Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns

Hông-Ân Cao, Christian Beckel, Thorsten Staake
Energieinformatik, Vienna, November 12-13, 2013

Outline

1. Motivation and Problem Statement
2. Related Work
3. Methods
4. Results
5. Summary
6. Future Work
7. Q&A

Outline

1. **Motivation and Problem Statement**
2. Related Work
3. Methods
4. Results
5. Summary
6. Future Work
7. Q&A

Motivation and Problem Statement

Bits to Energy Lab

- **Power plants (over-) dimensioned for peak demand**
  - 15% of generation in Massachusetts used less than 88h/year
- **Potential for utility companies:**
  - Use fine-grained data acquired by smart-meters
  - Replace synthetic load profiles
  - Determine characteristic load profiles
  - Identify costly behaviors through consumption peaks
  - Target and address relevant segments of households

Motivation and Problem Statement

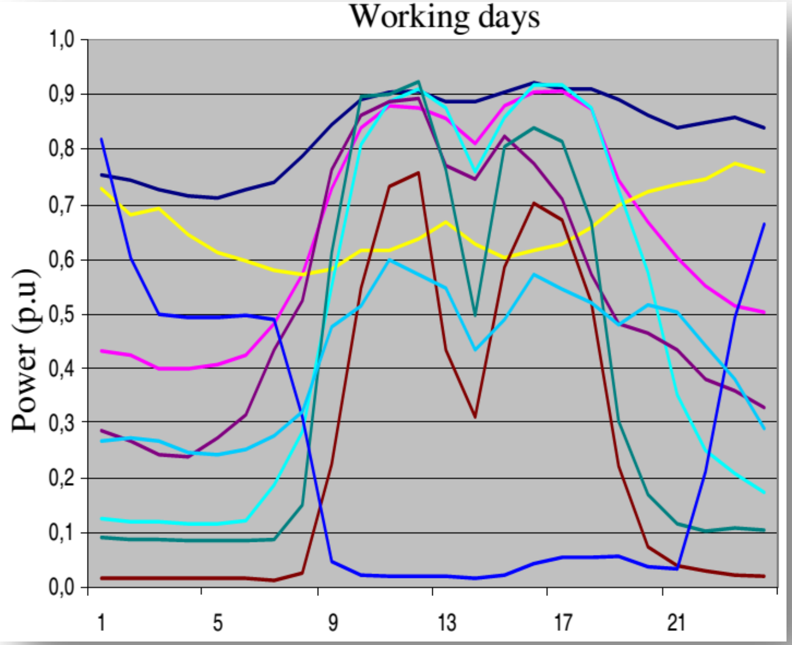
- **Goal:**
  - Evaluate the range of consumption profiles amongst the households (characteristics, distribution of each profile, etc.)
  - Identify "hurtful" (i.e. peak demand) consumption patterns
  - Measures to mitigate peak consumption through load shifting initiatives on targeted segments of customers:
    - Information on utility bills
    - Offer different tariffs
    - Apply behavioral cues

- **Only use load curves → clustering**
- **Set up clustering framework**
- **Evaluate precision of clustering to identify peak consumption**
- **Usability (integration in web portal) requires on-the-fly cluster membership decision**

Outline

Related Work

- **Load forecasting**
- **Supervised learning to extract side information:**
  - Classification of features (size of dwelling, etc.)
- **Clustering of consumption data:**
  - No focus on distinctiveness of clusters obtained
  - No focus on peaks
  - Smaller datasets used

