

Universidade Federal Fluminense

THIAGO GOMES LEAL GANHADEIRO

Avaliação alternativa de eficiência das distribuidoras
de energia elétrica brasileiras

Volta Redonda

2018

THIAGO GOMES LEAL GANHADEIRO

Avaliação alternativa de eficiência das distribuidoras de energia elétrica brasileiras

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Computacional em Ciência e Tecnologia da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional em Ciência e Tecnologia. Área de Concentração: Pesquisa Operacional.

Orientador:

Eliane da Silva Christo, D.Sc.

Coorientador:

Lidia Angulo Meza, Ph.D.

UNIVERSIDADE FEDERAL FLUMINENSE

Volta Redonda

2018

Ficha catalográfica automática - SDC/BEM

G195a Ganhadeiro, Thiago Gomes Leal
Avaliação alternativa de eficiência das distribuidoras de energia elétrica brasileiras / Thiago Gomes Leal Ganhadeiro ; Eliane da Silva Christo, orientadora ; Lidia Angulo-Meza, coorientadora. Volta Redonda, 2018.
77 p. : il.

Dissertação (mestrado)-Universidade Federal Fluminense, Volta Redonda, 2018.

DOI: <http://dx.doi.org/10.22409/PPG-MCCT.2018.m.14424327728>

1. Análise Envoltória de Dados (DEA). 2. Eficiência. 3. Regulação (Economia) . 4. Concessão de energia elétrica. 5. Produção intelectual. I. Título II. Christo, Eliane da Silva, orientadora. III. Angulo-Meza, Lidia, coorientadora. IV. Universidade Federal Fluminense. Escola de Engenharia Industrial e Metalúrgica de Volta Redonda.

CDD -

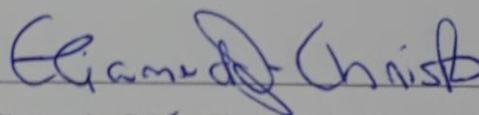
Bibliotecária responsável: Helania Oliveira Madureira - CRB7/4250

Avaliação alternativa de eficiência das distribuidoras de energia elétrica
brasileiras

Thiago Gomes Leal Ganhadeiro

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Computacional em Ciência e Tecnologia da Universidade Federal Fluminense, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional em Ciência e Tecnologia. Área de Concentração: Pesquisa operacional.

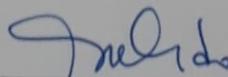
Aprovada por:



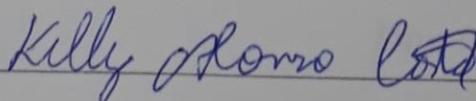
Prof. Eliane da Silva Christo, D.Sc. / MCCT-UFF
(Orientador)



Prof. Thiago Araújo Neves, D.Sc. / MCCT-UFF



Prof. Cecilia Toledo Hernandez, D.Sc. / MCCT-UFF



Prof. Kelly Alonso Costa, D.Sc. / VEP-UFF



Prof. Francisco Santos Sabbadini, D.Sc. / DEP-UERJ

Para todos que me ajudaram.

Resumo

Este trabalho tem por objetivo apresentar um método de cálculo das eficiências das distribuidoras brasileiras de energia elétrica. Tal avaliação é feita atualmente pela Agência Nacional de Energia Elétrica (ANEEL), de forma a regular os preços do setor. Isto é realizado através do cálculo do fator X , que representa a evolução de eficiência no modelo price-cap. O método proposto utiliza a extensão da rede como variável intermediária em um modelo network DEA e regressão múltipla para analisar quais variáveis ambientais podem afetar a eficiência de modo significativo. A partir dessas variáveis, é utilizado mapas auto organizáveis de Kohonen para agrupar as distribuidoras por semelhanças ambientais, normalizando as eficiências por cluster. Foram encontrados 4 agrupamentos baseados no ambiente de atuação de cada distribuidora. Cada distribuidora recebeu um valor de eficiência, que pode ser comparado às distribuidoras do mesmo grupo ou à totalidade do setor.

