

## Técnicas de Classificação para Caracterização da Curva de Carga de Empresas de Distribuição de Energia - Um Estudo Comparativo

Roberto M.G.Velasquez<sup>1</sup>, José F.M.Pessanha<sup>1</sup>,

Débora L.D.D.Jardim<sup>1</sup>, Saul L.Melo<sup>1</sup>, Albert C.G.Melo<sup>1</sup>

<sup>1</sup>CEPEL – Centro de Pesquisas de Energia Elétrica, Rio de Janeiro - RJ

E-mails: rmgv@cepel.br, francisc@cepel.br, debora@cepel.br, saul@cepel.br, albert@cepel.br

### Abstract

*This work presents the application of clustering techniques in the load curve characterization studies. The methodology is commonly used by the major utilities in Brazil with the purpose of identifying consumer demand profiles. The results obtained by different clustering techniques (cluster analysis, neural networks and fuzzy logic), applied on a sample of load curves of low voltage consumers, are presented and compared.*

### 1. Introdução

O comportamento da demanda de energia elétrica nas horas do dia e nas estações do ano é uma informação vital nas atividades de planejamento e operação de sistemas elétricos.

No caso de sistemas de distribuição, esta informação é obtida através da caracterização da carga, um procedimento cuja finalidade consiste em encontrar curvas de cargas típicas, representativas do perfil da demanda dos consumidores atendidos por uma dada distribuidora de energia elétrica. As tipologias de curvas de carga encontram inúmeras aplicações, das quais podem ser destacadas :

- Cálculo dos custos marginais e das tarifas de fornecimento
- Gerenciamento da demanda
- Políticas de conservação de energia
- Planejamento da operação
- Dimensionamento de redes
- Avaliação das perdas de energia

Em geral, quando o fornecimento de energia é realizado em tensões mais elevadas (>34.5kV), o levantamento das curvas de carga é mais simples: o número de consumidores é pequeno, o perfil de consumo tende a ser uniforme ao longo do ano, os medidores são eletrônicos e possuem capacidade de armazenamento de dados (memória de massa).

Entretanto, na média (13.8kV-34.5kV) e baixa (<13.8kV) tensão o número de consumidores é elevado

resultando em uma maior diversidade de usos e hábitos e, adicionalmente, os medidores são usualmente do tipo eletromecânico sem capacidade de armazenamento de dados. Assim nestes níveis de tensão o levantamento das curvas de carga é feito por amostragem, através das chamadas “campanhas de medidas”.

A construção das curvas de cargas típicas é uma das etapas mais importantes do processo de caracterização da curva de carga, pois requer a análise de um grande conjunto de medições. Ela é realizada em duas etapas :

Etapa 1 - Análise visual das medições de cada consumidor, para identificar uma curva de carga típica para o dia útil, o sábado e o domingo.

Etapa 2 – Aplicação de técnicas de classificação para agrupar as curvas de carga do dia útil identificadas na etapa 1, em grupos homogêneos. Após a formação dos grupos, as curvas de carga dentro de cada grupo são somadas, resultando nas tipologias.

Neste trabalho as técnicas de classificação foram aplicadas em uma amostra com 157 medições reais de curvas de carga de consumidores de média tensão, com intervalo de integração de 15 minutos (a curva de carga diária tem 96 pontos).

Para comparar as diferentes técnicas de classificação foi fixado um número de 10 classes e o critério de avaliação usado foi baseado na soma dos quadrados *intra e inter-classes*.

Na seção 2 são descritas as técnicas de agrupamento estatístico [1], tradicionalmente utilizadas em programas para caracterização da carga e cálculo de tarifas de uso dos sistemas de distribuição [2,3,4]. Nas seções 3 e 4 são apresentadas as técnicas de *redes neurais* [5,6] e *lógica difusa* [7], respectivamente, e as aplicações em problemas de classificação. Finalmente na seção 5 é feita uma comparação dos resultados obtidos pelas diferentes técnicas.

