

ANATIPO – Sistema Computacional para Construção de Tipologias de Curvas de Carga

A.L.A. Araújo, COSERN, J.F.M. Pessanha, V.L.O. Castellani e P.M. Ribeiro, CEPEL

Resumo – Os perfis típicos da demanda horária de eletricidade são informações fundamentais no cálculo das tarifas que remuneram o serviço de distribuição de energia elétrica. Estes perfis são obtidos por meio de um estudo de caracterização da carga, onde são utilizados métodos de análise de agrupamentos para identificar os perfis típicos ou tipologias, a partir de uma amostra de medições de curvas de carga de clientes e pontos da rede de distribuição. Visando agilizar a caracterização da carga, de forma a beneficiar particularmente os estudos tarifários, o CEPEL, em conjunto com a COSERN, desenvolveu o ANATIPO, um sistema computacional dotado de diversas funcionalidades para auxiliar a construção das tipologias. As tipologias obtidas pelo programa são automaticamente ajustadas ao mercado e gravadas em planilhas eletrônicas, prontas para serem incluídas nas propostas tarifárias a serem enviadas a ANEEL. Enfim, este artigo tem como objetivo apresentar o sistema ANATIPO e as suas principais funcionalidades.

Palavras-chave — curvas de carga, sistema computacional, tipologias.

I. INTRODUÇÃO

Os perfis típicos da demanda horária de eletricidade, também conhecidos como tipologias, são informações fundamentais no cálculo das tarifas [1-3] que remuneram o serviço de distribuição de energia elétrica e úteis na indicação de alternativas para o gerenciamento pelo lado da demanda [4,5].

As tipologias de clientes são definidas para cada classe de consumo e representam a diversidade do comportamento dos consumidores da classe. Da mesma forma são definidas as tipologias de redes, que representam os perfis típicos das solicitações de potência em pontos selecionados da rede de distribuição. Estas informações, em conjunto com os custos marginais por nível de tensão, permitem elaborar a estrutura de custos marginais de fornecimento, aberta nos postos ponta e fora de ponta, que serve de base para o cálculo das tarifas de uso dos sistemas de distribuição [3].

Basicamente, a construção das tipologias de um determinado segmento (nível de tensão, classe de consumo ou tipo de rede) inicia com a coleta de uma amostra de medições de curvas de carga de clientes ou redes [1,2,6]. Em seguida, aplica-se alguma técnica de análise de agrupamentos (*cluster analysis*) [7-10] na amostra de medições para particioná-la em agrupamentos (*clusters*) mutuamente exclusivos, de tal forma que as medições em um mesmo *cluster* sejam semelhantes entre si, porém diferentes das curvas pertencentes aos outros *clusters*. Os perfis horários típicos existentes na amostra são os perfis médios dos *clusters*.

No setor elétrico brasileiro, o SNACC¹ foi o primeiro sistema computacional para construção de tipologias de curvas de carga. Neste sistema, desenvolvido em meados da década de 1980, os *clusters* são obtidos pelo uso combinado de dois métodos estatísticos para análise de agrupamentos: nuvens dinâmicas – NUDYC [1,8] e método de descrição e classificação ascendente hierárquico – DESCR2 [1].

Os dois métodos são executados em sequência: primeiro o método das nuvens dinâmicas identifica os *clusters* iniciais (formas fortes) e em seguida estes *clusters* são agrupados pelo DESCR2, um método hierárquico semelhante ao de Ward [9]. Apesar da maneira sofisticada de formar os *clusters*, o SNACC não é um sistema amigável e possui limitações que dificultam o processo de construção das tipologias, entre as quais destaca-se o fato de não disponibilizar saídas gráficas para visualização dos agrupamentos.

Nos sistemas desenvolvidos mais recentemente, a disponibilidade de várias saídas gráficas representou um grande avanço. Entretanto, mesmo nestes sistemas, a construção das tipologias ainda é pouco confortável e pode ser melhorada, por meio da incorporação de novas rotinas que facilitem e tornem mais produtivo o processo de construção das tipologias.

Visando agilizar o estudo de caracterização da carga, de forma a beneficiar particularmente os estudos tarifários, o Centro de Pesquisas de Energia Elétrica (CEPEL), em conjunto com a Companhia Energética do Rio Grande do Norte (COSERN), desenvolveu o ANATIPO, um sistema computacional dotado de diversas funcionalidades. Este artigo tem como objetivo apresentar o sistema computacional ANATIPO e as suas principais funcionalidades.

Dentre estas funcionalidades destacam-se as rotinas para leitura de medições provenientes de medidores de diferentes fabricantes (arquivos em formato público ou em planilhas eletrônicas), rotinas gráficas que facilitam a identificação visual das curvas características dos dias úteis, sábado e domingo de cada medição, além de uma variedade de técnicas de agrupamento disponibilizadas pelo programa: *K-Means* [7,8], *Ward* [9] e *FCM* [10].

Um item inovador neste sistema computacional é a rotina para identificação automática das curvas características, capaz de acelerar substancialmente a análise das medições.

As tipologias, principais resultados obtidos pelo programa, são automaticamente ajustadas ao mercado e gravadas

¹ Sistema Nacional de Análise de Curvas de Carga - SNACC, desenvolvido pelo Departamento Nacional de Águas e Energia Elétrica (DNAEE).

em planilhas eletrônicas, prontas para serem incluídas nas propostas tarifárias a serem enviadas à Agência Nacional de Energia Elétrica – ANEEL.

