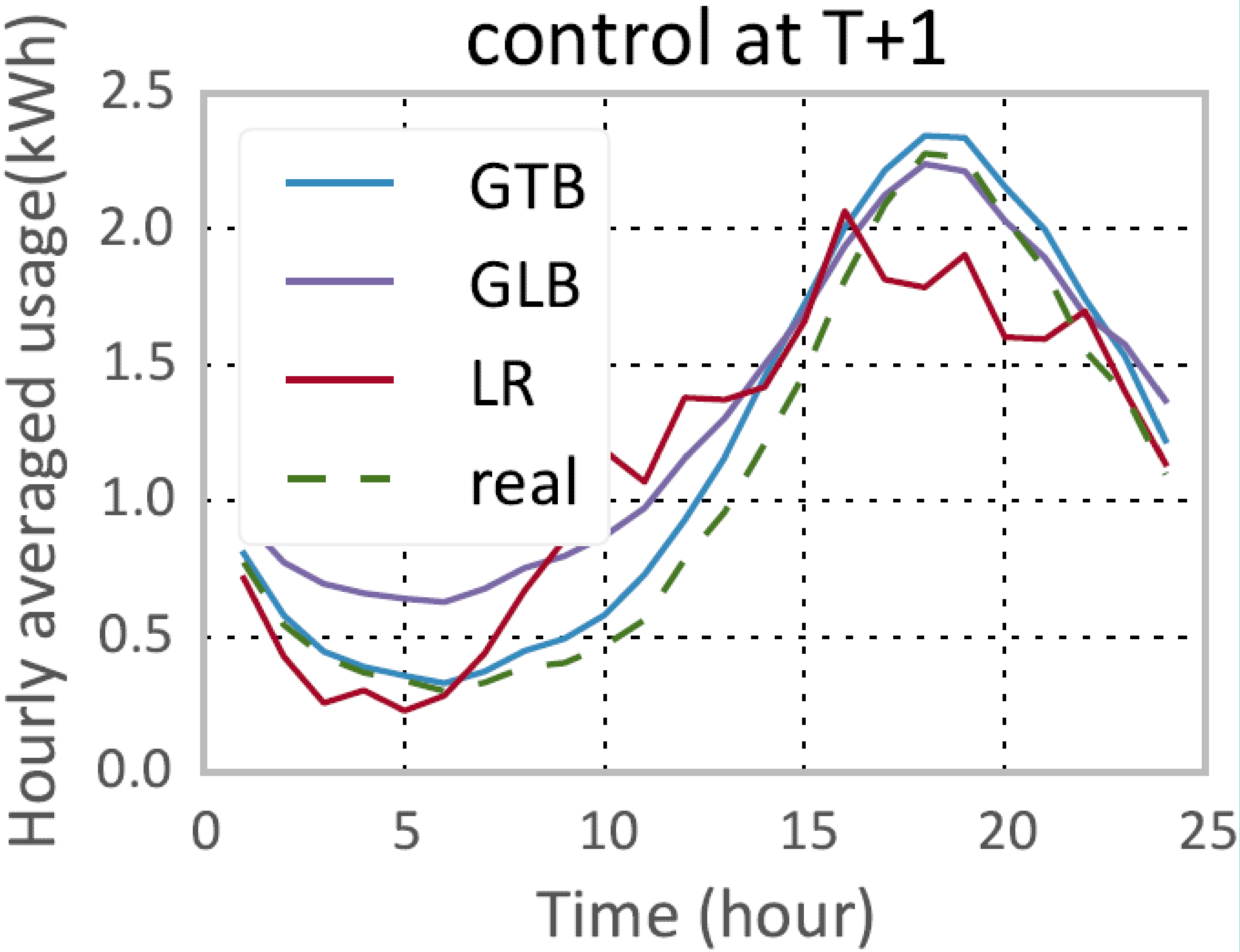
→ More accurate customer settlement payments & penalties

Better Baseline Methods are Possible



- Green dotted line is actual usage
- Red line is a typical prediction method
- Blue and purple lines are different machine learning "gradient tree" methods

Better Baseline Methods are Possible



- Machine learning methods do a better job at **predicting real usage**
- Better prediction of usage → better baselines for EM&V and customer settlements