

Clusterização de Alimentadores para Proposição de Obras Utilizando o Método de Ward

João Castilho Neto, Jonatas Pulz, Dário Takahata,
Tiago Milagres Miranda, Alden Uehara Antunes,
André Meffe

Daimon Engenharia e Sistemas
joao.castilho@daimon.com.br, jona-tas.pulz@daimon.com.br, dario@daimon.com.br, tiago.milagres@daimon.com.br, alden@daimon.com.br,
andre.meffe@daimon.com.br

Gustavo Paiva Guedes, Marcelo Renato de C. Paes
Junior, Ademálio de Assis Cordeiro

Grupo Energisa
gustavo.guedes@energisa.com.br, mcerqueira@energisa.com.br, ademalio@energisa.com.br

Resumo — O projeto apresentado neste artigo descreve a utilização do algoritmo de clusterização denominado Método de Ward aplicado ao estudo de alimentadores de média tensão. O intuito é separar os circuitos de forma que aqueles que estiverem presentes no mesmo agrupamento possuam características semelhantes. Assim, é facilitada a tomada de decisão na proposição de obras com o objetivo de melhorar os indicadores de continuidade da rede em estudo, impactando diretamente na qualidade do serviço oferecido. Além disso, a empresa também contará com retornos financeiros devido aos incentivos dados pela ANEEL e à redução das compensações por violação desses limites. Uma simulação feita com dados reais de alimentadores é apresentada para mostrar os resultados que podem ser obtidos pela metodologia.

Palavras-chaves — Clusterização, diagnóstico da rede, indicadores de continuidade, Método de Ward, proposição de obras.

I. INTRODUÇÃO

O estudo apresentado neste artigo faz parte do programa de Pesquisa e Desenvolvimento (P&D) da ANEEL e está sendo desenvolvido em conjunto pelo Grupo Energisa e a Daimon Engenharia e Sistemas. As análises estão sendo feitas, mais especificamente, nos alimentadores (circuitos) presentes na distribuidora Energisa-EBO, na região de Borborema, localizada no estado da Paraíba, com 208.415 clientes em novembro de 2016.

O problema central abordado se faz presente em todas as distribuidoras do País: melhorar os índices de continuidade estabelecidos pela ANEEL – DEC e FEC, descritos no Módulo 8 do PRODIST [1]. Esta medida garante às distribuidoras, benefícios econômicos, além de fornecerem um serviço de maior qualidade a seus clientes. Tais benefícios econômicos são previsto pela própria ANEEL de forma a incentivar as empresas a tomarem medidas adicionais para a melhoria sistemática dos níveis de continuidade.

Objetivamente, a melhoria desses índices impacta na porção técnica da componente Q (qualidade do serviço) do Fator X, que representa a eficiência da distribuidora, afetando diretamente o respectivo reajuste tarifário anual [2], [3]. Além disso, as compensações por violação dos limites de continuidade dos indicadores DIC, FIC, DMIC e DICRI serão reduzidas [1].

No cenário atual, as abordagens comumente empregadas pelas distribuidoras para a redução dos indicadores de continuidade são realizações de grandes investimentos no sistema, alterando os padrões construtivos, como, por exemplo, passando de redes aéreas com cabos nus para cabos cobertos / protegidos ou redes subterrâneas. Complementarmente, aumenta-se toda a estrutura de atendimento, com mais equipes e recursos. Essas opções, entretanto, representam aumentos substanciais nas despesas de investimento e de operação, resultado conflitante com o atual modelo regulatório brasileiro, pautado na busca incessante da modicidade tarifária.

Dessa forma, a utilização de recursos de automação, como chaves telecomandadas e com recursos de redes inteligentes é uma via que possibilita uma solução de compromisso. Ou seja, essa alternativa pode trazer ganhos substanciais aos índices de qualidade com custos mais aderentes à realidade do setor elétrico. Sendo assim, este projeto propõe uma ferramenta computacional para automação e reformulação inteligente da rede de distribuição para enfrentar o desafio da melhoria dos indicadores de continuidade.

