



XIX Seminário Nacional de Distribuição de Energia Elétrica

SENDI 2010 – 22 a 26 de novembro

São Paulo - SP - Brasil

Projeção de Demanda de Barramentos Elétricos da Elektro sem Medição Permanente através da Metodologia TwoStep Cluster

Rodrigo Manfredini	Márcia Inoue	Robinson Semolini
Elektro	Elektro	Elektro
rodrigo.manfredini@elektro.com.br	marcia.inoue@elektro.com.br	robinson.semolini@elektro.com.br

Palavras-chave

Agrupamento
Barramentos Elétricos
PRODIST
Projeção de Demanda
TwoStep Cluster

Resumo

O presente trabalho expõe uma metodologia de projeção de demanda máxima dos barramentos elétricos da distribuidora quando apenas alguns dos barramentos possuem medição permanente de demanda de potência. Este método foi empregado na projeção de demanda máxima por barramentos da Elektro, ciclo 2009-2014. A metodologia consiste em agrupar os barramentos de acordo com o consumo total, a participação de cada classe principal e informações geográficas do barramento, através da metodologia TwoStep cluster. Em seguida, é observado o crescimento dos barramentos com medição dentro de cada cluster e, desse modo, faz-se inferências sobre as demandas máximas dos barramentos elétricos sem medição permanente.

1. Introdução

No modelo atual do sistema elétrico brasileiro, a assertividade das projeções de demanda de potência máxima por barramento elétrico é de fundamental importância para o planejamento adequado e prudente da expansão do sistema elétrico.

É obrigação das distribuidoras fornecerem energia com qualidade e segurança a todo o seu mercado de concessão e garantir que os investimentos são adequados e necessários. Para tal, é necessário se prever de forma precisa os aumentos de demandas que necessitem de obras de expansão, pois estas obras podem demorar alguns anos para se concretizarem.

Outro fator importante é que, o cálculo do IAS (índice de aproveitamento para subestação), índice utilizado para avaliação da remuneração dos ativos durante as revisões tarifárias, considera o fator de

utilização da subestação e a expectativa de crescimento da carga da subestação pelos próximos 10 anos. Então, necessita-se que os desvios de projeções da demanda de potência máxima por barramento elétrico sejam os menores possíveis para que o planejamento de expansão do sistema seja feito de maneira efetiva, sem excessos de investimentos e nem atrasos na execução de obras.

Além disso, toda distribuidora contrata anualmente, junto ao ONS, os montantes de uso do sistema de transmissão (MUST) para o próximo triênio. Esta contratação tem que ser de maneira bastante precisa, pois as distribuidoras estão sujeitas a penalidades por subcontratação ou sobrecontratação. Também há o fato do Módulo 2 do PRODIST, que trata do Planejamento de Expansão do Sistema de Distribuição, afirmar que as distribuidoras devem realizar projeções por barramento com o intuito de planejar seu sistema de distribuição.

A utilização de modelos estatísticos/econômicos para realizar projeções de demanda de potência máxima anual por barramento elétrico é uma ferramenta que pode ser muito útil para inferir a demanda máxima por barramento elétrico. Contudo, nem sempre a utilização destes modelos é viável. Para que se possa fazer um modelo adequado, há a necessidade de um número razoável de observações históricas de demanda máxima. Wei (1990) recomenda, por exemplo, que necessita-se de, no mínimo 50 observações para que um modelo SARIMA (Box & Jenkins) seja realizado.

Além da necessidade de um número razoável de observações para se realizar modelos estatísticos, há também o problema de algumas empresas do setor ainda não possuem medição permanente de toda a sua rede, inviabilizando também a projeção por métodos estatísticos/econômicos devido ao fato de não possuírem observações históricas.

Nesse contexto, o presente trabalho propõe uma alternativa para a projeção da demanda de potência máxima para os barramentos elétricos que não possuem medição permanente através da metodologia TwoStep cluster. A metodologia de clusterização por subestação através do método K-Means, proposta por Seixas, J., Menezes, M. & Correa, C., 2008, pode apresentar instabilidade nos resultados, visto que o primeiro passo do algoritmo é escolher aleatoriamente os k clusters nas n observações, o que pode deixar a clusterização muito dependente desse número aleatório escolhido.

O trabalho está dividido da seguinte forma: Primeiramente serão descritos os atributos e o conjunto de dados utilizados na clusterização. Em seguida, é apresentada a metodologia TwoStep cluster. Após, é feita uma breve descrição dos modelos estatísticos ADL-SARIMA, utilizados para a projeção de demanda dos barramentos com medição. Por fim, a metodologia de projeção dos barramentos sem medição é apresentada.

