CEPEL 35 anos
Sistema Eletrobrás

França.Br 2009

Clustering Electric Load Curves: The Brazilian Experience

Workshop Franco-Brasileiro sobre Mineração de Dados

Workshop Franco-Brésilien sur la Fouille des Données

José Francisco Moreira Pessanha (CEPEL / UERJ) francisc@cepel.br

Luiz da Costa Laurencel (UERJ / UFF) getlecl@vm.uff.br

Introduction

CEPEL
Grupo Eletrobrás

- The Brazilian electric power sector adopts tariffs based on marginal cost pricing since 1980.
- The electricity tariffs are calculated by a methodology, whose origin is the Electricité de France (EDF) and the French 'marginaliste' economists like Allais and Boiteux.
- In this methodology, an important step is the identification of a few typical daily load profiles from a set of electric load curves measurements on a sample of customers.
- These profiles represent patterns of energy use of different class of customers e.g. residential, commercial, industrial, rural, public lighting, public administration etc.
- The standard way to identify the typical load profile from a sample of load curves is to perform a clustering of the representative workday load curves. The centroid of each cluster defines a typical load profile.
- This work presents a brief history about the softwares for identifying of typical daily load profiles developed in the Brazilian electric power sector.
- In order to tell this history we present the features of three softwares that have been used by the Brazilian electric distribution utilities.

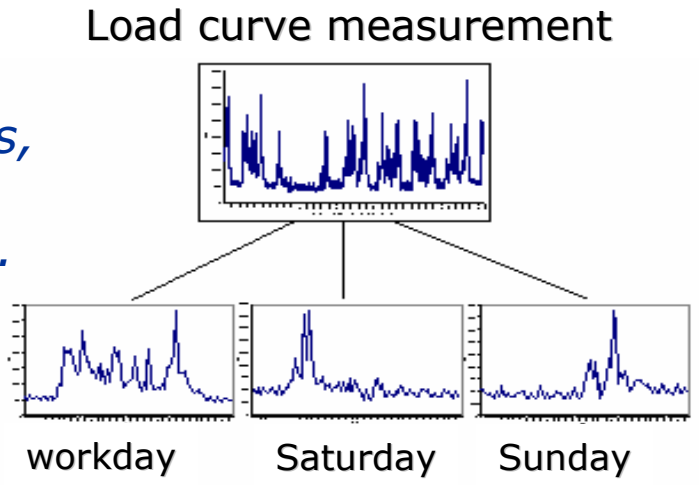
Procedure to identify typical load profiles

CEPEL
Grupo Eletrobrás

1) Select a sample of customers (or power transformers)

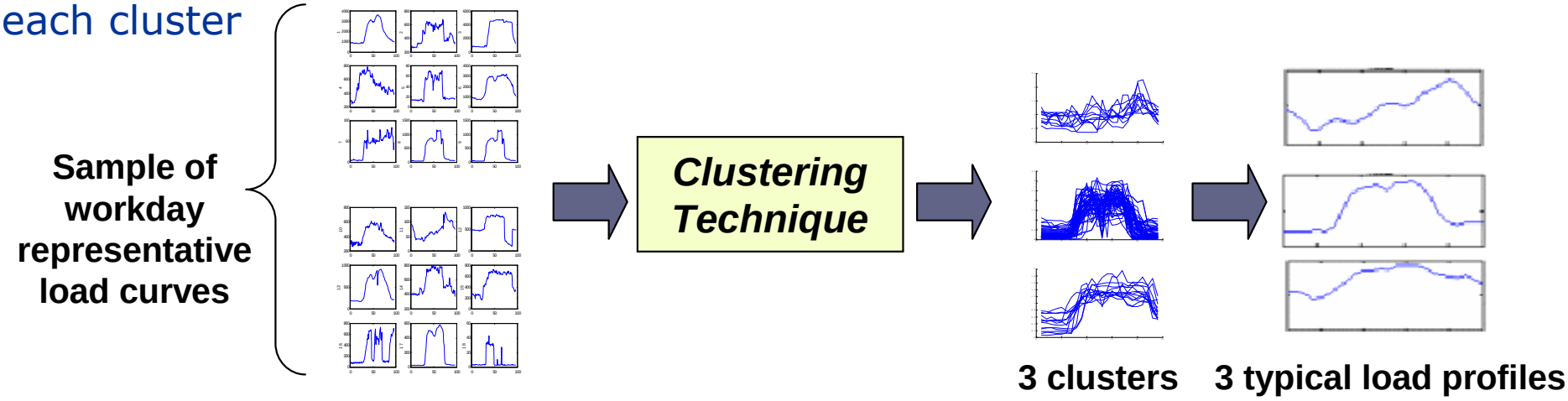
2) Get a load curve measurement from each customer

The load curves measures cover a period of two weeks, where the demand is recorded every 15 minutes by recording meters installed at each point in the sample.



3) Examine the load curve measurement (kW) of each customer in order to identify its three representative load curves (workday, Saturday and Sunday)

4) Apply a clustering technique to group customers with similar workday load curves (each daily curve has 96 points). After that, take the typical load profiles from each cluster



5) Derive typical load profiles for the entire population

Softwares for identifying of typical load profiles developed in the Brazilian electric power sector

CEPEL
Grupo Eletrobrás

We select a representative sample of three softwares for identifying of typical load profiles used in the Brazilian electric power sector.

- **SNACC (1991)**

Sistema Nacional de Avaliação do Comportamento da Carga

National system of evaluation of the load behavior

- **TARDIST (1998)**

Programa para cálculo dos custos marginais de fornecimento e tarifas de uso da distribuição

Computational program for computing the supply marginal costs and distribution tariffs

- **ANATIPO (2005)**

Sistema computacional para construção de tipologias de curvas de carga

Software for building typical load curves

SNACC Program (1991)

CEPEL
Grupo Eletrobrás

It was the first software for clustering load curves developed in the Brazilian power sector.

It was developed by the National Department of Waters and Electric Power (DNAEE).

DNAEE was officially closed upon the establishment of the Brazilian Electricity Regulatory Agency (ANEEL).

The SNACC employs the methods of cluster analysis programmed in two computational routines in Fortran brought from the Electricité de France (EDF): NUDYC (*nuées dynamique*) and DESCR2 (Ward method).

• *Molliere, M. Um ensemble de modules de classification automatique et de modules explicatifs associes, Note EDF, Direction des etudes et Recherches nº HI 2818/02, 1978.*

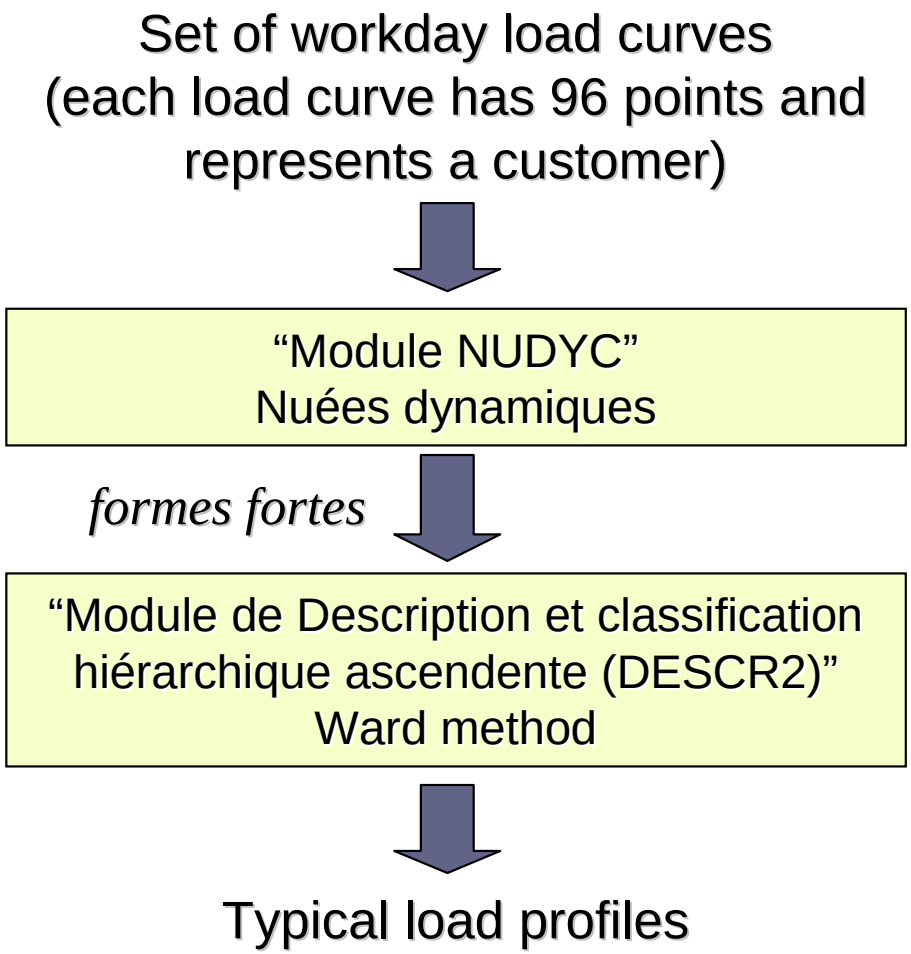
• *BRASIL, Ministério das Minas e Energia, DNAEE, Eletrobrás, Empresas Concessionárias de Energia Elétrica, Nova Tarifa de Energia Elétrica: metodologia e aplicação, DNAEE, Brasília,1985.*

SNACC Program (1991)

CEPEL
Grupo Eletrobrás

The NUDYC and DESCR2 routines are executed sequentially:

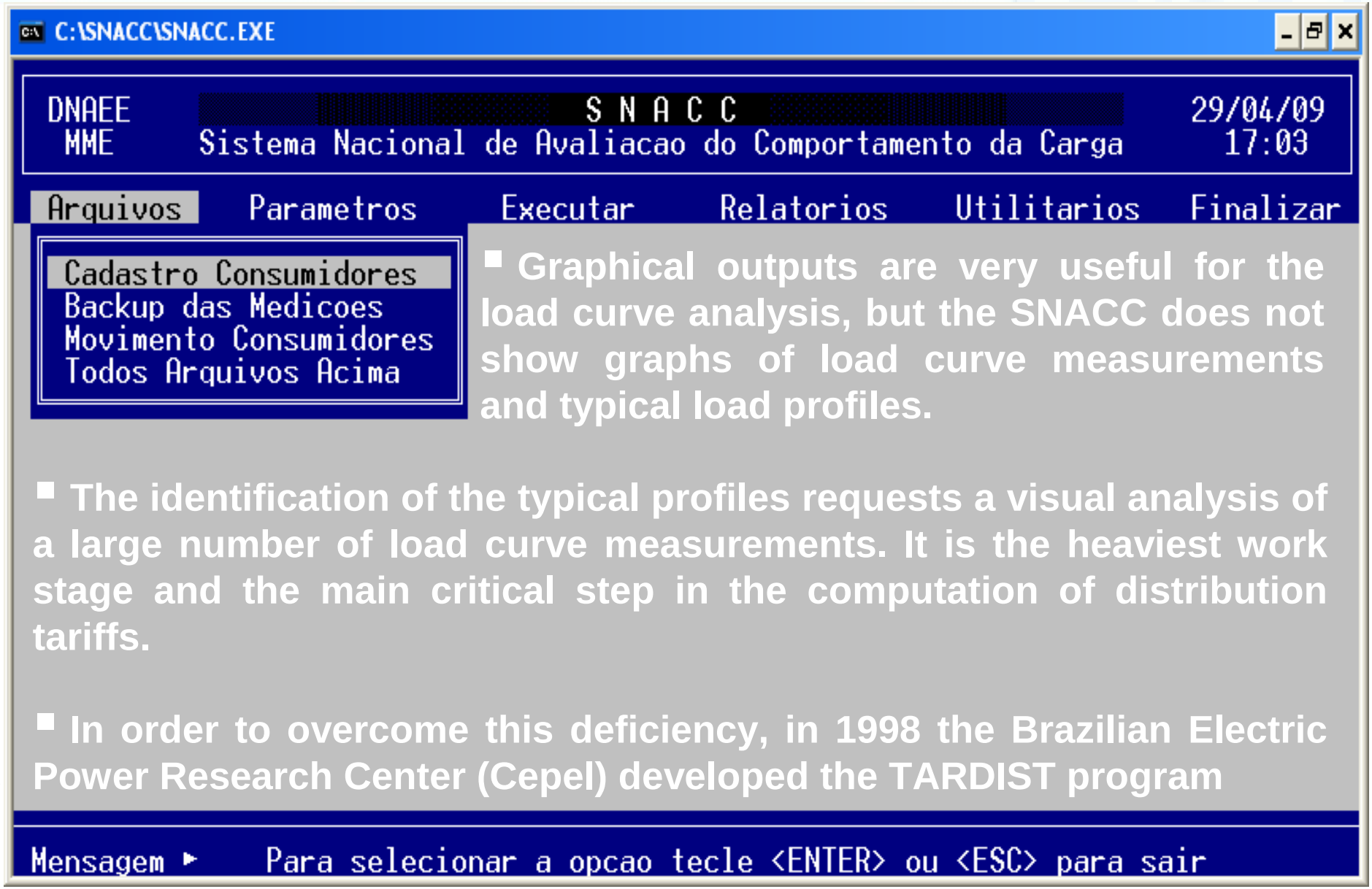
first the typical workday load curves are clustered by the "*nuées dynamique*" programmed in the NUDYC routine, then the *formes fortes* are clustered by the Ward method programmed in the DESCR2 routine.



Diday, E. Une nouvelle méthode em classification automatique et reconnaissance des formes. La méthode des nuées dynamiques. Revue de statistique Appliquée, 1971, vol. XIV nº 2. Institut de Statistique. Université de Paris.

SNACC Program (1991)

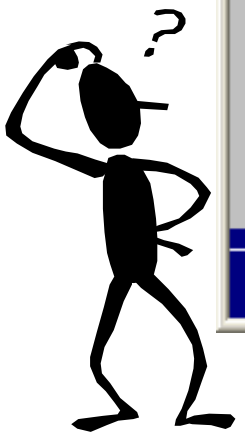
CEPEL
Grupo Eletrobrás



- Graphical outputs are very useful for the load curve analysis, but the SNACC does not show graphs of load curve measurements and typical load profiles.

- The identification of the typical profiles requests a visual analysis of a large number of load curve measurements. It is the heaviest work stage and the main critical step in the computation of distribution tariffs.

- In order to overcome this deficiency, in 1998 the Brazilian Electric Power Research Center (Cepel) developed the TARDIST program



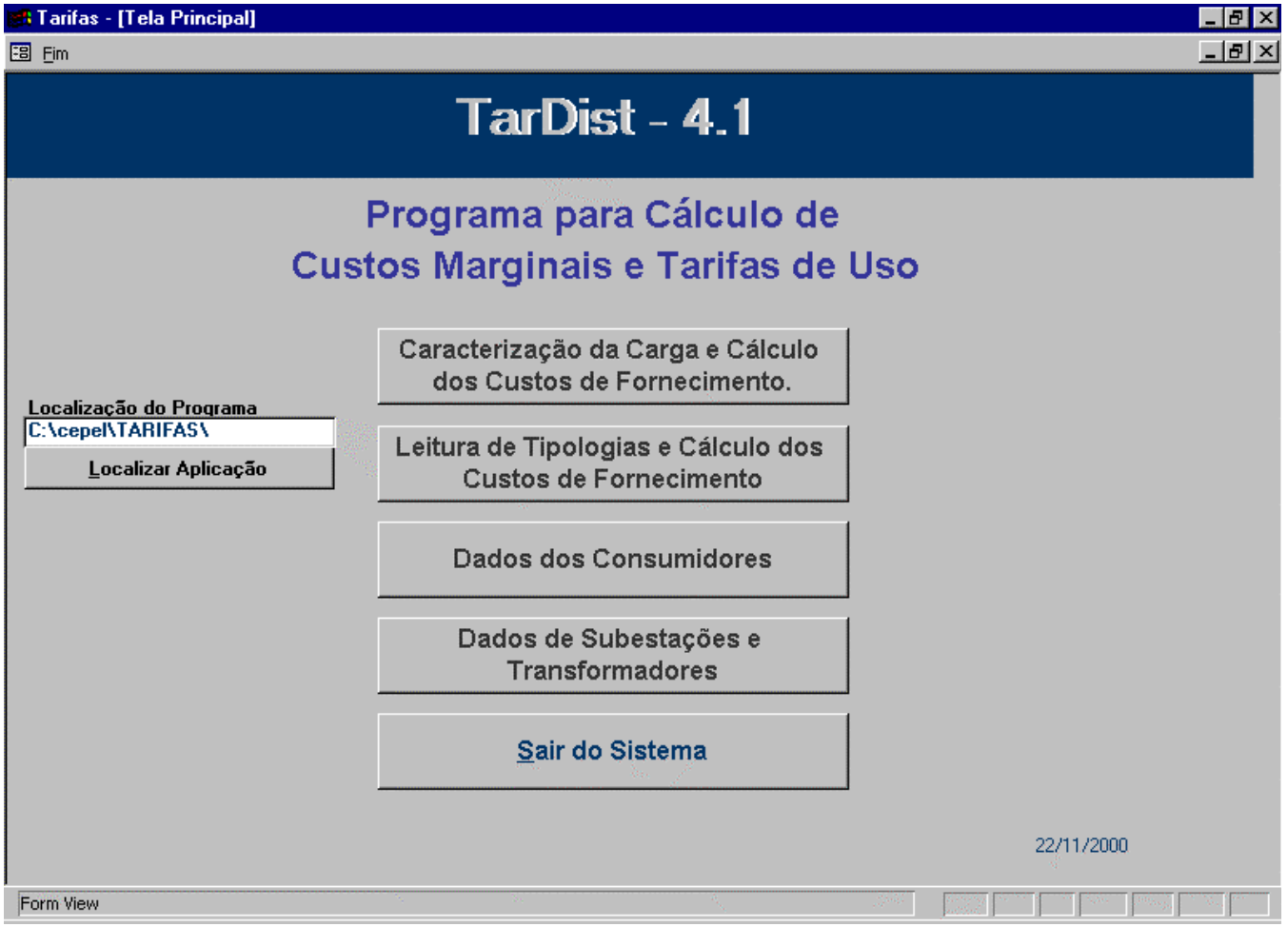
TARDIST program (1998)

CEPEL
Grupo Eletrobrás

Software developed by the Brazilian Electric Power Research Center (Cepel) to compute the distribution tariff framework based on marginal cost.

TARDIST also has a module to build typical load profiles, but it has a friendly user interface.

The software is used by the Brazilian Electricity Regulatory Agency (ANEEL) to set the distribution tariff.



Pessanha, J.F.M., Huang, J.L.C., Pereira, L.A.C., Passos Júnior, R., Castellani, V.L.O. Metodologia e sistema computacional para cálculo das tarifas de uso dos sistemas de distribui»cão, XXXVI SBPO, São João del Rey - MG,2004.

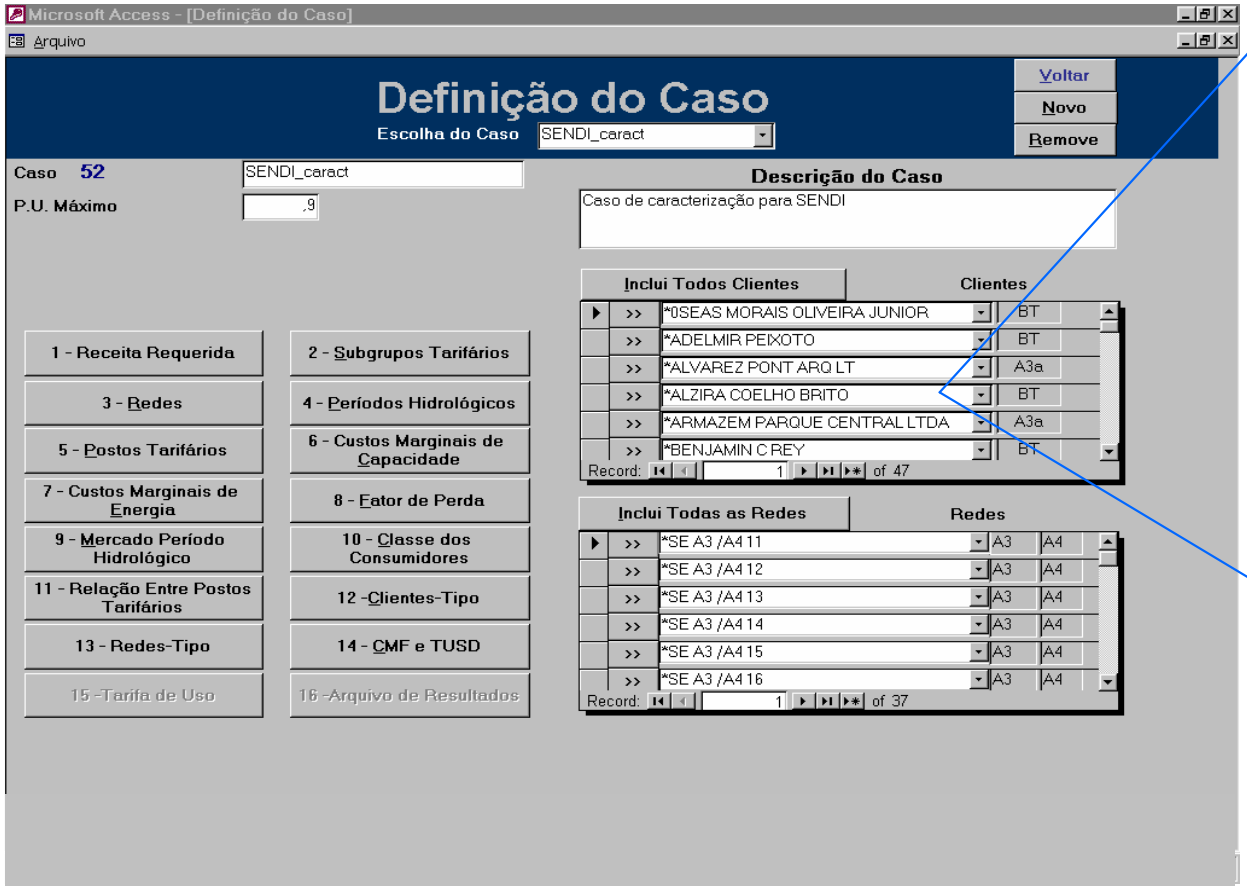
TARDIST program (1998)

