Regression Models for Demand Reduction based on Cluster Analysis of Load Profiles

Nobuyuki Yamaguchi, Central Research Institute of Electric Power Industry
Junqiao Han, Girish Ghatikar, Sila Kiliccote, Mary Ann Piette, Lawrence Berkeley National Laboratory
Hiroshi Asano, Central Research Institute of Electric Power Industry

June 2009

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Regression Models for Demand Reduction based on Cluster Analysis of Load Profiles

Nobuyuki Yamaguchi, *Member, IEEE*, Junqiao Han, Girish Ghatikar, Sila Kiliccote, Mary Ann Piette, *Non-Member*, and Hiroshi Asano, *Member, IEEE*

***Abstract*—This paper provides new regression models for demand reduction of Demand Response programs for the purpose of ex ante evaluation of the programs and screening for recruiting customer enrollment into the programs. The proposed regression models employ load sensitivity to outside air temperature and representative load pattern derived from cluster analysis of customer baseline load as explanatory variables.**

The proposed models examined their performances from the viewpoint of validity of explanatory variables and fitness of regressions, using actual load profile data of Pacific Gas and Electric Company's commercial and industrial customers who participated in the 2008 Critical Peak Pricing program including Manual and Automated Demand Response.

***Index Terms*—Automated Demand Response, Critical Peak Pricing, Cluster Analysis, Demand Reduction, K-means, Regression Model, Sensitivity to Outside Air Temperature**

I. Introduction

According to the 2008 Federal Energy Regulatory Commission Survey on Demand Response (DR) in the U.S. [1], customer enrollment and the number of entities that offered DR programs have increased since 2006. These circumstances would provide informative experiences of DR deployment for utilities and independent systems operators which are considering the introduction of DR programs and expecting to conduct better ex ante evaluation.

There are several approaches to understanding demand reduction of DR programs [2]. One of these methods use customer segmentation which classifies customers according to business types and/or customer size. This approach is suitable for ex ante evaluation [3]. On the other hand, it is not easy to make a proper combination of the segments to meet this assumption though this approach assumes that every customer in a same segment has a similar characteristic of demand reduction. Another considers DR strategies based on the each facility's end-use equipments [4], [5]. In this approach, analysts use more detailed information about customers, which is cost and time intensive.

N. Yamaguchi is a Research Scientist with Central Research Institute of Electric Power Industry, 2-11-1, Iwado Kita, Komae-shi, Tokyo 201-8511 Japan (e-mail: nyamaguchi@criepi.denken.or.jp). He is also a Visiting Researcher with Demand Response Research Center, Lawrence Berkeley National Laboratory.

J. Han, G. Ghatikar, Sila Kiliccote, and Mary Ann Piette are with Demand Response Research Center, Lawrence Berkeley National Laboratory, One Cyclotron Road, MS: 90-3111, Berkeley, CA 94720 USA.

H. Asano is with Central Research Institute of Electric Power Industry.

Meanwhile, load profiles of each customer have been recently available because of a spread of interval meters and information systems. Since the load profile represent each customer's load characteristic, acceptable estimates of demand reduction would be achieved by using the load profile without the segmentation or the information in the conventional studies.

The purpose of this study is for ex ante evaluation of the demand reduction of DR programs to examine whether the demand reduction can be explained by regression models, which employ the actual load profile and weather information as customer attributes. In the proposed clustering method, each load profile is classified according to of its shape. Then, the type is used as one of the explanatory variables of the regression models. Another explanatory variable is load sensitivity to outside air temperature (OAT). In order to verify the proposed method, we apply the regression models to demand reduction performed in 2008 summer by customers who participated in Critical Peak Pricing (CPP) program, one of the Pacific Gas and Electric Company(PG&E)'s price-based DR tariffs.

II. Demand Response Programs and Demand Reduction

A. PG&E's Critical Peak Pricing Program

In PG&E's CPP program [6], customers sign up on a tariff where all electric usage during summer on-peak and partial-peak hours is discounted on days when no CPP events are called. Contrarily, on CPP event days, maximum of 12 times during the summer season, higher "critical peak" energy charges are imposed for all electric usage that occurs weekdays, excluding holidays, between noon and 6 p.m. as follows:

-Moderate Price Period (MPP): Noon to 3 p.m. customers are charged approximately three times the partial-peak energy rate shown on their otherwise applicable rate schedule.

-High Price Period (HPP): 3 p.m. to 6 p.m. customers are charged approximately five times the on-peak energy rate shown on their otherwise applicable rate schedule.

Commercial and industrial customers must have peak demand of 200 kW or greater and equipped with an interval meter provided free of charge by PG&E when they sign up on the CPP tariff.