### 2. Análise de conglomerados

Diversas técnicas da estatística multivariada [1] são usadas para organizar um conjunto de  $N$  objetos em subconjuntos mutuamente exclusivos, denominados conglomerados ou *clusters*, de tal forma que os objetos em

um mesmo subconjunto sejam semelhantes entre si, mas ao mesmo tempo diferentes dos objetos pertencentes aos outros subconjuntos. Basicamente, estas técnicas procuram no conjunto de dados uma estrutura natural de agrupamento dos objetos.

Todos os algoritmos de análise de conglomerados são baseados em uma medida de similaridade ou, ao contrário de distância, que procuram expressar o grau de semelhança entre os objetos. Uma medida de distância muito utilizada quando os atributos são de natureza quantitativa é a distância euclidiana. Sejam  $x = (x_1, \dots, x_p)$  e  $y = (y_1, \dots, y_p)$  dois objetos caracterizados por  $p$  atributos. Então a distância euclidiana entre estes dois objetos é :

$$d_{x,y} = \sqrt{(x-y) \cdot (x-y)^T} \quad (1)$$

As distâncias entre os objetos podem ser organizadas na forma de uma matriz conhecida como matriz de distâncias, uma matriz simétrica, de ordem igual ao número de objetos,  $N$ , onde o elemento  $d_{ij}$  é a medida de distância entre os objetos  $i$  e  $j$ .

As técnicas de análise de conglomerados podem ser divididas em duas classes :

- *Métodos não hierárquicos* : particionam o conjunto de objetos em um número fixado de  $k$  classes
- *Métodos hierárquicos* : particionam o conjunto de objetos sequencialmente em 1,2,3,4 até  $k$  classes, obtendo no final uma estrutura em árvore, semelhante às classificações zoológicas (espécies, gêneros, famílias, ordem, etc.).

## 2.1 Métodos hierárquicos

Os métodos hierárquicos usam a matriz de similaridade para construir as classes através de um processo iterativo, que pode ser divisivo ou aglomerativo. No divisivo, inicialmente há apenas uma classe formada por todos os objetos e que é dividido sucessivamente até que no final cada classe contenha apenas um objeto. Já no aglomerativo, no início cada objeto é uma classe que sucessivamente sofre uma série de fusões com as outras classes até que no final todos os objetos estejam em um único conjunto.

Neste trabalho foram considerados apenas os métodos aglomerativos, em particular, métodos de encadeamento e o método de Ward. A seguir é apresentado o algoritmo do método aglomerativo para formar classes em um conjunto com  $N$  objetos :

- 1) Inicie com  $N$  classes, cada uma contendo apenas um objeto e construa a matriz de distâncias de ordem  $N$ .

- 2) Identifique o menor elemento da matriz de distâncias para encontrar o par de classes mais próxima ou similar.
- 3) Reúna as duas classes identificadas na etapa 2 em uma única classe e atualize a matriz de distâncias, retirando as linhas e colunas relativas às duas classes identificadas em 2 e incluindo a linha e coluna com as distâncias entre as demais classes e a nova classe formada. Note que a ordem da matriz de distâncias diminui de uma unidade a cada vez que a etapa 3 é executada
- 4) Repita os passos 2 e 3 até que reste apenas uma classe. A cada repetição guarde a identificação das classes que foram fundidas e também a distância entre elas. Estas informações serão utilizadas na montagem de um gráfico conhecido como dendrograma que mostra a sequência de aglomeração.

No início do algoritmo quando cada classe tem apenas um objeto, as distâncias euclidianas entre as classes são facilmente determinadas, pois correspondem exatamente às distâncias euclidianas entre os objetos. Entretanto, a partir do momento em que as classes são fundidas e passam a conter mais de um objeto, deve-se estabelecer um critério para determinar as distâncias entre elas. Nos métodos de encadeamento três critérios são utilizados: distância mínima (*single linkage*), distância máxima (*maximum linkage*) e distância média (*average linkage*).