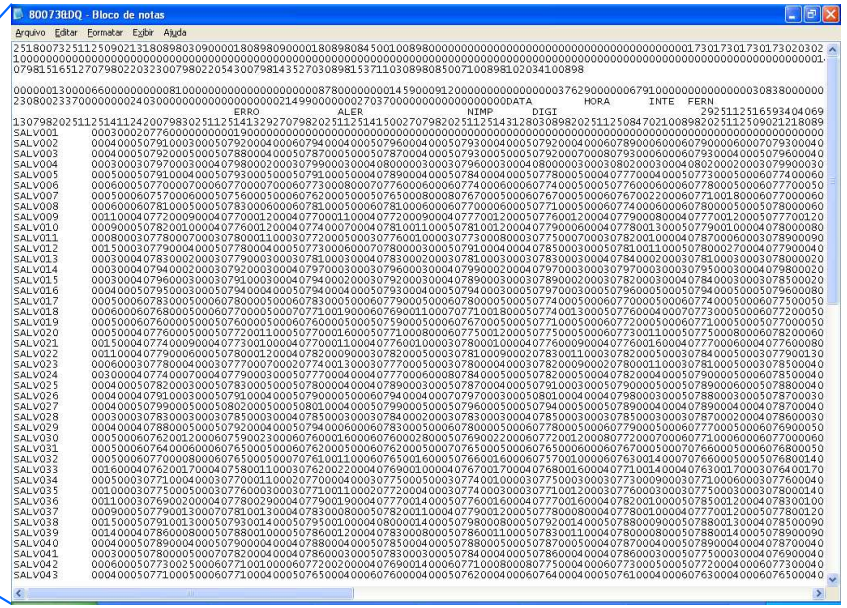
CEPEL
Grupo Eletrobrás

Input data:

- List of costumers and electric power transformers in the sample.
- Load curves measurements files (Excel or text format) for each sample point
- Annual energy demand (MWh) for each consumption class and energy flows (MWh) among the voltage levels.

load curve measurement file



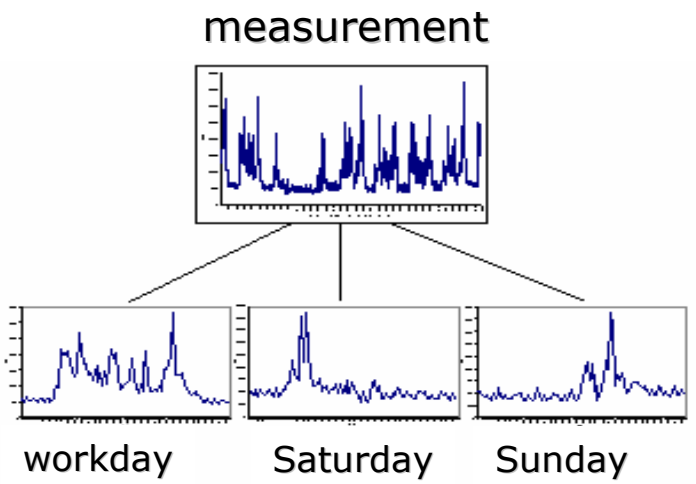


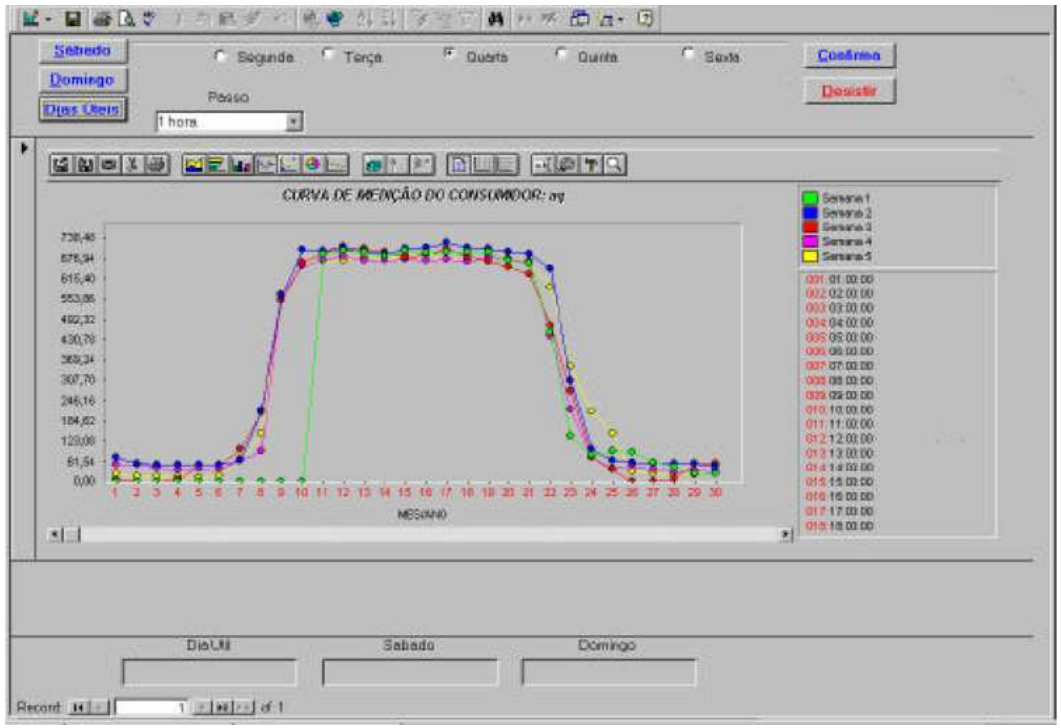
TARDIST program (1998)

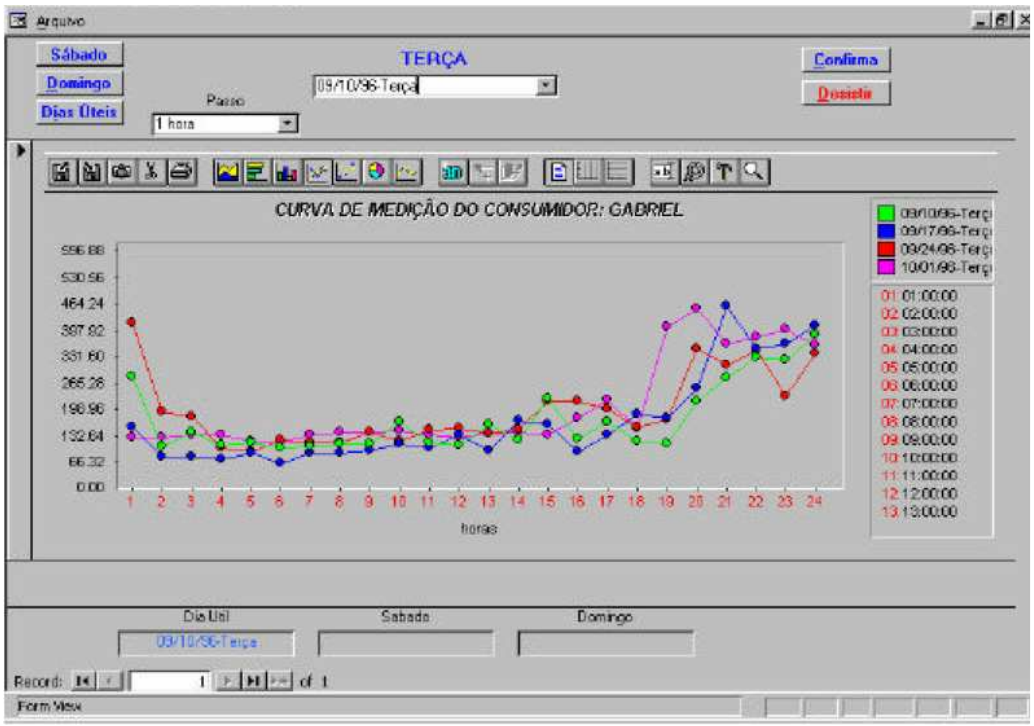
CEPEL
Grupo Eletrobrás

The friendly user interface shows the load curves registered in each measurement file.

Based on a pictorial analysis any user can select the three typical days (a workday, a Saturday and a Sunday) of each measurement file.





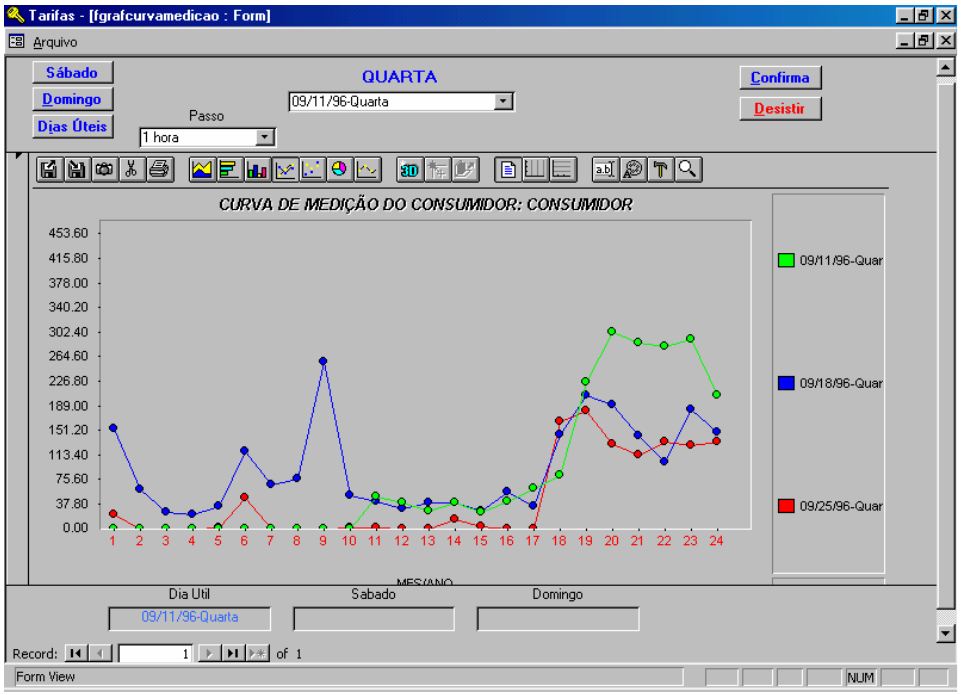


TARDIST program (1998)

CEPEL
Grupo Eletrobrás

TARDIST employs only the Ward method to cluster the workdays load curves.

