Anomaly Detection in Premise Energy Consumption Data

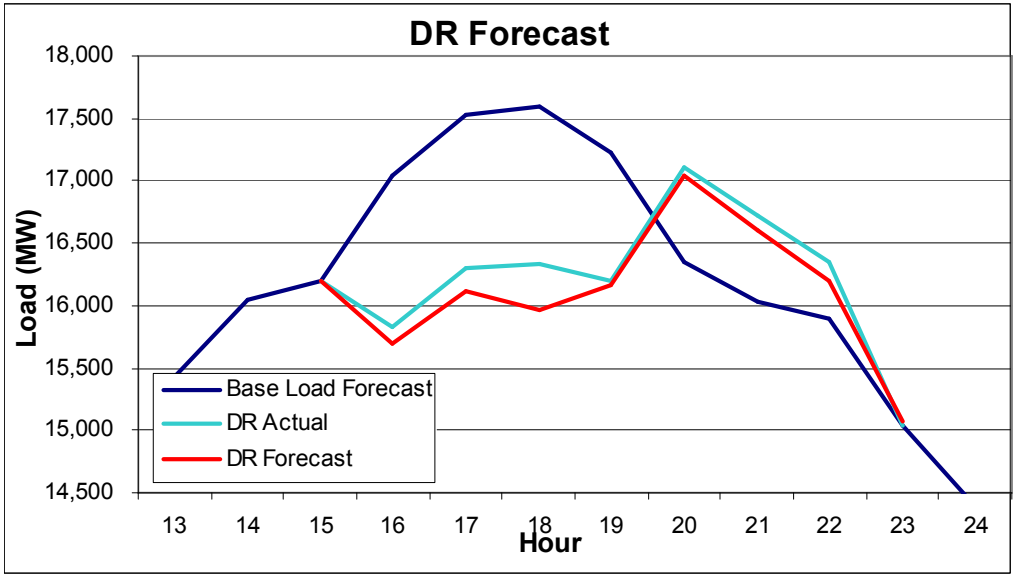
Jason W Black (Energy Systems, Battelle)
Yi Zhang, Weiwei Chen (MSL, GE Global Research Center)

CMU Electricity Conf
March 2012

GE imagination at work

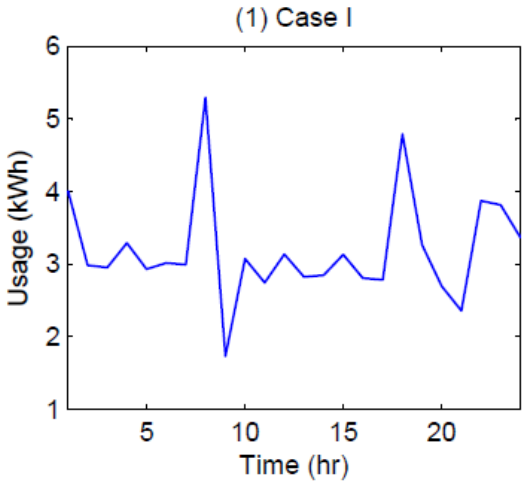
Demand Response Estimation

- **Critical X's:** weather, Temperature, seasons, days of week, devices, time…
- **X's are gathered from:**
  - AMI data
  - Real time weather forecasts
  - Historical data
- **Methods:**
  - Sampling Methods
  - Categorization Techniques
  - Empirical Methods
  - Regression techniques
  - Bayesian methods
  - Artificial Intelligence – Neural Nets, SVM,…

