



A ANÁLISE ENVOLTÓRIA DE DADOS COMO FERRAMENTA DE AVALIAÇÃO DE EFICIÊNCIA DE DISTRIBUIDORAS DE ENERGIA ELÉTRICA: UMA ABORDAGEM VIA ANÁLISE DE *CLUSTER*

Salvador Pereira de Carvalho Junior
Universidade Federal de Ouro Preto (UFOP)
Departamento de Engenharia de Produção
salvadorjuniorcp@gmail.com

Lásara Fabrícia Rodrigues
Universidade Federal de Ouro Preto (UFOP)
Departamento de Engenharia de Produção
lasara@em.ufop.br

RESUMO

Dentre as diversas mudanças ocorridas no setor elétrico brasileiro na última década, está uma nova abordagem para o cálculo de eficiência das concessionárias de distribuição de energia elétrica. Nesse sentido, este estudo apresenta uma abordagem baseada em Análise de *Cluster* e Análise Envoltória de Dados (DEA) para o cálculo de eficiência das 61 concessionárias de distribuição no Brasil. Essas concessionárias são segmentadas em 3 *clusters* de acordo com seu grau de similaridade utilizando o método hierárquico de clusterização. Os resultados mostram que o *cluster 2*, representado pelas empresas de médio porte, apresentaram os melhores resultados frente aos demais *clusters* (*clusters 1 e 3*). Entretanto, uma parcela significativa das empresas, incluindo do *cluster 2*, apresentam índices de eficiência insatisfatórios, sugerindo que essas empresas estão com problemas quanto ao gerenciamento de seus *inputs*. Medidas corretivas no curto, médio e longo prazo, devem ser tomadas para a reversão desse quadro.

PALAVRAS CHAVE. Análise Envoltória de Dados. *Clusters*. Método Hierárquico. Setor elétrico.

Paper topics Análise Envoltória de Dados, PO na Área de Energia

ABSTRACT

Among the considerable changes that have occurred in the Brazilian electricity sector in the last decade, a new approach on measuring the efficiency of electricity distribution utilities stands out. In this sense, this study presents an approach based on Cluster Analysis and Data Envelopment Analysis (DEA) to estimate the efficiency of 61 distribution utilities in Brazil. These utilities are segmented into 3 clusters according to their degree of similarity using the hierarchical method of clustering. The results show that cluster 2, represented by medium-sized companies, presented the best results in comparison to other clusters (clusters 1 and 3). However, a significant number of companies, including from Cluster 2, have earned low efficiency scores, suggesting that these companies are having problems managing their inputs. Corrective measures in the short, medium and long term should be pursued in order to reverse this scenario.

KEYWORDS. Data Envelopment Analysis. Cluster. Hierarchical Method. Electric Sector.
Paper topics Data Envelopment Analysis, OR in Energy



1. Introdução

O setor elétrico brasileiro (SEB) passou por grandes transformações durante os mais de 100 anos de sua existência. Segundo [Barbosa et al. 2016], o SEB, até a década de 90, era sustentado, do ponto de vista institucional, pelas empresas estatais, com uma visão predominantemente monopolística. A partir da década de 90, o cenário mundial de energia elétrica passou por reformas estruturais e institucionais influenciadas pelo pensamento liberal e uma visão de estado mínimo.

A partir dessas reformas, o SEB foi subdividido em quatro segmentos: geração, transmissão, distribuição e comercialização, onde a livre concorrência deveria prevalecer cabendo ao estado o papel da regulação, nesse caso, representado pela Agência Nacional de Energia Elétrica (ANEEL).

Diferentemente do segmento de geração, a transmissão e a distribuição de energia, no Brasil, tem seus preços regulados periodicamente pela ANEEL. No primeiro ciclo de revisão tarifária, em 2003, os contratos de concessão deveriam utilizar o regime *price cap*, isto é, o regulador define uma tarifa inicial e o mecanismo de reajuste anual conforme a inflação e ganhos de produtividade [Pessanha et al. 2007]. Ainda nessa primeira revisão e complementarmente na segunda revisão tarifária em 2007, a ANEEL instituiu o chamado *benchmark* regulatório. Nessas revisões, de acordo com [Rezende et al. 2014], uma distribuidora virtual referência era criada e, a partir dessa distribuidora virtual, as prestações de serviços de distribuição de energia elétrica eram simuladas nas mesmas condições vivenciadas pelas distribuidoras reais, com a diferença de que essa simulação prevê que a empresa utiliza seus recursos de forma eficiente e, assim, estabelece um *benchmark* para as empresas ineficientes.

Na terceira revisão, a ANEEL aplicou um novo modelo para o *benchmark* regulatório, definido através de uma fronteira eficiente através da metodologia *Data Envelopment Analysis* (DEA), uma técnica não paramétrica, baseada na programação linear [Coelli et al. 2005], já usada em muitas outras agências reguladoras de outros países [Haney e Pollitt 2009; Plagnet 2006]. Nessa revisão, as 64 distribuidoras de energia elétrica eram subdivididas em dois grupos de acordo com o consumo de energia elétrica (TWh) e então o modelo de eficiência era aplicado. Entretanto, como colocado por [Rezende et al. 2014], somente essa variável não é suficiente para agrupar empresas com características similares, além de não considerar outros fatores igualmente importantes, como por exemplo, o número de consumidores, a extensão da rede e o custo operacional (OPEX) de cada distribuidora.