2. Desenvolvimento

2.1 Conjunto de Dados Utilizados na Clusterização

Para o planejamento de expansão do sistema elétrico da Elektro, ciclo 2009-2014, foram necessárias projeções da demanda máxima de potência anual de 175 barramentos. Destes 175 barramentos, apenas 97 deles possuem medição de fronteira através do sistema Telemedição e, portanto, 78 barramentos não possuem medição permanente ou foram digitalizados recentemente. Portanto, para grande parte deles não foi possível a realização de modelos estatísticos de séries temporais.

A figura 1 mostra o fluxograma de como foram projetadas as demandas máximas anuais dos barramentos da Elektro.

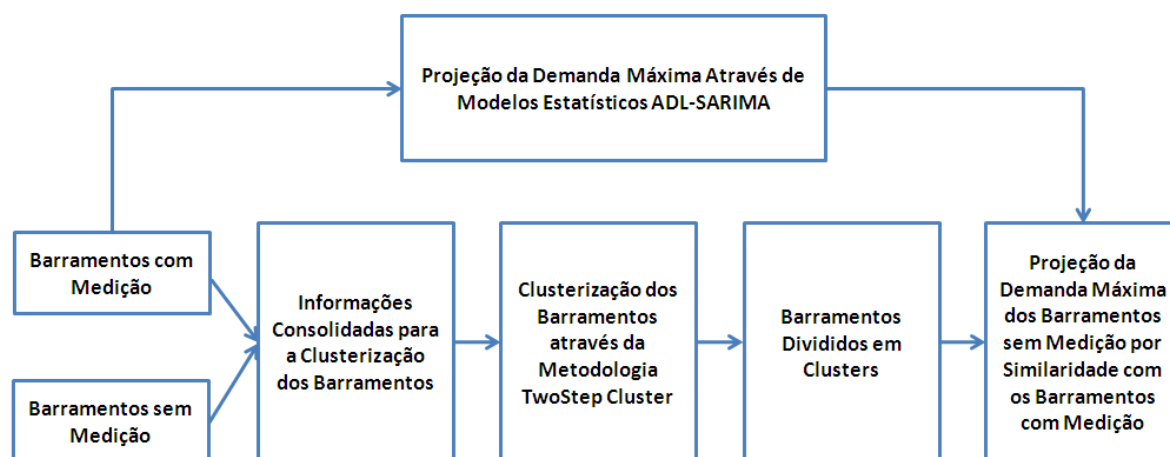


Figura 1: Fluxograma das projeções por barramento da Elektro

As informações disponíveis para a clusterização dos 175 barramentos são provenientes do sistema comercial da Elektro. São elas:

- **Consumo total:** Somatória do consumo medido no ano de 2008 das unidades consumidoras para cada barramento em estudo;
- **Percentual do consumo residencial:** Percentual do consumo residencial do barramento com relação ao consumo total;
- **Percentual do consumo industrial:** Percentual do consumo industrial do barramento com relação ao consumo total;
- **Percentual do consumo comercial:** Percentual do consumo comercial do barramento com relação ao consumo total;
- **Percentual do consumo rural:** Percentual do consumo rural do barramento com relação ao consumo total;
- **Percentual do consumo das demais classes:** Percentual do consumo dos serviços públicos, iluminação pública e poder público com relação ao consumo total;
- **Região:** Oito regiões distintas foram utilizadas. A figura 2 mostra-as de forma detalhada.

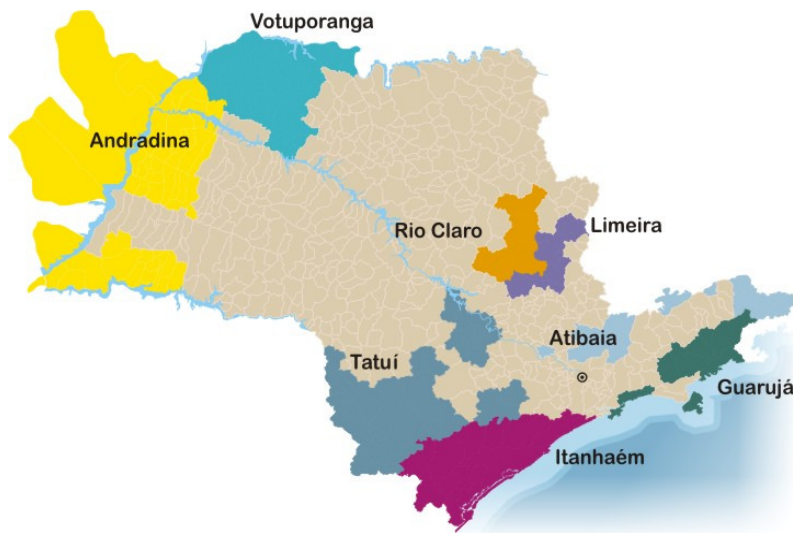


Figura 2: Regiões da Elektro

2.2 A Metodologia TwoStep Cluster

A metodologia TwoStep cluster é dividida em duas etapas: A primeira consiste em uma pré-clusterização dos dados e a segunda consiste em uma clusterização hierárquica dessa pré-clusterização realizada na etapa 1.