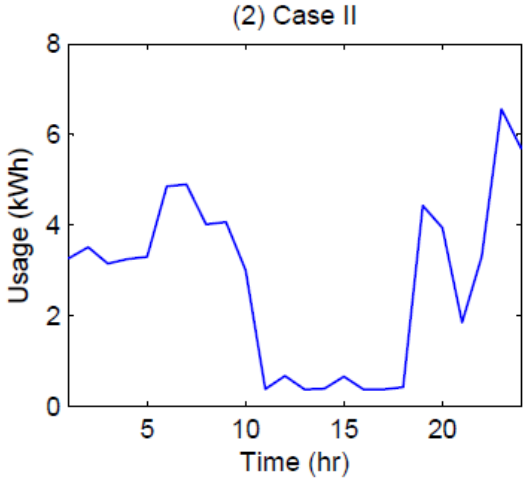


- 2 Error Sources:
  - Base Load Forecast
  - Response Estimate

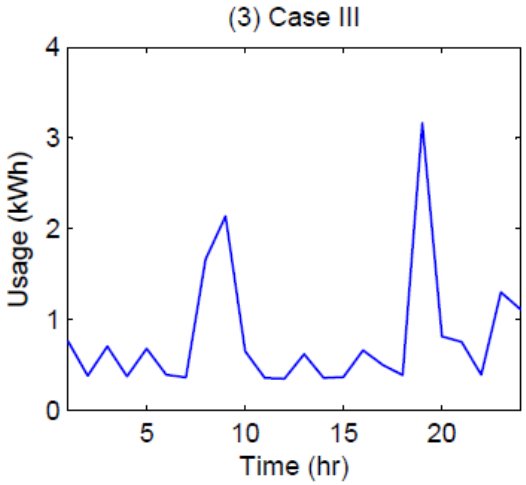
Typical Household Load Patterns



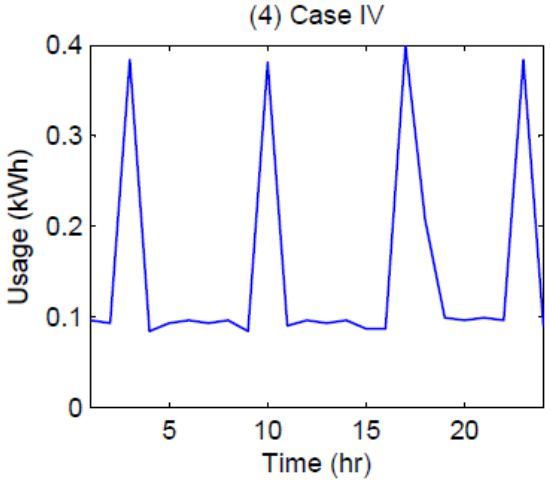
Weekend with heating/AC



Weekday with heating/AC



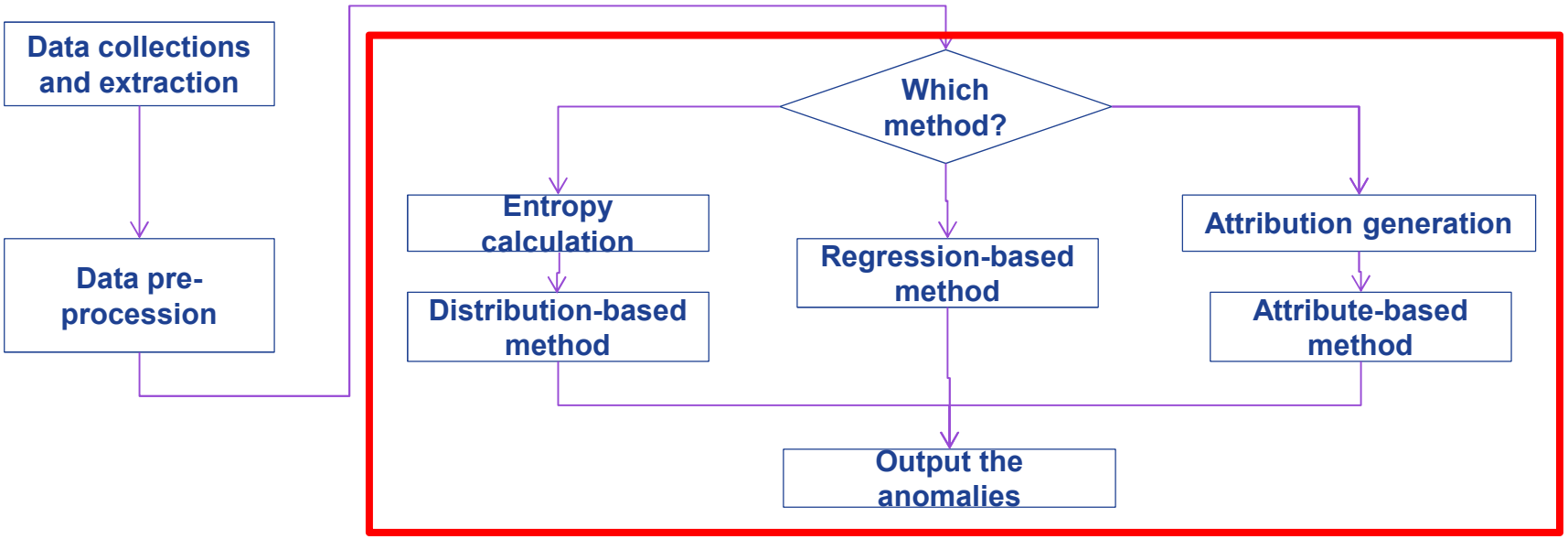
Weekday without heating/AC



Anomaly

Proposed Methods for Anomaly Detection

- **Regression-Based Anomaly Detection**
- **Distribution-Based Anomaly Detection**
- **Attribute-Based Anomaly Detection**