No critério de distância mínima, a distância entre duas classes é dada pela distância entre os dois objetos, um em cada grupo, mais próximos. Enquanto que no critério de distância máxima, ocorre o contrário, a distância entre duas classes é igual a distância entre os dois objetos, um em cada grupo, mais distantes. Por sua vez no critério de distância média a distância entre duas classes é dada pela média das distâncias entre os pares de objetos, um em cada classe.

Outro importante método hierárquico é o método de Ward. Este método se baseia na minimização da “perda de informação”, dada pelo acréscimo na soma dos quadrados dos desvios (WSS), quando se funde duas classes:

$$WSS = \sum_{i=1}^K \sum_{j=1}^{NOi} (x_j - \bar{x}_i)^T (x_j - \bar{x}_i) \quad (2)$$

onde  $K$  é o número de classes e  $NOi$  é o número de objetos na classe  $i$  e  $\bar{x}_i$  é o centróide da classe  $i$ .

Primeiramente, são testadas todas as possibilidades de junção dos grupos, mas somente irá se efetivar aquela que produzir o menor acréscimo na WSS, i.e., a menor perda de informação.

Nos métodos hierárquicos o número de classes não é fixado a priori, mas resulta da visualização do dendrograma, um gráfico que mostra a sequência das fusões ou divisões ao longo do processo iterativo.

## 2.2 Métodos não hierárquicos

Nestes métodos, a matriz de similaridade não é construída e o número de classes,  $k$ , é fixado a priori, ao contrário do que ocorre nos métodos hierárquicos. Um método não hierárquico muito popular e disponível na maioria dos pacotes estatísticos é o  $k$ -means, cujo algoritmo é dado a seguir :

- 1) Escolha um conjunto de  $k$  centróides iniciais. A escolha dos centróides iniciais pode ser aleatória, e.g., pontos quaisquer no espaço  $p$ -dimensional, ou baseada em algum outro critério.
- 2) Percorra a lista de objetos e aloque cada objeto ao centróide mais próximo. Note que após alocar todos os objetos, foram formados  $k$  classes.
- 3) Atualize o centróide de cada uma das  $k$  classes e volte para o passo 2. O algoritmo deve continuar até o momento que não haja realocação de objetos de uma classe para outra.

## 3. Redes Neurais Artificiais

A classificação e o reconhecimento de padrões são áreas onde a aplicação das redes neurais artificiais (RNA) [4,5] tem obtido um grande sucesso. Daí surgiu a motivação para aplicar a técnica de RNA na caracterização da curva de carga de sistemas de distribuição. Da mesma forma que os humanos, as RNA são capazes de aprender a reconhecer padrões através da apresentação de exemplos destes padrões. Após um período de treinamento as RNA são capazes de generalizar e reconhecer padrões de ocorrências novas.

Dada as características do problema de classificação em estudo, optou-se, dentre os diversos métodos de construção de topologias de RNA, pela utilização do Mapa de Kohonen, descrito a seguir.

### 3.1 Mapa de Kohonen

A camada de Kohonen é tal que, para uma dada entrada ( uma curva de carga com 96 pontos ) somente um neurônio apresenta na sua saída o nível lógico “um”, enquanto os demais estarão em nível “zero”. A associação entre cada neurônio de saída com um dado vetor de entrada é feita pelo vetor de pesos correspondente. A Figura 1 ilustra a estrutura:

Neste caso, a saída do  $m$ -ésimo neurônio  $s_m$ ,  $m=1,k$ , pode ser expressa como :

$$s_m = \sum_{i=1}^{96} w_{i,m} x_i \quad (3)$$

onde  $w_{i,m}$  denota o peso sináptico do  $i$ -ésimo elemento do vetor de entrada,  $x_i$ , no  $m$ -ésimo neurônio de saída.