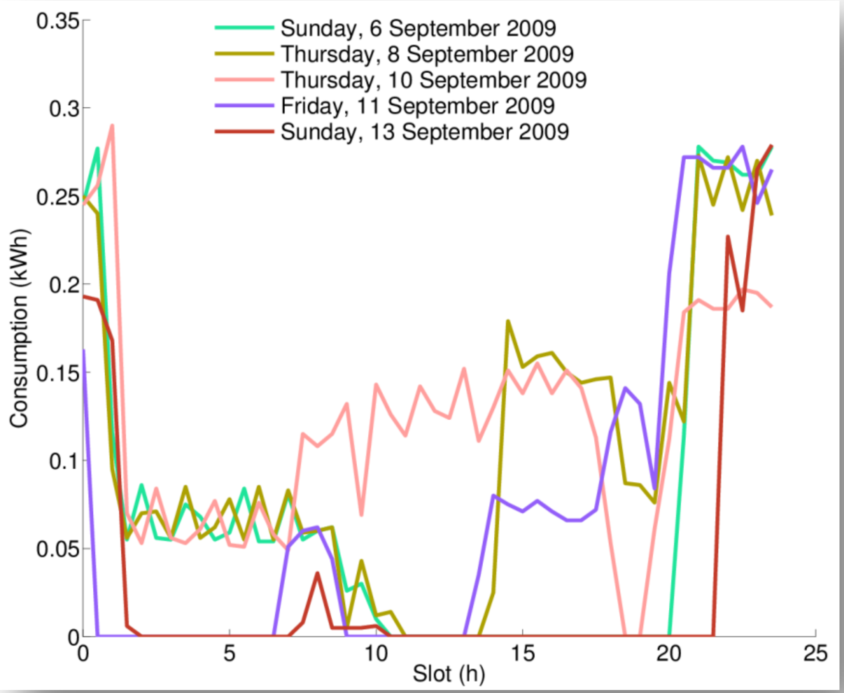


Outline

1. Motivation and Problem Statement
2. Related Work
3. **Methods**
4. Results
5. Summary
6. Future Work
7. Q&A

Methods: Data Mining Best Practices

- **Data analysis on large dataset:**
  - Collected in Ireland
  - Over 4000 households
  - 18 months of data
  - 1 sample every 30 minutes
- **Data cleaning**

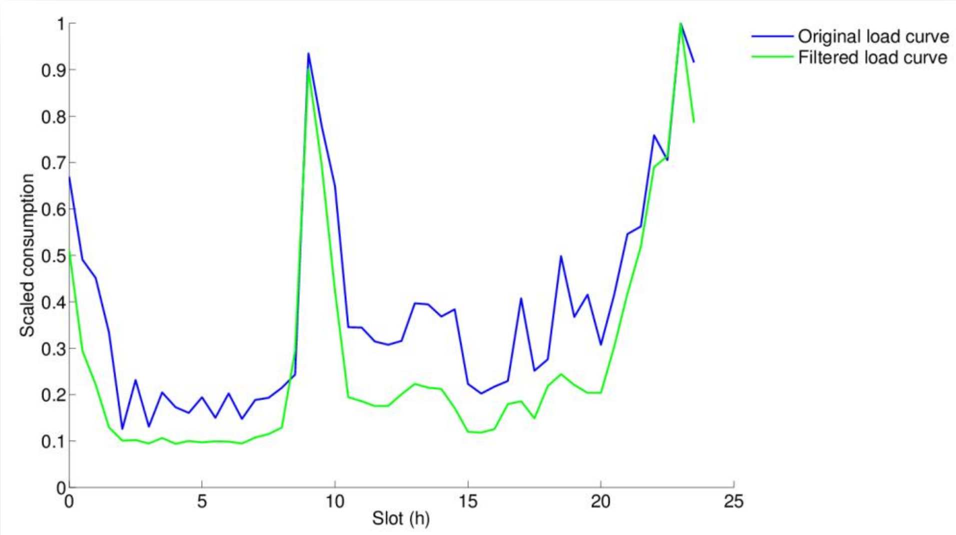


Start date	End date	# Days	# Weeks	Removed	Total
08/17/09	09/13/09	28	4	651	118271
08/17/09	10/31/10	287	41	720	118300
10/26/09	11/22/09	28	4	7153	1212138
10/26/09	12/31/10	215	31	7257	908177