Quatro diretrizes metodológicas foram propostas para atingir os objetivos estabelecidos no projeto. São elas:

- Módulo de Diagnóstico: nesta etapa, será feita uma avaliação dos circuitos e das redes de média tensão que os compõem utilizando insumos variados, como características físicas da rede, tipo de área atendida, DEC, FEC etc.;
- Módulo de Proposição de Obras: a partir do resultado do módulo anterior, permite a proposição automática e manual de novas obras, como construção de interligação ou trecho de alimentador, instalação de religador ou chave remota, recondutoramento, manutenções preventivas etc.;
- Módulo de Priorização: estabelece, a partir alternativas definidas na etapa anterior, aquelas que irão para a etapa final do estudo; e
- Módulo de Previsão dos Indicadores: realiza a previsão dos indicadores com ou sem a implementação das obras prioritárias.

Este artigo tem o objetivo de descrever o primeiro módulo do projeto. Nele, como dito anteriormente, serão avaliadas diversas informações disponíveis de alimentadores da rede de forma a separá-los em conjuntos. Cada agrupamento conterá

alimentadores com características semelhantes, facilitando as tomadas de decisão para intervenções de melhoria nos circuitos em questão. Da mesma forma, trechos com desempenhos satisfatórios não precisarão de alterações.

A separação dos alimentadores será feita por um algoritmo de clusterização denominado Método de Ward [4]. Nesse procedimento, as observações são agrupadas em *clusters* de forma que uma função objetivo pré-determinada seja minimizada.

Na próxima seção, será explicado, de forma mais detalhada, o funcionamento desse algoritmo. Em seguida, na seção III será descrita a metodologia do módulo de diagnóstico e como ele atinge os objetivos previstos. Na seção IV, os resultados obtidos pela metodologia aplicada a redes de média tensão reais serão apresentados. Por fim, a última seção encerra o artigo com conclusões e comentários finais.

II. CLUSTERIZAÇÃO E O MÉTODO DE WARD

A clusterização é uma técnica de análise de dados que consiste no agrupamento automático de dados / informações segundo algum critério de semelhança, visando dividir um conjunto de observações em grupos homogêneos. Assim, os elementos pertencentes a um mesmo grupo (*cluster*) serão similares entre si, com respeito às suas características de interesse medidas, enquanto que os pertencentes a grupos distintos deverão ser diferentes entre si em relação às mesmas características [5], [6].

Existem inúmeros algoritmos de clusterização disponíveis na literatura [7]-[9]. Os métodos hierárquicos envolvem a construção de uma hierarquia utilizando uma estrutura do tipo árvore. Podem ainda ser classificados de acordo com a abordagem empregada: aglomerativa (*bottom-up*) ou divisiva (*top-down*). A principal característica dos métodos hierárquicos é que os resultados de um estágio anterior são sempre aninhados com os resultados de um estágio posterior, justificando a estrutura em árvore. Um algoritmo clássico dessa metodologia é o Método de Ward [7], [9].

Desenvolvido por Joe Ward e publicado pela primeira vez em 1963, esse algoritmo tem o objetivo de realizar a clusterização de um grande número de elementos e possui uma abordagem aglomerativa. Dadas n amostras, esse procedimento permite a criação de $n - 1$ agrupamentos. Para isso, são feitas comparações entre todos os $n * (n - 1) / 2$ pares possíveis, verificando-se qual deles resultará no menor aumento do valor da função objetivo utilizada, que possui o objetivo de medir a similaridade entre as amostras (quanto menor o valor, maior a similaridade das amostras presentes no *cluster*). A função objetivo pode ser, por exemplo, a variância entre as amostras de um *cluster*, sendo que o *cluster* escolhido deve ser aquele cujas amostras tenham menor variância entre si [4].