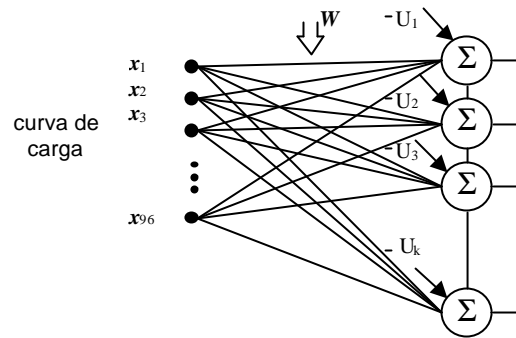


Figura 1 – Camada de Kohonen

Os pesos da camada de Kohonen são ajustados de tal forma que vetores de entrada similares ativem sempre o mesmo neurônio de saída, ou seja, pertençam à mesma classe. Sempre que um vetor de entrada é apresentado à rede, calcula-se a saída de cada um dos neurônios, através da equação 2, e o neurônio ativado ou vencedor é o que apresentar a maior saída. Em seguida os pesos relativos ao neurônio vencedor são atualizados através da equação :

$$w_{new} = w_{old} + \mathbf{a}(x - w_{old}) \quad (4)$$

onde  $w_{new}$  é o novo vetor de pesos,  $w_{old}$  é o vetor de pesos anterior e  $\mathbf{a}$  o coeficiente de treinamento.

### 3.2 Treinamento

Usualmente, o treinamento da camada de Kohonen é não supervisionado. Portanto, não se determina, *a priori*, qual neurônio será ativado para uma dada entrada. Inicialmente, cria-se um neurônio com o vetor de sinapses  $w^1$  igual ao primeiro vetor de entrada :

$$w^1 = x^1 \quad (5)$$

A seguir, apresenta-se o segundo vetor  $x^2$  à rede e compara-se o produto escalar  $(x^2)^T \cdot w^1$  com a constante  $U$ , dada por:

$$U = 1 - r^2 / 2 \quad (6)$$

onde  $r$  é uma constante conhecida como raio máximo e deve ser escolhida cuidadosamente, pois é através dela que será possível ajustar o grau de similaridade admissível para que a entrada seja classificada na classe  $w^1$ .

Caso  $(x^2)^T \cdot w^1 > U$ , efetua-se a atualização :

$$w_{new}^1 = w_{old}^1 + \mathbf{a}(x - w_{old}^1) \quad (7)$$

senão, cria-se um novo neurônio, cujo vetor de pesos é  $w^2 = x^2$ .

Assim para cada nova entrada apresentada à rede, verifica-se se algum dos neurônios já existentes foi ativado e atualiza-se os respectivos pesos, ou então, cria-se um novo neurônio.

Na prática, este processo de treinamento pode gerar um número elevado de neurônios de saída, sendo usual o emprego de métodos de poda *a posteriori*, quer seja desprezando-se neurônios poucas vezes ativados, ou fundindo-se neurônios próximos com um incremento do raio  $r$ . Neste trabalho com um raio de 0.62, foi possível obter uma rede com 10 neurônios de saída (*classes*).

#### 4. Lógica Fuzzy

A teoria de conjuntos nebulosos (*fuzzy logic*) [7] trata o aspecto impreciso e vago da informação. Baseado nesta teoria, é possível desenvolver modelos computacionais que emulem a habilidade do cérebro humano de lidar com um modo de raciocínio aproximado ou com informações ambíguas, e mesmo assim, encontrar soluções adequadas. Devido a grande capacidade de realizar inferências, são encontradas muitas aplicações de lógica difusa no desenvolvimento de sistemas especialistas e controladores e também no reconhecimento de padrão.

*C-Means Fuzzy* é uma técnica de agrupamento difuso onde cada *objeto* é alocado em diferentes classes em função de um certo grau de pertinência. Esta técnica foi originalmente introduzida como uma melhoria em relação as técnicas de agrupamento já existentes [7]. Ela fornece um método de agrupamento de objetos num espaço multidimensional em um número pré-definido de classes.