Methods: Data Mining Best Practices

Bits to Energy Lab

- **Data formatting:**
  - Average weekday data
  - Normalize the data
- **Clustering:**
  - 4 weeks of training data
  - Filtering the curve
  - Clustering 48-dim vectors vs. dimension reduction (22 features: mainly statistical data, peak information)
  - Different algorithms and distance measures
  - Variation of the number of clusters (5 to 14)



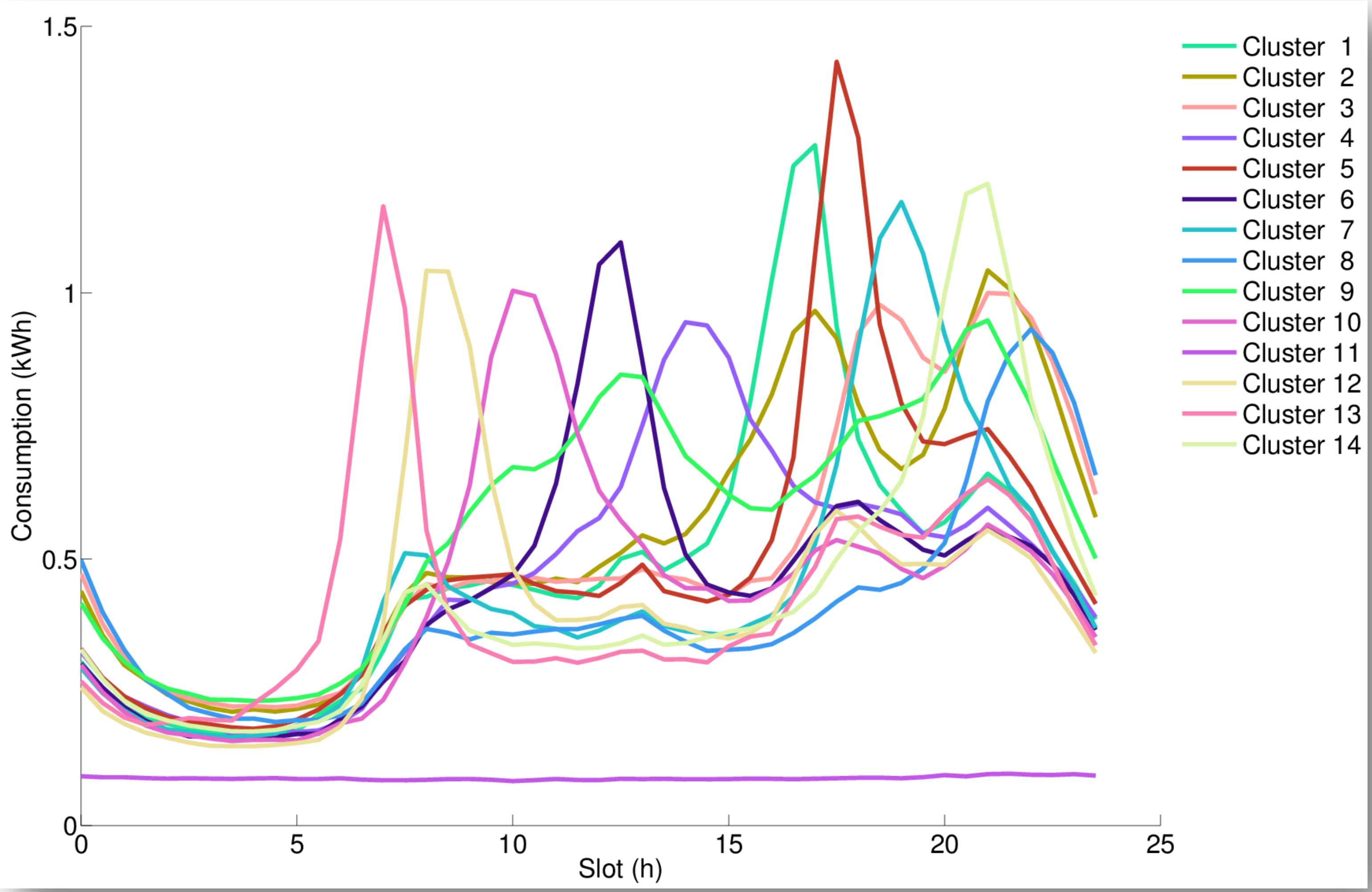
Clustering technique	Distance
SOM + K-Means	Manhattan
SOM + K-Means	Euclidean
K-Means	Manhattan
K-Means	Euclidean
K-Means	Correlation
K-Means	Cosine
Hierarchical	Manhattan
Hierarchical	Euclidean
Hierarchical	Correlation
Hierarchical	Cosine