Para contornar esse problema buscando levar em conta as variáveis citadas acima, uma técnica multivariada de análise de *cluster* é utilizada. Essa ferramenta é empregada para que o método de comparação utilizado, DEA, seja mais aderente e que forneça uma comparação justa entre as distribuidoras. Com as distribuidoras subdivididas em seus respectivos grupos, pode-se aplicar o modelo DEA e então analisar seus índices de eficiências.

O presente estudo se encontra organizado em 5 seções incluindo esta introdução. A Seção 2 apresenta a técnica DEA e seus modelos. Na seção 3 é discutida a técnica de análise de *cluster*. Os resultados obtidos encontram-se na Seção 4. Uma conclusão encerra o artigo na Seção 5.

2. Análise Envoltória de Dados

De acordo com [Çelen 2013], [Boles 1966] foi o primeiro autor a sugerir o uso da programação linear para estimar uma fronteira eficiente convexa de unidades produtivas, chamadas de DMU's (*Decision making units*). Entretanto essa técnica ganhou popularidade após a publicação do trabalho de [Charnes et al. 1978] ficando conhecida como Análise Envoltória de Dados (DEA).

Desde então, a utilização da análise envoltória de dados para calcular a eficiência técnica em aplicações diversas tem se tornado bastante comum [Cooper et al. 2007]. De acordo



com [Souza et al. 2014], uma das razões para esse fato é que a DEA opera de maneira satisfatória com múltiplas entradas (*inputs*) e múltiplas saídas (*outputs*) sem a necessidade de informações usuais de preço de mercado.

Segundo [Azadeh 2009], a técnica DEA identifica uma fronteira eficiente de DMU's (firmas onde ocorre a transformação *input-output*) com as melhores práticas e então utiliza esses dados para calcular o índice de eficiência de DMU's menos eficientes. Nesse sentido, [Giannakis et al. 2005] complementa que uma das principais vantagens dessa técnica é que as DMU's ineficientes são comparadas com DMU's eficientes ao invés de alguma análise estatística.

A respeito da seleção das DMU's a serem analisadas, deve se certificar que as unidades são homogêneas, isto é, devem produzir os mesmos produtos utilizando os mesmos insumos. Com relação ao número de DMU's a serem inseridas na análise, uma regra aproximada define que o número de DMU's n a serem escolhidos deve ser maior ou igual a $\max [i \times j, 3 \times (i + j)]$, onde i refere-se ao número de *inputs* e j ao número de *outputs* [Cooper et al. 2007].

2.1 Modelos DEA

Como mencionado anteriormente, o primeiro modelo DEA definido na literatura, conhecido como CCR, foi desenvolvido por [Charnes et al. 1978] considerando retornos constantes de escala (CRS). O modelo CCR com orientação a *input* é mostrado na Figura 1. Posteriormente, [Banker et al. 1984] relaxaram a premissa de retornos constantes e estenderam o modelo para retornos variáveis de escala (VRS) que também ficou conhecido como modelo BCC.

Para o cálculo de eficiência, o modelo CCR considera que todas as DMU's em análise são homogêneas, isto é, quando comparadas, o modelo não leva em conta os critérios de escala e dimensões (ambientais ou estruturais) de cada DMU considerando que todas as DMU's operem em escala ótima [Moura et al. 2016]. De maneira oposta, o modelo BCC, através da flexibilização do tipo de retorno de escala, permite relacionar DMU's não-homogêneas comparando apenas DMU's que operem em escala semelhante.

Com relação a fronteira de eficiência gerada pelo modelo CCR e BCC, o primeiro exibe uma reta com um ângulo de 45 graus, como pode ser visualizado na Figura 1. Já no modelo BCC, a eficiência de uma DMU é obtida dividindo-se sua produtividade pela maior produtividade dentre as DMU's que apresentam o mesmo tipo de retorno a escala. Assim, a fronteira BCC apresenta retas de ângulos variados o que caracteriza uma fronteira linear por partes (*pice-wise linear*) [Mariano et al. 2006].

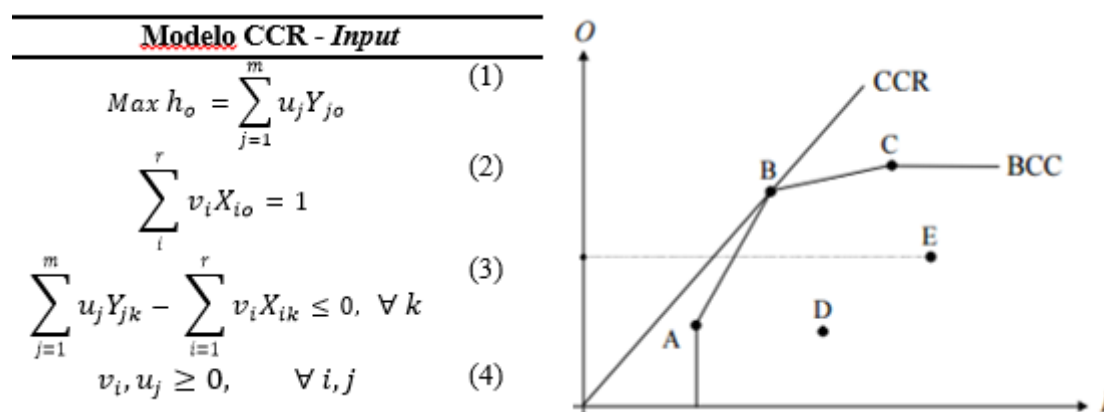


Figura 1: Modelo CCR – I (esquerda) e Fronteira de eficiência CCR e BCC (direita)