Dessa forma, repete-se sucessivamente o processo até que se reste apenas um *cluster*. Forma-se, assim, uma estrutura hierárquica completa [4].

A estrutura em árvore é facilmente representada por um dendrograma. A Fig. 1, a seguir, mostra um conjunto de amostras sendo agrupadas em *clusters*, de forma a minimizar a variância entre os elementos de um mesmo *cluster*. A cada passo, tem-se um grupo a menos e o valor da função objetivo aumenta. Na primeira etapa do agrupamento, as amostras 1 e 9 são unidas. Em seguida, as amostras 4 e 5, 3 e 8 e depois 2

e 7. Posteriormente, o *cluster* formado pela junção das amostras 1 e 9 e das amostras 3 e 8 são agrupados formando outro *cluster*. Esse processo continua até que, por fim, a amostra 6 se junta ao outro *cluster* existente, compondo um único *cluster* final. Observa-se que os *clusters* não contêm, necessariamente, o mesmo número de amostras.

Ao término da sequência, fica a critério do investigador definir a quantidade de clusters mais apropriada para a sua aplicação. Esse número pode ser escolhido, por exemplo, de acordo com um valor de tolerância máximo do resultado da função objetivo utilizada ou pode simplesmente ser um número pré-determinado. No exemplo da Fig. 1, três clusters foram escolhidos: o primeiro contendo as amostras 1, 2, 3, 7, 8 e 9; o segundo contendo as amostras 4 e 5; e o terceiro formado apenas pela amostra 6.

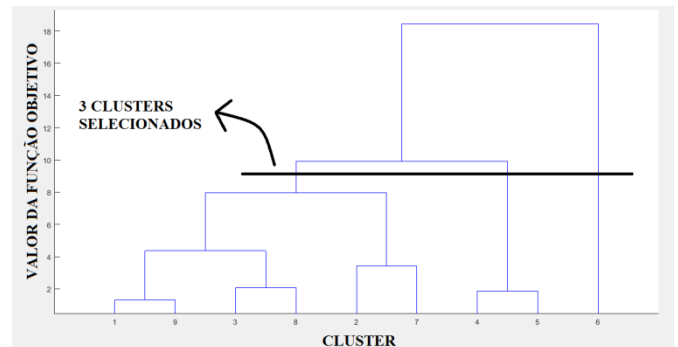


Fig. 1. Dendrograma apresentando a estrutura em árvore da clusterização feita pelo Método de Ward.

III. METODOLOGIA

O primeiro passo da metodologia consiste em definir quais variáveis serão utilizadas na avaliação dos alimentadores em estudo. Elas ajudarão na escolha do tipo de intervenção necessária nos circuitos de cada *cluster* obtido ao final do processo. Os seguintes parâmetros serão analisados:

- **Comprimento:** comprimento total do alimentador;
- **Urbanização:** porcentagem de consumidores do alimentador localizados na área urbana;
- **Número de consumidores:** número de consumidores presentes no alimentador;
- **Energia por consumidor:** energia média consumida por cada cliente dentro do período de análise no alimentador estudado;
- **Número de religadores:** número de religadores presentes no alimentador;
- **FEC:** FEC do alimentador em relação a toda a empresa no período de análise, de origem interna não programada e não expurgável;
- **DEC:** DEC do alimentador em relação a toda a empresa no período de análise, de origem interna não programada e não expurgável; e
- **Compensações:** compensações, em unidade monetária, realizadas devido à violação dos limites dos indicadores de continuidade dentro do período de análise.

Após a obtenção desses dados para cada alimentador, é necessário normalizá-los para que fiquem na mesma ordem de grandeza e que uma variável não se sobressaia à outra.