Esta metodologia é bastante rápida quando queremos agrupar um número grande de observações devido à pré-clusterização realizada na etapa 1. Também é permitido que se use variáveis categóricas e contínuas no processo. Outra vantagem é que há a possibilidade de escolha automática do número de clusters de acordo com o BIC (Schwarz's Bayesian Criterion) ou o AIC (Akaike Information Criterion).

2.2.1 Medida de Distância Utilizada

As medidas de distância possuem a função de determinar a similaridade de um cluster ao outro. Elas são utilizadas tanto na primeira quanto na segunda etapa da metodologia.

Duas medidas de distância são permitidas no método: A distância por log-verossimilhança e a distância Euclidiana. Quando há apenas variáveis contínuas no processo, é possível utilizar as duas distâncias, porém quando há variáveis categóricas envolvidas, somente é possível utilizar a distância por log-verossimilhança. Quando se usa esta última distância, deve-se assumir que as variáveis no modelo sejam independentes. Também deve-se assumir que cada variável contínua seja normalmente distribuída e que cada variável categórica tenha distribuição multinomial. Simulações empíricas demonstraram que o algoritmo funciona bem mesmo quando algumas dessas suposições anteriormente citadas são violadas, pois não há, por exemplo, testes de hipóteses durante o processo de agrupamento. No caso da clusterização dos barramentos, como utilizamos a variável categórica região, a distância utilizada será a distância por log-verossimilhança.

A distância por log-verossimilhança entre dois clusters é relacionada com o decréscimo na função log-verossimilhança que ocorre quando junta-se estes clusters em somente um. A distância entre o cluster j e s é definida por:

$$d(j, s) = \xi_j + \xi_s - \xi_{\langle j, s \rangle}, \quad (1)$$

onde

$$\xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right), \quad (2)$$

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}. \quad (3)$$

K^A	Número de variáveis contínuas utilizadas no processo.
K^B	Número de variáveis categóricas utilizadas no processo.
$\hat{\sigma}_k^2$	Variância estimada da k-ésima variável contínua no conjunto de dados total.
$\hat{\sigma}_{jk}^2$	Variância estimada da k-ésima variável contínua no cluster j .
N_k	Número de dados no cluster k .
N_{jkl}	Número de dados no cluster j cuja k-ésima variável categórica toma a l-ésima categoria.
L_k	Número de categorias da k-ésima variável categórica.
$\langle j, s \rangle$	Índice que representa o cluster formado quando é combinado o cluster j e s .

Se $\hat{\sigma}_k^2$ é ignorado na equação (2), a distância entre os clusters j e s é exatamente o decréscimo na função log-verossimilhança quando os dois clusters são combinados. O $\hat{\sigma}_k^2$ é adicionado à equação em virtude do problema causado quando $\hat{\sigma}_{jk}^2 = 0$, como ocorre no caso em que só temos uma observação no j -ésimo cluster.

2.2.2 Etapa 1: Pré-clusterização

O objetivo desta etapa é reduzir o tamanho da matriz de distâncias entre todos os possíveis pares de casos. Nesta etapa, o algoritmo decide, baseado na medida de distância escolhida, se a observação será colocada em um pré-cluster já formado ou se ela fará parte de um novo pré-cluster. O método de pré-clusterização é o método BIRCH (ver ZHANG, 1996) que consiste em construir uma árvore de clusterização modificada. A construção desta árvore de clusterização depende da ordem das observações. Para minimizar este efeito, sugere-se que se ordene os casos de maneira aleatória.

Quando a pré-clusterização se completa, todas as observações no mesmo pré-cluster são tratadas como uma mesma entidade. Logo, o tamanho da matriz de distâncias não dependerá mais do número de observações, mas sim do número de pré-clusters formados.

2.2.3 Etapa 2 : Clusterização Hierárquica

O segundo passo do algoritmo consiste em realizar uma clusterização hierárquica dos pré-clusters formados no passo 1. Como o número de pré-clusters formados é menor do que o número de observações para clusterizar, o método realizará a clusterização de forma mais rápida do que realizar uma clusterização hierárquica das observações sem a pré-clusterização.

Quanto maior o número de subclusters formados na pré-clusterização, melhores serão os resultados finais. Contudo, quanto mais subclusters tivermos, mais devagar será o a etapa 2. Portanto, é necessário que se encontre um número adequado de subclusters que não deixe o processo muito lento e que produza resultados finais com qualidade.

2.2.4 A escolha do número de clusters: A autoclusterização

Na metodologia TwoStep cluster, há duas possibilidades de se escolher o número de clusters. A primeira opção é escolher o número de clusters previamente e a segunda opção é utilizar a autoclusterização para escolher o número de clusters baseado no AIC ou BIC.

