DATA MINING TECHNIQUES IN ELECTRICITY PRICING

László Varga
E.ON Hungária Group

Workshop on Deregulated European Energy Market
September 24-25, 2009
Collegium Budapest, HUNGARY

Competitive Electricity Markets

The electricity business was born in 1882 when T. Edison built the first commercial power generation and distribution system in New York City.

At first, the electric systems were local affairs providing power only for neighborhood areas.

However, these territories began to overlap, and that was the first time for the competition in electricity supply.

Infrastructure grew in scale and as consolidation in ownership occured, the industry began to take on the characteristics of a „natural monopoly”.

Competitive Electricity Markets

The last three decades have seen radical changes in the electricity market structure around the world.

At the beginning there are widespread concern about competition and power system security are mutually exclusive.

A number of unique features in comparison with any other commodity.

- large volatility in electric prices as a consequences of the high variability in demand and large differences in the cost of electricity production.
- Electricity cannot be stored transmission congestion can lead to a temporary geographical fragmentation of the market

Competitive Electricity Markets

The electricity business, competitive or otherwise, comprises five more or less mutually exclusive services

- generation: production of wholesale quantities of power;
- transmission: the transportation of wholesale power over large distance using high voltage cable networks;
- Ancillary services: products to balance supply and demand in real time and maintain overall system security;
- distribution: the transportation of power from the transmission system to the consumer;
- wholesale/retail supply: services to facilitate the purchase and sale of the physical commodity (marketing/supply, metering, billing);

Customer characterization

The market liberalization has been given a greater degree of freedom in fixing tariff rates, respecting a set of regulatory rules.

Detailed knowledge of the customer behavior is thus essential for designing specific tariff options supplying the various types of customers.

Customer characterization: adequate information on the consumption patterns of customers.

Load profiling: tools used to perform customer characterization forming customer classes based on consumer load curves.

Load classification: models which assign different consumers to existing classes.

The main objectives of the presentation

1. Comparisons of several data mining techniques for consumer load characterization

- classical multivariate statistical methods

SPSS 15.0 for Windows
(SPSS Inc., Chicago, USA)

- computational intelligence approaches

NeuralWorks Predict
(NeuralWare Inc., Pittsburg, USA)

2. Application of load characterization for procurement cost calculations

Data mining techniques for consumer load characterization

Consumer load profiling

- cluster analysis
- self-organizing maps

Consumer load classification

- discriminant analysis
- three-layer feed forward neural networks

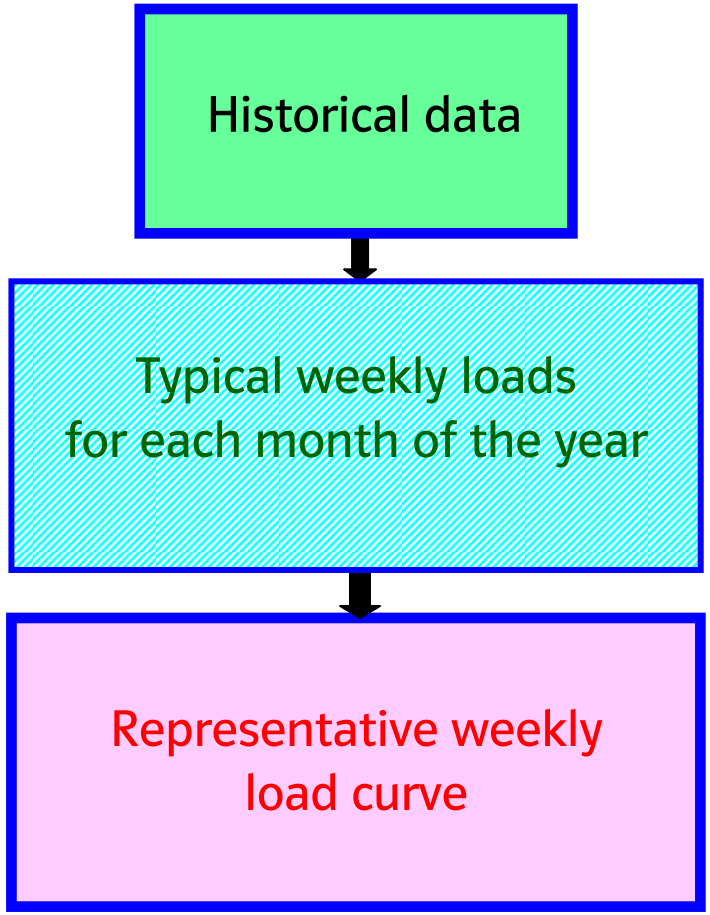
Measurement data

All metered historical data for consumers of E.ON Energy Service.

Quarter hourly data from January to December
in the years of 2005, 2006, 2007.