Outline

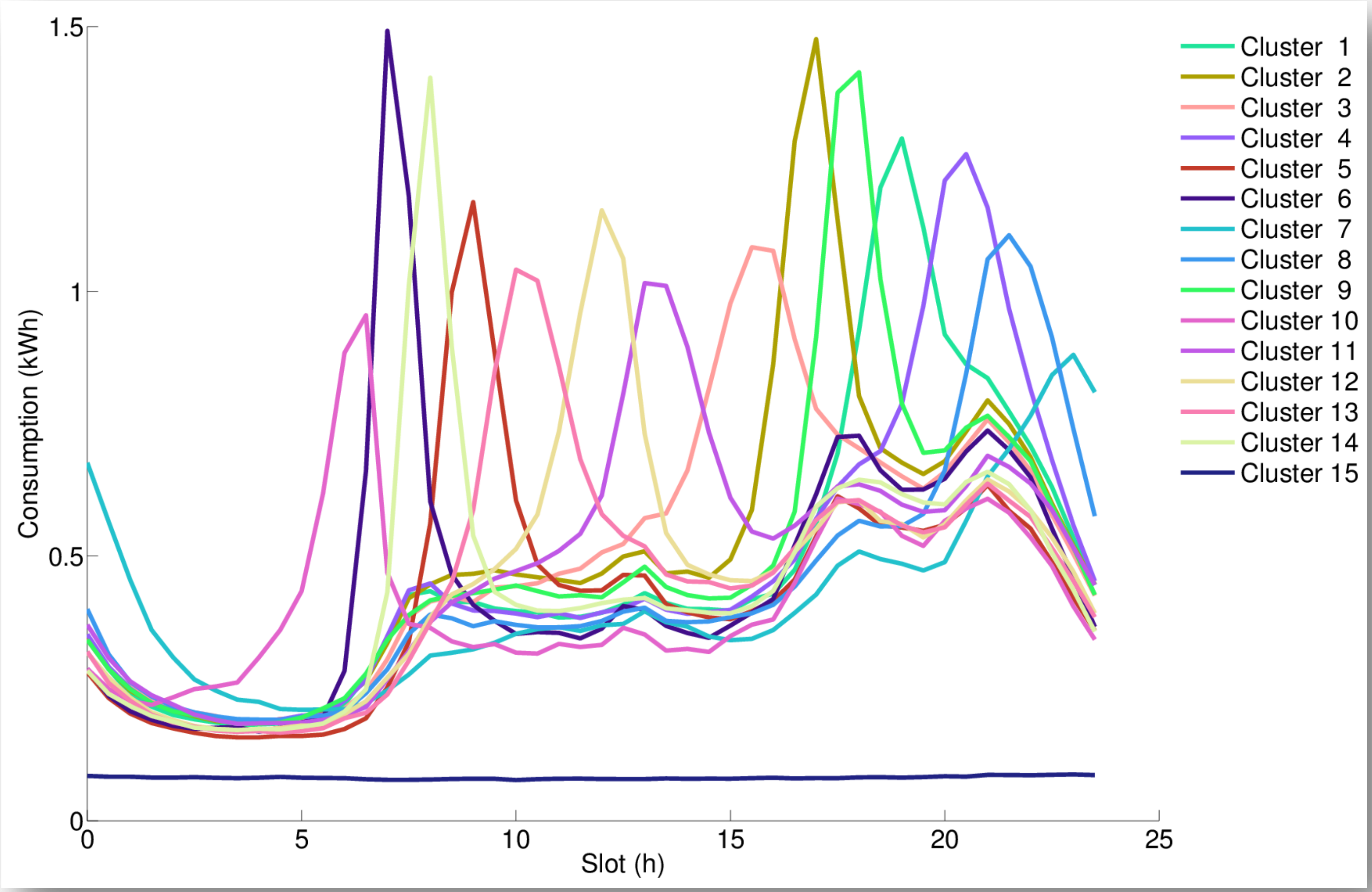
Results: Quantitative and Qualitative Evaluation