Fonte: [Charnes et al. 1978; Mello et al. 2004]

Considerando uma DMU k , de um universo de n unidades, produzindo m quantidade de produtos Y , a partir de r quantidades de insumo X . Uma unidade k qualquer produz Y_{jo} quantidades de produtos, utilizando X_{io} insumos. Segundo [Mello et al. 2005], o modelo mostrado na Figura 1 é conhecido como modelo dos multiplicadores devido a incorporação de



pesos (v_i, u_j) ao modelo. Dessa forma, cada DMU, atribui pesos a seus *inputs/outputs* da maneira que melhor lhe servir de forma que maximize sua eficiência (Equação 1). Uma DMU k é considerada eficiente se e somente se $h_o = 1$. Nesse modelo CCR, as variáveis de decisão são v_i e u_j . As restrições (2) e (3) garantem que o modelo não tenha infinitas soluções. A restrição (4) assegura a não negatividade dos pesos.

3. Análise de Cluster

Classificar algo ou alguém em grupos de acordo com seu grau de similaridade é uma tarefa comum em nosso dia-dia. [Gore 2000] ressalta que classificar é uma arte que visa tentar simplificar o mundo complexo em que vivemos, tentando organizar aspectos do nosso ambiente em unidades significantes como sexo, posições políticas, espécies, etc.

Para [Aldenderfer e Blashfield 1984], análise de *cluster* é uma denominação genérica para um grande grupo de técnicas estatísticas que podem ser utilizadas para criar uma classificação. Esses procedimentos formam empiricamente *clusters* ou grupos.

Os primeiros trabalhos na área de análise de *cluster* tem sua origem na antropologia [Driver e kroeber 1932] e na psicologia [Zubin 1938; Tryon 1939]. Segundo [Aldenderfer e Blashfield 1984], uma importante contribuição ao desenvolvimento das técnicas de *clustering* foi feita a partir do livro “*Principles of Numerical Taxonomy*” de [Sokal e Sneath 1963]. No entanto, este mesmo autor afirma que devido à complexidade matemática, esses métodos ficaram restritos por muito tempo.

Hoje em dia, em um cenário de avanço computacional, a propagação de técnicas de clusterização tem se irradiado para as mais diversas áreas de conhecimento [Dai e Kuosmanen 2014; Guimarães et al. 2013], incluindo aplicações utilizando a Análise Envolvória de Dados como método de cálculo de eficiência de distribuidoras de energia elétrica [Zanini 2004; Souza et al. 2010; Dai e Kuosmanen 2014; Andrade et al. 2014; Machado et al. 2016]. Dessa forma, há na literatura uma infinidade de métodos diferentes que se propõe a agrupar entidades de acordo com seu grau de similaridade, dentre os quais pode-se citar o método de clusterização hierárquico, que será utilizado nesse estudo.

3.1 Método hierárquico

Os Métodos Hierárquicos, indicados para pequenas amostras ($n \leq 250$), são técnicas simples onde os dados são particionados sucessivamente produzindo uma representação hierárquica dos agrupamentos [Everitt 2011].

Esse tipo de método pode ser subdividido em duas categorias: aglomerativo e divisivo. A abordagem aglomerativa inicia-se com todos os objetos (*clusters*) separados e os combina até formarem um único *cluster* no final. A segunda tem seu ponto de partida quando todos os objetos estão juntos e em cada passo subsequente, esses *clusters* vão sendo classificados até restarem n *clusters*.

Segundo metodologia adotada por [Vale 2006], no método aglomerativo, o primeiro passo é definir as medidas de proximidade para o cálculo da matriz de similaridade. Em seguida, o próximo passo é encontrar o menor valor na matriz de similaridade usando um método de ligação e então agrupa-los. Por fim, a matriz de similaridade é atualizada até restar um único *cluster*.

Medidas de proximidade são representadas pelas matrizes de similaridade e dissimilaridades. De acordo com [Everitt 2011], dois indivíduos estão “perto” quando suas dissimilaridades ou distância são pequenas ou sua similaridade é grande. Essas medidas são essenciais na organização dos objetos em *clusters*. A principal, segundo [Vale 2006], é a distância euclidiana (Equação 5). Onde x_i e y_i são, respectivamente, os valores do i -ésimo atributo para os dados x e y .



$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

Após o cálculo da matriz de similaridades entre as observações, define-se como os agrupamentos serão ligados. Os métodos mais comuns são: Ligação Simples, Ligação Completa, Centroides, Ligação Média, Ligação Mediana e método Ward.

O método Ward, que leva o nome de seu criador [Ward 1963], é fundamentado na mudança de variação entre os grupos e dentro dos grupos que estão sendo formados em cada passo do agrupamento. A Equação 6 ilustra o método Ward onde cada elemento é considerado como um único conglomerado, e em cada passo do algoritmo de agrupamento é calculado a soma de quadrados dentro de cada conglomerado.

$$W = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (6)$$

Onde i representa o agrupamento em questão, n o número total de objetos do agrupamento i e x_i o i -ésimo objeto do agrupamento i .