After data compression and filtering we have hourly normalized weekly load curves for 1231 consumers.
(representative weekly loads)

Representative weekly load curves



Typical weekly load:
Average value of the quarter hourly loads for each day of the week (Monday, Tuesday,…,Saturday)

Normalized weekly load:
Typical weekly load was normalized by the average weekly load on hourly basis

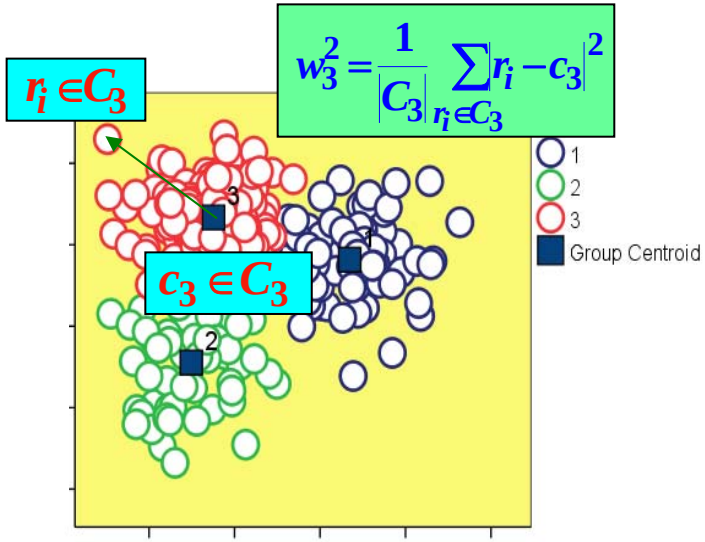
Representative weekly load:
Average value of the monthly normalized weekly loads

Consumer Load Profiling

Let $w_j^2 = \frac{1}{|C_j|} \sum_{r_i \in C_j} |r_i - c_j|^2$ be the average of mean square distance between weekly loads in j-th group and group center c_j, where $|C_j|$ denotes the number of patterns in j-th group. $(j = 1, 2, \ldots, M)$.

Mean Index Adequacy (*MIA*)

$$MIA = \sqrt{\frac{1}{M} \sum_{j=1}^{M} w_j^2}$$



Consumer Load Profiling

Classical multivariate statistical methods

- K-mean (*MacQeen*)
- hierarchical

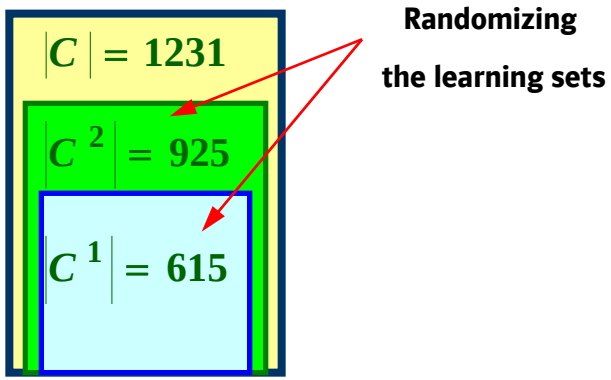
Computational intelligence approaches

- self-organizing map (*Kohonen*)

	Adequacy Indices		
Method	MSE	MIA	CDI
K-Mean	5.579	2.844	0.118
Hierarchical	5.938	2.844	0.114
SOM	3.890	2.341	0.095

Consumer Classification

$$C^2 \subset C^1 \subset C$$



Classical multivariate statistical method

- discriminant analysis
  - linear,
  - quadratic.