Na autoclusterização, primeiramente o BIC ou o AIC é calculado para cada número de clusters e estes são utilizados para propor uma estimativa inicial do número de clusters. Após isso, a estimativa inicial é refinada de modo a encontrar o maior acréscimo na distância entre os dois clusters mais próximos em cada estágio da clusterização hierárquica.

O BIC e o AIC para J clusters é definido como:

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N), \quad (4)$$

$$AIC(J) = -2 \sum_{j=1}^J \xi_j + 2m_j, \quad (5)$$

onde

$$m_j = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\} \quad (6)$$

e

N Número total de observações.

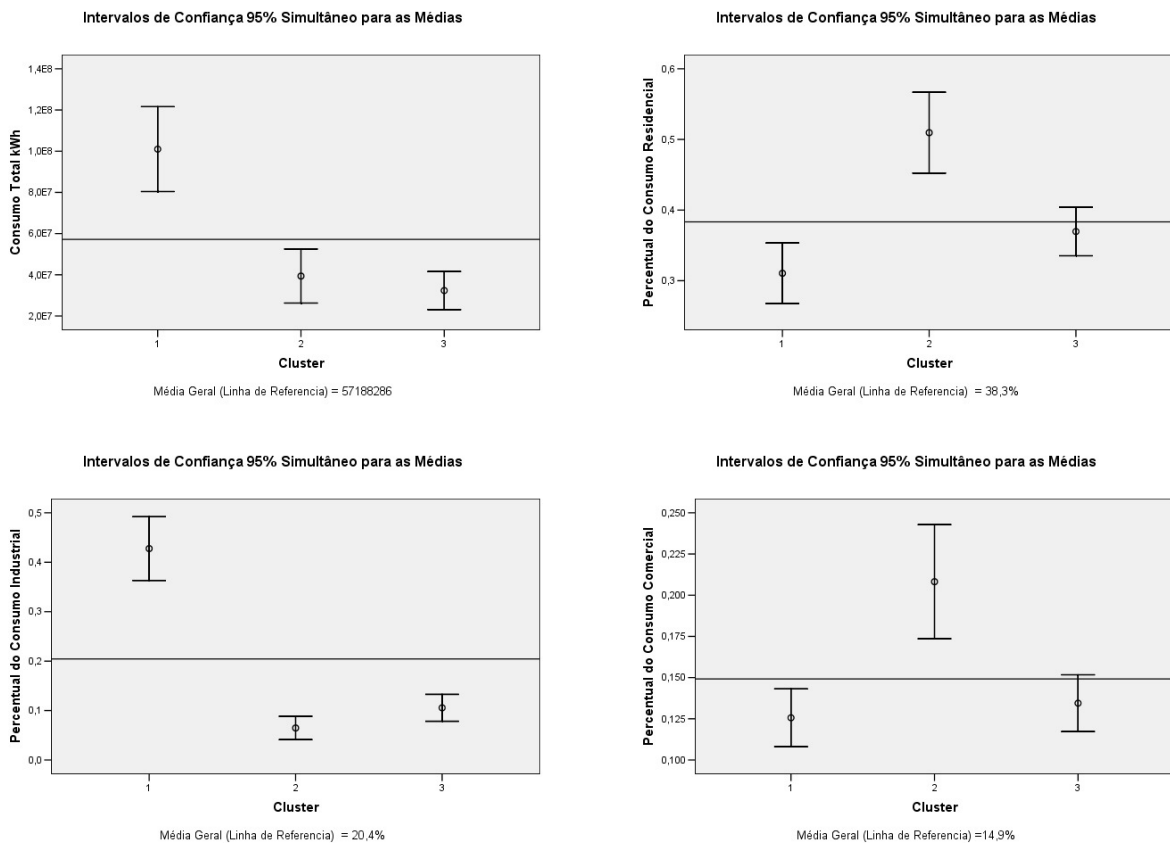
2.2.5 Resultados obtidos na Clusterização

Os 175 barramentos elétricos foram clusterizados de acordo com seis variáveis contínuas (percentual das classes residencial, industrial, comercial, rural, demais classes e consumo total) e uma variável categórica (região). A distância utilizada foi a distância por log-verossimilhança e o critério utilizado foi o BIC para a escolha do número de clusters. Desse modo, verifica-se na tabela 1 que três clusters foram obtidos.

Tabela 1: Distribuição dos clusters

		Número de Barramentos	% do Total
Cluster	1	59	33,7%
	2	42	24,0%
	3	74	42,3%
Total		175	100,0%

A figura 3 mostra os intervalos de confiança simultâneos para as médias das variáveis contínuas dos três clusters. As linhas horizontais correspondem às médias gerais. Observando a figura 1, vemos que os barramentos elétricos do cluster 1 possuem valores maiores para o consumo total e para o percentual industrial do consumo. Já os barramentos do cluster 2 possuem um percentual do consumo residencial e industrial acima da média. Finalmente, os barramentos do cluster 3 possuem uma alta participação do consumo rural.