Generally, each participant in CPP tariff reduces and/or shifts their electricity demand by controlling their Heating, Ventilating, and Air Conditioning (HVAC) system, lighting,

TABLE I
CPP EVENT DAYS AND OUTSIDE AIR TEMPERATURE IN 2008

	CPP Event Day	Max. OAT on CPP Day	Average of Max. OAT on baseline days
1	May 15	93 °F (34 °C)	63 °F (17 °C)
2	May 16	91 °F (33 °C)	63 °F (17 °C)
3	June 13	73 °F (23 °C)	69 °F (21 °C)
4	June 20	92 °F (33 °C)	74 °F (23 °C)
5	July 6	70 °F (26 °C)	68 °F (20 °C)
6	July 7	79 °F (31 °C)	68 °F (20 °C)
7	July 9	82 °F (28 °C)	68 °F (20 °C)
8	Aug 14	68 °F (20 °C)	68 °F (20 °C)
9	Aug 27	88 °F (31 °C)	72 °F (22 °C)
10	Aug 28	90 °F (32 °C)	72 °F (22 °C)
11	Aug 29	82 °F (28 °C)	72 °F (22 °C)
12	Sep 4	88 °F (31 °C)	75 °F (24 °C)

and miscellaneous equipment such as industrial process loads, cold storage, and irrigation water pumps [5]. There are several ways to control the demand on the customer's side. The most primitive way is a manual control. The second semi-automated control involves a pre-programmed demand reduction strategy initiated by a person via centralized control system. One of the most advanced measures is "Automated Demand Response (Auto-DR)." Demand Response Research Center at the Lawrence Berkeley National Laboratory (LBNL) has developed the Auto-DR and analyzed effectiveness of its deployment in PG&E Company's service territory [7], [8]. Auto-DR does not involve human intervention and pre-programmed DR strategies are executed by energy management control systems in a building or facility through receipt of an external communications signal. LBNL is currently is in process of standardizing these communications data models, called as "Open Auto-DR Communications Specification," in soon to be released report [9].

Table I shows CPP event days in 2008. In the table, maximum OAT on CPP event days and the average of the maximum OAT on baseline days at the San Francisco International Airport are depicted. The baseline days consist of 10 previous business days of CPP event day. The definition of the customer baseline load is described in the next section. As the table indicates, the average of the maximum OAT on the baseline days tends to be lower than the maximum OAT on CPP event days.

B. Calculation of Demand Reduction

Demand reduction is the difference between an actual load and a customer baseline load (CBL).

The CBL $L_{i,d,t}$ in this paper is defined as follows:

$$L_{i,d,t} = L'_{i,d,t} + \Delta_{i,d} \quad (1),$$

Where,

$$L'_{i,d,t} = C'_{i,t} + \mathrm{A}_{i,t} \cdot T_{i,d,t} + \varepsilon'_{i,d,t} \quad (2),$$

$L'_{i,d,t}$: the CBL i at period t on CPP event day d before a morning adjustment [10], [11],

TABLE II
DESCRIPTIVE STATISTICS

	Maximum Demand (kW)	Demand Reduction Rate in MPP *	Demand Reduction Rate in HPP *	Load Sensitivity to OAT *
Num. of Obs.	99	1182	1182	1188
Mean	960.61	0.024	0.0278	0.0057
Max.	10864.7	0.714	0.875	0.025
Min.	150.46	-0.516	-0.571	-0.013
Variance	2164756	0.0115	0.172	0.26E-4
Skew.	22.583	0.716	0.362	0.178
Kurtosis	4.265	6.530	5.759	0.822

*: dimensionless parameter

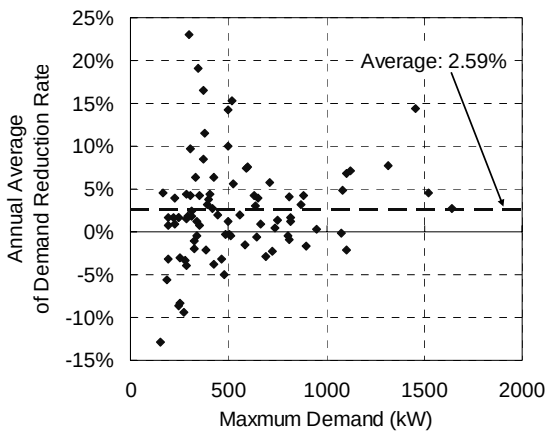


Fig. 1. A Scatter chart of annual average of demand reduction rate and customer's maximum demand.

$C'_{i,t}$: a constant,
$\mathrm{A}_{i,t}$: a load sensitivity to OAT at period t on event day d,
$T_{i,d,t}$: OAT,
$\varepsilon'_{i,d,t}$: an error term, and
$\Delta_{i,d}$: a shift term for the morning adjustment.

The morning adjustment is to mitigate a difference between actual demand and the CBL before the morning adjustment. The shift term is calculated from the data from 9am to 12noon as follows:

$$\Delta_{i,d} = \sum_{t \in 9am\,to\,12noon} AL_{i,d,t} - \sum_{t \in 9am\,to\,12noon} L_{i,d,t} \quad (3).$$

Where, $AL_{i,d,t}$ represents the actual demand on event day d.