Dessa forma, é possível aplicar o Método de Ward, agrupando os alimentadores em função das diversas características estudadas (mencionadas anteriormente). A formação

desses grupos é muito importante, pois as características típicas de cada grupo serão a base para os estudos subsequentes, em que serão feitas escolhas sobre as intervenções a serem realizadas em cada grupo. Por exemplo, circuitos compostos por redes rurais e isoladas (sem possibilidade de manobras com alimentadores vizinhos) deverão ter um tratamento distinto do de circuitos em áreas urbanas compostos por alimentadores com maior flexibilidade operativa. Da mesma forma, redes extensas são bem distintas de redes mais curtas; circuitos próximos de suas metas devem ser separados de circuitos com desempenho muito distante de suas metas; e assim sucessivamente. Além disso, em princípio, os circuitos que formarem grupos cujo desempenho já seja adequado não precisarão de intervenções.

O fluxograma da Fig. 2 mostra o caminho das informações. As características da rede, da área atendida, desempenho etc. são as entradas para o algoritmo de clusterização. Por fim, como resposta, os alimentadores são separados em *clusters* de acordo com suas características apresentadas. No exemplo da Fig. 2, tem-se evidenciadas as diferentes condições de urbanização dos alimentadores.

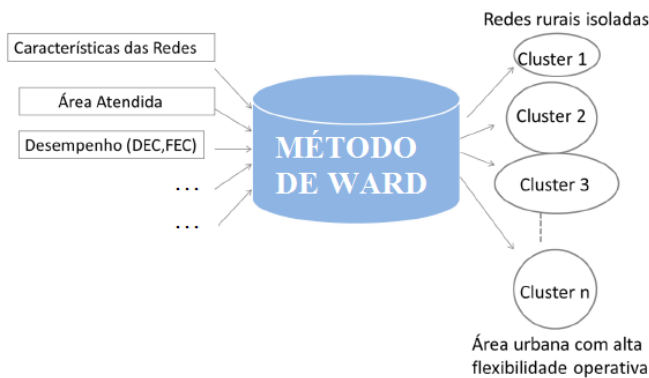


Fig. 2. Fluxograma da metodologia.

IV. RESULTADOS

A fim de mostrar os resultados capazes de serem produzidos pela metodologia apresentada, foram selecionados 26 alimentadores da distribuidora Energisa-EBO, na região de Borborema, no estado da Paraíba. A Tabela I apresenta os circuitos com as suas respectivas características.

Em seguida, os valores da Tabela I foram normalizados de forma que o maior valor de cada variável seja 1, e o menor valor possível seja 0. Assim, todas as variáveis se enquadram na mesma ordem de grandeza.

A definição do número de *clusters* a ser escolhido para separar os circuitos pode ser definida a critério do planejador, dependendo das amostras em análise. Porém, um bom valor costuma ser o maior possível de forma que a variância do último *cluster* gerado seja menor que 30% da variância de todas as amostras juntas, ou seja, do *cluster* final.

O dendrograma da Fig. 3 mostra os agrupamentos sucessivos feitos pelo Método de Ward durante a clusterização.

O *cluster* final possui variância de 10,0, então se escolhe um limite superior para a variância de 30% desse valor, que resulta em 3,0. Dessa forma, o número de *clusters* escolhido é cinco e a variância máxima obtida é de 2,6.

A Tabela II mostra os alimentadores divididos nos cinco *clusters* (A, B, C, D e E). Nela são apresentados os centroides de cada *cluster* para cada variável. O centroide nada mais é

que a média de determinada variável, considerando as amostras presentes no mesmo *cluster*.

Observando a Tabela II, e verificando os alimentadores de cada *cluster* assim como seu centroide, é possível notar algumas características marcantes em cada um deles.

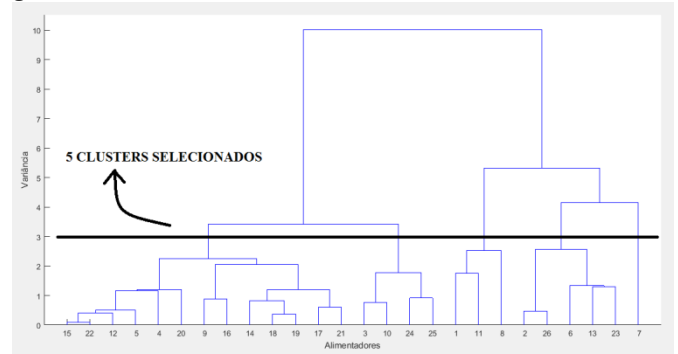


Fig. 3. Dendrograma da clusterização dos alimentadores da Energisa-EBO pelo Método de Ward.