4. Estudo de caso

O presente estudo tem como objetivo analisar a eficiência das 61 concessionárias de distribuição de energia elétrica do Brasil no ano de 2014. Para este fim, este estudo foi subdividido em 3 etapas principais (Figura 2), a saber: Seleção das variáveis, clusterização e aplicação do modelo de eficiência. Os dados utilizados foram coletados através do portal online da Agência Nacional de Energia Elétrica [Aneel 2017a; Aneel 2017b; Aneel 2017c].

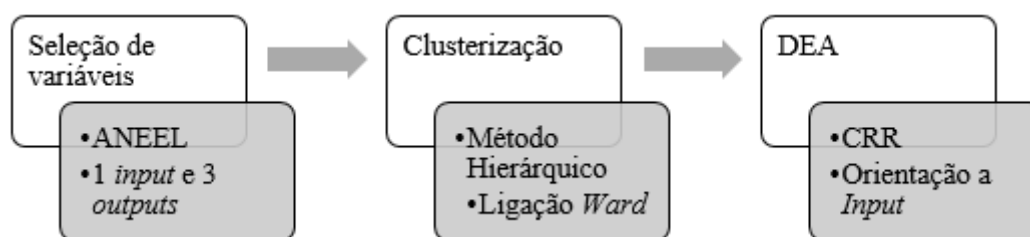


Figura 2: Procedimento metodológico

4.1 Seleção de variáveis

Segundo [Giannakis et al. 2005], pelo fato do índice de eficiência de DMU ineficiente ser obtido através de uma comparação com uma DMU eficiente, a escolha das variáveis é um processo sensível e de extrema importância para se obter índices de eficiências coerentes. Dessa forma, uma pré-seleção de candidatas a *inputs* e *outputs* se deu através de um levantamento bibliográfico objetivando encontrar as variáveis mais comumente utilizadas em análise de eficiência de distribuidoras de energia elétrica (ver, por exemplo [Souza et al. 2014; Rezende et al. 2014; Dai e Kuosmanen 2014]).

Nesta etapa foram identificadas cinco variáveis candidatas a *output*: Extensão da rede (Km), Número de consumidores, Energia distribuída (Mwh), Duração equivalente de interrupção por unidade consumidora (DEC) e Frequência equivalente de interrupção por unidade consumidora (FEC) e uma variável candidata a *input*: Custo operacional – OPEX (R\$).



Os modelos DEA admitem que o conjunto de produção seja convexo e a tecnologia seja monotônica, isto é, um aumento na quantidade de algum *input* não deve reduzir a quantidade dos *outputs* [Rezende et al. 2014]. Uma correlação positiva entre os *inputs* e *outputs* satisfaz essa condição.

A Tabela 1 apresenta a matriz de correlação entre as variáveis coletadas. As variáveis Extensão da rede, Número de consumidores, Energia distribuída apresentaram correlação positiva significativa com a variável de Custo operacional – OPEX, qualificando-as como *outputs* do modelo. De maneira oposta, as variáveis DEC e FEC são fracamente relacionadas com o OPEX.

É importante ressaltar que os indicadores de qualidade DEC e FEC são do tipo quanto menor, melhor, ou seja, quanto menor for a duração de interrupção de uma DMU, melhor a empresa. Dessa forma, essas variáveis são consideradas *outputs* indesejados [Gomes 2003; Tschaffon e Meza 2014] e por questões de limitação de escopo não serão tratadas nesse estudo.

Variável	Extensão da rede	Nº de consumidores	Energia distribuída	OPEX	DEC
Nº de consumidores	0,754				
Energia distribuída	0,609	0,950			
OPEX	0,535	0,653	0,801		
DEC	0,214	0,068	-0,036	-0,030	
FEC	0,075	-0,104	-0,112	0,087	0,910

Tabela 1: Matriz de correlação entre as variáveis

A Tabela 2 traz as estatísticas descritivas das variáveis utilizadas nesse estudo.

Variável	Média	Mínimo	Máximo	Desvio padrão
Input				
OPEX (R\$)	220.082	2.763	2.674.182	442.417
Output				
Extensão da rede (km)	41.304	137	410.220	67.024
Energia distribuída (MWh)	5.613.704	16.379	37.873.729	7.839.885
Nº de consumidores	1.249.085	3.555	8.136.003	1.735.897

Tabela 2: Estatísticas descritivas

4.2 Análise de *cluster*

Após a seleção das variáveis para a aplicação do modelo de eficiência, o próximo passo é o agrupamento das DMU que compartilham características semelhantes em *clusters*. Isso é importante, pois, apesar do ramo de atuação ser o mesmo e estarem sujeitas as mesmas regras de concessão, as concessionárias são bastante heterogêneas. Questões relacionadas a fatores geográficos e econômicos de cada região, o volume de energia gerado bem como a concentração espacial do tipo de consumidor (residencial, rural ou industrial) são alguns pontos a serem levados em conta em relação a diferenciação entre as empresas.

Na Tabela 2 é possível observar a discrepância que existe entre as concessionárias, tendo, por exemplo, distribuidoras com cerca de 3,5 mil consumidores e empresas com mais de 8 milhões de consumidores. Nesse sentido, a análise de *cluster* se mostra uma ferramenta útil para que se possa efetuar uma análise de eficiência mais justa e aderente a realidade.