O método é iniciado com o sorteio dos centróides de cada classe, onde o centróide determina a posição de referência da classe. Paralelamente o método atribui para cada objeto um grau de pertinência ( $MF_{ij}$ ) em relação a cada classe através do cálculo de uma medida de distância.

$$MF_{ij} = \frac{1}{dist_{ij}^{0.1}} \quad (8)$$

onde  $dist_{ij}$  é a distância Euclidiana entre os objeto  $i$  e centróide da classe  $j$ .

Mediante um processo iterativo são atualizados os centróides das classes e os graus de pertinência de cada objeto. Este processo iterativo é baseado na minimização de uma função objetivo (FO) que representa a distância de um objeto qualquer (ponderado pelo seu grau de pertinência) ao centróide de uma classe.

$$FO = \sum_{j=1}^k \sum_{i=1}^N (dist_{ij}^2 \cdot MF_{ij}) \quad (9)$$

onde  $k$  é o número de classes a serem formadas e  $N$  o número de objetos. Neste trabalho  $k=10$  e  $N=157$ .

#### 5. Comparação dos resultados

As técnicas de agrupamento descritas anteriormente foram aplicadas a uma amostra com 157 medições reais de carga de consumidores de média tensão, com intervalo de integração de 15 minutos (a curva de carga diária tem 96 pontos). Para facilidade de comparação, o número de classes foi fixado em 10 classes.

As Figuras 2-5 apresentam, respectivamente, os resultados da classificação obtida pelos métodos C-Means Fuzzy, RNA Kohonen, Hierárquico Ward e Não-hierárquico K-Means.

A partir de uma análise pictória verifica-se na Figura 5 que das 10 classes formadas pela técnica de lógica difusa, 5 delas são semelhantes (classes 3,6,7,9 e 10). O mesmo acontece com o método de Ward (classes 3,4,9 e 10 na Figura 2) e no método não-hierárquico K-Means (classes 6 e 8 na Figura 3). Isto sugere que o conjunto pode ser dividido em um menor número de classes.

Por outro lado, verificou-se que a RNA é mais sensível à diversidade, pois produziu 6 classes cada um com 1, 2 ou 3 curvas (classes 2,5,6,8,9,10 na Figura 4). A classificação pela RNA se concentrou na separação das curvas com comportamento bastante diferente. Esta característica pode algumas vezes dificultar a caracterização da carga. Por exemplo, na Figura 4, a classe 4 contém um grande número de curvas e poderia ter sido subdividido.

Entretanto, uma maneira complementar de se comparar o desempenho das diversas técnicas de classificação consiste em analisar os valores dos somatórios dos quadrados intra-classe (WSS) e inter-classe (BSS), definidos a seguir :

$$WSS_J = \sum_{X_i \in \text{classe } J} (X_i - \bar{X}_J)(X_i - \bar{X}_J)^T \quad (10)$$

$$WSS = \sum_{J=1}^K WSS_J \quad (11)$$

$$BSS = \sum_{J=1}^K N_J (\bar{X}_J - \bar{\bar{X}})(\bar{X}_J - \bar{\bar{X}})^T \quad (12)$$

onde  $k$  é o número de classes ( $k=10$ ),  $X_i$  é um vetor linha com a  $i$ -ésima curva de carga,  $\bar{\bar{X}}$  é a curva de carga média da amostra,  $\bar{X}_J$  é a curva de carga média na classe  $J$  e  $N_J$  é o número de curvas de carga na classe  $J$ , tal que,  $N_1 + N_2 + \dots + N_k = N$ , o total de curvas de carga ( $N=157$ ).

A Tabela 1 apresenta os valores de BSS e WSS obtidos para os métodos hierárquicos (Ward, Avarege e Single), não-hierárquicos, RNA e lógica difusa. Observa-se nesta tabela que o melhor desempenho foi alcançado pela técnica de lógica difusa, seguida de perto pelos métodos hierárquico Ward e não-hierárquico K-Means. Posteriormente aparece o método hierárquico Average com desempenho similar ao da RNA Kohonen. Finalmente, o pior desempenho foi obtido pelo método hierárquico *Single*, que aparece de forma isolada na Tabela.