A seguir, a seção 2 apresenta o processo de construção das tipologias. A seção 3 apresenta a heurística adotada para fazer a seleção automática das curvas características. Os algoritmos dos métodos de análise de agrupamentos implementados no ANATIPO são apresentados na seção 4. A seção 5 apresenta algumas saídas gráficas do sistema ANATIPO e, finalmente, as principais conclusões do trabalho são resumidas na seção 6.

II. CONSTRUÇÃO DAS TIPOLOGIAS

A construção das tipologias inicia com a obtenção das medições de curvas de carga de clientes e redes. Na baixa tensão o número de consumidores é bastante elevado, os medidores instalados são do tipo eletromecânico e não possuem memória de massa para armazenar os registros de demanda horária. Neste caso, deve-se realizar uma amostragem, ou melhor, uma campanha de medidas, para coletar as medições de curva de carga de clientes e transformadores de distribuição.

Por outro lado, na alta tensão o número de pontos a serem medidos é menor e os medidores instalados nos clientes são eletrônicos, com capacidade para armazenar os registros de demanda horária. Neste caso, a coleta das medições é direta, bastando apenas recuperar as medições armazenadas na memória de massa dos medidores. Nas duas situações, recomenda-se que cada medição tenha uma abrangência de no mínimo cinco dias úteis, um sábado e um domingo completos.

Após a coleta das medições deve-se fazer uma inspeção visual de cada medição, com o objetivo de identificar três curvas diárias características do ponto de medição: uma curva típica para o dia útil, uma para o sábado e uma para o domingo, conforme apresentado na Figura 1.

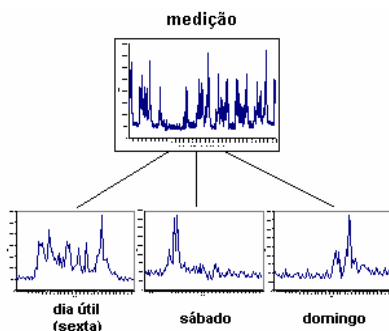


Figura 1. Seleção das curvas características.

A identificação das tipologias de um determinado segmento (nível de tensão, classe de consumo ou tipo de rede) envolve a aplicação de técnicas de análise de agrupamentos ou *cluster analysis*, com a finalidade de dividir o conjunto de curvas características do dia útil em agrupamentos internamente homogêneos (*clusters*), a partir dos quais são extraídas as tipologias. Este processo é resumido na Figura 2.

Curvas características do dia útil

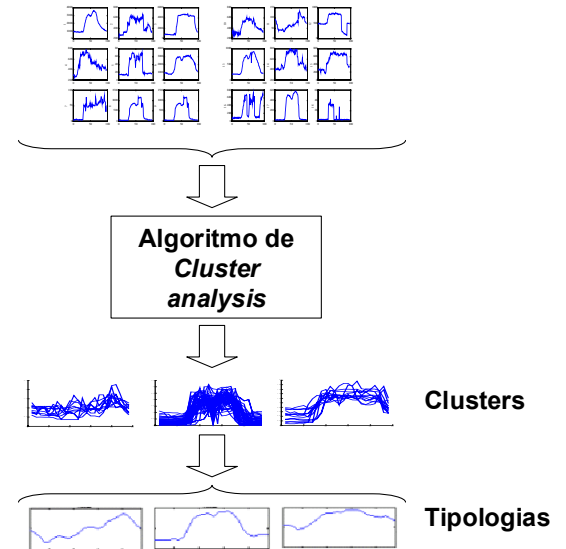


Figura 2. K tipologias obtidas a partir de N medições ($K < N$)

As curvas características do dia útil formam o conjunto de curvas a serem classificadas, que irão definir os perfis das tipologias. As curvas de sábado e domingo servem apenas para calcular fatores de ponderação usados no ajuste [11] dos perfis típicos ao montante de energia anual (MWh) do segmento que eles representam. Este ajuste é conduzido de acordo com o procedimento a seguir:

1) Em cada *cluster* i , $i=1, k$, some separadamente as curvas do dia útil (U), do sábado (S) e do domingo (D), obtendo três curvas agregadas (perfis típicos) para cada *cluster* i :

$$U_h^i, S_h^i, D_h^i, \forall h = 1, 24 \text{ horas.}$$

2) Totalize as curvas horárias do dia útil, sábado e domingo em cada *cluster* i :

$$E_u^i = \sum_{h=1}^{24} U_h^i \quad E_s^i = \sum_{h=1}^{24} S_h^i \quad E_d^i = \sum_{h=1}^{24} D_h^i \quad (1)$$

3) Calcule os fatores de ponderação do sábado (P_s^i) e do domingo (P_d^i) de cada *cluster* i :

$$P_s^i = \frac{E_s^i}{E_u^i} \quad P_d^i = \frac{E_d^i}{E_u^i} \quad (2)$$

4) Calcule a energia anual da curva agregada do *cluster* i :

$$C_i^{anual} = E_u^i \cdot 24 (N_u + N_s \cdot P_s^i + N_d \cdot P_d^i) \quad (3)$$

onde N_u , N_s e N_d são, respectivamente, o número de dias úteis, sábados e domingos mais feriados do ano.