In order to estimate $\mathrm{A}_{i,t}$ and $C'_{i,t}$ in (2), we used 10 previous business days' data of event day d. We use OAT $T_{i,d,t}$ which is measured at the nearest weather station chose from 15 weather stations of National Oceanic and Atmospheric Administration (NOAA). The average of distance from sample participants to the nearest weather station is 11 mile (18 km).

The demand reduction rate $R_{i,d,xPP}$ is defined in this paper as follows:

$$R_{i,d,xPP} = \frac{1}{N_{xPP}} \cdot \frac{1}{PL_i} \cdot \sum_{t \in xPP} \left(L_{i,d,t} - AL_{i,d,t} \right) \quad (4).$$

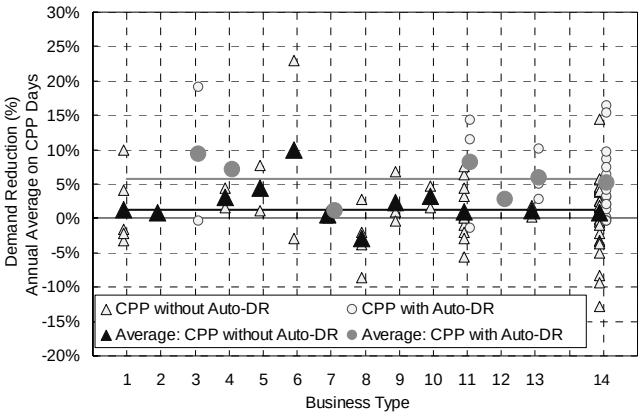


Fig. 2. Annual average of demand reduction of each industry. The indices of Business Type in this figure represent:

1. Support Activities for Crop Production,
2. Utilities,
3. Food Manufacturing, Wineries,
4. Computer and Electronic Product Manufacturing,
5. Wholesale Trade, Retail Trade,
6. Warehousing and Storage,
7. Internet Service Providers, Web Search Portals, and Data Processing Services,
8. Lessors of Nonresidential Buildings (except Miniwarehouses),
9. Professional, Scientific, and Technical Services,
10. Corporate, Subsidiary, and Regional Managing Offices,
11. Educational Services,
12. Health Care and Social Assistance,
13. Public Administration, and
14. Others or Unknown.

Where, *xPP* represents HPP or MPP. N_{xPP} means the number of period t in the *xPP*.

The table II illustrates the descriptive statistics of demand reduction rate of sample participants in the CPP program. The number of sample participants is 99 accounts which include 27 Auto-DR accounts. As figure 1 depicts, there is no significant relationship between the demand reduction rate and customer's maximum demand.

The load sensitivity to OAT α_i in table II is the mean value of $A_{i,t}$ and defined as follows:

$$\alpha_i = \frac{1}{PL_i} \cdot \frac{1}{6 \times 4} \cdot \sum_{t \in PP}^{6\times 4} A_{i,t} \quad (5),$$

Where,

PL_i: the hourly maximum demand for typical summer season,

PP: a set of observations from noon to 6pm (in HPP and MPP). It is 6 hours an event day and each hour has 4 observations because of a 15-minute interval meter.

Figure 2 illustrates the annual average percent demand reduction of the sample. In this figure, every account is classified into 14 business categories according to the North American Industry Classification System (NAICS) code. This example of classification would not explain the characteristics of the every category's demand reduction because of two reasons at least. One is that discrepancies exist among accounts in the same category. In order to reduce these discrepancies, for instance, the categories should be more finely classified. The other is that we can not understand whether these demand reductions are features of every category or account because some categories include very few accounts. The number of the accounts is shown in Table III. One of the counter plans against this problem is increase in the number of observations. The two reasons mentioned above are not easy to overcome because there is a trade off between increase in the number of categories and increase in the number of observation if we can not add the new sample accounts.

TABLE III
THE NUMBER OF ACCOUNTS IN EVERY CATEGORY

Index of Business Type		1	2	3	4	5	6	7
Num. of Accounts	CPP without Auto-DR	6	1	0	2	3	2	2
	CPP with Auto-DR	0	0	2	1	0	0	1
	Total	6	1	2	3	3	2	3

Index of Business Type		8	9	10	11	12	13	14
Num. of Accounts	CPP without Auto-DR	4	4	2	11	0	2	33
	CPP with Auto-DR	0	0	0	3	1	3	16
	Total	4	4	2	14	1	5	49

III. REGRESSION MODELS OF DEMAND REDUCTION

For the purpose of ex ante estimate of the demand reduction of CPP programs, the authors consider to develop regression models which do not use any data from site audits nor expert judgment. The load profiles of each customer reflect each customer's characteristic.