The user can change the clusters' composition, in order to correct any misclassification made during the clustering process.



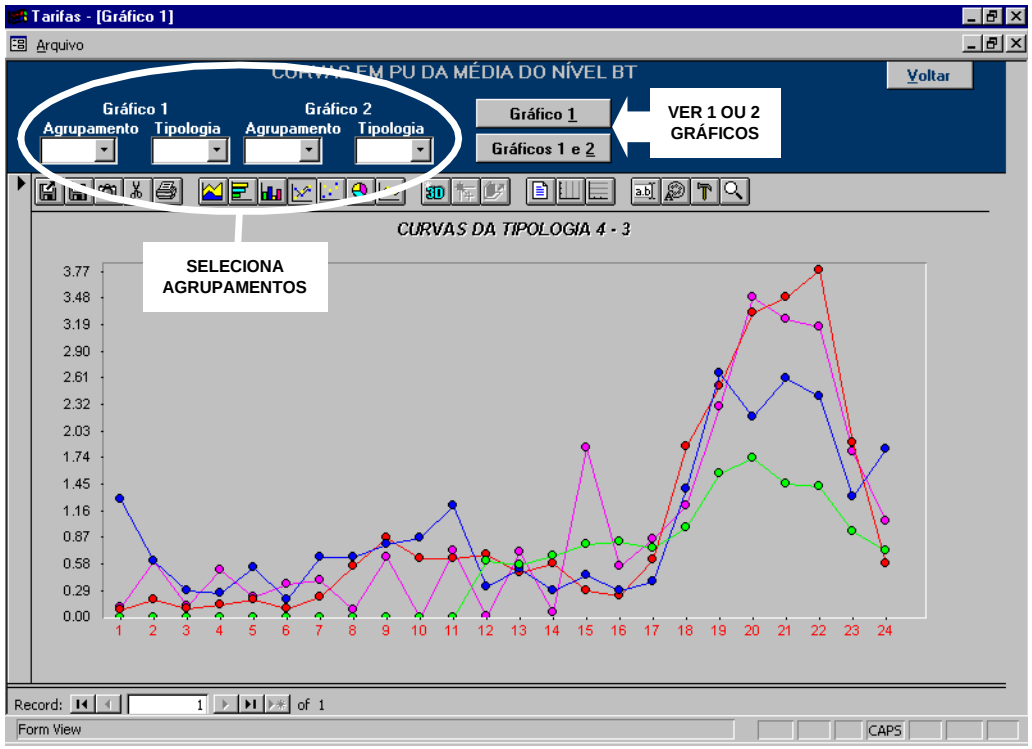
- Load profiles plots for each cluster
- Share (%) of each cluster in the energy consumption
- Within Sum Squares (WSS) and Between Sum Square (BSS).

TARDIST program (1998)

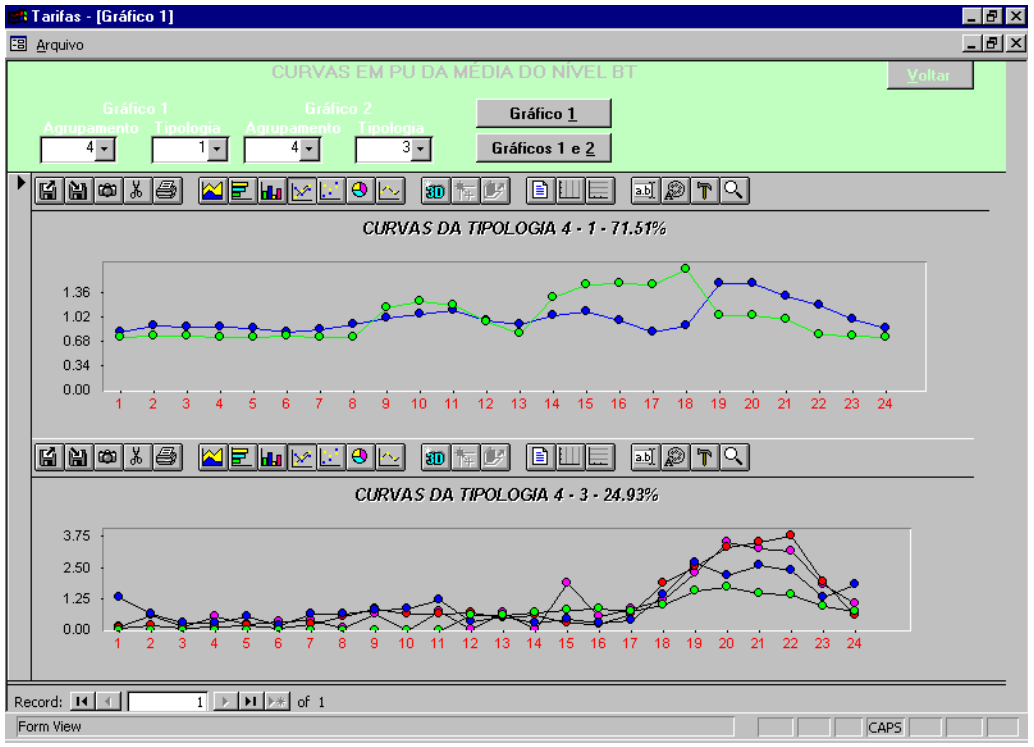
CEPEL
Grupo Eletrobrás

The software shows the load curves classified in each cluster.

One cluster



Two clusters



ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

Software developed by Cepel to Cosern (Companhia Energética do Rio Grande do Norte) in its R&D Program.

Purpose: To identify typical load profiles from a sample of load curves measurements.

Main characteristics:

- **Import load curve measurements (kW) from files in text format or Excel worksheet**
- **Routine for automatic identification of the typical workday, Saturday and Sunday curves of each measurement file**
- **Allows data analysis through graphics and reports**
- **Three clustering techniques: Ward, k-Means and Fuzzy Clustering Method (FCM)**
- **User friendly graphical interface**
- **Output reports (text format and Excel)**

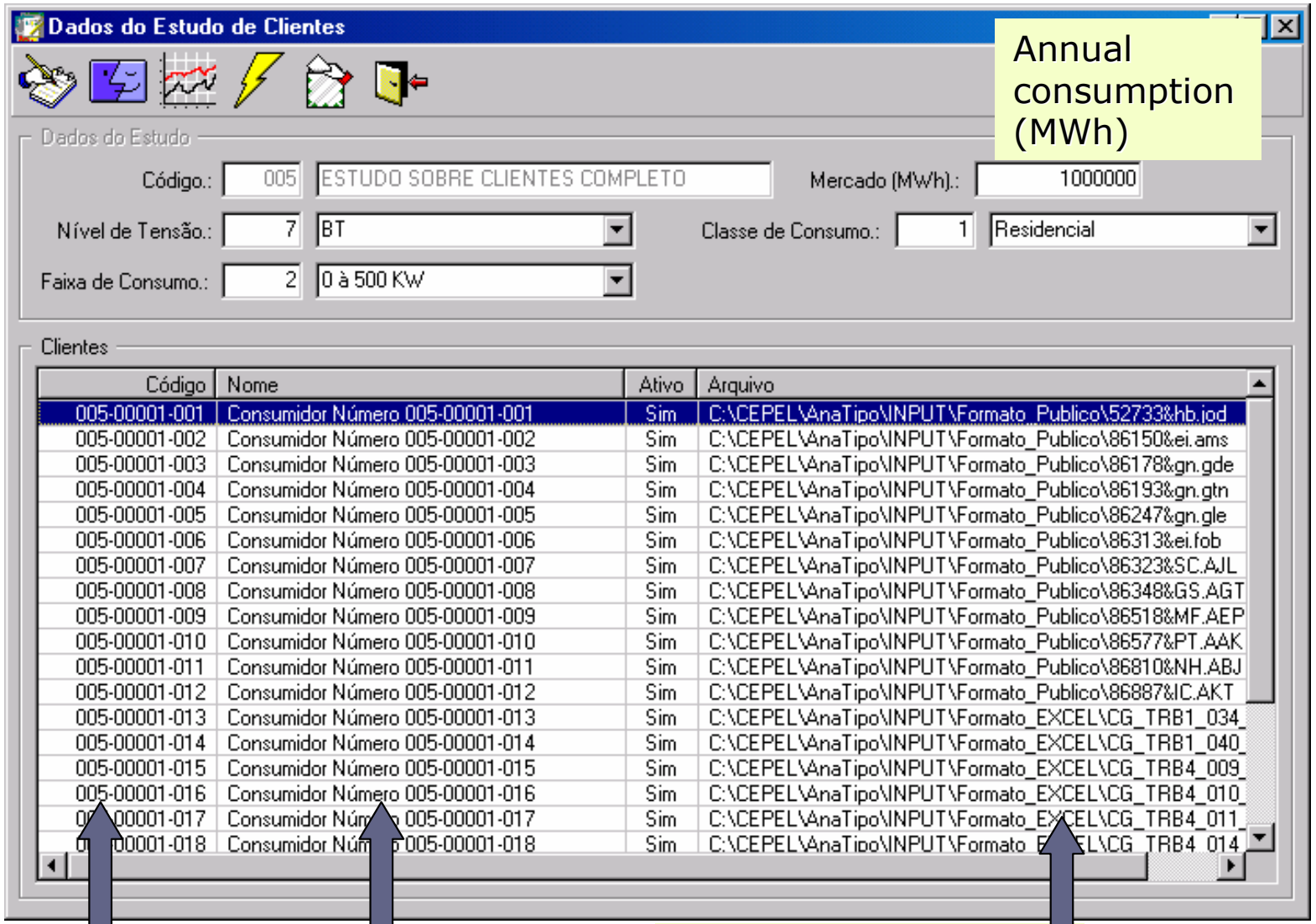
Pessanha, J.F.M., Castellani, V.L.O., Araújo, A.L.A. Uma nova ferramenta computacional para construção de tipologias de curva de carga, X SEPOPE, Florianópolis - SC,2006.

ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

Sample of customers or electric power transformers

Data input form



Customer code

Customer name

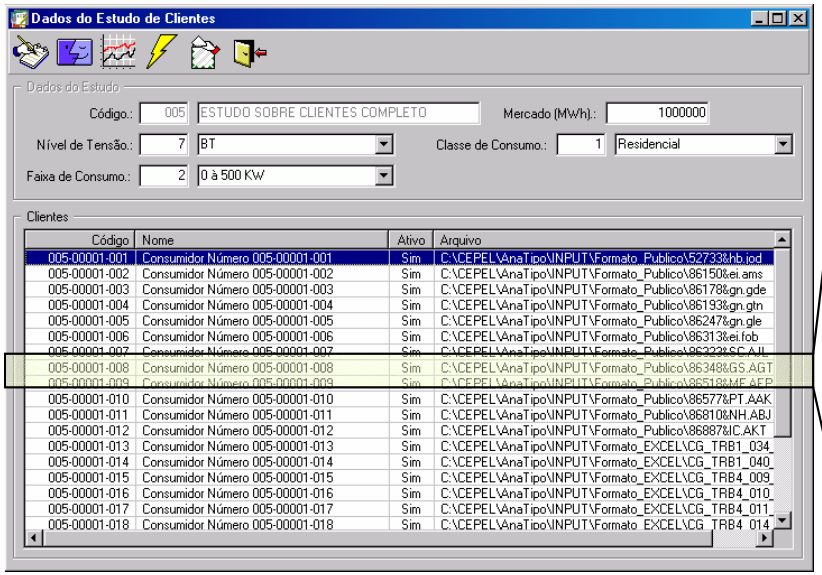
File with load curve measurement

ANATIPO Program (2005)

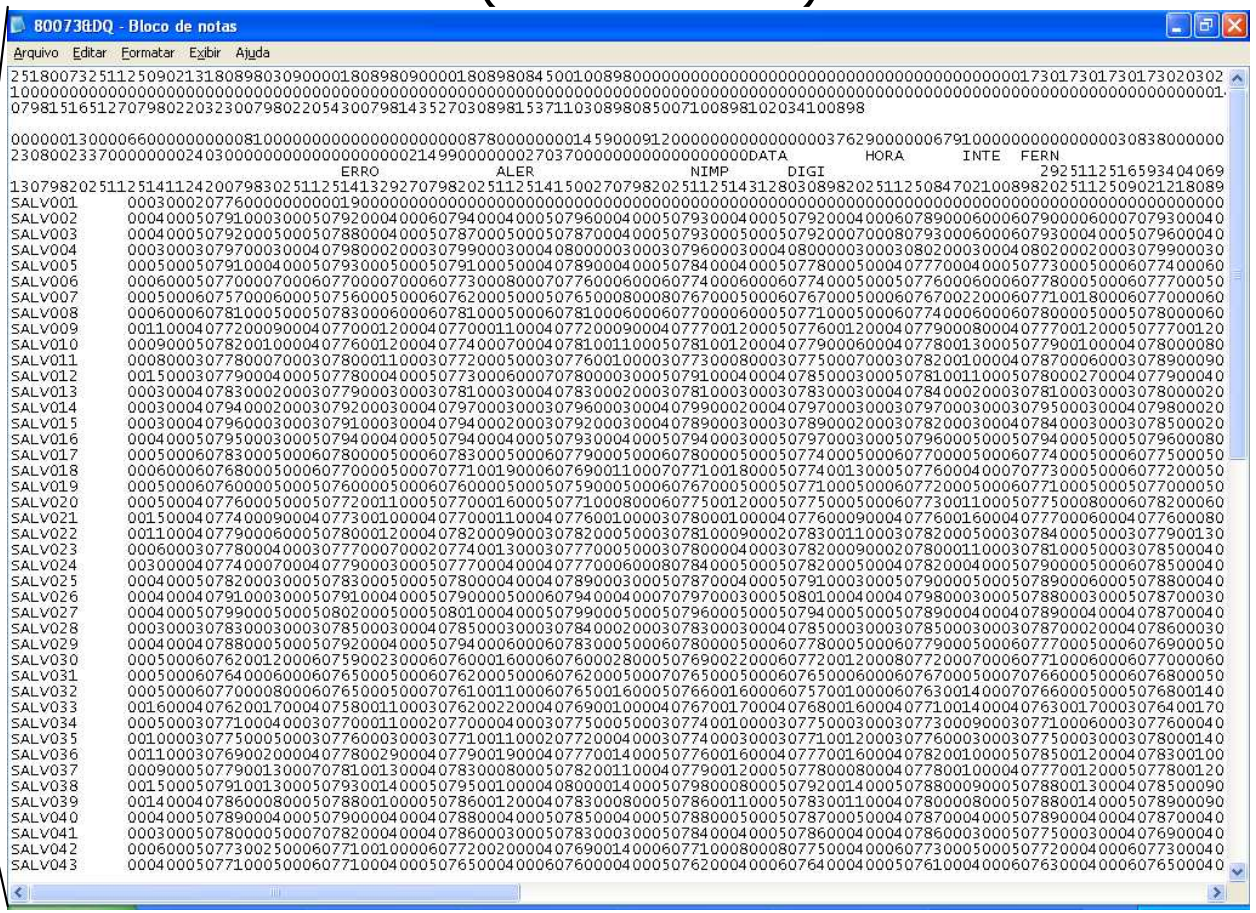
CEPEL
Grupo Eletrobrás

Reading load curve measurements

Sample points



load curve measurement file (text format)