Abstract

The objective of this work is to present a method for calculating the efficiencies of Brazilian electricity distributors. Such an assessment is currently made by the National Electric Energy Agency (ANEEL), in order to regulate the prices of the sector. This is done by calculating the factor X , which represents the evolution of efficiency in the price-cap model. The proposed method uses network extension as an intermediate variable in a DEA network model and multiple regression to analyze which environmental variables can significantly affect efficiency. From these variables, Kohonen's self-organizing maps are used to group distributors by environmental similarities, normalizing cluster efficiencies. Four clusters were found, based on the operating environment of each distributor. Each distributor achieved an efficiency value, which can be compared to distributors in the same group or the entire industry.

Palavras-chave

1. Energia
2. DEA
3. SOM
4. ANEEL
5. Eficiência

Sumário

1	Introdução	8
1.1	Motivação	9
1.2	Objetivo	11
1.2.1	Objetivo Geral	11
1.2.2	Objetivo específico	11
1.3	Estruturação do trabalho	11
2	Fundamentação Teórica	13
2.1	Regulação	13
2.1.1	Mercados competitivos e falhas de mercado	13
2.1.2	Teorias da Regulação	16
2.1.3	Regulação Tarifária	18
2.2	Análise Envoltória de Dados(DEA)	21
2.2.1	Conceituação	21
2.2.2	Eficiência de Escala	22
2.2.3	Modelo CCR	24
2.2.4	Modelo BCC	27
2.2.5	Restrição aos Pesos	28
2.2.6	Modelos Network DEA	28
2.2.7	Modelo Network DEA Aditivo	29
2.2.8	Modelo Network DEA Aditivo Não Decrescente	31
2.3	Redes Neurais	33

2.3.1	Conceituação	33
2.3.2	Mapas auto organizáveis de Kohonen	34
2.3.3	Definição do número de clusters	38
2.3.4	Utilização com variáveis ambientais	38
2.4	Análise Multivariada	39
2.4.1	Conceituação	39
2.4.2	Regressão Múltipla	39
2.4.3	Análise de Fatores	42
3	Revisão da Literatura	46
4	Metodologia	51
5	Resultados	60
6	Conclusão	69
7	Trabalhos Futuros	71
	Referências	72

Capítulo 1

Introdução

O mercado de distribuição de energia elétrica no Brasil, em boa parte de sua extensão territorial, não é controlado diretamente pelo governo ou por alguma instituição governamental. Tendo em vista que esse setor pode ser compreendido como um monopólio natural, é necessário haver acompanhamento por parte de alguma agência reguladora de forma a impedir eventuais abusos de poder de mercado, no caso, a ANEEL, Agência Nacional de Energia Elétrica. De forma a minimizar os riscos inerentes a essa condição, um dos modelos de regulação econômica utilizados atualmente é o modelo price-cap, que corrige a remuneração obtida em função da inflação, de um fator relacionado a qualidade do serviço, denominado fator Q, e de um fator relacionado à produtividade da distribuidora de energia, denominado fator X.

O modelo utilizado atualmente pela ANEEL para o cálculo do fator X é um modelo DEA seguido de regressão múltipla, com a eficiência como variável dependente e os fatores ambientais como variáveis independentes. Isso pode implicar o entendimento de que a eficiência de uma distribuidora é dada, em grande parte, pelo ambiente em que ela se situa, gerando o entendimento de que os fatores inerentes à própria administração das empresas estão sendo excluídos da análise .

Além disso, como entrada para o modelo DEA usado pela ANEEL é utilizado o custo operacional (OPEX), e como saídas são utilizadas a dimensão da rede de distribuição, o número de consumidores e o consumo. Porém, é discutível o fato de a rede ser utilizada como output, haja vista o fato de ela ser gerada pelo custo operacional, mas utilizada para gerar o consumo e alcançar o número de consumidores.

Dessa forma, pode-se argumentar que há alguns pontos sujeitos a mudanças no método atualmente utilizado pela ANEEL, para os quais este trabalho aponta alternativas de

soluções.

Nesse sentido, o objetivo deste trabalho é a elaboração de um método de cálculo da eficiência a ser utilizada para a definição do fator X para as distribuidoras brasileiras de energia elétrica. Propõe-se que a rede seja tratada como variável intermediária e que as variáveis ambientais sejam utilizadas para gerar agrupamentos em mapas de Kohonen, de forma a possibilitar normalizar as eficiências encontradas por grupo.