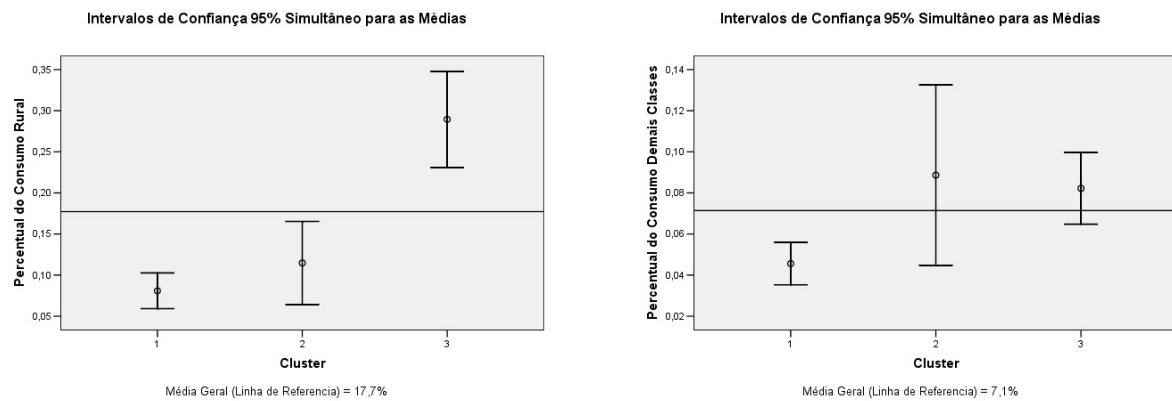


Figura 3: Intervalos de Confiança Simultâneos para as Médias das Variáveis Contínuas.

Com o intuito de mensurar a importância da variável categórica região nos clusters, teste qui-quadrados foram realizados, em que a distribuição observada do cluster é comparada à distribuição esperada de todos os barramentos. Valores grandes da estatística indicam que a distribuição da região do cluster difere da distribuição de todos os barramentos. A figura 4 mostra que a variável região foi importante para os três clusters, pois o valor da estatística qui-quadrado supera a linha do valor crítico. Esta variável é muito importante para a projeção de demanda máxima, pois locais próximos uns aos outros possuem crescimento semelhante de demanda.

Para as variáveis contínuas, a importância de cada variável no cluster é medida através de testes t, em que é comparada a média do cluster com a média geral. Quanto maior for a estatística t (em módulo) mais importante é a variável para o cluster. Quando a estatística t supera o valor crítico, ela foi importante para a classificação. Valores positivos dela indicam que a variável está acima da média geral, enquanto que valores negativos da estatística são indicativos que a variável está abaixo da média.

A figura 5 mostra o gráfico das estatísticas t. Pode-se observar que, no cluster 1, a variável mais importante para o cluster foi o baixo percentual do consumo rural com relação a média geral. Também podemos notar baixo percentual do consumo das demais classes e um alto percentual do consumo industrial e do consumo total. Neste cluster estão os maiores barramentos, que geralmente atendem municípios bastante industrializados.

No cluster 2, nota-se que os barramentos elétricos pertencentes a este cluster possuem um percentual do consumo industrial abaixo da média, enquanto que possuem um percentual do consumo residencial e comercial acima da média. Neste cluster estão muitos barramentos que atendem municípios do litoral.

No cluster 3, observa-se que há barramentos com baixo percentual de consumo industrial e baixo consumo total. Também nota-se que há grande participação do consumo rural. Neste cluster estão os barramentos menores, que atendem, basicamente, municípios rurais.