Type of clustering	Algorithm	Distance	Filt. Window	# Clusters	Peak Match Score	Distinctiveness Score
Whole clust.	K-Means	Correlation	5	14	0.2199	290
Whole clust.	K-Means	Correlation	5	13	0.21554	236
Whole clust.	K-Means	Correlation	4	14	0.21425	290
Whole clust.	K-Means	Correlation	5	12	0.21182	190
Whole clust.	K-Means	Correlation	4	13	0.20963	236
Whole clust.	K-Means	Correlation	5	11	0.2059	162
Whole clust.	K-Means	Correlation	4	12	0.20321	204
Whole clust.	SOM + K-Means	Euclidean	5	14	0.20179	273
Whole clust.	K-Means	Cosine	5	14	0.20156	260
Whole clust.	SOM + K-Means	Manhattan	5	14	0.19824	259
Whole clust.	K-Means	Correlation	4	11	0.19778	174
Whole clust.	K-Means	Euclidean	5	13	0.19673	192
Whole clust.	K-Means	Correlation	3	14	0.19624	290
Whole clust.	SOM + K-Means	Manhattan	5	13	0.19614	230
Whole clust.	K-Means	Cosine	5	13	0.19597	210
Whole clust.	K-Means	Correlation	2	14	0.19548	276
Whole clust.	K-Means	Cosine	5	12	0.1951	192
Whole clust.	SOM + K-Means	Euclidean	5	13	0.19502	230
Whole clust.	K-Means	Correlation	5	10	0.1946	136
Whole clust.	K-Means	Correlation	2	13	0.19417	238