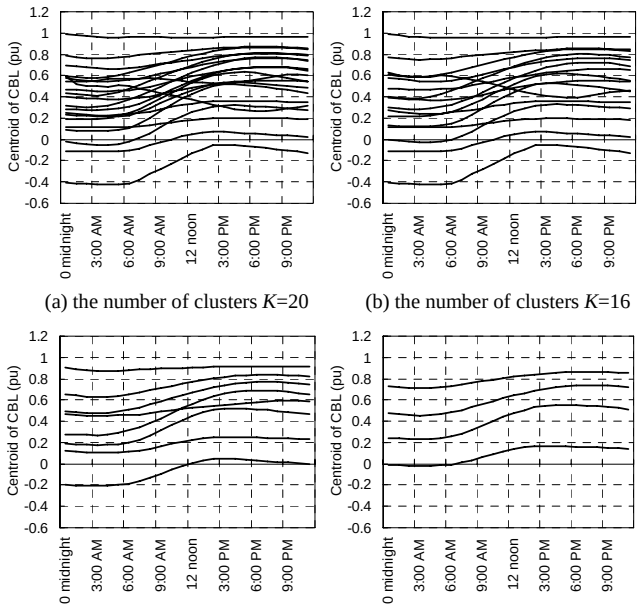
The proposed method includes two steps. First, explanatory variables are derived from load profiles and OAT. In other words, the observed CBLs are classified into some categories and the indices of categories that are used as explanatory variables in the proposed regression models. So as to classify CBLs, cluster analysis is examined in section IV. The load sensitivity to OAT is also applied as explanatory variable.

The next step is to fit the regression models. Some combinations of explanatory variables are tried in section V.

IV. CUSTOMER'S ATTRIBUTES BASED ON CLUSTER ANALYSIS OF LOAD PROFILES

A. K-means Clustering

Cluster analysis have been recently studied with the object of selecting DR policies [12], improving electric rate schedules [13], [14]. Reference [15] is one of the detailed literature surveys on comparisons among customer clustering methods such as modified follow-the-leader, hierarchical clustering, k-means and fuzzy k-means algorithm based partitional clustering, and the Kohonen self-organizing map (SOM). In [15], the results of the clustering validity assessment has shown that the modified follow-the leader and



(a) the number of clusters K=20 (b) the number of clusters K=16

(c) the number of clusters K=8 (d) the number of clusters K=4

Fig. 3. Centroids in the cases where the numbers of clusters are 8, 12, 16 and 20, respectively.

the hierarchical clustering emerge as the most effective ones. However, both these methods are not applied in this paper because they have implementation difficulties. The modified follow-the-leader requires a trial-and-error approach so as to determine the number of the categories and analysts can not set the number of the categories explicitly. A hierarchical cluster structure, which is built in the hierarchical clustering, is not used in this paper. Therefore, the k-means or partitional clustering, which is not hierarchical, instead partitions the data set into K number of categories or clusters designated beforehand, are applied. According to [16], a classical k-means algorithm has the following structure:

1. Assign each object randomly to one of the clusters k=1,2, …, K.
2. Compute the means for each of the clusters:

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{z_i \in C_k} \boldsymbol{z}_i \quad (6).$$

   Where, $\boldsymbol{z}_i$ represents a feature vector as an individual object. C_k expresses a subset which consists of objects in cluster k. N_k means the number of objects in C_k.
3. Reassign each object $\boldsymbol{z}_i$ to the cluster with the closest mean $\boldsymbol{\mu}_k$ which is called "centroid".
4. Return to step 2 until the means of the clusters do not change anymore or the criteria is met.

In order to choose the closest centroid for each cluster in step 2, distance function $d(\boldsymbol{z}, \boldsymbol{\mu})$ is defined as follows:

$$d(\boldsymbol{z}, \boldsymbol{\mu}) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(z_h - \mu_h)^2} \quad (7).$$

Where, H indicates the size of the feature vector. z_h and μ_h represent the h th elements of vector $\boldsymbol{z}$ and $\boldsymbol{\mu}$, respectively.