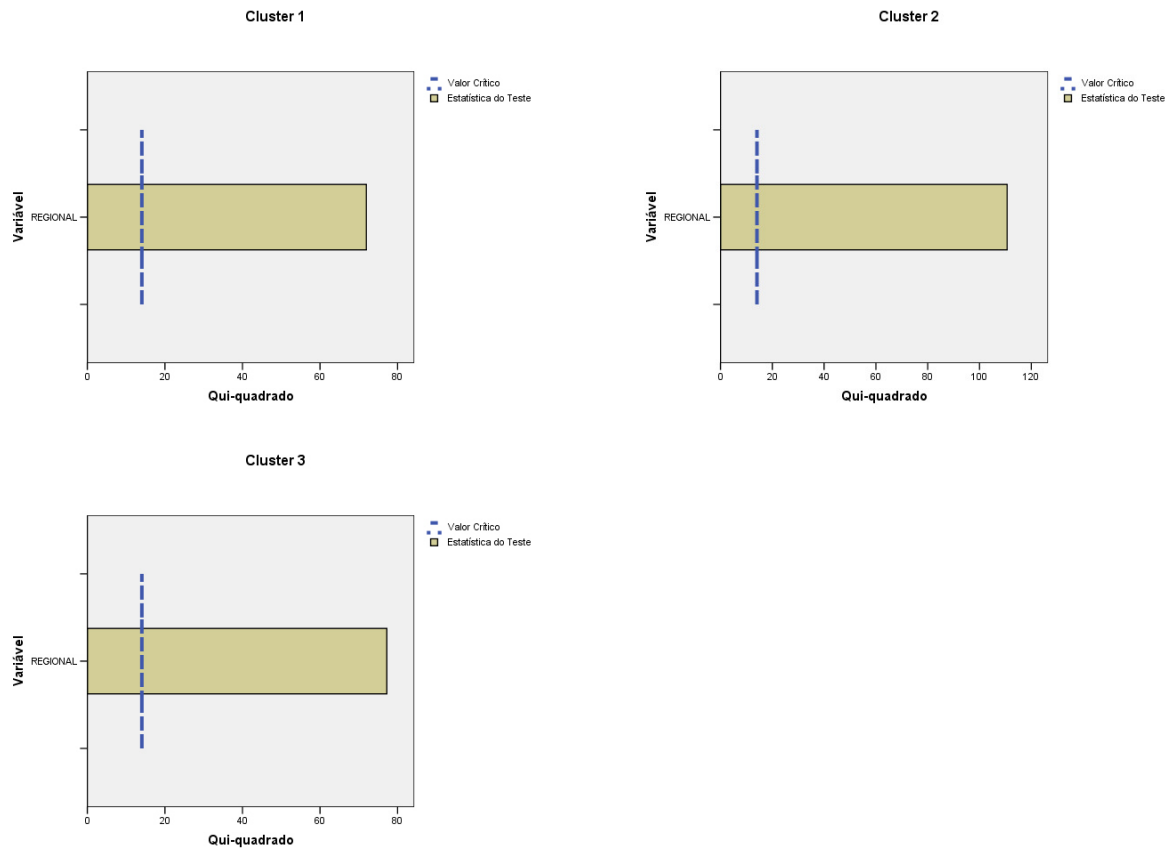


Figura 4: Importância da Variável Categórica Região por Cluster.

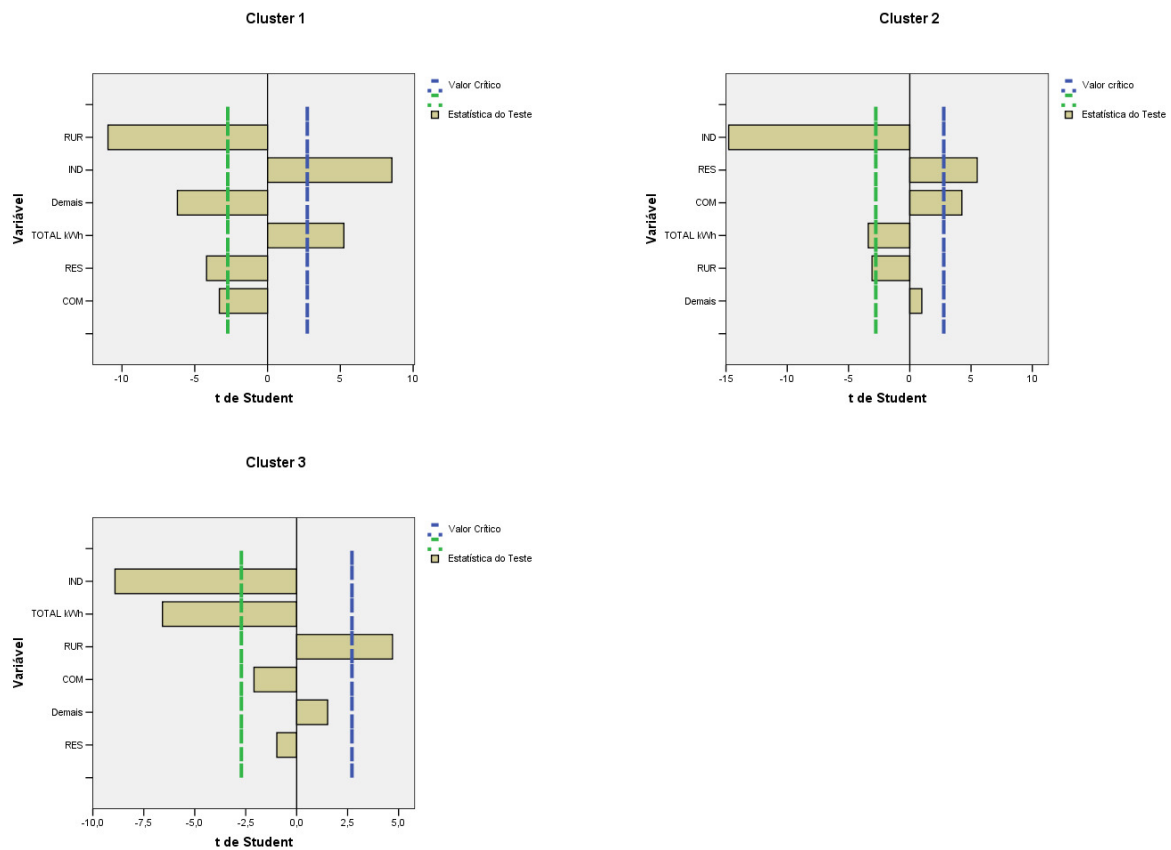


Figura 5: Importância das Variáveis Contínuas por Cluster.