5) Calcule a participação do *cluster* i no mercado total do segmento que ele representa:

$$\alpha_i = \frac{C_i^{anual}}{\sum_{i=1}^k C_i^{anual}} \quad (4)$$

6) Obtenha a parcela do mercado associado ao *cluster* i :

$$ET_i = \alpha_i \cdot M \quad (5)$$

onde M é o mercado anual em MWh do segmento que ele representa.

7) Calcule a demanda média (MW) da curva agregada do *cluster* i :

$$D_i^{média} = \frac{ET_i}{24 \cdot (N_u + N_s P_s^i + N_d P_d^i)} \quad (6)$$

8) Finalmente, ajuste a tipologia i ao mercado, multiplicando a demanda média pela demanda horária em p.u. da média:

$$T_h^i = D_i^{média} \cdot \frac{U_h^i}{\left(\frac{\sum_{h=1}^{24} U_h^i}{24} \right)} \quad (7)$$

III. IDENTIFICAÇÃO AUTOMÁTICA DAS CURVAS CARACTERÍSTICAS

A identificação das curvas características pode ser realizada por meio de uma inspeção visual de cada arquivo de medição. Entretanto, este procedimento consome muito tempo do analista conhecedor do mercado. Com o intuito de acelerar esta identificação, introduziu-se no ANATIPO uma rotina que seleciona automaticamente as curvas características do dia útil, do sábado e do domingo.

Esta rotina não elimina o trabalho do analista, mas visa auxiliá-lo. A idéia é fornecer três curvas ao analista durante no momento da inspeção visual, deixando para o analista a decisão final de aceitar a sugestão do programa ou selecionar manualmente outras curvas.

O critério implementado na rotina de seleção automática é bastante simples: dado um arquivo de medição, calculam-se as três curvas diárias médias, uma para sábado, uma para domingo e outra para o dia útil. Em seguida identificam-se para cada um destes dias quais as medições diárias registradas mais próximas das respectivas curvas médias.

Como a média não é uma medida robusta [12], as medições que possuam valores nulos e as realizadas em dias de instalação e retirada do medidor não participarão do cálculo das curvas médias. Para o caso das curvas do dia útil adota-se um critério adicional, com o objetivo de identificar curvas com energias atípicas. Neste critério calculam-se inicialmente os quartis da distribuição de energias das medições do dia útil e, a partir deles, definem-se os seguintes limites ou cercas:

$$CS = Q3 + \alpha(Q3 - Q1) \quad (8)$$

$$CI = Q1 - \alpha(Q3 - Q1) \quad (9)$$

onde $Q1$ e $Q3$ representam, respectivamente, o primeiro e terceiro quartis da distribuição de energias das medições do dia útil e α representa uma constante arbitrária, usualmente igual a 1,5 [12].

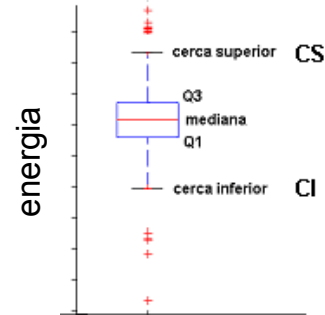


Figura 3. Quartis, mediana e cercas

Todas as curvas que possuam energias abaixo da cerca inferior (CI) ou acima da cerca superior (CS) não participam do cálculo da média. Na figura 3 estas curvas estão indicadas pelo símbolo +.

A seguir, a figura 4 ilustra o processo de seleção automática das curvas características a partir de um arquivo de medição com 40 dias registrados.

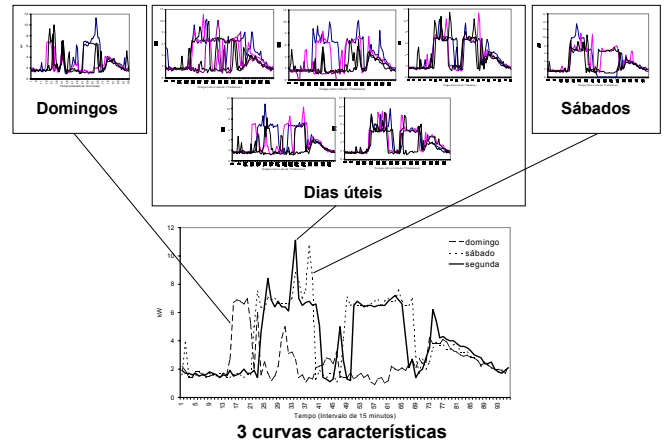


Figura 4. Quartis, mediana e cercas

IV MÉTODOS DE ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos denota uma ampla variedade de métodos de classificação não supervisionados [13], utilizados para encontrar uma estrutura natural de agrupamentos em um conjunto de objetos multidimensionais, baseados em uma medida de similaridade ou distância entre os objetos. Os vários métodos de análise de agrupamentos visam resolver o problema de como dividir um conjunto com n objetos em k *clusters* mutuamente exclusivos, de tal forma que os objetos em um mesmo *cluster* sejam semelhantes entre si, mas diferentes dos objetos pertencentes aos outros *clusters*.

Admitindo que os n objetos sejam caracterizados por p atributos de natureza quantitativa, cada objeto pode ser representado por um vetor x_i , $i=1,2,\dots,n \in \mathfrak{R}^p$. Neste caso, a distância euclidiana é usualmente a métrica adotada para avaliar o grau de semelhança entre os objetos.