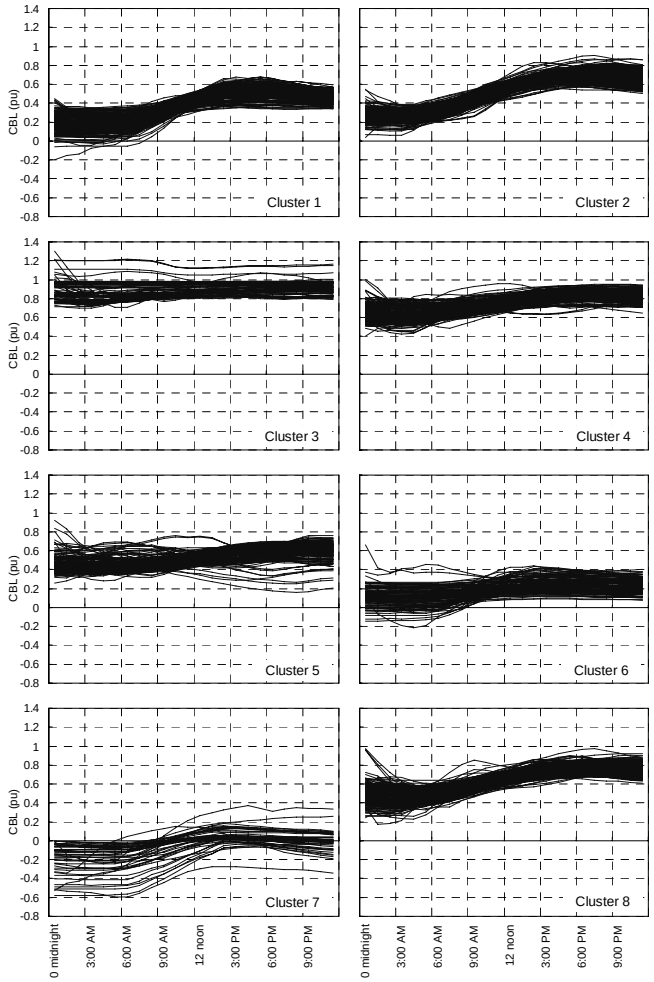


Fig. 4. Individual CBLs in every cluster in the case where the number of clusters is 8. Every cluster's centroid is plotted in fig. 3. (c).

B. *Cluster Analysis of Customer Baseline Load*

In this paper, the CBLs of every customer on every CPP day are applied as objects in k-means clustering. The number of clusters is examined in the cases where K=4, 8, 16 and 20 because Chicco et al [14] insists that an indicative number of clusters not higher than 15-20 could fit the electricity supplier's needs. In order to avoid to be trapped into local minima, the authors repeated the calculation in 100 iterations with different initial sets of the objects which are chosen randomly.

Figure 3 draws sets of the centroids in the cases where K=4, 8, 16 and 20. From the viewpoint of application of regression models, it would be better that the number of clusters is smaller. On the other hand, a rich expression of the feathers of CBL in detail needs the large number of clusters. The cases of K=16 and 20 describe various load shapes very well including a night-peak load shape. On the contrary, the case of K=4 can not depict so richly. The case of K=8 explains various shapes by a small number of clusters Taking account of a trade-off of the number of categories and individuals, this paper focus on the case of K=8 in those that follow.

TABLE IV
THE NUMBER OF CBLS IN EVERY CLUSTER AND INDUSTRIAL CATEGORY

The Num. of CBLs		Index of Cluster								Sum.
		1	2	3	4	5	6	7	8	
Index of Business Type	1	18	1	12	5	9	6	5	16	72
	2	7	3	0	0	0	2	0	0	12
	3	1	1	0	2	11	0	0	9	24
	4	10	17	0	0	0	1	4	4	36
	5	0	12	8	14	0	0	0	2	36
	6	5	12	0	0	1	1	4	1	24
	7	0	0	16	7	9	0	0	3	35
	8	10	30	0	0	1	0	0	7	48
	9	10	5	0	15	0	3	2	13	48
	10	9	2	3	1	2	0	3	4	24
	11	28	27	3	11	25	25	6	40	165
	12	0	0	0	2	2	0	0	8	12
	13	3	8	2	18	0	0	0	29	60
	14	76	79	25	132	74	63	19	118	586
Sum.		177	197	71	216	123	101	43	254	1182

TABLE V
THE NUMBER OF CBLS IN EVERY CLUSTER AND CPP EVENT DAY

The Num. of CBLs		Index of Cluster								Sum.
		1	2	3	4	5	6	7	8	
Index of CPP Event Day	1	20	19	6	14	6	4	2	26	97
	2	20	20	7	14	10	3	0	23	97
	3	17	8	4	13	17	12	9	17	97
	4	5	11	6	24	7	15	10	21	99
	5	6	13	6	14	10	19	7	24	99
	6	5	10	8	23	11	15	7	20	99
	7	5	15	8	24	12	13	8	14	99
	8	21	16	6	10	16	11	0	19	99
	9	24	18	4	16	10	2	0	25	99
	10	16	24	5	19	12	2	0	21	99
	11	10	24	4	21	13	4	0	23	99
	12	28	19	5	15	10	1	0	21	99
Sum.		177	197	71	216	123	101	43	254	1182

Figure 4 plots the individual CBLs in every cluster in the case of K=8. Cluster 7 has negative load during nighttime because of the morning adjustment. Since the negateive CBLs emerge when electricity usage in morning period is large in the baseline days and very low on the CPP event days ,this means that the customers in Cluster 7 use not on the event days but usually use electricity. Cluster 6, which represents low-level loads, is also affected by the same action of morning adjustment.

Table IV shows distribution of CBLs in every cluster and industrial category. The distribution of the business type into the clusters has obvious characteristics. The industry of Internet Services Providers, Web Search Portals, and Data Processing Services (Business Type 7) gathers about their 46% for Cluster 3 which has flat and high level load shape. Cluster 5, which contains night-peak loads, also holds a large part of Business Type 7. Cluster 7, which has negative load during nighttime because of the morning adjustment, includes many CBLs which are facilities at school districts though categorized into Business type 14 (Others or Unknown). This suggests that these facilities consume electricity before summer and are closed in summer vacations. About 20-30% of CBLs of Public Administration and Others (Business Type 13 and 14) is classified into cluster 4. It is the advantage of cluster analysis that CBLs of not only some specified business types but also Business Type 14 (Others and Unknown).