São encontrados 4 agrupamentos de distribuidoras, e é encontrada uma eficiência por distribuidora, de forma que pode ser realizada comparação entre distribuidoras do mesmo agrupamento, reduzindo as interferências ambientais no cálculo da eficiência. Entretanto, ainda é possível a comparação entre agrupamentos diferentes, de forma a manter certa proporcionalidade entre as correções aplicadas ao fator X .

1.1 Motivação

Uma das falhas de mercado existentes na literatura é a presença de monopólios naturais, cuja explicação mais comum é a de que esses mercados correspondem a bens cuja exploração torna-se mais eficiente e competitiva contando com a exploração de uma única empresa. Isso geralmente ocorre devido à presença de altos investimentos de entrada ou altos custos fixos, de modo que, para que mais de uma empresa possa se manter no mercado, o preço de ambas aumentaria de forma a elevar o preço para o consumidor, embora não elevando o ganho das empresas. Isso, portanto, gera uma grande perda de bem-estar social. Portanto, é justificável a presença de um único fornecedor desse bem. Entretanto, essa situação é uma situação de monopólio, e, portanto, pode gerar perdas de peso-morto, gerando uma situação favorável ao monopolista que gera um prejuízo muito maior aos consumidores.

Uma das maneiras de se lidar com monopólios naturais é a exploração pelo próprio Estado, que, em tese, buscaria maximizar o bem-estar social, de maneira que buscaria ajustar o preço do bem à situação de concorrência perfeita. Outra maneira é por meio de regulação econômica, quando o serviço é prestado por entes privados.

No quesito de regulação econômica, existem dois modelos usuais: o modelo cost-plus [10], que visa restituir os custos operacionais e remunerar o custo de oportunidade e o modelo price-cap [10], que institui uma tarifa máxima a ser cobrada. A fórmula geral do modelo price-cap é dada por 2.1, presente no capítulo 2.

Um dos desafios do modelo price-cap é a metodologia de cálculo do fator X , que reflete a produtividade da empresa. Em geral, é bem complicado determinar a produtividade de uma empresa, tanto por conta da assimetria de informação quanto pelo próprio conceito de produtividade.

Para o cálculo do fator X , a ANEEL utiliza um modelo DEA não decrescente, corrigido por uma regressão múltipla com as variáveis ambientais. Compreende-se por variáveis ambientais aquelas externas à gestão da empresa que podem afetar a eficiência estimada [17].

O modelo utilizado pela ANEEL permite uma avaliação objetiva das eficiências de cada empresa, pois há uma comparação com as demais, e considera a influência do ambiente onde as distribuidoras operam, incluindo fatores que podem afetar as eficiências das mesmas.

Entretanto, existem alguns pontos da análise feita pela ANEEL que merecem atenção. Primeiramente, no modelo DEA utilizado, considera-se que o custo operacional, OPEX é uma entrada do processo, enquanto a extensão da rede, o número de consumidores e o consumo são saídas do processo. Entretanto, a variável extensão da rede é peculiar: dependendo da análise, ela pode se comportar tanto quanto input, pois é utilizada para gerar consumo e atender os consumidores, quanto como output, pois para ser mantida demanda custo operacional.

Outro ponto de interesse é o fato de se utilizar regressão múltipla como forma de correção das eficiências encontradas. Pode-se argumentar que ao se utilizar regressão múltipla, cria-se o entendimento de que apenas os fatores ambientais influenciam na eficiência da distribuidora, de forma que os fatores intrínsecos, relacionados à própria operação e administração das distribuidoras não seriam, de fato, essenciais para a eficiência, o que contraria o próprio propósito do modelo price-cap.

Dessa forma, pode ser interessante analisar novos métodos de cálculo para o cálculo do fator X , que considere esses pontos de interesse. Essa metodologia é o propósito deste trabalho.

1.2 Objetivo

1.2.1 Objetivo Geral

Este trabalho tem por objetivo elaborar um método alternativo ao utilizado atualmente pela ANEEL para avaliar as eficiências das distribuidoras de energia, que considere as variáveis ambientais na análise.