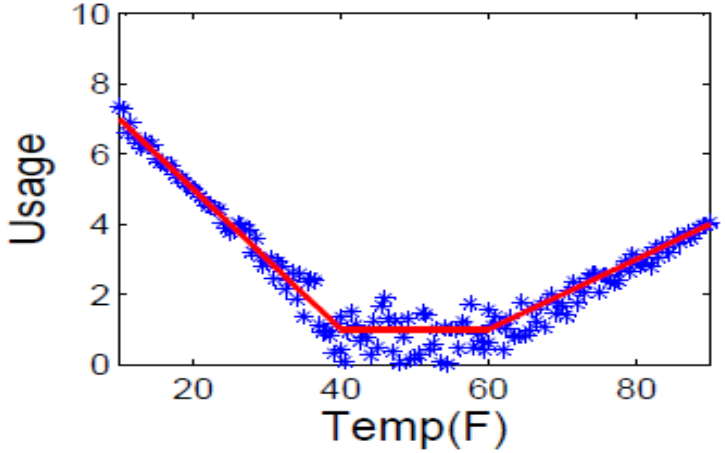


Regression-based Anomaly Detection

Battelle
The Business of Innovation

Algorithm:

1. Piecewise regression: 1, 2, and 3 piece linear functions
2. F test: determine the best model $f(X)$
3. Estimate energy consumption $\hat{Y} = f(X)$
4. Calculate ratio $R = \frac{\hat{Y}}{\cdot\cdot}$, mean $\bar{R}$, and standard



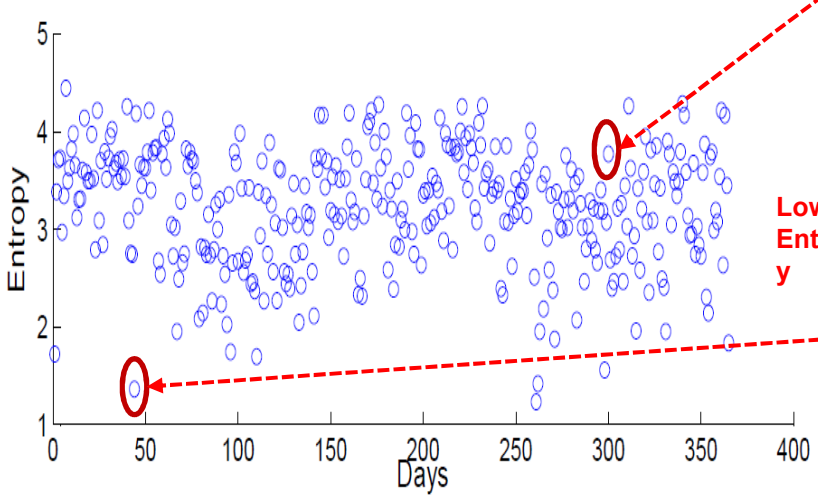
GE imagination at work

Distribution-based Anomaly Detection

In information theory, entropy is a measure of the uncertainty associated with a random variable.

$$H(x) = \sum_{i=1}^{n} p(x_i) log_b p(x_i)$$

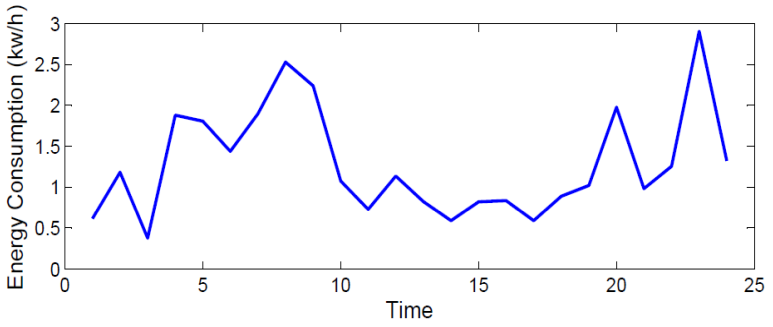
E(Abnormal) < E(Normal)