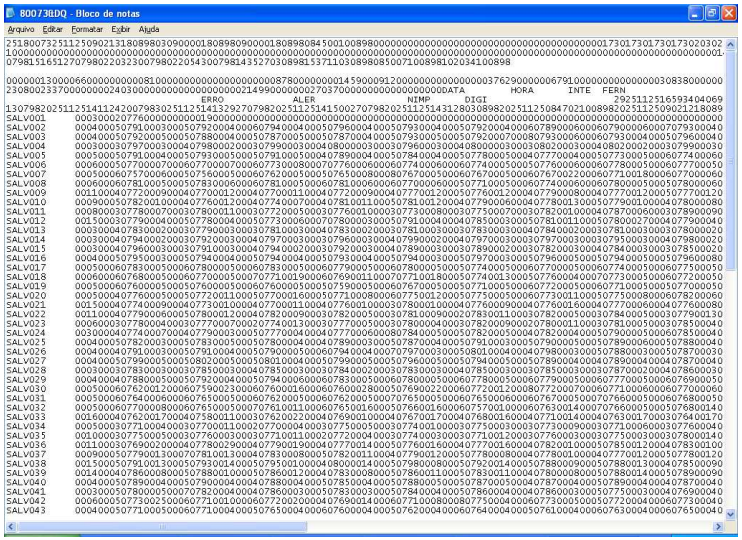


ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

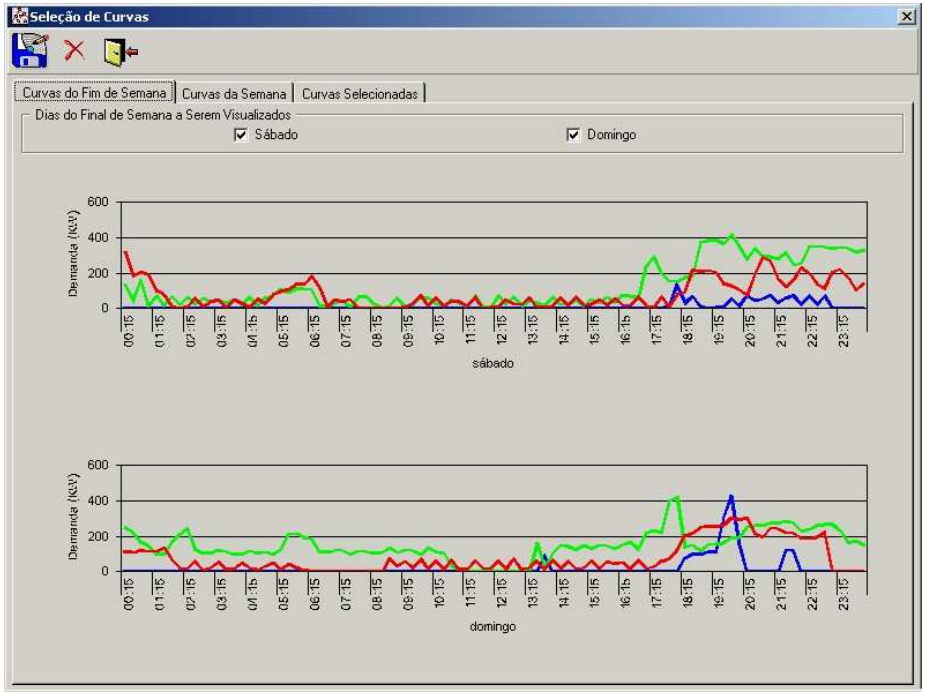
Reading load curve measurements

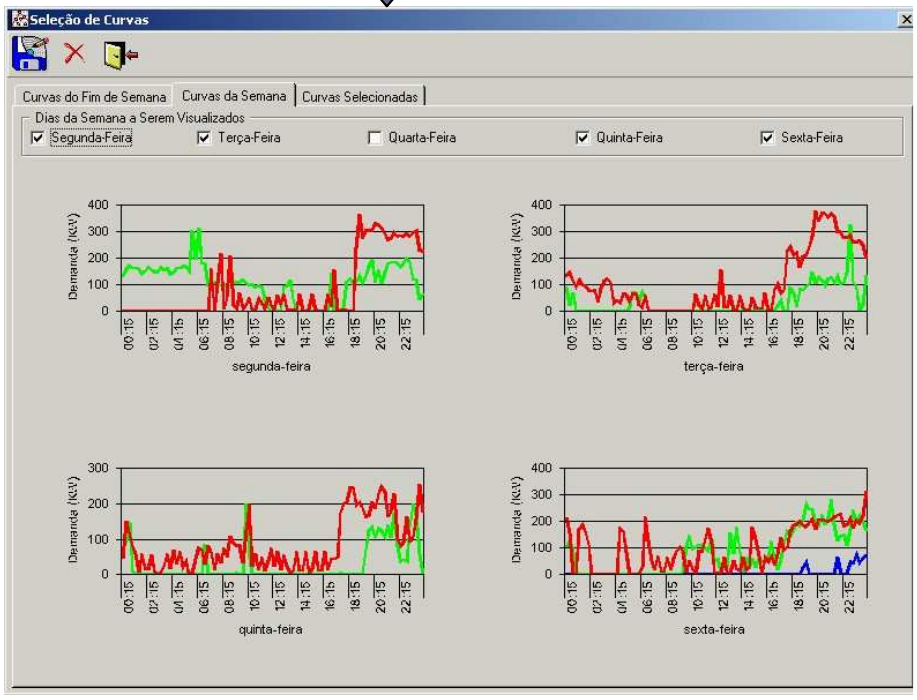
load curve measurement file



Saturday and Sunday measurements

workdays measurements

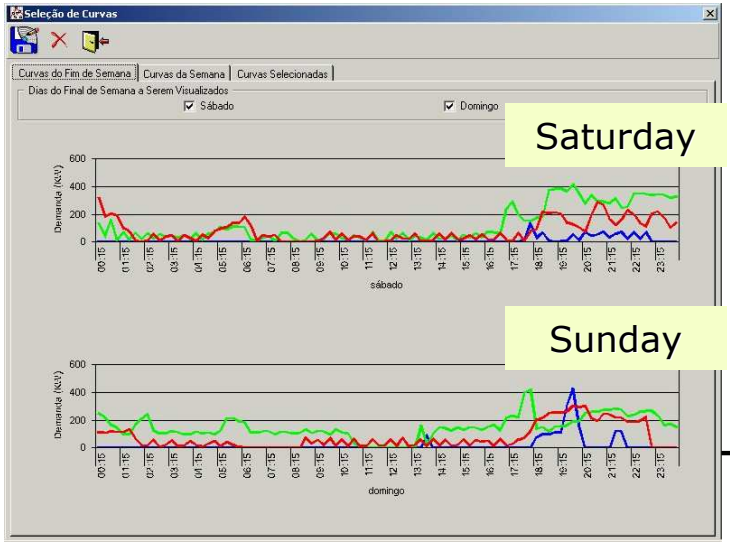




ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

Saturday and Sunday measurements



Saturday

Sunday

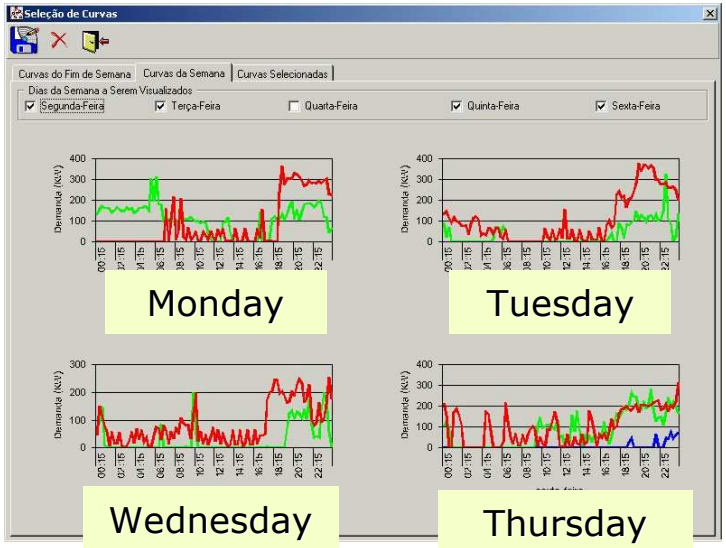
Identify the typical Saturday load curve

Identify the typical Sunday load curve

Two modes to select the three representative load curves from measurement files:

- ***non-automatic (pictorial analysis of each measurement)***
- ***automatic***

Workday measurements



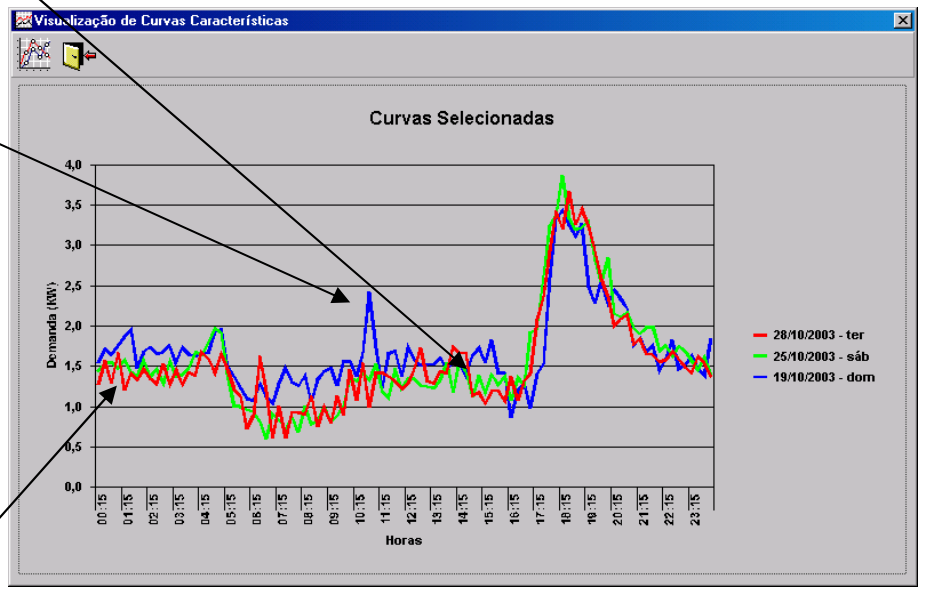
Monday

Tuesday

Wednesday

Thursday

Identify the typical workday load curve

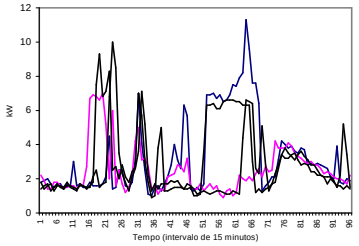


The three representative load curves : workday, Saturday and Sunday

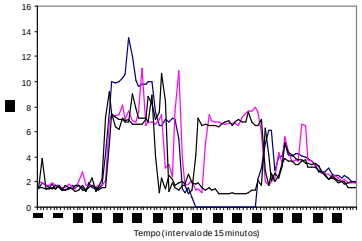
ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

Automatic selection of the representative load curves of each measurement file



Sunday



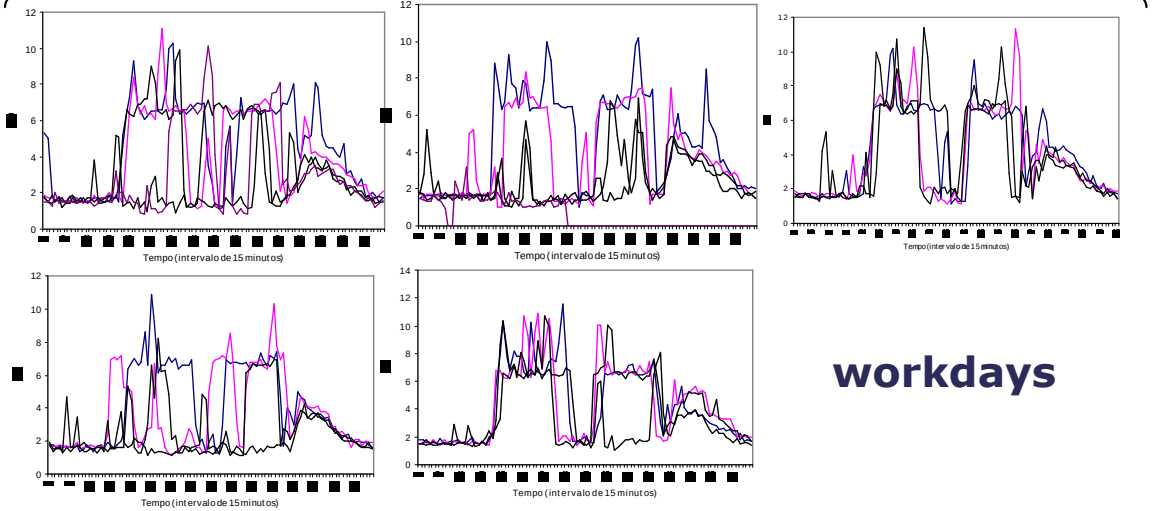
Saturday

Automatic selection

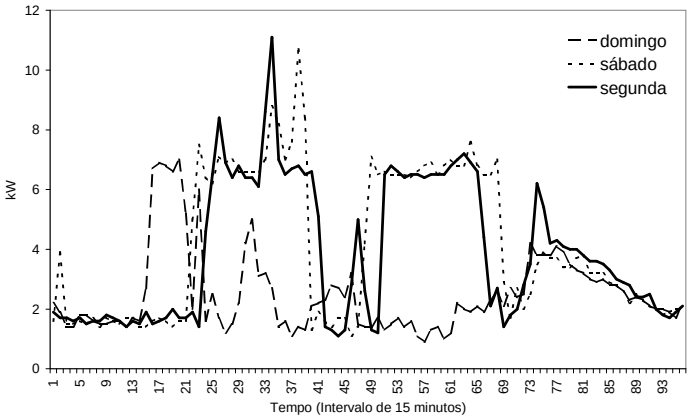
1 - Delete the bad data (outage service and abnormal power consumption)

2 - Compute three average curves (Sunday, Saturday and workday)

3 – Choice the three curves near to the average curves



workdays



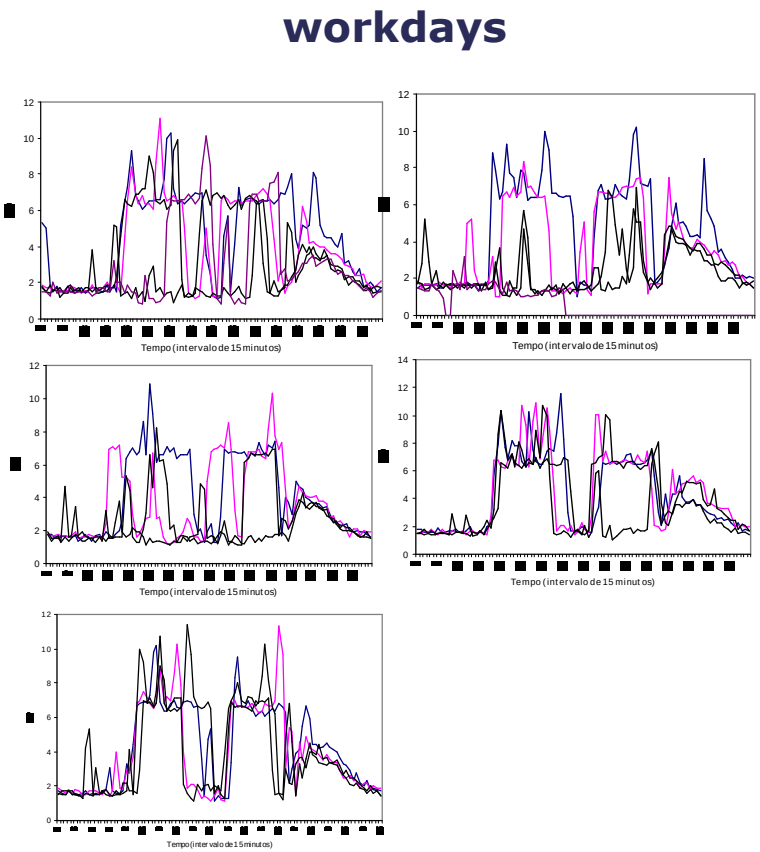
The three representative load curves for a customer: workday, Saturday and Sunday