1.2.2 Objetivo específico

- Modificar o modelo DEA, deixando de utilizar um modelo composto por uma entrada e três saídas, e passando a utilizar um modelo network DEA, tornando a rede variável intermediária, mantendo o OPEX como entrada e o consumo e número de consumidores como saídas;
- Utilizar a regressão para verificar quais variáveis ambientais realmente influenciam na eficiência das distribuidoras;
- Com base nessas variáveis significativas encontradas, agrupar as distribuidoras por similaridade ambiental, utilizando mapas auto organizáveis de Kohonen
- Normalizar as eficiências por grupo.

Vale notar que existe uma limitação neste trabalho: por divergências sobre qual seria a metodologia mais apropriada para a correção da inflação, de forma a obter uma análise histórica mais precisa, preferiu-se não incluir os efeitos da inflação no OPEX dos anos anteriores, de modo que pode haver distorção dos resultados nesse sentido. Isso pode afetar as variáveis ambientais que devem ser tratadas como importantes para a eficiência.

1.3 Estruturação do trabalho

Este trabalho possui mais sete capítulos, além desta introdução. No Capítulo 2 será explicada a fundamentação teórica do trabalho, explicando os conceitos relacionados à regulação, incluindo falhas de mercado, as teorias da regulação e os preceitos de regulação tarifária, Análise Envoltória de Dados (DEA), incluindo eficiência de escala, modelos clássicos, modelos network DEA, com foco no modelo aditivo, e o modelo network DEA aditivo não-decrescente. Ainda no Capítulo 2 serão explicados os conceitos de redes

neurais, com enfoque especial aos mapas auto organizáveis de Kohonen; e será dada uma explicação sobre análise multivariada.

No capítulo 3 será feita uma revisão da literatura, de modo a observar os pontos analisados por outros autores e permitir visualizar a diferença deste trabalho com relação aos demais.

No capítulo 4 será exposta a situação analisada com maiores detalhes e será explicada a metodologia proposta para a avaliação de eficiência das distribuidoras.

No capítulo 5 serão expostos e discutidos os resultados obtidos da aplicação da metodologia supracitada.

O capítulo 6 apresenta as conclusões obtidas por este trabalho.

Por fim, o capítulo 7 apresenta sugestões para trabalhos futuros a serem desenvolvidos na área deste trabalho.

Capítulo 2

Fundamentação Teórica

2.1 Regulação

De forma a compreender o completo intuito deste trabalho, é preciso compreender regulação. Desta forma, essa seção é voltada a explicar os principais conceitos relacionados à regulação, como os motivos que levam a sua necessidade.

É possível entender regulação como um conjunto específico de comandos, uma interferência estatal deliberada sobre uma determinada atividade econômica [4]. Dessa forma, compreende-se regulação como o Estado atuando sobre a economia, impondo regras a serem seguidas ou um incentivando determinado comportamento desejado dos entes regulados. Além disso, diferentemente da legislação, que possui caráter fixo, a regulação acompanha as mudanças de mercado e procura adequar-se a elas.

Portanto, entende-se o mercado de distribuição de energia elétrica, no Brasil, como um mercado regulado, pois há interferência de um ente regulador, que serve ao Estado, no caso a ANEEL, Agência Nacional de Energial Elétrica, sobre o mercado, buscando modificar ou controlar tendências, entre elas o preço.

Para compreender as razões que levam à necessidade de regulação e o modo como funciona a regulação econômica no setor de distribuição de energia elétrica, é necessário explicar os conceitos de mercado competitivo e falhas de mercado.

2.1.1 Mercados competitivos e falhas de mercado

Entende-se por mercado perfeitamente competitivo aquele em que todos os participantes do mercado, tanto consumidores quanto produtores, são tomadores de preço [29].

Isso significa que nenhum consumidor ou fornecedor possui, individualmente, a capacidade de gerar mudanças no preço do produto. O preço é definido pelo mercado, sendo aquele que iguala a receita aos custos, considerando o custo de oportunidade, que, basicamente, é uma mensuração do quanto se deixa de ganhar por não investir o capital em outra oportunidade. Em um mercado competitivo, caso um fornecedor deseje cobrar mais barato, todo o contingente de consumidores tenderia a comprar com esse fornecedor, que seria incapaz de atender a todos e forçado a subir o preço. Caso desejasse cobrar mais caro, não existiriam consumidores dispostos a comprar, pois sempre haveria um produto idêntico a um preço mais baixo.