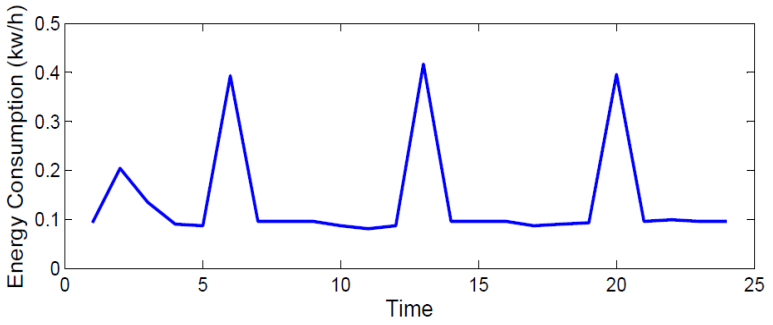


High Entropy

Low Entropy

Normal Weekday





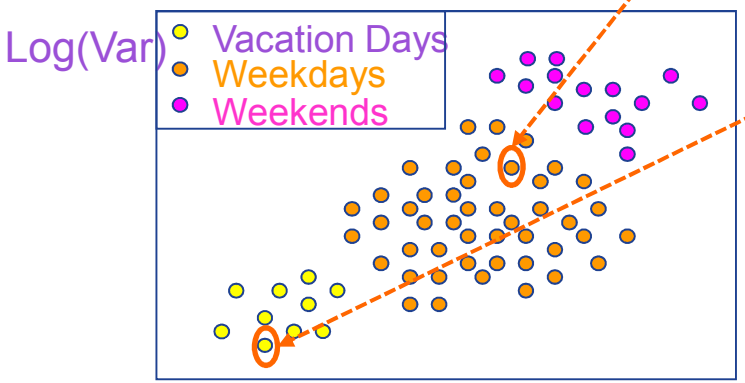
Abnormal Weekday

GE imagination at work

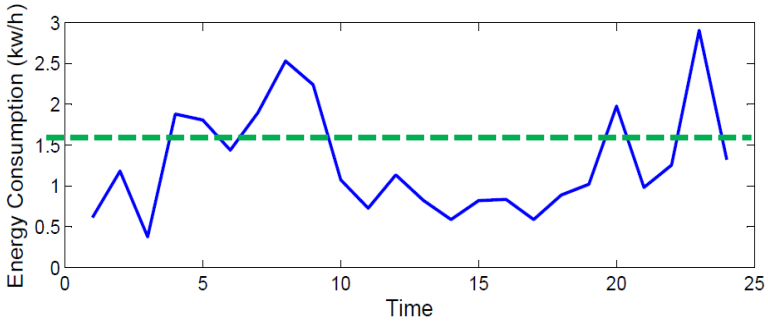
Attribute-based Anomaly Detection

Attributes generated from daily consumption data are:

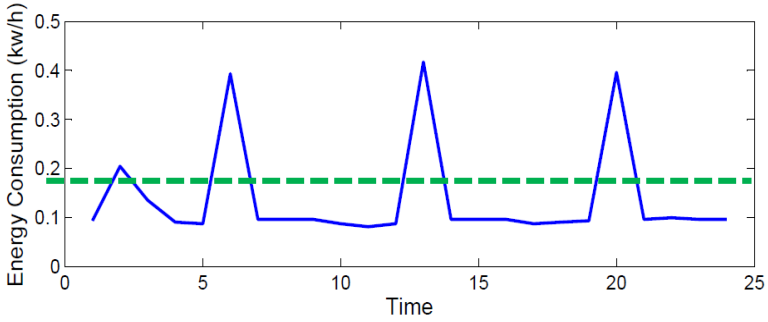
- mean
- Variance
- Maximum
- range (=maximum - minimum)
- ratio (=maximum/minimum).



Normal Weekday



High usage mean and peak



Low usage mean and peak

Vacation Weekday

GE imagination at work

Raw Data

2 sets of raw data

- Single customer data
  - Detailed usage
  - 8 months

Time	A/C Down Condensing Unit	A/C Up Condensing Unit	Air Handler Downstairs	Air Handler Upstairs	Computer Room	Dishwasher	Dryer(Gas)	Entertainment Center	Microwave	Other	Range	Pool Pump	Refrigerator	Main	Washer	Indoor Temperature Downstairs	Indoor Temperature Upstairs	Weather Temp
1/1/2010 0:00	0	0	0	0	0.03	0	0	0.01	0	0.01	0	0	0.06	0.1	0	281.97	283.64	53.6
1/1/2010 0:15	0	0	0	0.06	0.03	0	0	0.01	0	0.01	0	0	0.01	0.1	0	282.07	283.53	53.6
1/1/2010 0:30		0	0	0	0.03	0	0	0.01	0	0	0	0	0.02	0.05	0	282.01	283.64	53.6
1/1/2010 0:45	0	0	0	0.05	0.03	0	0	0.01	0	0.02	0	0	0.07	0.18	0	282.14	283.71	53.6