As distribuidoras foram classificadas utilizando-se o Método Hierárquico, tendo o método *Ward* como método de ligação e a distância Euclidiana com medida de similaridade. Para realizar o agrupamento foram utilizadas as quatro variáveis mostradas na Tabela 2.

Três *clusters* foram criados, como pode ser observado na Figura 3. No *Cluster* 1, em verde, com 30 distribuidoras, estão predominam as empresas localizadas no nordeste, centro-



oeste e no norte do país as quais se caracterizam por uma grande distribuição espacial de rede, porém com baixo número de unidades consumidoras e pouca energia distribuída como consequência. No *Cluster 2*, em vermelho, com 20 empresas, estão principalmente empresas do sul, sudeste e nordeste que apresentam mercados espacialmente concentrados com consumo médio de energia e com um mercado relativamente grande. Por fim, o *Cluster 3*, em azul, com 11 distribuidoras representam as empresas localizadas nos grandes centros industriais, principalmente do sudeste do país.

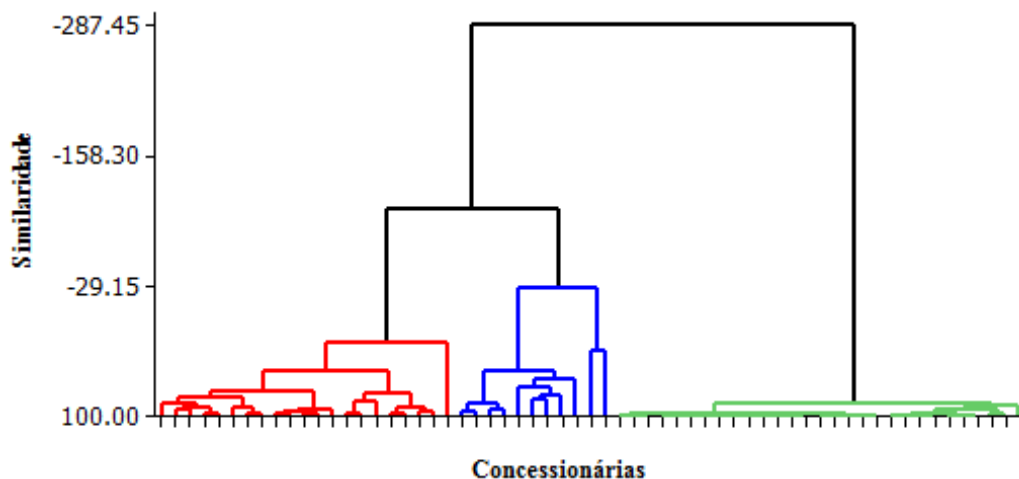


Figura 3: Dendrograma obtido pelo Método Hierárquico

4.3 Aplicação do modelo DEA

Neste estudo de caso o modelo CCR com orientação a *input* foi escolhido. Segundo [Giannakis et al. 2005], esse tipo de orientação é a mais utilizada na análise de eficiência de distribuidoras de energia elétrica. Em relação ao tipo de retorno de escala, a premissa de retornos constantes de escala foi utilizada.

4.3.1 Análise e discussão dos resultados

O Modelo DEA, CCR orientado ao *input*, foi aplicado em cada um dos três *clusters*. A Tabela 3 apresenta os resultados obtidos no *cluster 2*. Tabelas de conteúdo semelhantes foram geradas para o *cluster 1* e 3, porém, por questões de limitação de espaço, serão omitidas.

Na Tabela 3, três empresas atingiram eficiência de 100%, São elas: CPFL Piratininga, ESCELSA e ETO. Isto significa que essas empresas estão operando em sua capacidade ótima, não apresentando *slacks* em seus *inputs* e *outputs*. Qualquer DMU com pelo menos um *slack* seja no *input* ou nos *outputs* é considerada ineficiente. Dessa forma, no *Cluster 2*, três empresas são eficientes e 17 são consideradas ineficientes.

As DMUs eficientes estão na fronteira de eficiência gerada pelo modelo CCR e todas as DMU's ineficientes encontram-se abaixo dessa reta. Para que uma DMU ineficiente nesse *cluster* passe a fazer parte da fronteira de eficiência é necessário que a empresa reduza (*input*) ou aumente (*outputs*) seus *slacks* tendo em vista qual DMU eficiente é sua referência (empresa com características semelhantes).

Tomando como exemplo a concessionária Energisa Mato Grosso (EMT), a empresa possui uma eficiência de 0,92, isto é, a empresa está convertendo seus *inputs* em *outputs* em uma taxa de eficiência de 92%. Para que a empresa opere com 100% de eficiência, é preciso que haja uma redução de R\$ 18.390,1 em seu custo operacional concomitante com um aumento de 425.945,9 no total de unidades consumidoras e mantendo o atual nível de energia distribuída e extensão de rede.