Um mercado competitivo, na teoria econômica, é aquele que produz melhor alocação de bens e recursos possível. De forma geral, diz-se que um mercado é perfeitamente competitivo se: o produto transacionado é homogêneo, a informação disponível for completa e simétrica, os custos de transação são nulos, a mobilidade de agentes é livre no mercado, os consumidores e vendedores são pequenos em relação ao mercado[32]. Isso significa que diversos produtores estão dispostos a vender o mesmo produto por um preço idêntico, há diversos consumidores buscando aquele determinado produto, aceitando pagar aquele preço, a informação chega a todos produtores e consumidores ao mesmo momento e a entrada ou saída do mercado é livre, não havendo custos associados. Como os preços são idênticos e os produtos também, não há preferências dos consumidores com relação aos produtores.

Nesse sentido, vale ressaltar que o lucro, considerando o custo de capital, seria nulo, ou seja, nenhum produtor seria capaz de obter lucros extraordinários e o preço de venda do bem seria igual ao custo marginal de produção.

Na prática, entretanto, a quase totalidade dos mercados não é competitiva. Isso ocorre porque existem falhas de mercado, que são fatores que afastam o mercado do modo de concorrência perfeita. Dentre as falhas de mercado mais comuns, pode-se citar: monopólio, monopólio natural, externalidades, bem público e assimetria de informação [29]. No contexto da distribuição de energia elétrica, são extremamente relevantes o conceito de monopólio natural e assimetria de informação.

O conceito de monopólio natural está intimamente ligado ao conceito de monopólio. Um produtor é monopolista quando é o único a ofertar um bem que não tem substitutos próximos [29]. Um monopólio ocorre quando um único produtor é responsável por atender toda a demanda do mercado. Isso costuma gerar grandes problemas, pois nessa situação é perfeitamente possível que o produtor defina o preço de sua mercadoria acima do preço

de concorrência perfeita e obtenha lucros extraordinários. O grande problema é que essa conduta retira o excedente do consumidor sem repor ao mesmo nível o excedente do produtor, gerando o que se chama de perdas de peso-morto e, conseqüentemente, perda de bem-estar social.

O excedente do consumidor é compreendido como “o ganho líquido na compra de um bem” [29], ou seja a diferença entre o que se estaria disposto a pagar e o valor efetivamente pago pelo bem. Semelhantemente, o excedente do produtor pode ser entendido como a diferença entre o que o produtor estaria disposto a receber pelo total dos bens que vende e o preço efetivamente recebido por esses bens. O bem estar social é a soma do excedente do consumidor e do produtor, e pode ser entendido como uma medida do quanto uma determinada sociedade está sendo beneficiada através da transação de um determinado bem [29]. Dessa forma, a perda de bem-estar social, acarretada pelo abuso do poder de mercado de um produtor monopolista acarreta em perdas para a sociedade como um todo.

Uma situação curiosa, entretanto, ocorre quando os custos fixos para a produção de um determinado bem são tão elevados que seria indesejável a presença de mais de um fornecedor naquele mercado, pois haveria elevação de preços muito acentuada, inclusive em situação de concorrência perfeita, de modo que os consumidores seriam fortemente prejudicados, ou seja, teriam seu excedente fortemente diminuído ao passo que o excedente do fornecedor também seria reduzido. Isso, por óbvio, reduz muito o bem estar social. Esse caso específico em que um produtor monopolista gera alocação de bem-estar social melhor do que a concorrência é denominado de monopólio natural [29].

Em geral, nos casos de monopólio natural, ou o serviço é fornecido diretamente pelo Estado, ou regulado por este. Um estudo prévio [27] demonstra que as distribuidoras de energia elétrica exibem características de monopólios naturais, como, por exemplo, grandes economias de escala. Além disso, essas características restringem a entrada no mercado e a regulação de preços torna-se necessária [42]. Dessa forma, essas distribuidoras, no Brasil, são reguladas por uma agência reguladora, a ANEEL, que é responsável por mitigar as falhas de mercado presentes no setor e proteger o desenvolvimento do setor. Cabe notar que o desenvolvimento do setor não significa total proteção ao consumidor, pois há casos em que as demandas do consumidor seriam responsáveis pela perda de bem estar social por causa da redução do excedente do fornecedor. Essa redução do excedente do fornecedor pode se tornar tão acentuada que nenhum fornecedor teria interesse em atuar no mercado, de forma que todos seriam prejudicados.