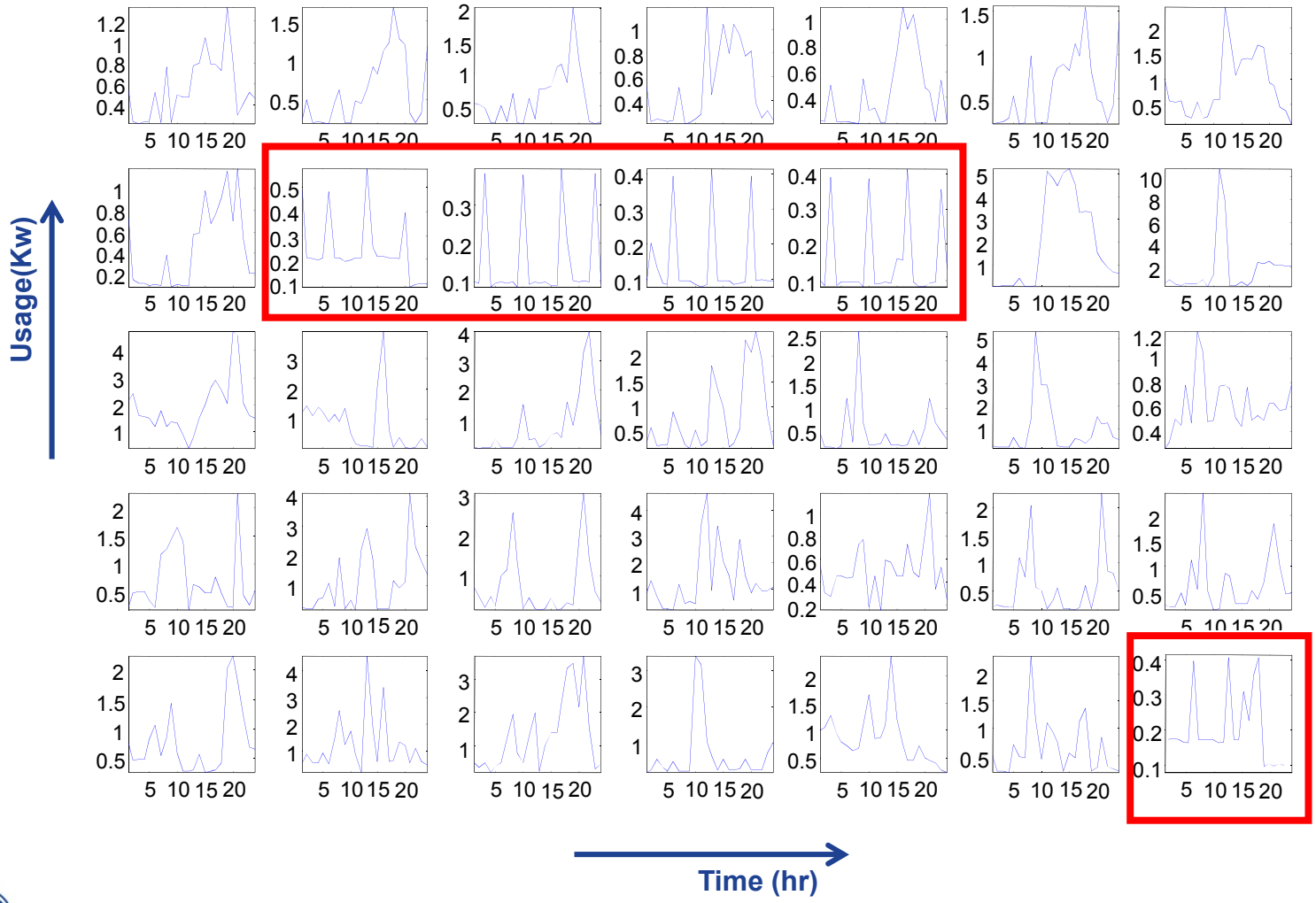
- 9-customer data
  - Some information is missing
  - 90 days

ab	name	'Laurel'		
	main	<2160x1 double>	0.0412	1.8768
	dryer	<2160x1 double>	0	0.3203
	microwave	<2160x1 double>	0	0.5240
	range	<2160x1 double>	0	0.5240
	refrigerator	<2160x1 double>	0.0053	0.0920
	washer	<2160x1 double>	0	0.5240
	waterheater	<2160x1 double>	0	0.5240
	others	<2160x1 double>	-0.3030	1.4485
ab	time	''		