O sistema ANATIPO disponibiliza ao analista três métodos de análise de agrupamentos: *K-Means*, *Ward* e *FCM*. Qualquer um dos métodos pode ser escolhido e uma comparação dos desempenhos destes métodos pode ser encontrada na referência 13, quando aplicados à classificação de curvas de carga.

A. *K-Means*

Um dos mais conhecidos métodos de *cluster analysis* é o *K-Means*. Este método divide um conjunto de n vetores em k *clusters*, $C_j, j=1,k$, de tal forma que os centros de cada *cluster*, $c_j, j=1,k$, minimizem a soma dos quadrados intra-*cluster* (SQIntra):

$$SQIntra = \sum_{j=1}^k \left(\sum_{i, x_i \in C_j} \|x_i - c_j\|^2 \right) \quad (10)$$

Na equação 1, $\|x_i - c_j\|^2$ é o quadrado da distância euclidiana entre o objeto x_i e o centro de gravidade c_j , do *cluster* C_j , onde o objeto foi alocado. Cada objeto pode pertencer a apenas um dos k *clusters*, portanto, a classificação dos n objetos pode ser representada por uma matriz binária U de dimensões $n \times k$, onde $u_{ij}=1$ se o i -ésimo objeto pertence ao j -ésimo *cluster* e $u_{ij}=0$ caso contrário. Se os centros dos k *clusters* são fixos, a partição ótima consiste em alocar cada objeto no *cluster* com o centro de gravidade mais próximo, ou seja, os valores de u_{ij} são definidos pela seguinte regra:

$$u_{ij} \begin{cases} 1 & \text{se } \|x_i - c_j\| \leq \|x_i - c_k\| \forall k \neq j \\ 0 & \text{caso contrário} \end{cases} \quad (11)$$

A matriz U tem as seguintes propriedades:

$$\sum_{j=1}^k u_{ij} = 1 \quad (12)$$

$$\sum_{i=1}^n u_{ij} = n_j \quad (13)$$

$$\sum_{j=1}^k \sum_{i=1}^n u_{ij} = n_1 + \dots + n_k = n \quad (14)$$

Por outro lado, para uma dada partição dos n objetos a matriz U é fixa e o centro do j -ésimo *cluster*, $j=1,k$, que minimiza a função 10 é a média dos objetos do *cluster*:

$$c_j = \frac{1}{n_j} \sum_{i, x_i \in C_j} x_i \quad (15)$$

Computacionalmente o método *K-Means* pode ser implementado segundo o algoritmo a seguir, onde a matriz U e os centros dos *clusters* são obtidos de forma iterativa:

Passo 1 - Inicialize os centros dos *clusters*. Usualmente a inicialização se dá pelo sorteio de k objetos entre os n a serem classificados.

Passo 2 - Determine a matriz U de acordo com o critério do centro mais próximo, conforme em 11.

Passo 3 - Calcule o valor da função SQIntra. Pare se o valor da função estiver abaixo de uma tolerância pré-especificada ou se a melhoria em relação à iteração anterior for desprezível.

Passo 4 - Atualize os centros dos *clusters* de acordo com a equação 15 e volte para o passo 2. (10)

Este algoritmo é computacionalmente eficiente e produz bons resultados se os *clusters* são compactos, esféricos e bem separados no espaço [14]. Entretanto, o algoritmo não garante a convergência para uma solução ótima e o seu desempenho depende dos centros iniciais escolhidos no passo 1.

B. *Fuzzy Clustering Method*

Para melhorar o desempenho do algoritmo *K-Means*, Bezdek em 1973 [15] propôs a incorporação do conceito *fuzzy* na função SQIntra, resultando no *Fuzzy C-Means Clustering* (FCM). Neste algoritmo, inicialmente cada objeto pertence a todos os *clusters*, porém com distintos graus de pertinência. Após a convergência do algoritmo, cada objeto é alocado no *cluster* onde apresenta maior grau de pertinência, ou seja, a *defuzzyficação* [10] é pelo máximo. O FCM é uma generalização do método *K-Means*, com partição binária ou rígida ($u_{ij}=1$ ou $u_{ij}=0$), para um método onde a partição é *fuzzy* ($0 \leq u_{ij} \leq 1$).

No FCM a partição ótima resulta da solução do seguinte problema de otimização não linear com restrições de soma unitária para os graus de pertinência de cada objeto:

$$\text{Min } J = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\|^2 \quad (16)$$

s.a.

$$\sum_{j=1}^k u_{1j} = 1$$

...

$$\sum_{j=1}^k u_{nj} = 1$$

No problema de otimização em 16, m é uma constante maior que 1 e reflete a natureza *fuzzy* da partição. Para $m \rightarrow 1$ a partição tende a ser rígida ($u_{ij} \rightarrow 0$ ou $u_{ij} \rightarrow 1$) e para $m \rightarrow \infty$ a partição é totalmente *fuzzy* ($u_{ij} \rightarrow 1/k$). Usualmente m é 1,25 ou 2 [16]. Com o auxílio da função Lagrangeana o problema de otimização em 16 pode ser escrito como:

$$\text{Min } J = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\| + \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^k u_{ij} - 1 \right) \quad (17)$$