Computational intelligence approach

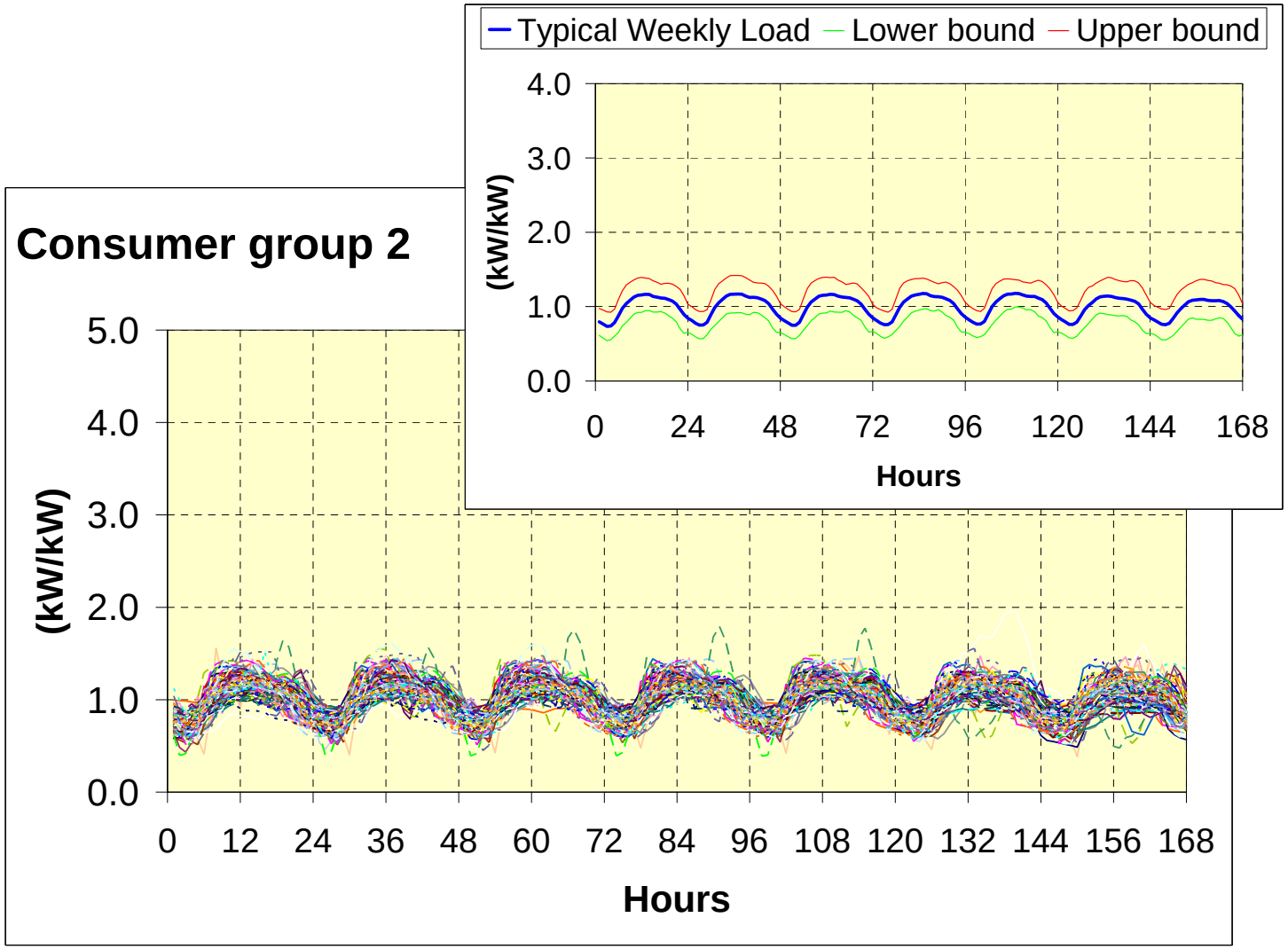
- Three-layer feed-forward neural networks

Method	Correct classification (%) Average	Max	Min
Linear	90.8	97.0	80.5
Quadratic	94.7	100.0	87.5
Neural Network	96.0	100.0	93.4

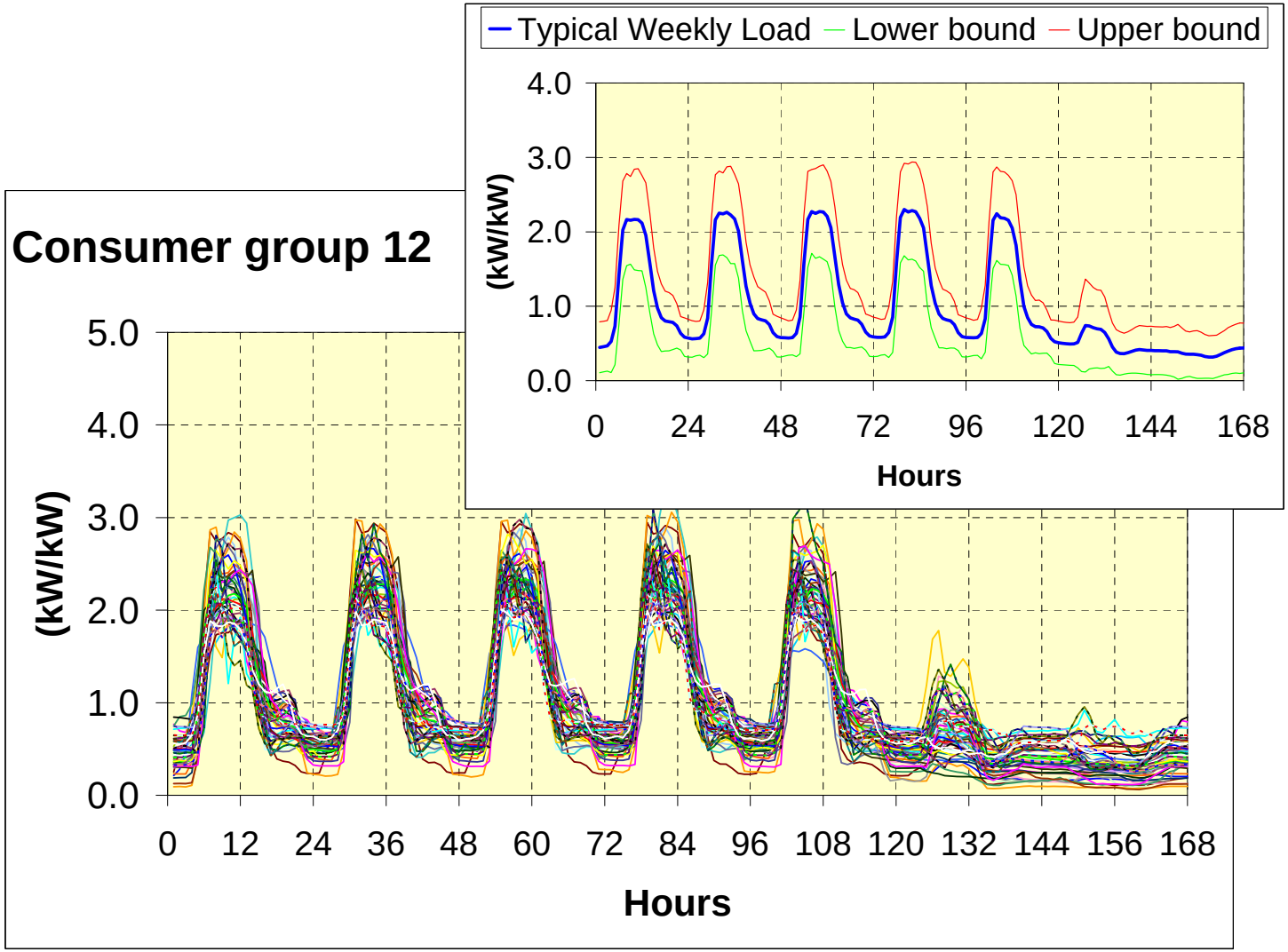
Neural network — Correct classification (%)