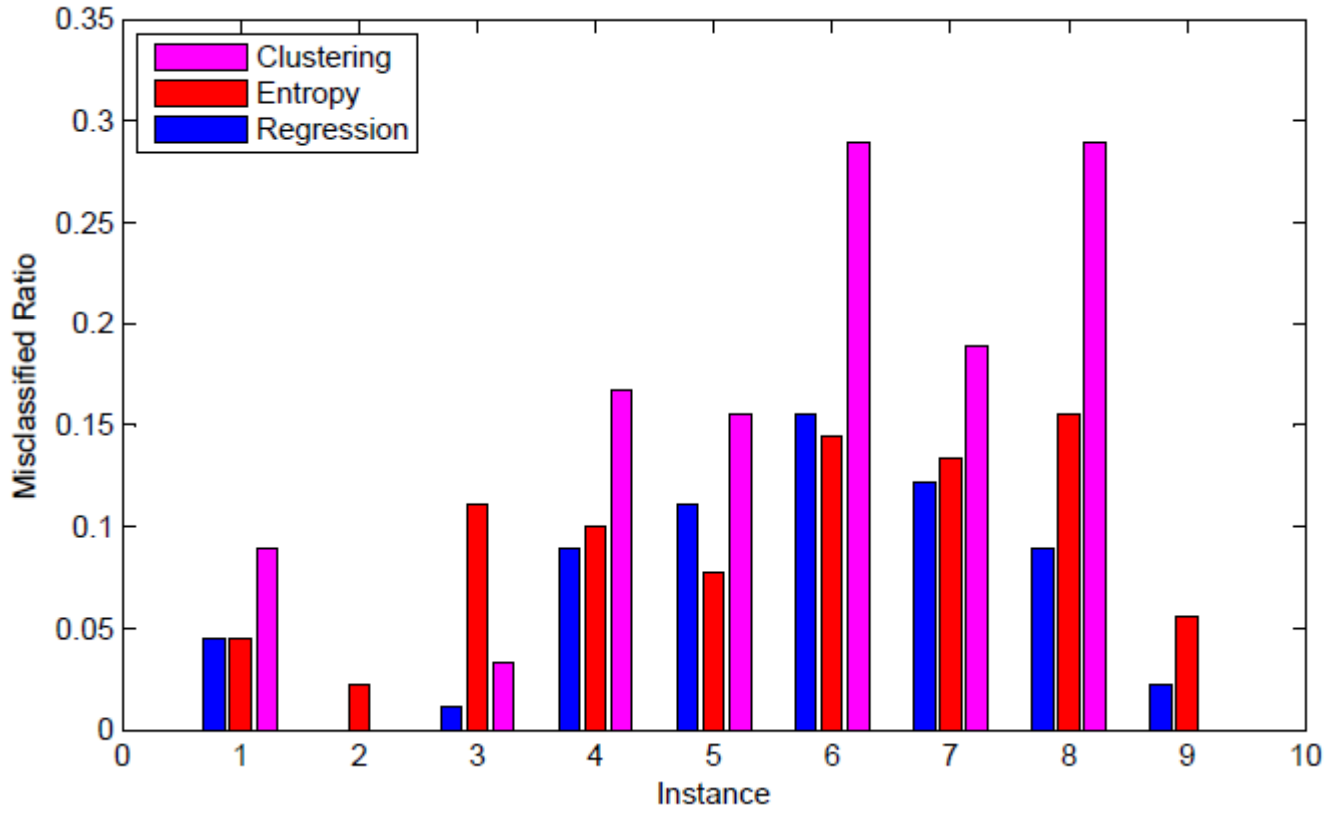
GE imagination at work

Abnormal Days Detection

(use statistics, machine learning, etc. to detect days with abnormal behavior/responsiveness in nonevent days or during DR events)



Simulation Results: 9-customer Data

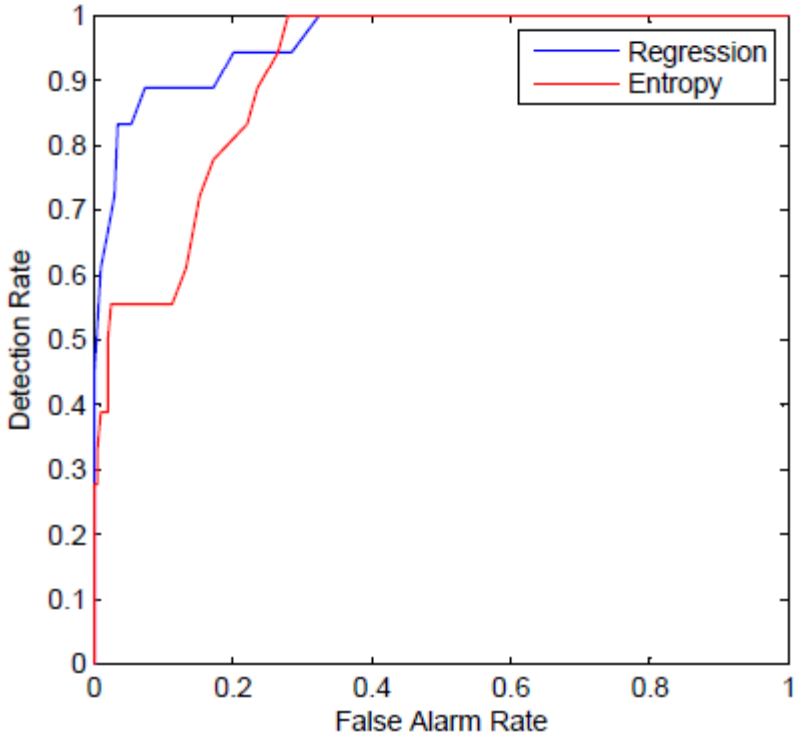


Simulation Results: Single-customer Data

Detection Results

	Detection Rate	False Alarm Rate
Regression	88.89%	7.35%
Entropy	55.56%	9.80%
Clustering	72.22%	30.88%