TABLE VI
ESTIMATE RESULTS OF MODEL 1 FOR MPP

Parameter	Estimate	Std. Error	t Statistic	P value
C_1	-0.76E-2	0.45E-2	-1.69748	0.090
β_1	4.40011	0.695980	6.32217	0.000 **
γ_1	7.37937	0.719017	10.2631	0.000 **
Adjusted R-squared: 0.085659				

TABLE VII
ESTIMATE RESULTS OF MODEL 1 FOR HPP

Parameter	Estimate	Std. Error	t Statistic	P value
C_1	0.23E-2	0.56E-2	0.405742	0.685
β_1	3.38002	0.871604	3.87793	0.000 **
γ_1	6.21092	0.900455	6.89754	0.000 **
Adjusted R-squared: 0.038731				

Table V describes distribution of CBLs in every cluster and every CPP event day. The number of CBLs in Cluster 1 on CPP Day 4, 5, 6, and 7 (on June 20 and in July) is smaller than than other CPP days of the same cluster. The main business type of Cluster 1 is Others and Unknown but this cluster includes Postharvest Crop Activities and Production. The CBL numbers in Cluster 6 and 7 on CPP days from 8 to 12 (in August and September) is also smaller than in the first half of the CPP season. This trend would come from compositions of Cluster 6 and 7 which include many facilities on Educational Services (Business Type 11).

V. NUMERICAL RESULTS

In this section, 3 combinations of explanatory variables mentioned at the previous sections are examined so as to explore for a better regression model. Throughout this section, C_x represents a constant in Model x, $\beta_{x,*}$ and $\gamma_{x,*}$ represent coefficients to be estimated, and ε_x represents a error term, respectively.

A. Model 1: Temperature Sensitivity Regression Model

This model consist of the load sensitivity to OAT α_i (5) which are expected to express a feature of cooling equipment such as HVAC, air-conditioners, refrigerators and freezers in a facility. The regression model is described as follows:

$$R_{i,d,xPP} = C_1 + \beta_1 \cdot \left(1 - D_{AutoDR,i}\right) \cdot \alpha_i + \gamma_1 \cdot D_{AutoDR,i} \cdot \alpha_i + \varepsilon_1 \quad (8).$$

Where, $D_{AutoDR,\,i}$ expresses a dummy variable, $D_{AutoDR,\,i}$ =1 if CBL i has Auto-DR equipment, otherwise $D_{AutoDR,\,i}$ =0.

Table VI and VII show estimate results of model 1 for both the MPP and HPP. Both the load sensitivity to OAT and Auto-DR dummy are effective as explanatory variables. As the result show, Auto-DR is more effective to reduce customers' load. In addition, demand reduction increase in proportion to the sensitivity to OAT. This result is consistent with the facts that global temperature adjustment of HVAC is one of the

TABLE VIII
ESTIMATE RESULTS OF MODEL 2 FOR MPP

Parameter	Estimate	Std. Error	t Statistic	P value
$\beta_{2,1}$	-0.022588	8.57E-03	-2.63651	0.008 **
$\beta_{2,2}$	0.030021	8.40E-03	3.57546	0.000 **
$\beta_{2,3}$	0.027409	0.013958	1.96364	0.050 *
$\beta_{2,4}$	0.022846	8.84E-03	2.58312	0.010 **
$\beta_{2,5}$	-0.025051	0.012647	-1.98079	0.048 *
$\beta_{2,6}$	-6.35E-03	0.011295	-0.56238	0.574
$\beta_{2,7}$	-3.74E-03	0.017018	-0.21962	0.826
$\beta_{2,8}$	0.033547	7.67E-03	4.37193	0.000 **
$\gamma_{2,1}$	0.019189	0.018592	1.03208	0.302
$\gamma_{2,2}$	0.073899	0.015432	4.78881	0.000 **
$\gamma_{2,3}$	2.93E-03	0.027667	0.105854	0.916
$\gamma_{2,4}$	0.057146	0.012373	4.61863	0.000 **
$\gamma_{2,5}$	0.073506	0.012647	5.81218	0.000 **
$\gamma_{2,6}$	0.024463	0.025107	0.974356	0.330
$\gamma_{2,7}$	0.10557	0.042261	2.49802	0.012 *
$\gamma_{2,8}$	0.065943	0.0122	5.40522	0.000 **
Adjusted R-squared: 0.071379				