2.3 Projeção de Demanda Máxima Anual para Barramentos com Medição Permanente através da metodologia ADL-SARIMA

Após a clusterização de todos os barramentos com e sem medição, foram elaborados modelos de séries temporais ADL-SARIMA para os 97 barramentos com medição. O histórico utilizado para a realização das projeções foram as medições de fronteira mensais por barramento de abril de 2005 a abril de 2009.

O modelo ADL-SARIMA consiste na realização de uma regressão linear múltipla, com a variável resposta sendo explicada por covariáveis diversas, defasadas ou não. Também é permitida a inserção da própria variável resposta defasada como covariável. Após a realização da regressão, os resíduos são modelados através do modelo SARIMA (Box & Jenkins) com a finalidade de que não se tenha mais autocorrelação entre eles. Assim, o modelo pode ser escrito da seguinte forma:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + u_t$$
$$\rho(B)u_t = \theta(B)\varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$
(7)

onde $\rho(B)$ e $\theta(B)$ são os polinômios de defasagem dos modelos ARMA.

Por exemplo, para o barramento de Artur Nogueira, o seguinte modelo foi obtido:

$$D_t = 0,173 D_{t-1} + 0,625 D_{t-12} + 24,501 \text{ PIB}_t + 73,924 \text{ Temperatura}_t + u_t$$
$$u_t = (1 - 0,836B)\varepsilon_t$$
(8)

onde D_t é a demanda máxima máxima de potência do mês t . Os modelos foram realizados no software Eviews 6.

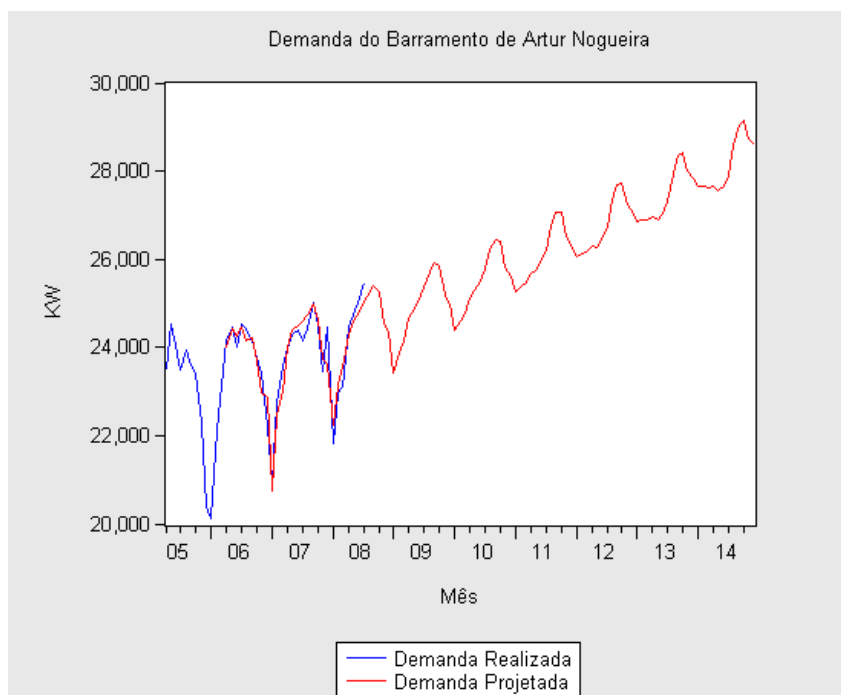


Figura 6: Demanda Projetada do Barramento de Artur Nogueira.

2.4 Projeção da Demanda Máxima dos Barramentos sem Medição por Similaridade com os Barramentos com Medição

Com todos os barramentos divididos em clusters e com os barramentos com medição projetados, é possível inferir a taxa de crescimento dos barramentos que não possuem medição através da similaridade dos barramentos pertencentes ao mesmo cluster.