ROC Curve



Conclusions

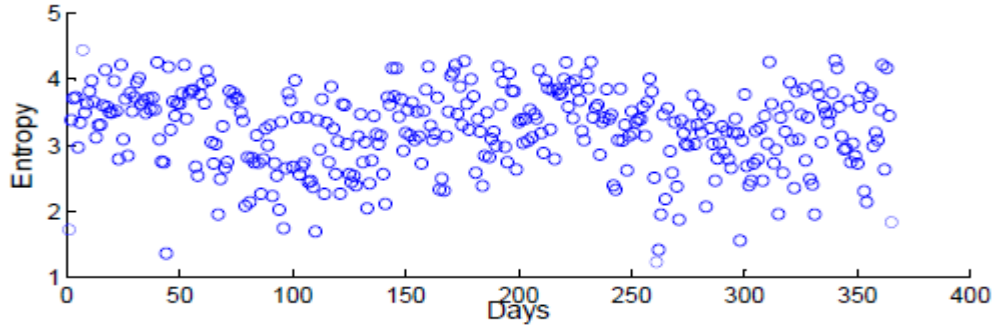
- Anomaly detection achievable in demand response estimation.
- Proposed Regression-based, distribution-based, and attribute based methods can be used.
- Data mining on each customer's power consumption profile is required for algorithm selection and parameter tuning

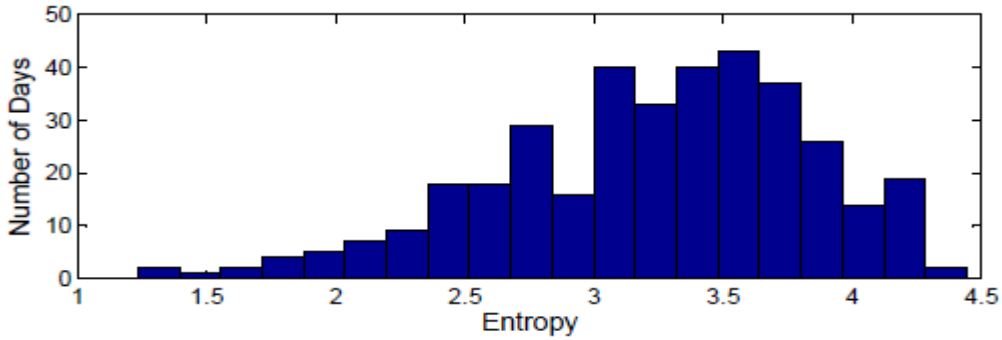
Backup

Distribution-based Anomaly Detection cont.

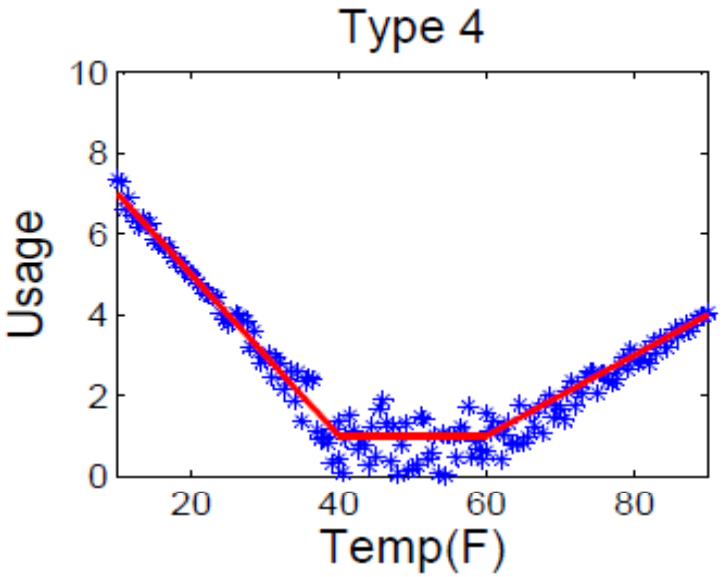
Algorithm:

1. Normalize energy consumption data
2. Calculate entropy H for each daily profile, mean $\bar{H}$, and standard deviation σ_H
3. Identify day i as an anomaly day if $H_i < \bar{H} - \alpha \cdot \sigma_H$





Piecewise Regression



3 Piece Regression

$$y = \begin{cases} a_1 x + b_1 & if\ x < c_1 \\ a_2 x + b_2 & if\ c_1 \le x \le c_2 \\ a_3 x + b_3 & if\ x > c_2 \end{cases}$$

$$a_1 c_1 + b_1 = a_2 c_1 + b_2$$

$$a_2 c_2 + b_2 = a_3 c_2 + b_3$$

Attribute-based Anomaly Detection cont.