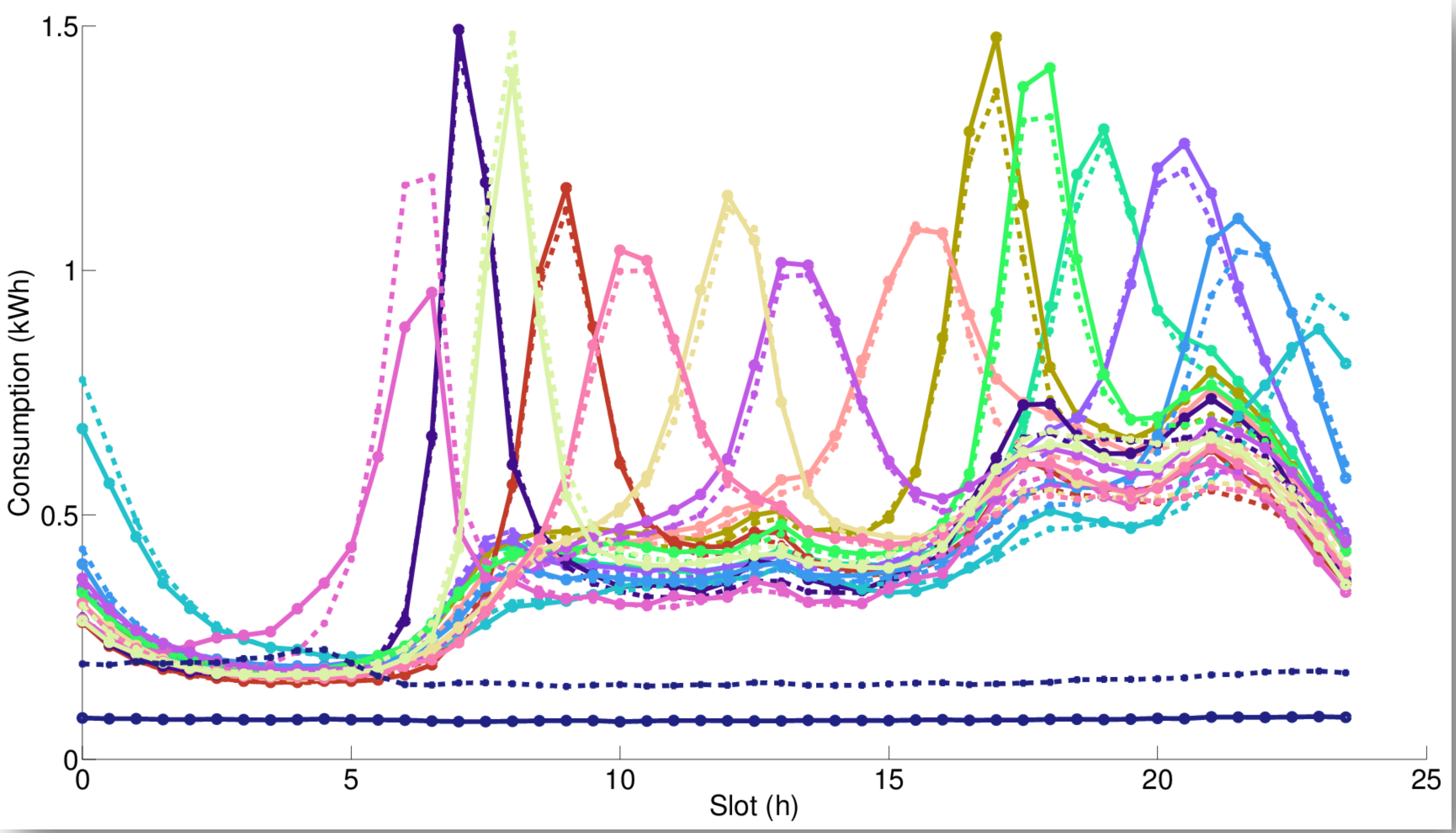
Results: The Aggregating Effect of the Euclidean Distance



Results: Distinct Peaks Throughout The Day



Results: Validation of The Results on the Test Set



Outline

Summary

- **Defined a pattern to be identified in the consumption pattern (i.e. peaks):**
  - Potential to target "hurtful" behaviors at different moments of the day
- **Evaluated :**
  - Most common clustering algorithms and distance measures
  - Use of all the available consumption data vs. aggregated information
- **Possible to establish distinctive "reference" consumption patterns**
- **Deciding cluster membership is a quick operation**
  - Can be performed on the fly

Outline

1. Motivation and Problem Statement
2. Related Work
3. Methods
4. Results
5. Summary
6. **Future Work**
7. Q&A

Future Work

- **Time series analysis of the household cluster membership:**
  - MCMC for missing data
  - Clustering of similar households (bag of words vs. time series approach)
    - Spectral clustering robustness
  - Markov chain modeling
  - Mapping survey data (household characteristics) to consumption pattern
- **In collaboration with** BEN Energy
  - Regional/cultural effect on characteristic load profiles through the usage of Swiss data
  - Portal sign-up probability using machine learning and publicly available data

Thank you for your attention

Contact

Hông-Ân Cao | Distributed Systems Group | Bits to Energy Lab | Department of Computer Science, ETH Zurich
Office: +41 44 632 02 73 | E-Mail: hong-an.cao@inf.ethz.ch

Team

Prof. Dr. Elgar Fleisch	Dr. Tobias Graml	Hông-Ân Cao
Prof. Dr. Friedemann Mattern	Vojkan Tasic	Felix Lossin
Prof. Dr. Thorsten Staake	Verena Tiefenbeck	Anna Kupfer
	Christian Beckel