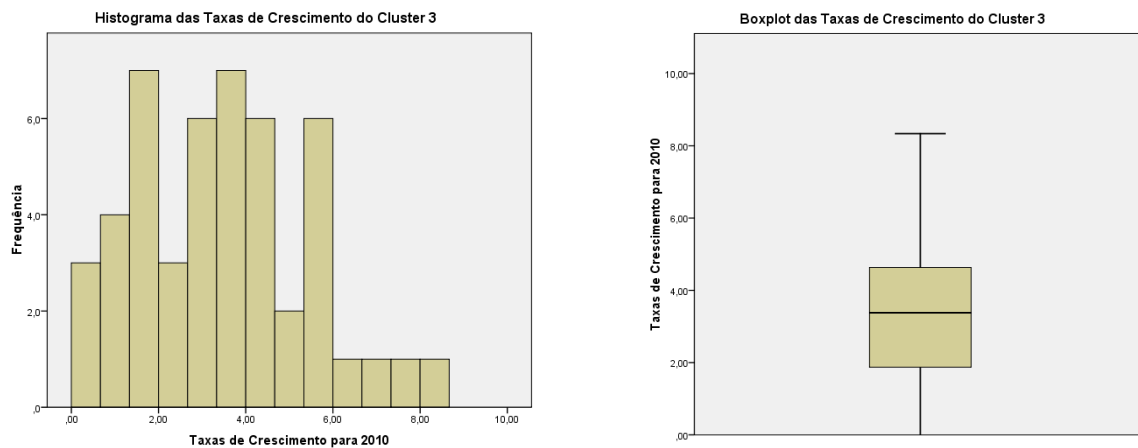


Figura 7: Histograma e Box-plot das Taxas de Crescimento do Cluster 3 para o ano de 2010.

Para cada um dos três clusters, foram observadas as distribuições das taxas de crescimento anuais dos barramentos com medição e, assim, foi possível retirar todos os crescimentos discrepantes. Em seguida, foi tomada a taxa média de crescimento dos barramentos com medição para cada cluster e essas taxas médias foram utilizadas para projetar a demanda máxima dos barramentos sem medição de modo que os barramentos sem medição do mesmo cluster tivessem a taxa média de crescimento do próprio cluster. Assim, as taxas de crescimento da demanda máxima dos barramentos sem medição foram inferidas.

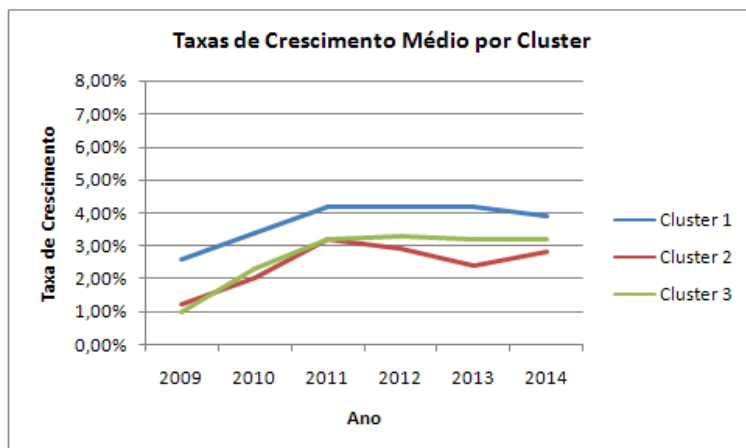


Figura 8: Taxas de Crescimento Médio por Cluster

3. Conclusões

A metodologia TwoStep cluster se mostrou bastante efetiva para agrupar os barramentos com a finalidade de se fazer projeções da demanda máxima de potência para os barramentos sem medição. Duas características da metodologia contribuíram significativamente para esta boa classificação:

- **Utilização de variável categórica região no processo de clusterização.** A metodologia permitiu o uso da medida de distância por log-verossimilhança e assim foi possível o uso de variáveis categóricas. Esta variável fez com que facilitasse a clusterização dos barramentos semelhantes que estão próximos geograficamente.
- **Utilização da autoclusterização.** Algumas vezes, é muito difícil determinar um número razoável de clusters para o problema em questão, e a autoclusterização vem para auxiliar na solução deste problema.

4. Referências Bibliográficas

CHIU, T., FANG, D., CHEN, J., WANG, Y., e JERIS, C. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, p. 263, 2001

ENDERS, W. *Applied Econometrics Time Series*. Wiley Series in Probability, 2004.

JOHNSON, R., WICHERN, D. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey, 1998.

SEIXAS, J., MENEZES, M., CORREA, C. Categorização, por Classes Predominantes, das subestações de MT da CEMIG, utilizando o método das k-Médias, Sendi, 2008.

WEI, W. S. *Time series analysis*. Addison-Wesley Publishing Company Inc., New York, p. 478,1990.

ZHANG, T., RAMAKRISHNON, R. e LIVNY, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. *Proceedings of the ACM SIGMOD Conference on Management of Data*, Montreal, p. 103-114, 1996.

Software Estatístico Utilizado para a Clusterização – SPSS versão 16.