Learning set	Average
1231 (100%)	96.0
925 (75%)	94.2
615 (50%)	91.3

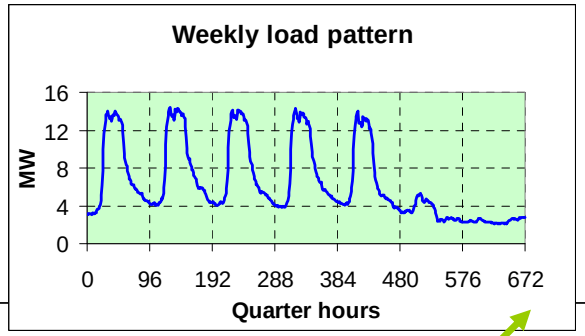
Consumer Load Profiling

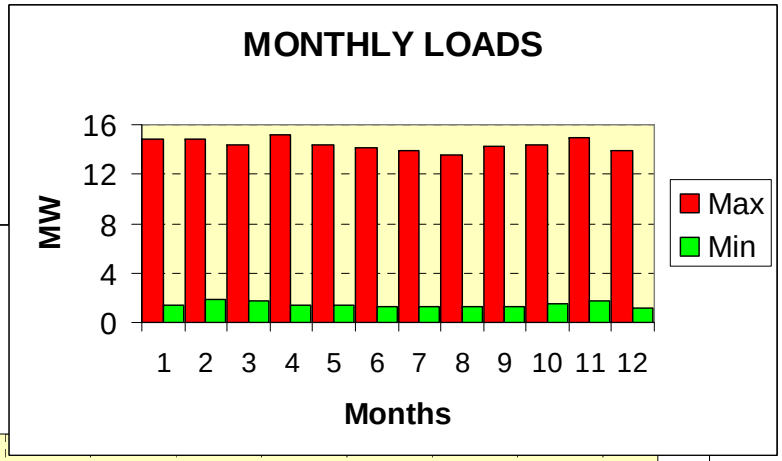


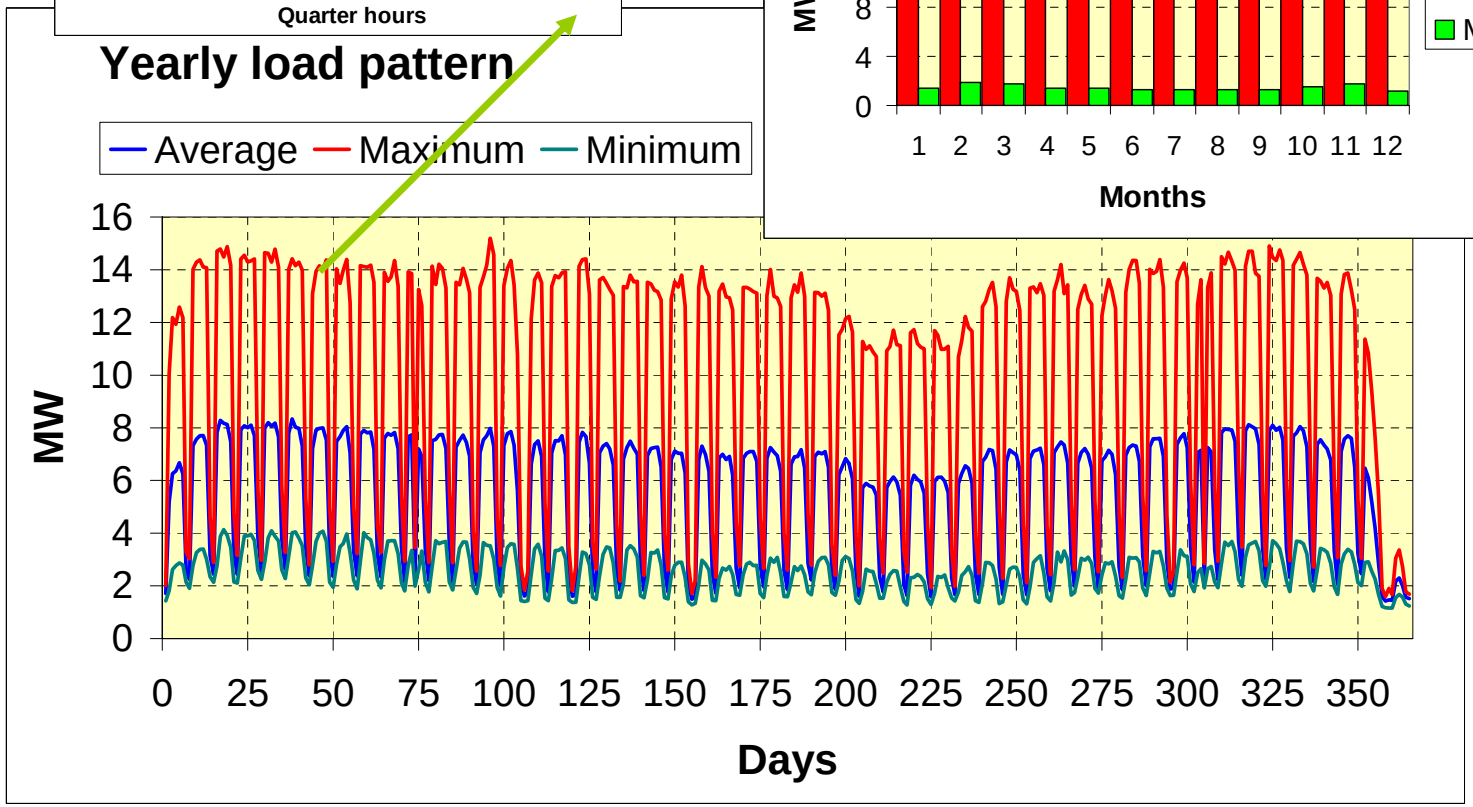
Consumer Load Profiling