Na equação 17, λ_i , $i=1, n$ são os multiplicadores de Lagrange para as n restrições de igualdade em 16. Derivando a equação 17 em relação a u_{ij} , c_j e λ_j obtém-se as seguintes condições necessárias:

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (18)$$

$$u_{ij} = \frac{1}{\sum_{t=1}^k \left(\frac{\|x_i - c_j\|}{\|x_i - c_t\|} \right)^{\frac{2}{m-1}}} \quad (19)$$

Usando as equações 18 e 19 pode-se implementar o algoritmo do FCM, onde a matriz U e os centros dos k clusters são obtidos iterativamente:

Passo 1 - Inicialize a matriz U com valores entre 0 e 1, observando que em cada linha da matriz a soma dos valores deve ser unitária. Esta etapa pode ser vista como a *fuzzificação* em um sistema de inferência fuzzy [10].

Passo 2 - Use a equação 18 para calcular as coordenadas dos k centros

Passo 3 - Calcule $J = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\|^2$. Pare se o valor

de J estiver abaixo de uma tolerância ou se a melhoria em relação à iteração anterior for desprezível.

Passo 4 - Use a equação 19 para atualizar a matriz U e volte para o passo 2

Após a convergência do algoritmo, cada objeto é alocado no cluster onde apresenta o maior grau de pertinência. Da mesma forma que no *K-Means*, o algoritmo FCM não garante a convergência para a solução ótima e a performance também depende das coordenadas dos centros iniciais.

C. Método de Ward

O método de *Ward* é um método hierárquico e aglomerativo, significando que ele particiona um conjunto de n objetos sucessivamente em $n-1, n-2, \dots, 2$ clusters. Por exemplo, para um conjunto com 5 objetos a, b, c, d, e , as possíveis partições geradas por um método hierárquico aglomerativo são:

5 clusters - a / b / c / d / e
 4 clusters - ab / c / d / e
 3 clusters - ab / cd / e

$$2 \text{ clusters - } ab / cde \quad (14)$$

A SQIntra indica o grau de homogeneidade interna dos clusters: quanto menor for esta soma, mais homogêneos são os clusters e melhor a qualidade da partição. A SQIntra cresce monotonicamente à medida que o número de clusters diminui, indicando que os clusters ficam maiores e cada vez mais heterogêneas. O exemplo anterior mostra que nos métodos hierárquicos, a cada iteração é feita a fusão de dois clusters e, por isso, é inevitável o acréscimo da SQIntra e a perda de qualidade da partição a cada iteração.

Para minimizar o acréscimo da SQIntra na passagem de k para $k-1$ clusters e não degenerar a qualidade da partição, o método de *Ward* agrupa os dois clusters, cuja agregação resulte no menor acréscimo da SQIntra. Este critério torna o método de *Ward* o mais eficiente dos métodos hierárquicos [7].

Para fazer a partição sucessiva dos n objetos em $n-1, n-2, \dots, 2$ clusters, os métodos hierárquicos necessitam de uma matriz simétrica de ordem n , onde cada elemento representa a distância entre dois objetos. À medida que os n objetos vão sendo agrupados, a ordem da matriz diminui e cada elemento passa a fornecer a distância entre os clusters. No exemplo dos objetos a, b, c, d, e a ordem da matriz de distância decai na sequência 5,4,3,2. A distância entre dois clusters a ser colocada na matriz de distâncias é dada por:

$$d_{ij} = \frac{p_i p_j}{p_i + p_j} \|c_i - c_j\|^2 \quad (20)$$

onde p_i e p_j são os pesos, ou melhor, o número de objetos nos clusters i e j respectivamente, e $\|c_i - c_j\|^2$ é o quadrado da distância euclidiana entre os centros dos clusters i e j .

V SISTEMA COMPUTACIONAL ANATIPO

Dotado de uma interface gráfica amigável, o ANATIPO possui uma variedade de funções que agilizam e tornam mais confortável o processo de construção das tipologias de clientes e redes.

Nesta seção procura-se apresentar algumas saídas gráficas e resultados obtidos pelo ANATIPO que evidenciam as vantagens da utilização de um sistema que integra todo o processo de construção das tipologias, desde a leitura das medições, passando pelos algoritmos de análise de agrupamentos, até o ajuste das tipologias ao mercado.

A seguir, a figura 5 ilustra uma tela do ANATIPO com a lista das medições de consumidores residenciais considerados em um estudo para construção de tipologias de um determinado segmento do mercado atendido pela distribuidora.