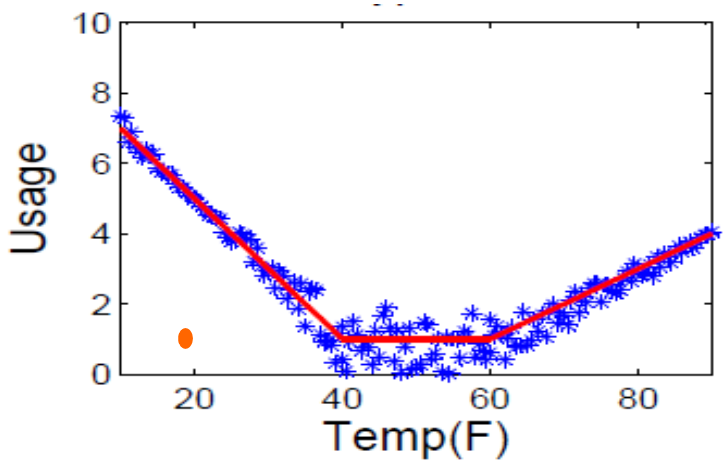
Algorithm:

1. Generate attributes from raw energy consumption data
2. Run k-Means clustering n times using random start point and select the best one output
3. Identify the cluster with the smallest mean usage as anomaly days

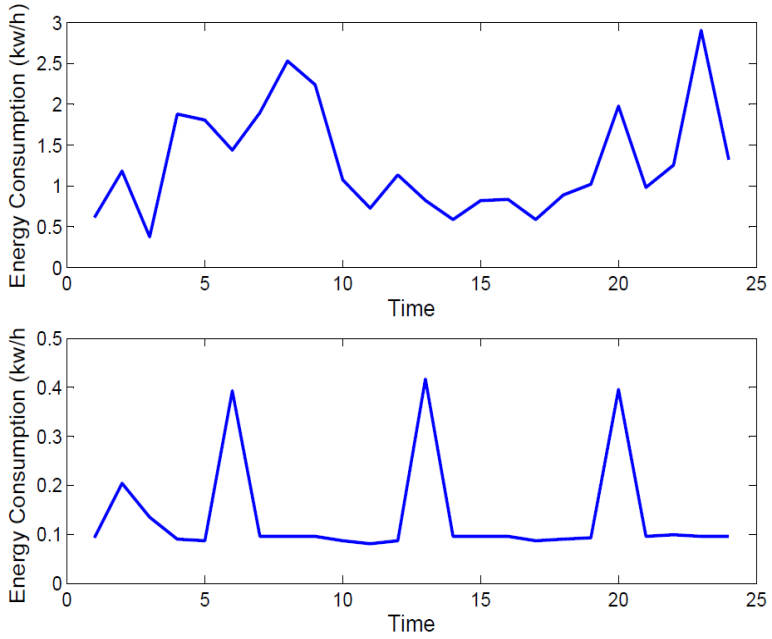
Different Anomaly Detection Methodologies

Battelle
The Business of Innovation

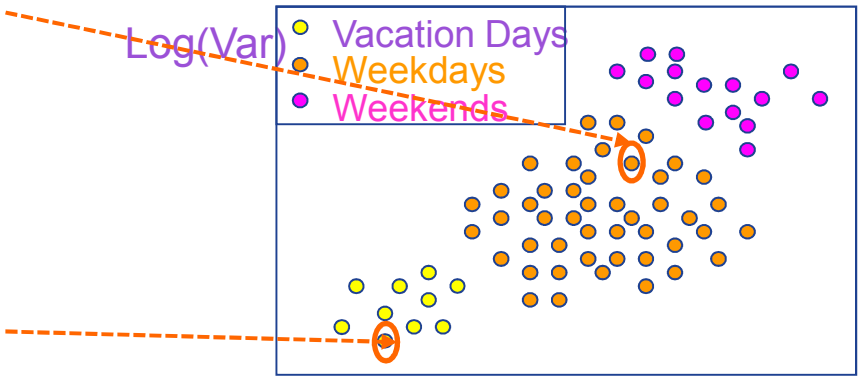
Regression based



Distribution based



Attribute based



Demand Response (DR) Process