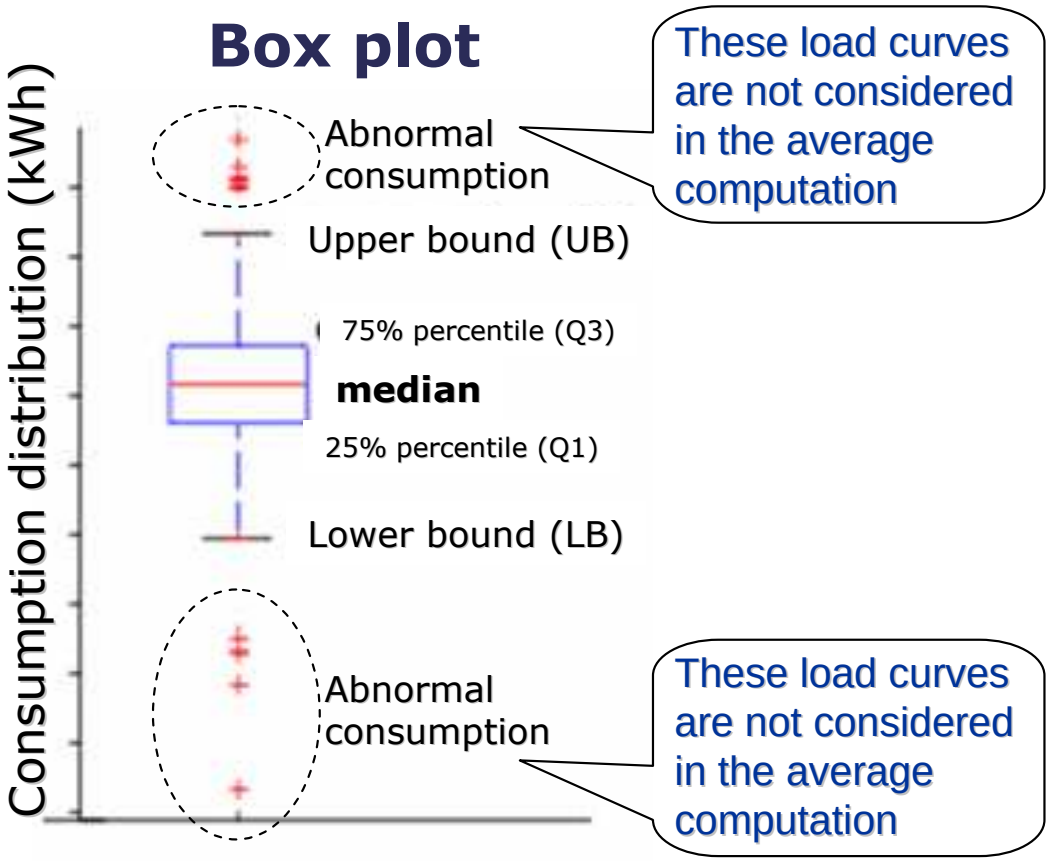
This procedure is applied automatically to each load curve measurement file

ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

Identification of the load curves with abnormal consumptions

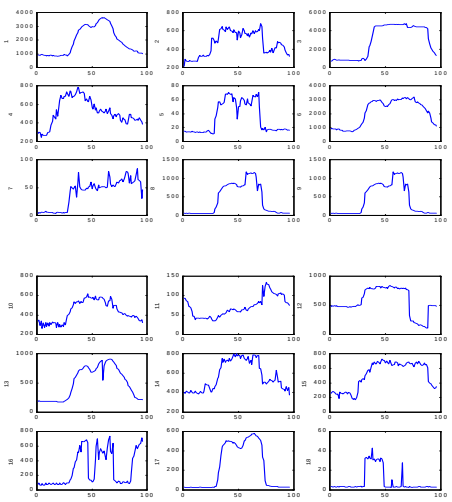




$$UB = Q3 + 1.5*(Q3-Q1)$$
$$LB = Q1 - 1.5*(Q3-Q1)$$

ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

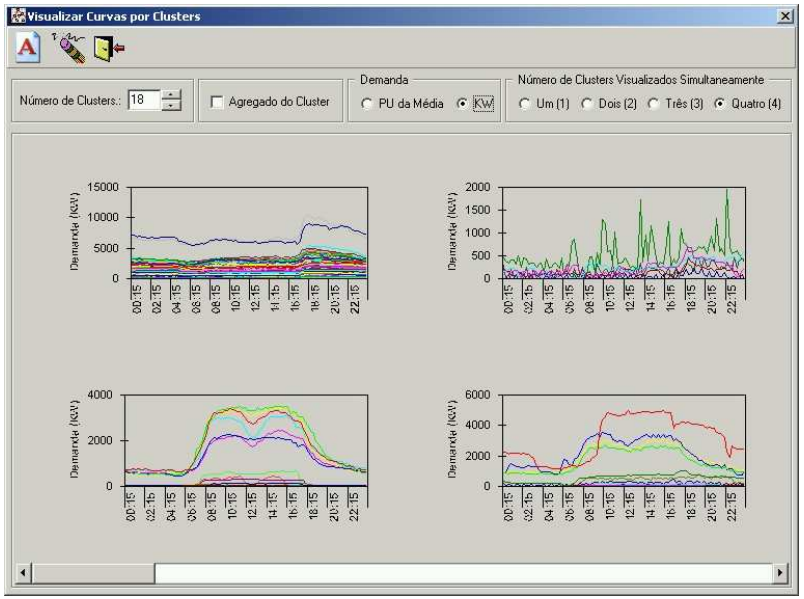


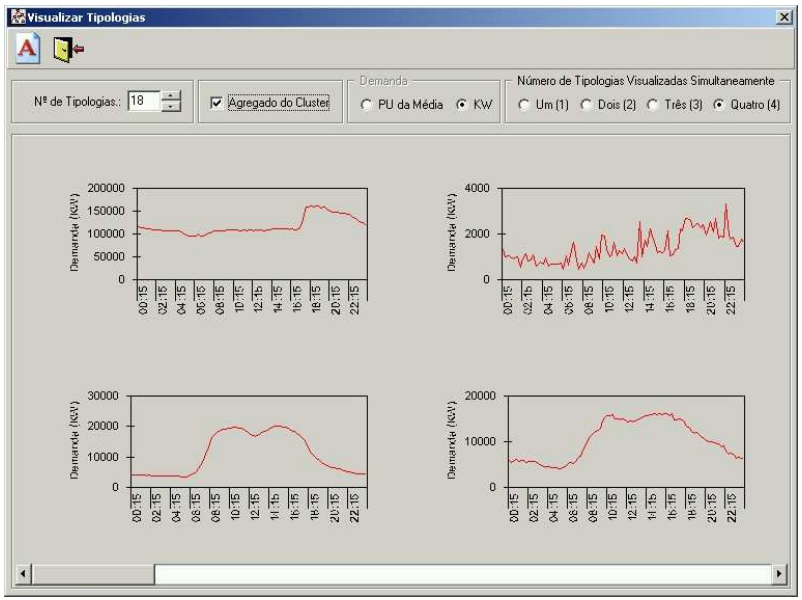
Set of workdays representative load curves