Cluster 2							
Cód	DMU	Eficiência	DMU referência	Slack			
				Extensão da rede	Output N° cons.	Energia distribuída	Input OPEX
31	CEB-DIS	0,61	32 e 42	-	85.833,3	-	54.596,5
32	CPFL- Piratininga	1,00	-	-	-	-	-
33	Bandeirantes	0,98	32 e 42	10.283	-	-	-2.901,8
34	ESSE	0,41	32	6.825	-	644.667,7	73.165,4
35	CEAL	0,45	32	13.042	-	1.172.057,3	90.070,5
36	AmE	0,70	32 e 42	-	354.914,6	-	-40.583,8
37	COSERN	0,97	32	14.874	-	1.104.936,8	-2.538,5
38	AMPLA	0,88	32	53.347	-	1.426.584,2	-25.983,2
39	EPB	0,30	32	6.229	-	1.781.703,6	213.580,4
40	CEEE-D	0,30	32 e 42	-	34.672,8	-	295.940,5
41	AES-SUL	0,19	32 e 42	-	399.495,3	-	543.314,3
42	ESCELSA	1,00	-	-	-	-	-
43	CERON	0,53	32 e 45	-	148.323,7	-	-68.463,8
44	RGE	0,71	32 e 45	-	143.967,2	-	-46.084,7
45	ETO	1,00	-	-	-	-	-
46	EMS	0,50	32 e 45	-	167.554,5	-	119.258,1
47	CEPISA	0,45	32 e 45	-	-	1.642.533,8	159.682,9
48	CEMAR	0,84	32 e 45	-	-	3.930.644,5	-34.432,9
49	CELPA	0,80	32 e 45	-	-	1.512.485,2	-43.893,4
50	EMT	0,92	32 e 45	-	425.945,9	-	-18.390,1

Tabela 3: Resultados obtidos pelo modelo CCR no Cluster 2

O Cluster 2, com 3 DMU's eficientes, apresentou eficiência média de 68% (Figura 4), a maior entre os 3 clusters. O Cluster 1, apesar de ter o maior número de DMU's eficientes, 4 empresas, apresentou eficiência média de 51% indicando que um grande número de DMU's estão convertendo seus *inputs* em *outputs* de maneira insatisfatória. As empresas pertencentes a esse cluster são consideradas de pequeno porte e representam apenas 3,7% de toda a energia distribuída no país. Já o Cluster 3, com 3 DMU's eficientes, representa a maior fatia do mercado nacional (61,7% da energia distribuída) obteve eficiência média de 60%. A Tabela 4 resume essas informações.

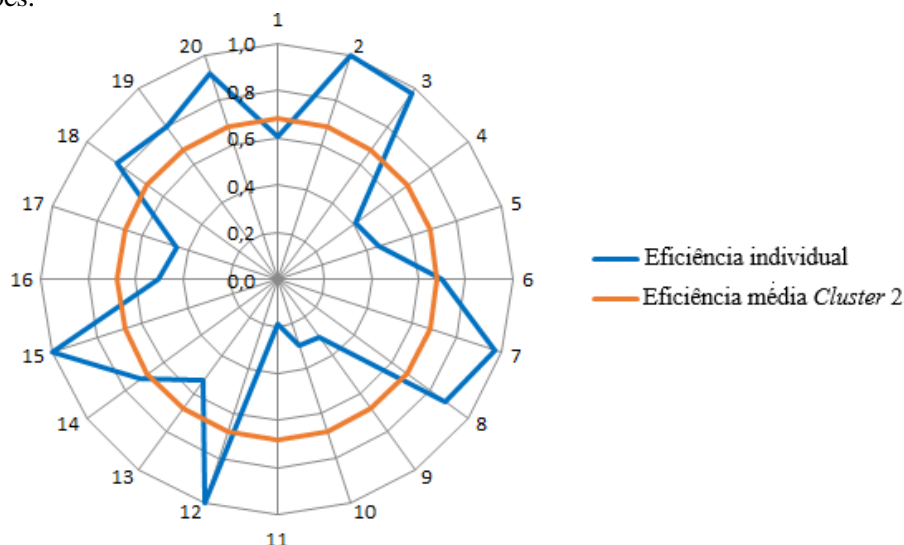


Figura 4: Eficiência individual versus eficiência média no Cluster 2



	Cluster 1	Cluster 2	Cluster 3
DMU's eficientes	4	3	3
Eficiência média	51%	68%	60%
Nº de concessionárias	30	20	11
Energia distribuída	3,7%	34,6%	61,7%

Tabela 4: Considerações sobre os 3 *clusters* analisados

Embora qualquer DMU que apresente coeficiente de eficiência abaixo de 1 seja considerada ineficiente, é importante ressaltar que há diferentes estratificações entre as DMU's ineficientes. Em função disso, [Norman e Stoker 1991] classificam as DMUs em:

- Eficiente: eficiência de 100%;
- Marginalmente ineficiente: Eficiência entre 90 e 99%; candidata a se tornar eficiente desde que uma pequena redução em seus *inputs* garanta uma melhora na eficiência;
- Eficiência mediana: Eficiência entre 70 e 90%; opera com baixa eficiência, usando *inputs* em excesso frente aos *outputs* apresentados;
- Ineficiente: Eficiência abaixo de 70%; indica que há um problema no gerenciamento de seus *inputs* para alcançar os *outputs* pretendidos.

Segundo Trindade et al. [2015], as medidas necessárias para que uma DMU se torne eficiente envolve três horizontes de tempo: curto, médio e longo prazo, DMU's marginalmente eficientes estão próximas a fronteira de eficiência e, portanto, o curto prazo é o horizonte de tempo adequado. Isso acontece pelo fato de não haver necessidade de ação corretiva muito incisiva para se tornar eficiente, DMU's com eficiência mediana necessitam de ações mais trabalhadas que incluem uma redução maior nos custos e por isso requerem um horizonte de médio prazo para que possam melhorar seus níveis de eficiência. Por fim, o horizonte de longo prazo é adequado as DMU's ineficientes cujos índices de eficiência são baixos e necessitam de um gerenciamento macro de seus recursos para que essas empresas possam maximizar sua eficiência.