Tabela 1. Comparação de Resultados

<b>Método</b>	<b>BSS</b>	<b>WSS</b>
Hierárquico Ward	15.9354	11.0614
Hierárquico Average	12.9990	13.9978
Hierárquico Single	4.8642	22.1326
Não Hier.(K-Means)	15.0810	11.9180
RNA Kohonen	12.7700	14.2268
Fuzzy C-Means	16.1263	10.8705

## 6. Conclusões

Além dos métodos de classificação tradicionais utilizados pelas concessionárias de energia, baseados em técnicas estatísticas, novos métodos podem ser utilizados na agregação de curva de carga de consumidores. Os resultados indicaram um bom desempenho dos métodos baseados na lógica difusa e em redes neurais artificiais. Este fato sugere a continuidade da aplicação destes métodos, utilizando, por exemplo, uma amostra maior de curvas de carga bem como fixando-se diferentes números de classes.

## 7. Referências

- [1] Johnson, Richard A., Winchern, Dean W, Applied Multivariate Statistical Analysis 4<sup>th</sup> edition, Editora Prentice-Hall, 1998.
- [2] BRASIL, Ministério das Minas e Energia, DNAEE, Eletrobrás , Empresas Concessionárias de Energia Elétrica, Nova Tarifa de Energia Elétrica; metodologia e aplicação, DNAEE, Brasília, 1985.
- [3] Proposta de Tarifas de Uso (Resolução N.º 286/99).Nota de Esclarecimento. ANEEL-Agencia Nacional de Energia Elétrica.Versão 19/06/00.
- [4] Pessanha,J.F.M., Velasquez,R.M.G., Passos Jr., R., Melo,A.C.G, Caldas, R.P., Tanure,J.E.P.S, Santos,P.E.S, Metodologia e Aplicação do Cálculo dos Custos Marginais de Fornecimento e das Tarifas de Uso dos Sistemas de Distribuição. Latin America Power & Gas, Brasil, 2001. (a ser apresentado)
- [5] Melo, S.L. , Caloba, L.P., Nadal, J., Classificação de Batimentos Cardíacos Utilizando Rede Neural com Treinamento Competitivo Supervisionado. Anais do II ENIA, Brasil, 1999.
- [6] Wasserman, P.D., Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.
- [7] Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