Aggregated electric loads for the consumer group 12

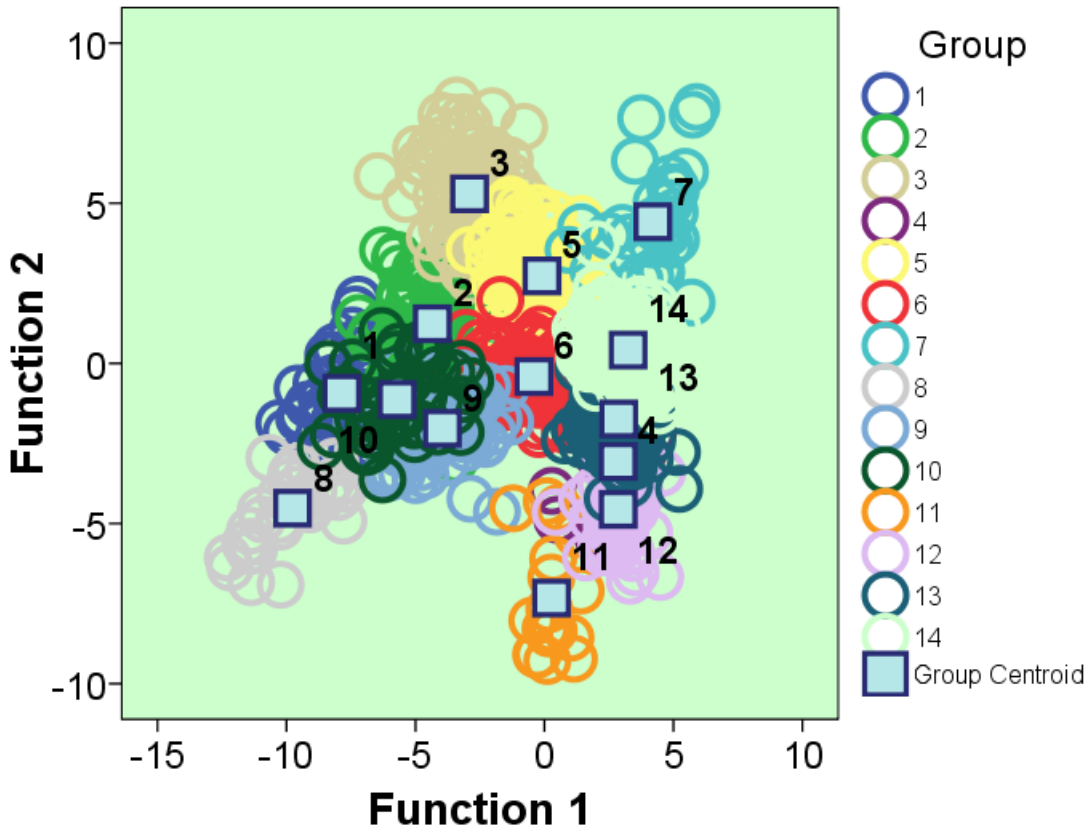






Consumer Load Profiling

Clustering results in the plain of the first two canonical variables



Consumer Load Profiling and procurement costs

European Energy Exchange (EEX)