No *cluster* A encontram-se circuitos com elevados valores de DEC e FEC, além de serem compridos. Uma intervenção apropriada seria uma obra de recondutoramento do alimentador. Dessa forma, o número de falhas diminuirá e, por consequência, do DEC, do FEC e das compensações. Além disso, esse grupo deve concentrar maiores esforços na redução dos índices de continuidade por apresentar os maiores valores de compensação da empresa.

O *cluster* B é composto por alimentadores também de grande comprimento, DEC elevado e taxa de urbanização média. Nos circuitos desse grupo, podem ser realizadas interligações entre os trechos urbanos, o que proporcionaria uma diminuição dos valores do DEC.

Em seguida, no *cluster* C, encontram-se alimentadores de porte médio, em relação aos demais circuitos da empresa, com elevado FEC e, consequentemente, DEC e compensações elevadas. Sendo assim, obras de recondutoramento a fim de reduzir o número de falhas surtirão bastante efeito nesses alimentadores.

O *cluster* D, por sua vez, é formado por alimentadores pequenos e altamente urbanos. Proporcionar interligações de socorro, tanto interno quanto externo, auxiliará na redução do DEC.

Por fim, o *cluster* E possui clientes de grande consumo de energia, representando risco de elevadas compensações em uma eventual violação dos limites dos índices de continuidade. Então, um investimento em manutenção preventiva é a intervenção mais indicada, ou eventualmente, circuitos de socorro exclusivo.

Dessa forma, em casos de restrições orçamentárias, o *cluster* A deve ser priorizado no momento de proposição de obras, pois apresenta grande quantidade de clientes e elevados valores de compensações.

Outra simulação foi feita, agora para 294 alimentadores da distribuidora Energisa-PB, da Paraíba. O dendrograma da clusterização é apresentado na Fig. 4.

Mais uma vez utilizou-se o critério dos 30%. A variância do conjunto final foi de 83,8, logo a variância máxima permitida é de 25,1. Portanto, foram escolhidos quatro *clusters*, sendo que a variância máxima obtida foi de 21,9.

Esse último teste confirma que o algoritmo é robusto, podendo ser utilizado para um grande número de observações (alimentadores).

TABELA I. ALIMENTADORES DA ENERGISA-EBO E SUAS CARACTERÍSTICAS.