Código	Nome	Ativo	Arquivo
005-00001-001	Consumador Número 005-00001-001	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\527338ab.pdf
005-00001-002	Consumador Número 005-00001-002	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\061506es.amr
005-00001-003	Consumador Número 005-00001-003	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\062178ign.gls
005-00001-004	Consumador Número 005-00001-004	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\061939ign.gls
005-00001-005	Consumador Número 005-00001-005	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\062471ign.gls
005-00001-006	Consumador Número 005-00001-006	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\062318em.sld
005-00001-007	Consumador Número 005-00001-007	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\063285C.AUL
005-00001-008	Consumador Número 005-00001-008	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\063486G3.AG1
005-00001-009	Consumador Número 005-00001-009	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\0652188AF.AEP
005-00001-010	Consumador Número 005-00001-010	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\065577LPT.AMK
005-00001-011	Consumador Número 005-00001-011	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\0680104NH.AB1
005-00001-012	Consumador Número 005-00001-012	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_Publico\0688050C.A1.F
005-00001-013	Consumador Número 005-00001-013	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_EXCEL\CD3_TR01_034
005-00001-014	Consumador Número 005-00001-014	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_EXCEL\CD3_TR01_040
005-00001-015	Consumador Número 005-00001-015	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_EXCEL\CD3_TR04_006
005-00001-016	Consumador Número 005-00001-016	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_EXCEL\CD3_TR04_010
005-00001-017	Consumador Número 005-00001-017	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_EXCEL\CD3_TR04_011
005-00001-018	Consumador Número 005-00001-018	Sim	C:\CEPEL\Arquivos\INPUT\Fomato_EXCEL\CD3_TR04_014

Figura 5. Lista de arquivos com medições de curva de carga

No sistema ANATIPO, o processo de seleção das curvas características ocorre em um ambiente gráfico amigável, onde é possível visualizar simultaneamente todas as curvas, classificadas por dia da semana (figuras 6 e 7), registradas em um arquivo de medição. Nesta saída gráfica pode-se comparar facilmente o perfil de carga em diferentes dias da semana e selecionar diretamente no gráfico as curvas do dia útil, sábado e domingo que melhor caracterizem a medição.

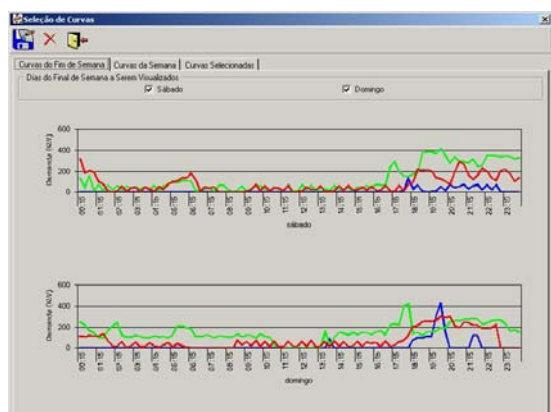


Figura 6 – Curvas do Fim de Semana

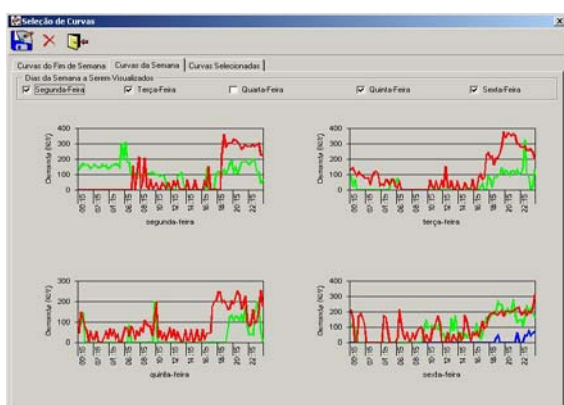


Figura 7 – Curvas dos dias úteis

Na figura 6 pode-se visualizar separadamente ou simultaneamente as curvas do sábado e do domingo, bastando clicar sobre o dia que se deseja visualizar.

De forma similar na tela da figura 7 pode-se visualizar e selecionar a curva característica do dia útil da semana. Para selecionar uma curva de carga nas telas das figuras 6 e 7 basta clicar sobre ela.

As curvas características selecionadas podem ser visualizadas na pasta *Curvas Selecionadas* (figura 8).

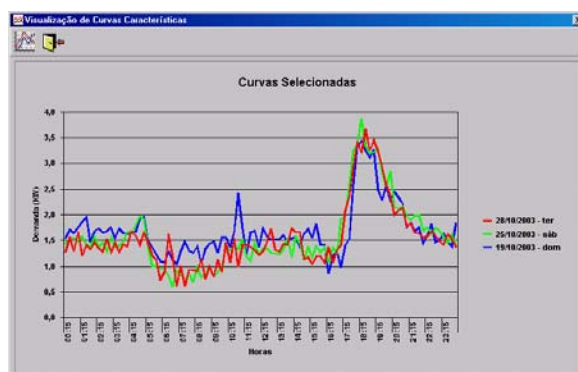


Figura 8. Tela para visualização de uma medição e seleção manual das respectivas curvas características

Ao invés de fazer a inspeção visual de cada arquivo de medição, um processo que pode consumir bastante tempo, a identificação das curvas características pode ser realizada de forma automática. Nesta opção o sistema importa e processa a identificação das curvas características de lotes de medição.

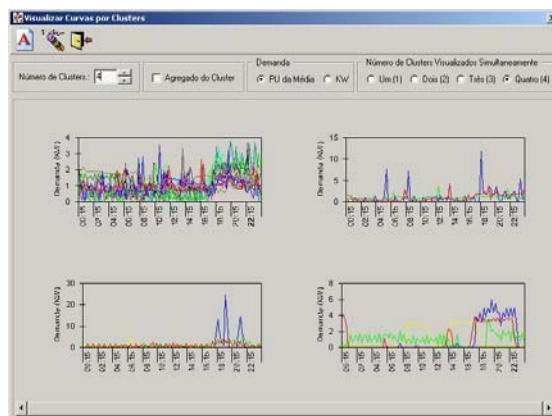


Figura 9 – Visualização das curvas em cada cluster

Os resultados, obtidos por qualquer um dos três métodos de *cluster analysis*, são apresentados em uma saída gráfica que permite visualizar as curvas de carga classificadas em cada agrupamento (figura 9) e a respectiva tipologia (figura 10).