Block products considering average of the hourly EEX spot prices over the time interval of blocks:

No.	PRODUCTS		Time period
1	Base-load		00-24
2	Weekdays	Peak	08-20
3	Holidays	Peak	08-20
4	Weekdays	Base-load	00-24
5	Holidays	Base-load	00-24
6	Weekdays	Off-peak	21-07
7	Holidays	Off-peak	21-07
8	Weekdays	Peak	06-22
9	Holidays	Peak	06-22
10	Weekdays	Morning	07-11
11	Weekdays	High Noon	12-14
12	Weekdays	Afternon	15-18

Consumer Load Profiling and procurement costs

The task is to cover the consumer load of a given consumer group using block products and balancing energy at minimum procurement cost.

Objective function (*non-linear and non-convex*):
(Block capacity x Time interval) x Product price.

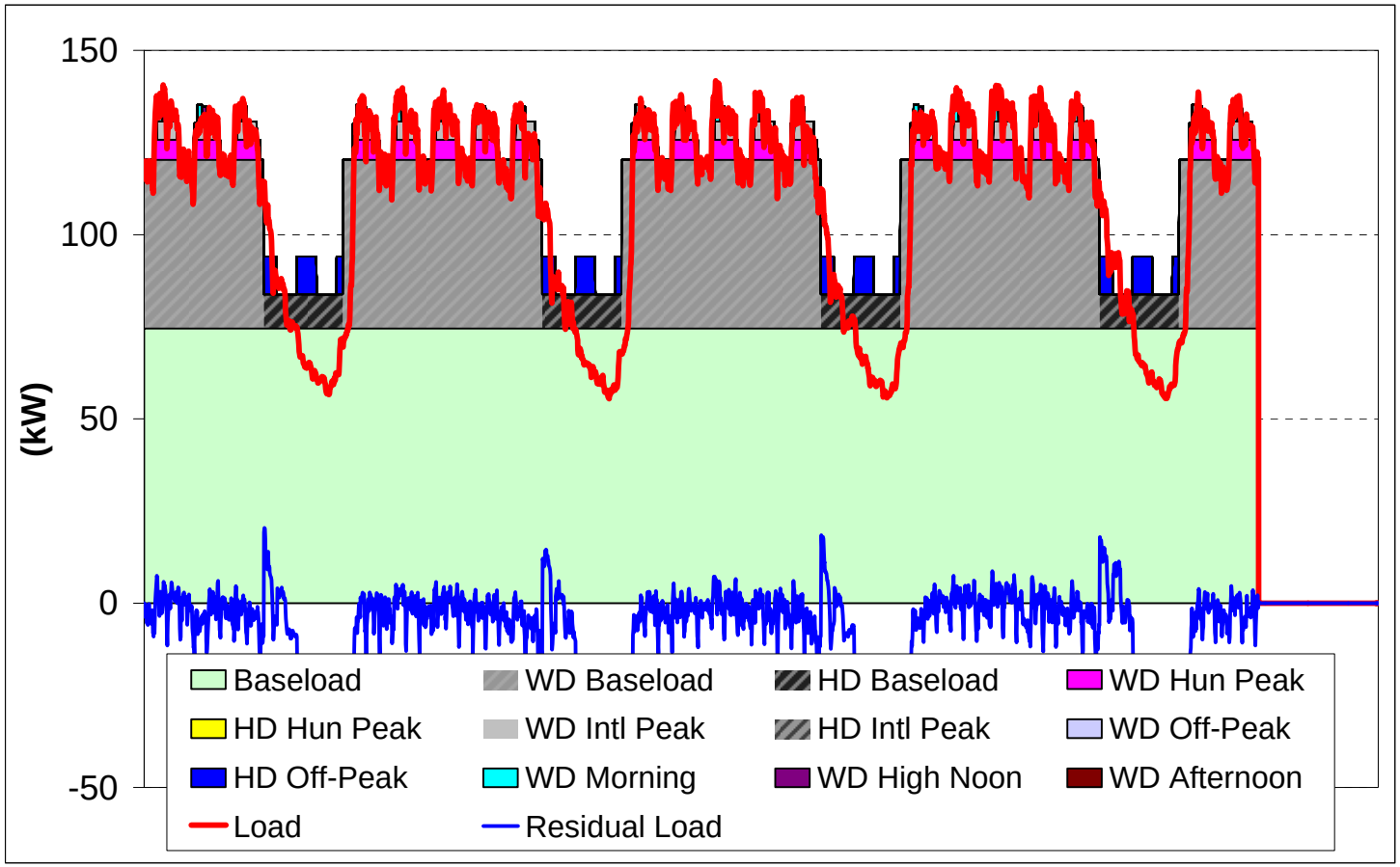
Independent variable (*continuous or integer*):
Block capacity.