TABLE IX
ESTIMATE RESULTS OF MODEL 2 FOR HPP

Parameter	Estimate	Std. Error	t Statistic	P value
$\beta_{2,1}$	7.77E-04	9.79E-03	0.079373	0.937
$\beta_{2,2}$	0.052401	9.59E-03	5.46345	0.000 **
$\beta_{2,3}$	0.027108	0.015944	1.70014	0.089
$\beta_{2,4}$	0.02307	0.010103	2.28353	0.022 *
$\beta_{2,5}$	7.75E-03	0.014446	0.536814	0.591
$\beta_{2,6}$	-0.031708	0.012902	-2.45766	0.014 *
$\beta_{2,7}$	-0.226432	0.01944	-11.6479	0.000 **
$\beta_{2,8}$	0.053596	8.77E-03	6.11466	0.000 **
$\gamma_{2,1}$	0.027924	0.021238	1.31484	0.189
$\gamma_{2,2}$	0.099106	0.017627	5.62227	0.000 **
$\gamma_{2,3}$	0.110763	0.031603	3.50482	0.000 **
$\gamma_{2,4}$	0.074	0.014133	5.23586	0.000 **
$\gamma_{2,5}$	0.05513	0.014446	3.8162	0.000 **
$\gamma_{2,6}$	0.02859	0.028679	0.996877	0.319
$\gamma_{2,7}$	-0.185117	0.048274	-3.83468	0.000 **
$\gamma_{2,8}$	0.079451	0.013936	5.70126	0.000 **
Adjusted R-squared: 0.187773				

most practical DR strategies [5]. Adjusted R-squared, which represents goodness of fit, means that this
model is not enough to explain the demand reduction.

B. Model 2: Cluster Regression Model

Model 2 includes a dummy variable for the index of the cluster.

TABLE X
ESTIMATE RESULTS OF MODEL 3 FOR MPP

Parameter	Estimate	Std. Error	t Statistic	P value
C_3	-8.21E-03	4.51E-03	-1.8178	0.069
$\beta_{3,1}$	1.11097	1.0883	1.02083	0.307
$\beta_{3,2}$	7.31323	1.23349	5.92891	0.000 **
$\beta_{3,3}$	8.48751	5.04037	1.68391	0.092
$\beta_{3,4}$	6.12828	1.6091	3.8085	0.000 **
$\beta_{3,5}$	2.02024	1.7163	1.17708	0.239
$\beta_{3,6}$	-4.41341	2.7046	-1.63182	0.103
$\beta_{3,7}$	8.32512	1.93993	4.29145	0.000 **
$\beta_{3,8}$	6.12399	1.20464	5.08367	0.000 **
$\gamma_{3,1}$	3.50279	1.72745	2.02772	0.043
$\gamma_{3,2}$	8.49616	1.42608	5.9577	0.000 **
$\gamma_{3,3}$	7.1369	11.164	0.639278	0.523
$\gamma_{3,4}$	12.3821	1.98754	6.22987	0.000 **
$\gamma_{3,5}$	6.5547	1.16515	5.62564	0.000 **
$\gamma_{3,6}$	5.90292	2.53714	2.3266	0.020 *
$\gamma_{3,7}$	12.1778	5.41974	2.24694	0.025 *
$\gamma_{3,8}$	7.77339	1.25607	6.18868	0.000 **
Adjusted R-squared: 0.115733				

$$R_{i,d,xPP} = \sum_{k=1}^{8} \beta_{2,k} \cdot \left(1 - D_{AutoDR,i}\right) \cdot D_{k,i} + \sum_{k=1}^{8} \beta_{2,k} \cdot D_{AutoDR,i} \cdot D_{k,i} + \varepsilon_2 \qquad (9)$$

Where, $D_{k,i}$ indicates a dummy variable, $D_{k,i}$ =1 if CBL i is in cluster k, otherwise $D_{k,i}$ =0.

Table VIII and IX illustrate estimate results for MPP and HPP respectively. The estimates without Auto-DR in cluster 2, 3, 4, and 8, which have relatively high level demand, have positive values. In all clusters, the statistically significant estimates with Auto-DR have positive value. The adjusted R squared is improved for HPP from model 1 but goes worse for MPP.

C. Model 3: Cluster-Temperature Sensitivity Regression Model

Model 3 includes the load sensitivity to OAT and a dummy variable for the index of the cluster as explanatory variables.

$$R_{i,d,xPP} = C_3 + \sum_{k=1}^{8} \beta_{3,k} \cdot \left(1 - D_{AutoDR,i}\right) \cdot \alpha_i \cdot D_{k,i} + \sum_{k=1}^{8} \beta_{3,k} \cdot D_{AutoDR,i} \cdot \alpha_i \cdot D_{k,i} + \varepsilon_3 \qquad (10)$$

Table X and XI describe estimate results of model 3. Similarly to model 1 and 2, Auto-DR participants perform better demand reduction then non-Auto-DR participants. Particularly, CBLs in cluster 2, 4, and 7 show better demand reduction. Because these estimates are statistically significant, validity of cluster analysis and load sensitivity to OAT in

order to explain demand reduction would be suggested by the estimate results of model 3. Adjusted R squared, however, are not so large though improved from model 1 and 2.