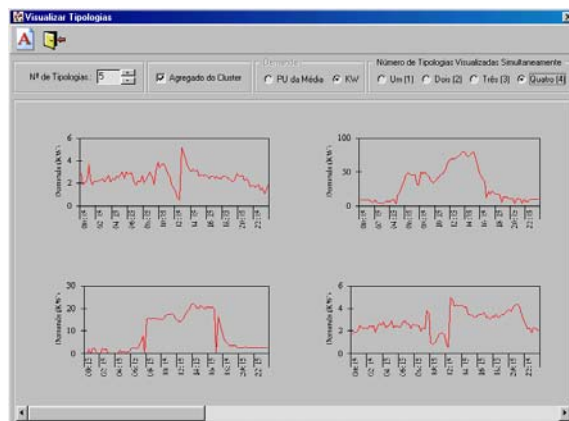


Figura 10. Tipologias

Na tela da figura 9 ainda é possível eliminar uma curva de carga de um agrupamento. Ao fazer a exclusão de uma curva de um determinado agrupamento, o sistema irá eliminá-la de todos os *clusters* gerados. Caso a eliminação de uma curva gere um agrupamento vazio, este também será eliminado.

Outra forma de alterar os resultados obtidos pelo algoritmo de análise de agrupamentos consiste em transferir curvas de carga de um *cluster* para outro.

Conforme indicado na figura 11 a transferência das curvas entre os dois *clusters* selecionados pode acontecer em qualquer sentido e para efetuar a transferência basta clicar na curva a ser transferida e logo em seguida clicar nas seta que indica o sentido da transferência. Caso seja transferida a última curva de um *cluster*, o sistema irá gerar uma nova seqüência para o número de *cluster*.

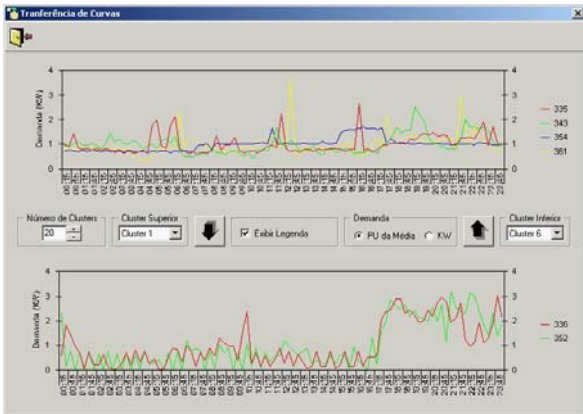


Figura 11 – Transferência de Curvas

O número adequado de *clusters*, ou melhor, de tipologias que caracterizam o conjunto de medições não é conhecido previamente. Para orientar a definição deste número o sistema disponibiliza um arquivo texto (figura 12) com indicadores de inércia inter classe e intra classe (soma dos quadrados) e o índice de compacidade e separação [16] que permitem avaliar o grau de homogeneidade dos agrupamentos formados.

NUMERO DE CLASSES	COMPACIDADE E SEPARACAO	SOMA DOS %INTER CLASSES	QUADRADOS %INTRA CLASSE	S. QUADRA INTER CLASSES
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.

Figura 12 – Análise da Variância

As estatísticas soma dos quadrados intra (*SQIntra*) e inter (*SQInter*) classes variam em função do número de *clusters* (*k*). Entretanto, a soma delas, denominada de inércia total, é constante e independe de *k*.

A partir destas estatísticas calcula-se a proporção da soma dos quadrados inter e intra classe na inércia total (*SQT*), índices úteis na avaliação da qualidade da partição da massa de dados em *k clusters*.

A razão *SQInter/SQT* varia entre 0 e 1. Quando o número de *clusters* aumenta (*k* ↑), os agrupamentos tornam-se mais homogêneos internamente (*SQIntra* ↓) e mais distintos entre si (*SQInter* ↑). No limite, quando cada *cluster* tem apenas uma curva de carga, *SQIntra* = 0 e *SQInter* = *SQT*, e portanto *SQInter/SQT* = 1 ou 100%.

De outra forma, quando o número de *clusters* diminui (*k* ↓), os agrupamentos se tornam mais heterogêneos (*SQIntra* ↑) e mais parecidos entre si (*SQInter* ↓). No limite, quando todos as curvas estão no mesmo *cluster*, *SQIntra* = *SQT* e *SQInter* = 0, resultando em *SQInter/SQT* = 0.

A partir destes resultados, pode-se aceitar a partição em um determinado número de *cluster* como sendo razoável se a razão (*SQInter/SQT*) > 80%, significando que os *clusters* formados são homogêneos e explicam a maior parte da variabilidade total dos dados.