Constraints (*linear*):
- for the products (capacity and time interval)
- lower and upper bound for under and over covering.

Premium Solver Platform 8.0
(Frontline Systems, Inc. Inline Village, NV, USA)

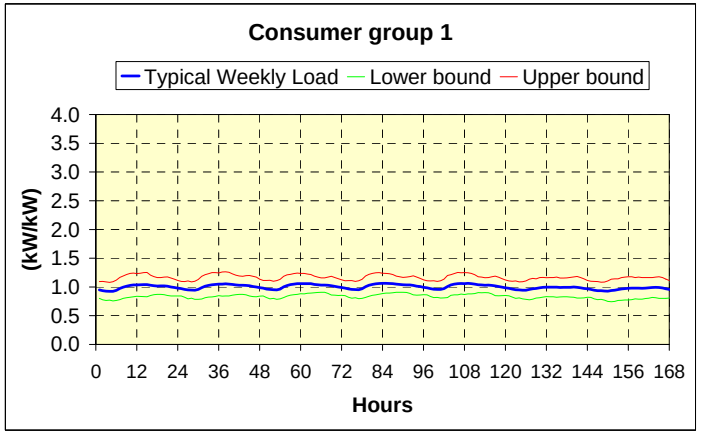
Consumer Load Profiling and procurement costs

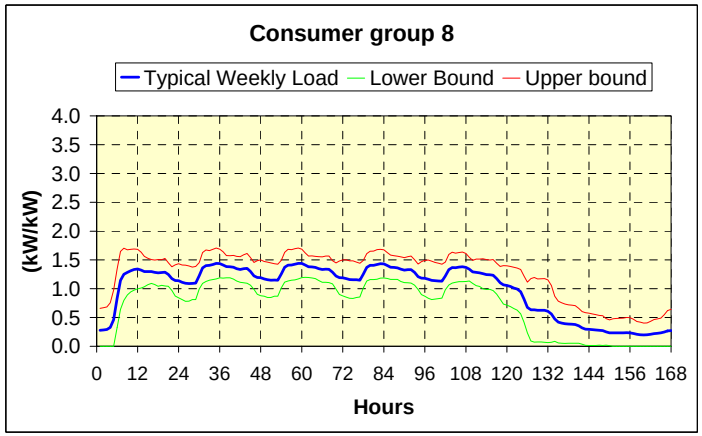
Result of the optimized purchase costs using 12 products

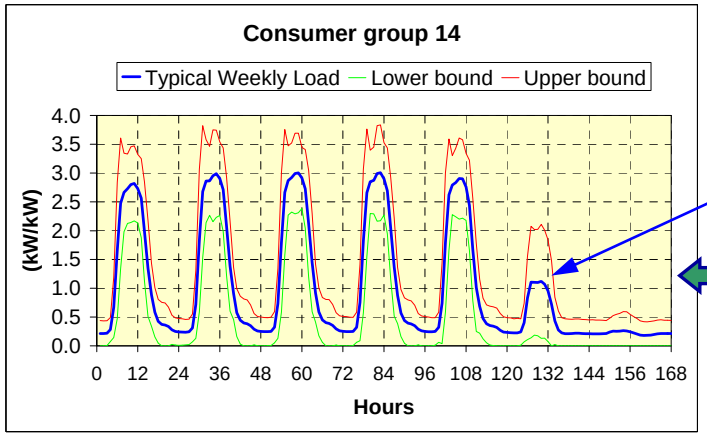


Consumer Load Profiling and procurement costs

Consumer group	Vol-INDEX	Proc. Cost
1	0.500	1.066
8	5.655	1.393
14	13.241	1.876







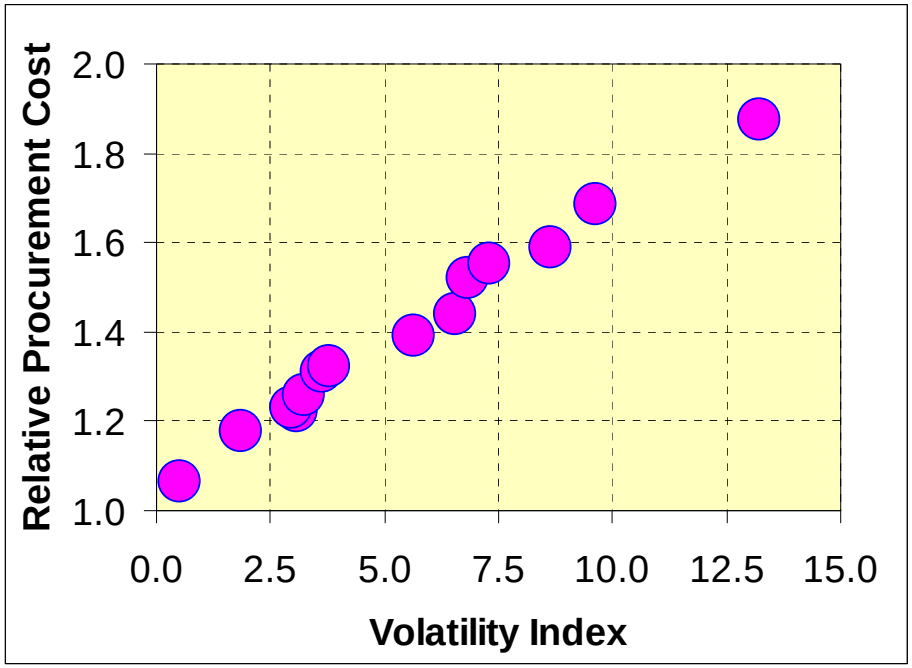
Volatility index

Euclidean distance between the normalized load curve and the constant load with value 1.0.