Amostra	Alimenta- dor	Com- primento (km)	Urbaniza- ção (%)	Número de consumido- res	Energia por consumidor (MWh/mês)	Número de religadores	DEC (horas)	FEC (inter- rupções)	Compensações (R\$)
1	ABR-V1	120,839	71,4751	7986	0,129042	2	0,548702	0,28295	9269,83
2	ABR-V2	90,5294	45,6646	3679	0,199784	1	0,162444	0,092782	912,16
3	ABR-V3	49,9407	94,5729	10595	0,129957	3	0,221516	0,137197	1820,06
4	AER-P1	7,93756	100	3537	0,344157	0	0,021845	0,017075	106,98
5	CGU-L4	11,4742	99,9763	4226	0,279627	3	0,05495	0,027253	1264,95
6	ABR-V6	159,885	72,63	5232	0,100339	1	0,156357	0,051651	3156,2
7	AER-P1	2,66157	100	22	60,355	0	0,000146	0,000102	0
8	BVT-M1	40,1304	99,989	18250	0,121718	8	0,327975	0,214354	4494,29
9	BVT-M3	18,2618	99,9856	6928	0,203933	6	0,052891	0,0327	316,21
10	BVT-M3	16,7742	99,9755	12249	0,151236	2	0,218065	0,196254	2797,51
11	CGU-L1	243,806	86,6215	16078	0,167367	5	0,462306	0,400043	2907,65
12	BVT-M5	9,36299	100	3628	0,401448	2	0,02811	0,014627	572,35
13	CGU-L1	200,736	88,2582	7435	0,189937	3	0,26989	0,10276	5370,79
14	CGU-L2	12,6722	100	5052	0,387839	2	0,102928	0,080142	931,31
15	CGU-L3	11,4827	99,9594	4924	0,366657	3	0,015955	0,010909	4,86
16	CGU-L4	10,6358	100	7215	0,265145	6	0,152841	0,144867	2071,5
17	CGU-L5	15,7729	99,8845	9525	0,109298	1	0,177385	0,108072	1370,52
18	CGU-L3	11,9289	99,9459	7395	0,177975	3	0,084785	0,052887	3371,31
19	CTL-C2	12,5691	99,9792	4809	0,254488	3	0,091136	0,123289	3521,99
20	CGU-L2	14,6256	99,4102	1526	0,977887	4	0,012317	0,013111	3695,15
21	BVT-M2	21,737	99,3025	5448	0,257011	0	0,121525	0,133456	1966,33
22	CTL-C6	9,77238	99,9534	4291	0,381792	3	0,010168	0,006332	210,17
23	CGD-Y2	249,51	72,0641	7238	0,090829	1	0,18768	0,07589	1975,55
24	QMD-Z2	128,427	81,4615	5734	0,084761	1	0,122892	0,111288	889,23
25	QMD-Z4	109,758	87,7491	8530	0,153965	3	0,207475	0,113424	1217,85
26	QMD-Z4	102,831	45,6159	1859	0,167863	0	0,183141	0,067463	2338,68

TABELA II. CLUSTERS COM SEUS RESPECTIVOS ALIMENTADORES E CENTROIDES.

Cluster	Centroides								
	Alimentadores (número da amostra)	Comprimento (km)	Urbanização (%)	Número de consumidores	Energia por consu- midor (MWh/mês)	Número de religadores	DEC (horas)	FEC (inter- rupções)	Compensa- ções (R\$)
A	1, 8 e 11	134,925	86,03	14104,67	0,139376	5	0,446328	0,299116	5557,26
B	2, 6, 13, 23 e 26	160,698	64,85	5088,60	0,149750	1,2	0,191902	0,078109	2750,68
C	3, 10, 24 e 25	76,225	90,94	9277,00	0,129980	2,25	0,192487	0,139541	1681,16
D	4, 5, 9, 12, 14, 15, 16, 17, 18, 19, 20, 21 e 22	12,941	99,88	5269,54	0,339020	2,77	0,071295	0,058825	1492,59
E	7	2,662	100,00	22,00	60,355000	0	0,000146	0,000102	0

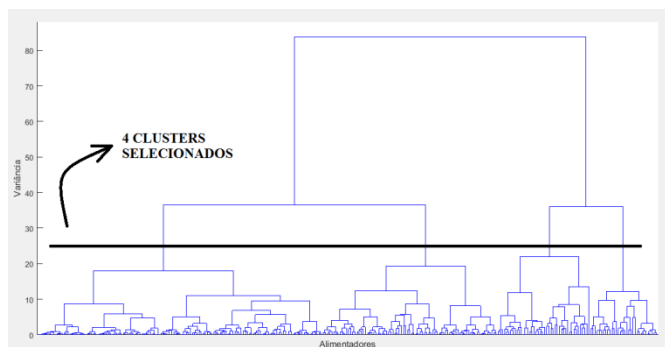


Fig. 4. Dendrograma da clusterização dos alimentadores da Energisa-PB pelo Método de Ward.

V. CONCLUSÕES

Este artigo apresentou uma nova metodologia para análise de intervenções (obras) a serem realizadas em alimentadores de média tensão.