Cluster analysis

The user can choice K-Means, Ward, or FCM

centroids

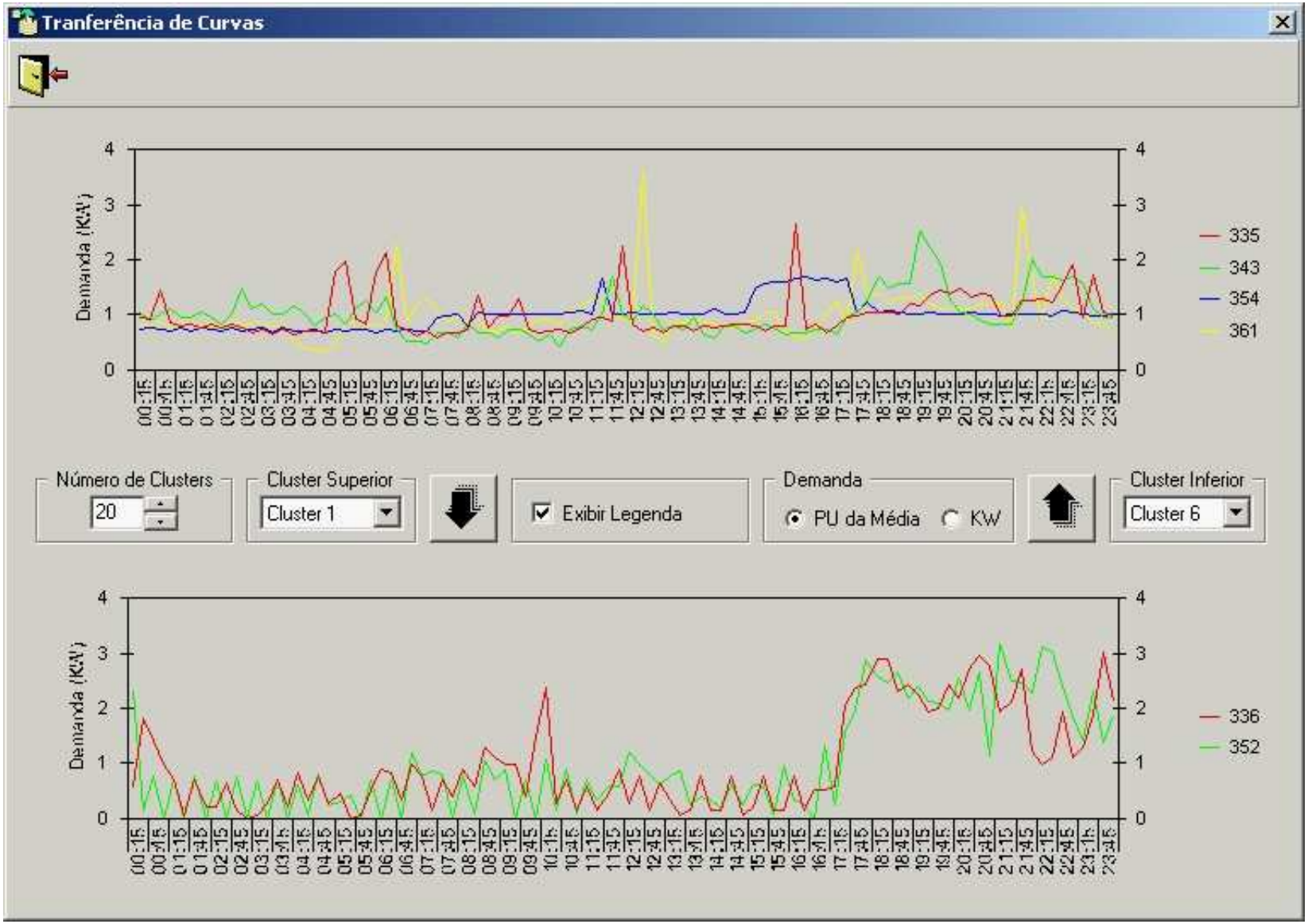




ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

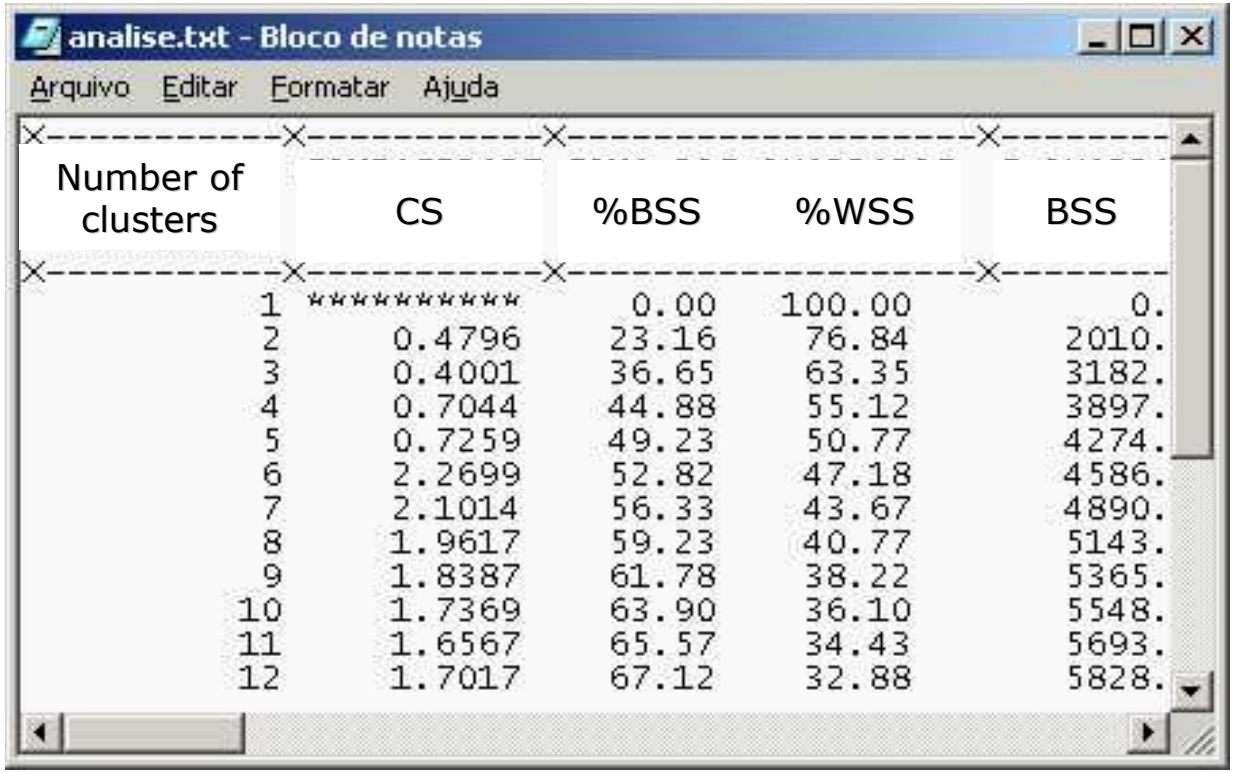
User can move a load curve from one cluster to another



ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

Output report with statistics to set the number of clusters


analise.txt - Bloco de notas

Arquivo Editar Formatar Ajuda

Number of clusters	CS	%BSS	%WSS	BSS
1	**********	0.00	100.00	0.
2	0.4796	23.16	76.84	2010.
3	0.4001	36.65	63.35	3182.
4	0.7044	44.88	55.12	3897.
5	0.7259	49.23	50.77	4274.
6	2.2699	52.82	47.18	4586.
7	2.1014	56.33	43.67	4890.
8	1.9617	59.23	40.77	5143.
9	1.8387	61.78	38.22	5365.
10	1.7369	63.90	36.10	5548.
11	1.6567	65.57	34.43	5693.
12	1.7017	67.12	32.88	5828.

BSS = Between sum squares

$$BSS = \sum_{j=1}^{k} \left\| c_j - \bar{c} \right\|^2$$

WSS = Within sum squares

$$WSS = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| x_i - c_j \right\|^2$$

%BSS = BSS / (WSS + BSS)

%WSS = WSS / (WSS + BSS)

The ideal number of clusters minimizes the compacity and separation measure (CS)

$$CS = \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \Big/ n \cdot \min \left\| c_{j1} - c_{j2} \right\|^2$$

k = number of clusters
x = load curve
c= cluster centroid
n = number of objects (load curves)
u = membership function
m = degree of fuzzyfication (m=1 for crisp methods k-Means and Ward)

ANATIPO Program (2005)

CEPEL
Grupo Eletrobrás

The main result is a worksheet with the typical load profiles

Microsoft Excel - PLANILHA_20040812141220.XLS

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda Tradução

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CEPEL - Centro de Pesquisas de Energia Eletrica												
2	*Sistema ANATIPO*												
3	*Tipologias em MW*												
4													
5	**INTERVALO**	**HORA**			**TIPO 1**			**TIPO 2**			**AGREGADO**		
6				**útil**	**sábado**	**domingo**	**útil**	**sábado**	**domingo**	**útil**	**sábado**	**domingo**	
7	00:00-01:00	1		15.253	18.113	15.600	0.953	0.867	0.780	16.207	18.980	16.38	
8	01:00-02:00	2		14.907	18.113	16.120	0.867	0.953	0.693	15.773	19.067	16.813	
9	02:00-03:00	3		15.167	17.940	15.947	0.780	0.780	0.780	15.947	18.720	16.727	
10	03:00-04:00	4		15.167	17.680	16.033	0.867	0.867	0.693	16.033	18.547	16.727	
11	04:00-05:00	5		15.253	17.767	15.600	0.693	0.953	0.780	15.947	18.720	16.38	
12	05:00-06:00	6		18.287	19.674	16.033	0.867	0.867	0.693	19.154	20.540	16.727	
13	06:00-07:00	7		32.760	35.707	16.380	1.040	1.127	0.953	33.800	36.834	17.333	
14	07:00-08:00	8		44.460	63.787	16.293	1.300	1.733	0.953	45.760	65.521	17.247	
15	08:00-09:00	9		142.568	123.588	16.207	1.820	1.387	0.693	144.388	124.974	16.9	
16	09:00-10:00	10		306.976	227.935	15.773	6.067	1.387	1.127	313.043	229.322	16.9	
17	10:00-11:00	11		317.896	244.055	16.293	6.413	1.213	0.693	324.309	245.269	16.987	
18	11:00-12:00	12		287.042	234.522	18.200	1.560	1.127	0.780	288.603	235.649	18.98	
19	12:00-13:00	13		296.749	225.855	17.680	4.680	1.040	0.607	301.429	226.895	18.287	
20	13:00-14:00	14		219.962	184.428	17.073	0.780	2.253	0.347	220.742	186.682	17.42	
21	14:00-15:00	15		245.442	204.275	16.293	6.153	1.560	0.867	251.596	205.835	17.16	
22	15:00-16:00	16		250.989	202.455	17.247	5.893	1.040	0.520	256.882	203.495	17.767	
23	16:00-17:00	17		232.875	209.475	16.467	5.807	0.693	0.520	238.682	210.169	16.987	
24	17:00-18:00	18		228.109	215.282	16.900	0.780	1.560	1.300	228.889	216.842	18.2	
25	18:00-19:00	19		178.622	131.474	16.380	1.127	0.867	0.780	179.748	132.341	17.16	
26	19:00-20:00	20		117.868	94.901	17.247	1.907	0.953	0.867	119.774	95.854	18.113	
27	20:00-21:00	21		111.974	47.927	17.593	1.387	1.647	0.953	113.361	49.574	18.547	
28	21:00-22:00	22		79.041	32.414	17.333	1.127	1.387	0.347	80.167	33.800	17.68	
29	22:00-23:00	23		44.114	25.827	16.900	0.780	1.213	0.433	44.894	27.040	17.333	
30	23:00-24:00	24		23.054	17.680	16.033	1.473	1.040	0.173	24.527	18.720	16.207	
31	**% mercado**			98.347			1.653			100			
32	**mercado MWh**			983472.4			16527.56			1000000			
33	**demanda média MW**			135.606			2.297			137.902			
34	**poderação sabado**			0.808			0.517			0.804			
35	**poderação domingo**			0.122			0.314			0.125			
36	**fator de carga %**			42.657			35.811			42.522			
37	**demanda máx MW Ponta**			178.622			1.907			179.748			
38	**demanda máx MW Fora_Ponta**			317.896			6.413			324.309			
39													

PLANILHA

Excel worksheet ready to be used in the distribution tariff computation

Conclusions

CEPEL
Grupo Eletrobrás

The softwares for identifying typical load profiles developed in the Brazilian electric power sector are based on the French tradition, specially from the EDF's experience.

Most of the studies carried out in the Brazilian electricity distribution utilities have been used statistics techniques (Ward method, k-Means or the "*Nuées dynamiques*") to obtain the typical load profiles. For example, the statistics techniques have been used in the tariff revision process.

However it is possible to find few studies that use Self-Organizing Map (SOM) in order to get the typical load profiles. Most of these studies have been used the Matlab Neural Network Toolbox. The challenge remains to develop a software based on SOM like "*Courboscope*" (Debregeas & Hebrail, 1998) developed by the EDF's R&D Division.

Debrégeas, A., Hébrail G. (1998). Interactive Interpretation of Kohonen Maps Applied To Curves, In KDD'98, Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New-York, pp.179-183, AAAI Press.

References

1) BRASIL, Ministério das Minas e Energia, DNAEE, Eletrobrás, Empresas Concessionárias de Energia Elétrica, Nova Tarifa de Energia Elétrica: metodologia e aplicação, DNAEE, Brasília,1985.

2) Boiteux, M. La tarification dês demandes en pointe: application de la théorie de la vente au coût marginal, Revue générale d l'electricité, 1949.

3) Bouroche, J.M, Saporta, G., L'analyse dês données, PUF, 9e édition, Paris, 2005.

4) Debrégeas, A., Hébrail G. (1998). Interactive Interpretation of Kohonen Maps Applied To Curves, In KDD'98, Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New-York, pp.179-183, AAAI Press.

5) Diday, E. Une nouvelle méthode em classification automatique et reconnaissance des formes. La méthode des nuées dynamiques. Revue de statistique Appliquée, 1971, vol. XIV nº 2. Institut de Statistique. Université de Paris.

6) Hébrail, G. Practical data mining in a large utility company, Revue Questiio (Quaderns d'Estadistica i Investigacio Operativa), Vol.25, N.3, pp.509-520, 2001.

7) Lebart, L.; Piron, M.; Morineau, A. Statistique exploratoire multidimensionnelle, 3e édition, DUNOD, Paris, 2000.

8) Jain, J.S.R., Sun C.T., Mizutani, E. Neuro-Fuzzy and Soft Computing: a computational approach to learning and machine intelligence, Prentice Hall Inc, 1997.

9) Molliere, M. Um ensemble de modules de classification automatique et de modules explicatifs associes, Note EDF, Direction des etudes et Recherches nº HI 2818/02, 1978.

10) Pessanha, J.F.M., Huang, J.L.C., Pereira, L.A.C., Passos Júnior, R., Castellani, V.L.O. Metodologia e sistema computacional para cálculo das tarifas de uso dos sistemas de distribui»cão, XXXVI SBPO, São João del Rey - MG,2004.

11) Pessanha, J.F.M., Castellani, V.L.O., Araújo, A.L.A. Uma nova ferramenta computacional para construção de tipologias de curva de carga, X SEPOPE, Florianópolis - SC,2006.

12) Pessanha, J.F.M., Laurencel, L.C., Souza, R.C. Kohonen Map to build load curve types, XXXVI SBPO, São João Del Rey, Brasil, 2004.

José Francisco Moreira Pessanha (francisc@cepel.br)

Luiz da Costa Laurencel (getlcl@vm.uff.br)