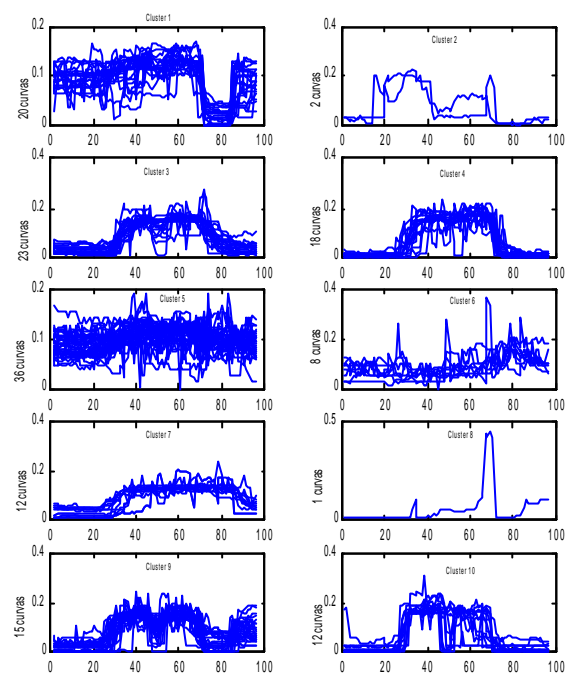


Figura 2 - Método Hierárquico Ward

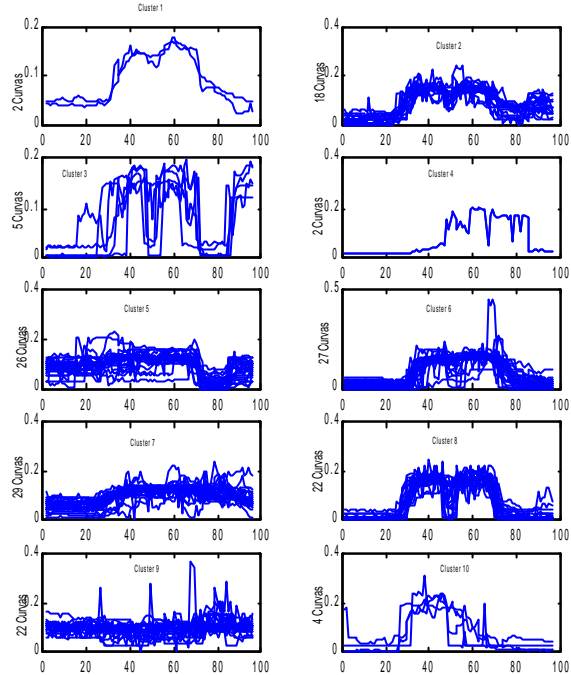


Figura 3 - Método Não-Hierárquico K-Means

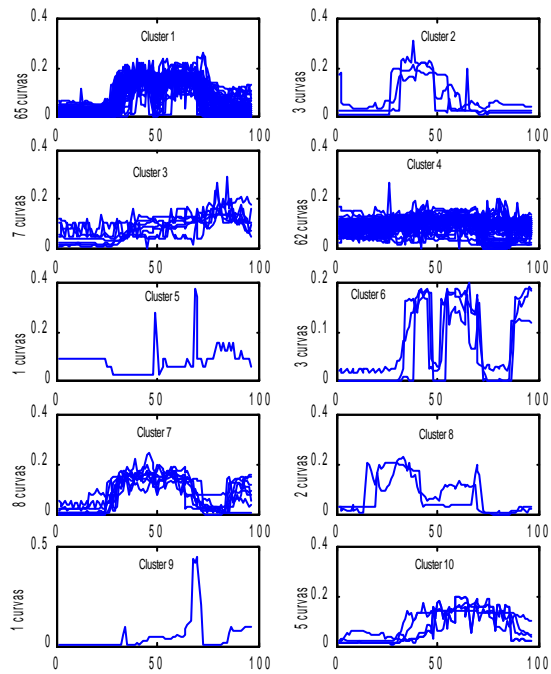


Figura 4 - Método RNA Kohonen

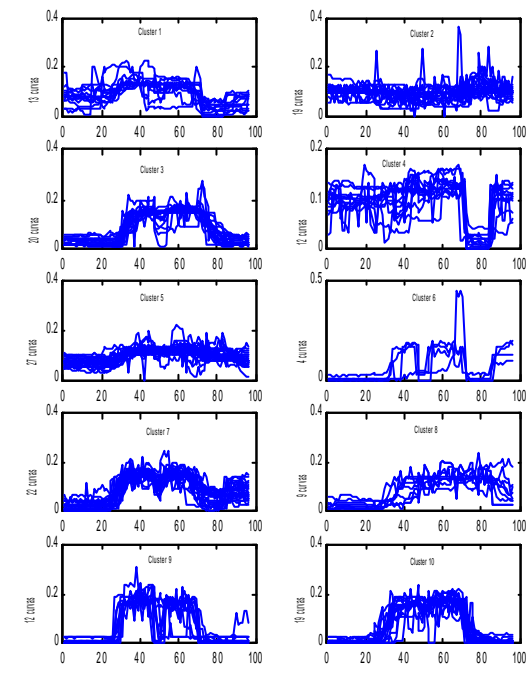


Figura 5 - Método C-Means Fuzzy