As melhorias sugeridas têm o intuito de reduzir os valores dos indicadores de qualidade DEC e FEC, gerando retorno financeiro para as distribuidoras, pois as intervenções propostas possuem características que melhoram a base de remuneração da empresa, incorporando-se novos ativos à rede de distribuição. A questão da melhora da alocação de despesas também é um produto que o *software*, após a conclusão de seus quatro módulos, poderia sugerir (incremento de plano de podas e melhoria das inspeções visuais e termográficas).

Entretanto, será dada ênfase à alocação otimizada de investimentos para a melhoria dos indicadores de continuidade.

Dessa forma, as propostas indicadas prezam pelo recondu-toramento e pela utilização de recursos de automação, como chaves telecomandadas e com recursos de redes inteligentes.

Para facilitar a tomada de decisão das obras a serem pro-postas, a metodologia utiliza o conceito de clusterização. Nele, um conjunto de amostras é agrupado em diferentes subconjuntos (*clusters*) de forma que as observações presen-tes no mesmo grupo sejam semelhantes entre si segundo algum critério. Da mesma forma, as observações que estive-rem em *clusters* diferentes serão distintas, segundo o mesmo critério. O Método de Ward é o algoritmo de clusterização utilizado na metodologia.

A fim de observar os resultados capazes de serem obtidos pelo projeto desenvolvido, sua metodologia foi implementada para 26 alimentadores da distribuidora Energisa-EBO. Após a clusterização, foram obtidos cinco *clusters*, cada um com suas características marcantes. Assim, foi possível selecionar, de forma mais fácil e eficaz, diferentes obras as serem reali-zadas em cada um deles com o objetivo de se diminuir os indicadores de continuidade dos circuitos, de acordo com suas principais características.

Uma segunda simulação foi feita, utilizando 294 alimenta-dores da Energisa-PB, reduzindo-os a quatro clusters. O teste confirma a robustez do algoritmo, que pode ser utilizado para uma grande quantidade de circuitos.

Além de auxiliar na proposição de obras em si, o projeto apresentado também é importante para a orientação do usuá-rio sobre quais alimentadores terão prioridade para receber intervenções perante uma grande quantidade de possibilida-des, visando à economia de recursos.

VI. REFERÊNCIAS

- [1] *Procedimento de Distribuição de Energia Elétrica no Sistema Elétrico Nacional, Módulo 8 - Qualidade da energia elétrica*. ANEEL-PRODIST Std., pp. 54-73, 2017.
- [2] ANEEL. “Entendendo a Tarifa.” Internet: http://www.aneel.gov.br/entendendo-a-tarifa/-/asset_publisher/uQ5pCGhnyj0y/content/parcela-b/654800?inheritRedirect=false, 12 de Março de 2016 [20/02/2017].
- [3] *Procedimento de Regulação Tarifária, Módulo 2: Revisão Tarifária de Concessionárias de Distribuição de Energia Elétrica*, Submódulo 2.5 (Revisão 2.0) – Fator X. ANEEL, 2015.
- [4] Joe H. Ward, Jr. Hierarchical Grouping to Optimize an Objective Func-tion, *Journal of the American Statistical Association*, Vol. 58, No. 301, Mar. 1963, pp. 236-244.
- [5] Mingoti, S. A. *Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada*. Editora UFMG, 2005.
- [6] Sharma, S. *Applied Multivariate Techniques*. New York: John Wiley & Sons, 1996.
- [7] Queiroz, L. M. O. – Tese de Doutorado: Estimacão e Análise das Perdas Técnicas na Distribuição de Energia Elétrica - Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e de Computação, Campi-nas, 2010.
- [8] Xu, R.; Wuncsh II, D. C. *Clustering*. Wiley, 2009.
- [9] Hair, J. F.; Anderson, R. E.; Tathan, R. L.; Black, W. C. *Análise Multi-variada de Dados*. 5ª Ed. Porto Alegre: Bookman, 2005.