TABLE XI
ESTIMATE RESULTS OF MODEL 3 FOR HPP

Parameter	Estimate	Std. Error	t Statistic	P value
C_3	-9.61E-04	5.82E-03	-0.16521	0.869
$\beta_{3,1}$	-0.02163	0.01017	-2.12644	0.033 *
$\beta_{3,2}$	6.54411	1.2915	5.06706	0.000 **
$\beta_{3,3}$	0.02837	0.014786	1.91874	0.055
$\beta_{3,4}$	5.23578	1.66863	3.13776	0.002 **
$\beta_{3,5}$	-0.02409	0.013621	-1.76862	0.077
$\beta_{3,6}$	-5.39E-03	0.012443	-0.43329	0.665
$\beta_{3,7}$	7.9612	1.94489	4.09338	0.000 **
$\beta_{3,8}$	5.27918	1.27649	4.13571	0.000 **
$\gamma_{3,1}$	2.92999	1.7484	1.67582	0.094
$\gamma_{3,2}$	7.90419	1.45465	5.43373	0.000 **
$\gamma_{3,3}$	3.89E-03	0.027565	0.1411	0.888
$\gamma_{3,4}$	11.4508	2.03919	5.6154	0.000 **
$\gamma_{3,5}$	5.95943	1.20148	4.96006	0.000 **
$\gamma_{3,6}$	5.31218	2.54983	2.08335	0.037 *
$\gamma_{3,7}$	11.3755	5.42434	2.09712	0.036 *
$\gamma_{3,8}$	7.12941	1.29552	5.50311	0.000 **
Adjusted R-squared: 0.144784				

VI. CONCLUSIONS

This paper presents the regression models which employ indices derived from cluster analysis and load sensitivity to OAT as explanatory variables. The numerical results suggest the following.

1) Auto-DR facilitates higher and reliable demand reduction than non-auto-DR participants.
2) Load sensitivity to OAT is suitable as explanatory variable of regression models for demand reduction.
3) Cluster analysis and its algorithms is one of the effective tools to estimate demand reduction.
4) In the case that business types are not available, cluster analysis might be useful.
5) Combination of load sensitivity and cluster analysis improve the performance of the regression models.

On the other hand, the goodness of fit of regression models, which is expressed by the coefficient of determination, adjusted R squared, is still not enough. Future study would include tests on other methods of clustering and other years' experiments of DR events.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions of Ingrid Bran and Claire Curtin at Global Energy Partner for their work on the data management. Thanks also Albert Chu at PG&E. This document was partially supported by the California Energy Commission Public Interest Energy Research Program through the Demand Response Research Center under Work for Others Contract No.500-03-026, Am #1 and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

[1] FERC, *2008 Assessment of Demand Response and Advanced Metering*, Staff Report, 2008.
[2] N. Yamaguchi, H. Asano, and M. Takahashi, "Preliminary Survey on Measurement of Market Potential of Demand Response Program", *IEEJ Technical Meeting*, PE-08-104/PSE-08-103, 2008.
[3] C. Goldman et al, "Estimating Demand Response Market Potential among Large Commercial and Industrial Customers: A Scoping Study", California Energy Commission, PIER program, LBNL 61498, 2007.
[4] H. Haeri and L. M. Gage, "Demand Response Proxy Supply Curves", Prepared for PacifiCorp, 2006.
[5] N. Motegi et al, "Introduction to Commercial Building Control Strategies and Techniques for Demand Response", California Energy Commission, PIER program, LBNL 59975, 2006.
[6] Critical Peak Pricing (CPP) Program, PG&E web site [Online].Available: http://www.pge.com/mybusiness/energysavingsrebates/demandresponse/cpp/
[7] M.A. Piette, D. Watson, N. Motegi, and S. Kiliccote, "Automated Critical Peak Pricing Field Tests: 2006 Pilot Program Description and Results", California Energy Commission, PIER program, LBNL 62218, 2007.
[8] G. Wikler et al, "Pacific Gas & Electric Company 2007 Auto-DR Program: Task 13 Deliverable: Auto-DR Assessment Study", 2008 [Online] Available: http://drrc.lbl.gov/pubs/pge-auto-dr-assessment-study.pdf
[9] M. A. Piette, G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky, and C. McParland,"Open Automated Demand Response Communications Specification (Version 1.0)". California Energy Commission, PIER Program, 2009.
[10] K. Coughlin et al, "Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California", CEC PIER, LBNL 63728, 2008.
[11] J. Han et al, "Field Test Results of Automated Demand Response in a Large Office Building", CEC PIER, LBNL 1131E, 2008.
[12] S. Valero, M. Ortiz, C. Senabre, C. Alvarez, F. J. G. Franco, and A. Gabaldon, "Methods for Customer and Demand Response Policies Selection in New Electricity Markets", *IET Gener. Trans. Distrib.*, Vol. 1, No. 1, Jan., 2007.
[13] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer Characterization Options for Improving the Tariff Offer", *IEEE Trans. on Power Systems*, Vol. 18, No. 1, Feb., 2003.
[14] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load Pattern-Based Classification of Electric Customers", *IEEE Trans. on Power Systems*, Vol. 19, No. 2, May., 2004.
[15] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Comparisons Among Clustering Techniques for Electrcity Customer Classification", *IEEE Trans. on Power Systems*, Vol. 21, No. 2, May., 2006.
[16] F. van der Heihden, R. P. W. Duin, D. De Ridder, and D. M. J. Tax, *Classification, Parameter Estimation and State Estimation*, UK: Wiley, 2004, p232.