A estratificação das DMU's, de acordo com seu grupo de eficiência, seguindo a classificação de [Norman e Stoker 1991] (Tabela 5) mostra que o *Cluster 2* obteve os melhores resultados frente aos demais *clusters*. No cluster 1 e 3, apenas 36% são consideradas eficientes, marginalmente eficientes ou com eficiência mediana, ao passo que, no cluster 2 são 50% das DMU's nessa faixa. Isto significa que, para as empresas marginalmente eficientes, um pequeno ajuste em seu custo operacional no curto prazo pode ser suficiente para alcançar o máximo de eficiência. Para as DMU's com eficiência mediana correções para alcançar a eficiência máxima podem ser realizadas num horizonte de médio prazo. Com relação as DMU's ineficientes, cujas ações de correções são voltadas ao longo prazo, a porcentagem de empresas do *cluster 1, 2 e 3* são 64%, 50% e 64% respectivamente.

Grupo de eficiência	Cluster 1	Cluster 2	Cluster 3
<70%	64%	50%	64%
70 - 90%	17%	25%	9%
90 - 99%	6%	10%	0%
100%	13%	15%	27%

Tabela 5: Grupos de eficiência por *cluster*

5. Conclusão

A Análise Envoltória de Dados foi utilizada para o cálculo de eficiência de 61 distribuidoras de energia elétrica. Para este fim, as DMU's foram segregadas em três grupos de acordo com seu grau de similaridade utilizando o método de clusterização hierárquico (tendo a distância Euclidiana método Ward como parâmetros) e então o modelo DEA CRS orientado a



input foi aplicado. Os *clusters* foram criados com objetivo de comparar apenas DMU's com características semelhantes entre si, a fim de evitar distorções nos índices de eficiência. Tais DMU's foram agrupadas utilizando 4 variáveis, sendo 1 *input* e 3 *outputs*, de acordo com revisão de literatura do assunto,

Os resultados mostraram que as empresas médias, pertencentes ao *cluster 2*, obtiveram os melhores resultados. Nesse grupo, 50% das DMU's encontram-se com eficiência maior ou igual a 70%. Isto indica que boa parte dessas empresas podem melhorar seus índices de eficiência em um curto ou médio período de tempo. Contrariamente, no *cluster 1*, representado pelas pequenas empresas, e no *cluster 3*, representado pelas maiores empresas, a maioria das DMU's apresentam índices de eficiência abaixo de 70%, indicando que boa parte delas não estão apresentando resultados satisfatórios na conversão *input-output*. Uma das limitações do estudo é a impossibilidade de se fazer uma comparação global entre as DMU's devido à restrição de homogeneidade. Uma possível alternativa é a aplicação do método BCC nesse caso, visto que este método apenas compara DMU's com retornos de escala semelhantes.

6. Referências Bibliográficas

Aldenderfer, M. S. e Blashfield, R. K. (1984). *Cluster Analysis*. Sage University Paper Series: Quantitative Applications in the Social Science, Newbury park, California.

Andrade, F. N., Alves, L. A., Silva, C. E. R. F., Mello, J. C. C. B. S. (2014). Evaluating Electricity Distributors Efficiency Using Self-Organizing Map and Data Envelopment Analyses. *Revista IEEE América Latina*, 12: 1464-1472.

ANEEL. Ministério de Minas e Energia. 2017a. Central de informações econômico-financeiras. Disponível em: <<http://www.aneel.gov.br/central-de-informacoes-economico-financeiras>>. Acessado: 09-01-2017.

_____. _____. 2017b. Indicadores coletivos de continuidade. Disponível em: <<http://www.aneel.gov.br/indicadores-coletivos-de-continuidade>>. Acessado: 09-01-2017.

_____. _____. 2017c. Atributos físicos-elétricos. Disponível em: <<http://www2.aneel.gov.br/aplicacoes/atributosfisicoeletricos>>. Acessado: 13-01-2017.

Azadeh, A., Ghaderi, S. F., Omrani, H. e Eivazy, H. (2009). An integrated DEA-COLS-SFA algorithm for optimization and policy making of electricity distribution units. *Energy Policy*, 37(7): 2605–2618.

Banker, R. D., Charnes, A. e Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9): 1078–1092.

Barbosa, K. D. S., Fernandez, R. N. e Gonçalves, M. T. (2016). Avaliando os aspectos institucionais do setor elétrico brasileiro por meio da teoria econômica de contratos. *Planejamento e políticas públicas*, 46: 73-107.

Boles, J. N. (1966). Efficiency squared—efficient computation of efficiency indexes. In: *Proceedings of Thirty Ninth Annual Meeting of the Western Farm Economics Association*, p.137–142.

Çelen, A. (2013). Efficiency and productivity (TFP) of the Turkish electricity Distribution companies: An application of two-stage (DEA&Tobit) analysis. *Energy Policy*, 63:300–310.