Consumer Load Profiling and procurement costs

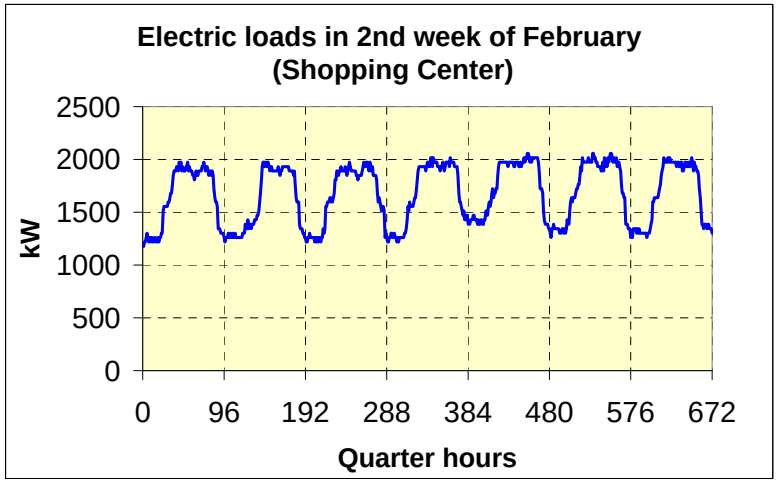
Consumer group	Vol-INDEX	Proc. Cost
1	0.500	1.066
8	5.655	1.393
14	13.241	1.876

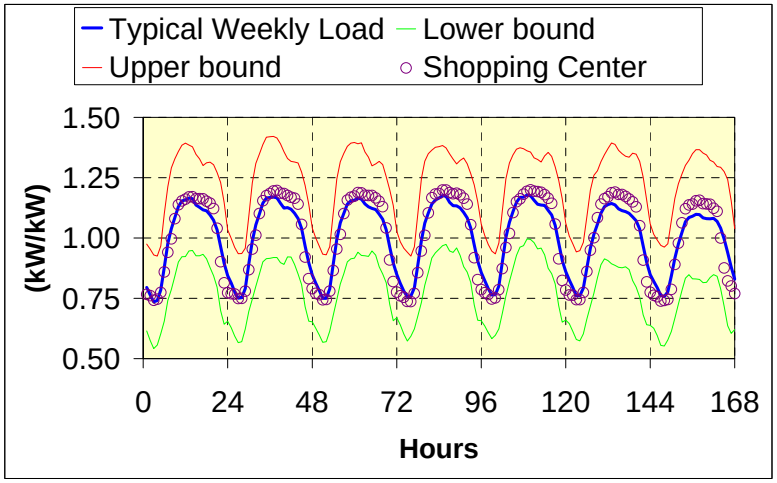
Procurement costs versus volatility index

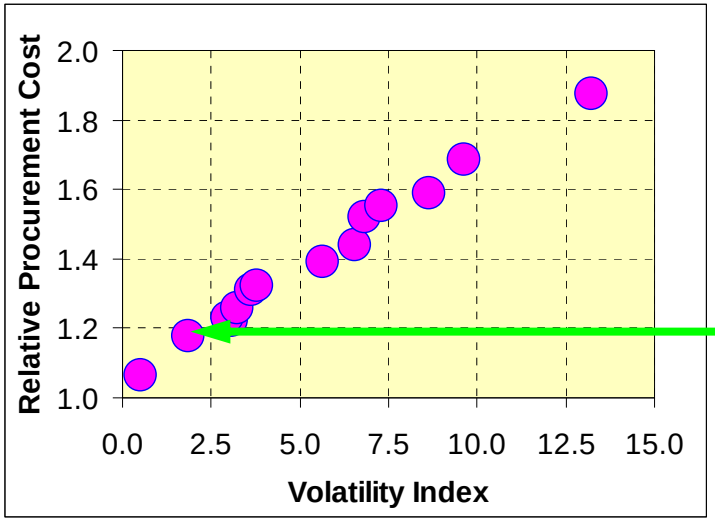


Electricity Pricing based on consumer classification

Data compression and normalization







Consumer classification

Consumer GROUP 2

Conclusions

Consumer load profiling

- Computational intelligence methods performed better in comparison with classical multivariate approaches

Consumer load classification

- Three-layer feed-forward neural network performed the best classification accuracy

Consumer load characterization and procurement cost

- There is a relatively simple relationship between the procurement cost and volatility index