O índice de compacidade e separação (*CS*) é definido como:

$$CS = \frac{\sum_{j=1}^k \sum_{i=1}^n u_{ij} \|x_i - c_j\|^2}{n \cdot \min \|c_{j1} - c_{j2}\|^2} \quad (21)$$

Em 21 o termo $\min \|c_{j1} - c_{j2}\|$ é a menor distância entre os centros de dois *clusters* e se o método usado for rígido, *u_{ij}* é uma variável binária (*u_{ij}* = 1 ou *u_{ij}* = 0). O numerador de *CS* mede a heterogeneidade interna dos *clusters*, enquanto o denominador mede a separação entre eles. Em uma partição ótima a heterogeneidade interna deve ser mínima e a separação entre os *clusters* deve ser máxima, portanto, o número ideal de *clusters* deve minimizar *CS*.

Após a formação dos *clusters* e a definição do número de tipologias, o ANATIPO ajusta, automaticamente, as tipologias ao mercado anual (MWh) do segmento que elas representam.

Como resultado principal, o ANATIPO gera uma planilha eletrônica (Figura 13) com as tipologias para os dias útil, sábado e domingo, ajustadas ao mercado e prontas para serem anexadas às propostas tarifárias da distribuidora.

Figura 13. Planilha com tipologias ajustadas ao mercado

Finalmente o sistema ANATIPO, cria um diretório Tar-Dist dentro do diretório do caso, com um arquivo texto, para que as tipologias possam ser carregadas no *TarDist* [3], um

sistema computacional desenvolvido também pelo CEPEL, para cálculo dos custos marginais de fornecimento e tarifas de uso dos sistemas de distribuição.

VI CONCLUSÕES

Neste artigo foi apresentado o ANATIPO, um sistema computacional desenvolvido pelo CEPEL em conjunto com a COSERN, por meio de um projeto de P&D ANEEL, com o objetivo de agilizar os estudos de caracterização da carga, beneficiando particularmente os estudos tarifários.

O sistema é dotado de interface gráfica amigável e conta com uma série de funcionalidades que tornam mais confortável a construção das tipologias de curvas de carga, desde a aquisição das medições, passando pela seleção das curvas características até a agregação e ajuste das tipologias ao mercado.

Dentre as diversas funcionalidades do programa destacam-se as rotinas para leitura de arquivos de medição em formato público e planilha eletrônica, o que confere uma grande flexibilidade no tratamento dos dados provenientes das campanhas de medidas.

Outras funcionalidades importantes do programa, apresentadas resumidamente neste trabalho, são as rotinas gráficas para identificação manual e automática das curvas características de cada medição, as rotinas gráficas para visualização dos *clusters* de curvas de carga, e os três métodos de análise de agrupamentos (*K-Means*, *Ward* e *FCM*) implementados.

Enfim, estas funcionalidades proporcionam maior agilidade na construção das tipologias de curvas, contribuindo para a elaboração das propostas tarifárias.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

- [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] BEBER, S. DANIEL “Cronograma de tarefas para realizar el estudio de un cuadro tarifario”, II Reúñion Internacional de Distribución y Comercialización de Energía Eléctrica, Lima, Peru, agosto de 2001.
- [3] 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 Distribuição”, XXXVI Simpósio Brasileiro de Pesquisa Operacional, São João del Rei – MG, 2004.
- [4] JANNUZI, GILBERTO M., SWISHER, JOEL N.P. *Planejamento Integrado de Recursos Energéticos*. Campinas, São Paulo, Editora Autores Associados, 1997.
- [5] HWANG, J.C. Assessment of Air Condition Load management by load Survey in Taipower. *IEEE Transactions on Power Systems*, Vol, 16 no 4, November 2001
- [6] CHEN, C.S., HWANG, J.C., HUANG, C.W. “Application of Load Survey Systems to Proper Tariff Design”, *IEEE Transactions on Power Systems*, Vol. 12, nº 4, November 1997
- [7] JOHNSON, RICHARD A., WINCHERN, DEAN W. *Applied Multivariate Statistical Analysis*, 4th edition, Prentice-Hall, 1998
- [8] BOUROCHE, JEAN-MARIE, SAPORTA, GILBERT *Análise de Dados*, Zahar Editores, Rio de Janeiro, 1980.
- [9] ALDENDERFER, M.S., BLASHFIELD, R.K. *Cluster Analysis*, Series: Quantitative Applications in the Social Sciences, SAGE publications
- [10] JANG, 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
- [11] BRASIL, Ministério das Minas e Energia, Secretaria de Energia, DNAEE; *Programa de Revisão Tarifária - PRT, Projeto 1 - Caracterização da Carga*, 1994.
- [12] MURTEIRA, BENTO J.F. *Análise Exploratória de Dados – Estatística Descritiva*, McGraw Hill, Lisboa, 1993.
- [13] PESSANHA, J.F.M., VELASQUEZ, R.M.G., MELO, A.C.G., CALDAS, R.P. “Técnicas de Cluster Analysis na Construção de Tipologias de Curvas de Carga”, XV Seminário Nacional de Distribuição de Energia Elétrica - SENDI, Salvador, 2002.
- [14] JAIN, A.K., DUIN, R.P.W., MAO, J. “Statistical Pattern Recognition: A Review”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, nº 1, January 2000.
- [15] BEZDEK, JIM C.; *Fuzzy Mathematics in Pattern Classification*, PhD thesis, Applied Math, Center, Cornell University, Ithaca, 1973.
- [16] FRANCO, CLÁUDIA R. “Categorização Nebulosa e Validação de Categorias Aplicadas a Problemas de Reconhecimento de Padrões”, Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, IM/NCE, 2002.