- Charnes, A., Cooper, W. e Rohdes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6):429-444.
- Coelli, T. J., Rao, D. S. P., O'donnell, C. J. e Battese, G. E. (2005). An Introduction to Efficiency and Productivity Analysis. Springer, 2nd ed. USA.
- Cooper, W. W., Seiford, L. M. e Tone, K. (2007). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Kluwer Academic Publishers, USA.
- Dai, X. e Kuosmanen, T. (2014). Best-practice benchmarking using clustering methods: Application to energy regulation. *Omega*, 42(1): 179-188.
- Driver, H. E. e Kroeber A. L. (1932). Quantitative expression of cultural relationships. University of California press, Berkeley, California.
- Everitt B. S. (2011). Cluster Analysis. 5a ed., John Wiley & Son, New York.
- Giannakis D., Jamasb T. e Pollitt M. (2005). Benchmarking and incentive regulation of quality of service: an application to the UK electricity distribution networks. *Energy Policy*, 33(17): 2256–2271.
- Gomes, E.G. (2003). Modelos de Análise de Envoltória de Dados com Ganhos de Soma Zero. Tese (Doutorado em Engenharia de Produção). COPPE – Universidade Federal do Rio de Janeiro. Rio de Janeiro.
- Gore, P. (2000). Cluster analysis, in Tinsley, H. & Brown, S. Handbook of applied multivariate statistics and mathematical modeling, Academic Press.
- Guimarães, R. M., Asmus, C. I. R. F. e Burdorf, A. (2013). Caracterização da exposição de população a organoclorados: uma aplicação da análise de cluster. *Revista Brasileira de Epidemiologia*, 16(2): 231-239.
- Haney, A. B. e Pollitt, M. G. (2009). Efficiency analysis of energy networks: An international survey of regulators. *Energy Policy*, 37: 5814-5830.
- Machado, L. G., Mello, J. C. C. B. S., Robofredo, M. C. (2016). Efficiency Evaluation of Brazilian Electrical Distributors Using DEA Game and Cluster Analysis. *Revista IEEE América Latina*, 14: 4499-4505.
- Mariano, E. B., Almeida M. R. e Rebelatto D. A. N. (2006). Princípios Básicos para uma proposta de ensino sobre análise por envoltória de dados. In: *Anais do XXXIV COBENGE*, Universidade de Passo Fundo – UPF.
- Mello, J. C. C. B. S., Gomes, E. G., Neto, L., e Lins, M. P. E. (2004). Suavização da fronteira DEA: o caso BCC tridimensional. *Investigação Operacional*, 24: p. 89-107.
- Mello, J. C. C. B. S., Meza, L. A., Gomes, E. G. e Neto, L. B. (2005). Curso de Análise envoltória de dados. In: *Anais do XXXVII SBPO*, Gramado.
- Moura, A. L., Rubem, A. P. S. e Mello, J. C. C. B. S. (2016). Avaliação de empresas aéreas com modelo de análise envoltória de dados clusterizado pela escala de operação. *Journal of Transport Literature*, 10(1): 20–24.



- Norman, M. e Stoker B. (1991). *Data envelopment analysis: the assessment of performance*. John Wiley & Sons Publishers, New York.
- Pessanha, J. F. M., Souza, R. C. e Laurencel, L. C. (2007). Um modelo de análise envoltória de dados para o estabelecimento de metas de continuidade do fornecimento de energia elétrica. *Pesquisa Operacional*, 27(1): 51-83.
- Plagnet, M. A. (2006). Use of benchmarking methods in Europe in the Electricity Distribution Sector. In Conference on Applied Infrastructure Research, Berlin.
- Rezende, S. M., Pessanha, J. F. M. e Amaral, R. M. (2014). Avaliação cruzada das distribuidoras de energia elétrica. *Production*, 24(4): 820-832.
- Sokal, R. R. e Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy*. San Francisco: W. H. Freeman.
- Souza, M., Souza, R., Pessanha, J., Oliveira, C. e Dialho, M. (2014). An application of data envelopment analysis to evaluate the efficiency level of the operational cost of Brazilian electricity distribution utilities. *Socio-Economic Planning Sciences*, 48(3): 1-6.
- Souza, M. V. P., Souza, R. C. e Pessanha, J. F. M. (2010). Custos operacionais eficientes das distribuidoras de energia elétrica: um estudo comparativo dos modelos DEA e SFA. *Gestão & Produção*, 17(4): 653-667.
- Trindade, D., Barroso, A. P. e Machado, V. H. (2015). Project management efficiency of a Portuguese electricity distribution utility using data envelopment analysis. *Procedia Computer Science*, 64: 674-682.
- Tryon, R. (1939). *Cluster Analysis*. McGraw-Hill, New York.
- Tschaffon, P. e Meza, L. A. (2014). Assessing the efficiency of the electric energy distribution using Data Envelopment Analysis with undesirable outputs. *Revista IEEE America Latina*, 12(6): 1027-1035.
- Vale, M. N. (2006). *Agrupamentos de dados: avaliação de métodos e desenvolvimento de aplicativo para análise de grupos*. Dissertação (mestrado em engenharia elétrica), PUC – Rio, Rio de Janeiro.
- Zanini, A. (2004). *Regulação econômica no setor elétrico brasileiro: uma metodologia para definição de fronteiras de eficiência e cálculo do fator X para empresas distribuidoras de energia elétrica*. Tese (Doutorado em Engenharia Elétrica), PUC – Rio, Rio de Janeiro.
- Zubin, J. A. (1938). A technique for measuring likemindedness. *Journal of Abnormal and Social Psychology*, 33: 508-516.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, 58(301): 234-244.