Outro conceito relevante para compreender a dinâmica da regulação econômica do se-

tor energético brasileiro é o de assimetria de informação. Assimetria de informação ocorre quando a informação não está completa ou disponível a todos os agentes envolvidos [39]. Esse fator em geral pode gerar grande desconfiança, pois pode haver grandes discrepâncias entre a conduta que se acredita ser correta pelo fornecedor e pelo consumidor. Além disso, com base na assimetria de informação podem surgir diversos questionamentos na relação regulado-regulador, tanto por mal interpretação dos desejos do regulador, o que pode acarretar em um não cumprimento ou cumprimento inadequado das normas elaboradas por este, quanto por observação alguma oportunidade de vantagem, por parte do ente regulado, ao valer-se de alguma informação ao qual o regulador não possui acesso ou não possui meios adequados de verificar a veracidade. Em uma situação hipotética, um regulado sujeito ao modelo de receita por preço máximo, por exemplo, poderia simular a venda de mais produtos do que os realmente vendidos de forma a ludibriar o ente regulador e receber um receita maior. Caso não haja fiscalização efetiva, por diversos fatores como proceso inadequado de fiscalização, custos de fiscalização elevados ou mesmo ausência de capital humano suficiente, seria possível que o regulado obtivesse receitas extraordinárias.

Ainda dentro do conceito de assimetria de informação, pode-se notar o contexto de qualidade do serviço prestado. Caso uma passagem de transporte seja metade do preço de outra, e esta seja a única informação disponível no mercado, por exemplo, é possível que todos os consumidores prefiram essa determinada passagem às concorrentes. Ao se adicionar a informação de que o lugar de transporte do passageiro é em um compartimento de carga, por exemplo, é possível que a situação se inverta.

Outra situação associada à assimetria de informação ocorre quando há margens de questionamento em determinada norma elaborada pelo ente regulador. É factível afirmar que caso um regulador opte por dar um bônus de receita a um regulado se este atingir uma meta de 90 por cento de alguma métrica, como tempo de disponibilidade de equipamentos, um regulado que cumpra 89,6 por cento dessa métrica pode tentar requerer ao ente regulador o bônus em questão, por não estar claro se devem ou não ser consideradas as casas decimais.

2.1.2 Teorias da Regulação

Compreendendo os conceitos de mercado competitivo e falhas de mercado e sabendo que uma das maneiras de correção dessas falhas é a utilização de regulação, é interessante notar que, em geral, não se assume automaticamente que a regulação surge como forma de se corrigir determinada falha de mercado. Existem, na realidade, diversos entendimentos

que explicam o surgimento de agências reguladoras e da regulação em si. Tais teorias se dividem em três grupos principais: as teorias do interesse público, as teorias do interesse privado e as teorias institucionalistas [39].

As teorias do interesse público partem do princípio de que existe um desejo de alcançar objetivos que levem à elevação do bem estar da sociedade. São teorias que assumem que os reguladores são bem intencionados e desejam o melhor para a sociedade. Existem três teorias principais dentro das teorias do interesse público: o da análise do bem-estar econômico, as abordagens políticas substantivas e as abordagens do procedimento político.

Na análise do bem estar econômico parte-se do princípio de que existe uma falha de mercado que deve ser corrigida. Dessa maneira, o órgão regulador surge objetivando corrigir tais falhas de modo a maximizar o bem estar social. As abordagens políticas substantivas seguem as diretrizes da análise do bem estar econômico, somadas à justiça social. Dessa forma, o regulador teria ainda preocupações como redistribuição de riqueza, potencialização de desejos coletivos e da diversidade, além da sustentabilidade.

As abordagens do procedimento político focam-se no processo em si, e não no bem estar social. Estas abordagens afirmam que o melhor resultado será aquele que possuir um processo aberto, amplo, justo e transparente, sendo este o objetivo ideal do órgão regulador.

Com relação às teorias do interesse privado, parte-se da premissa de que indivíduos ou grupos de indivíduos buscam a maximização de seus próprios interesses junto ao regulador. Há duas abordagens primárias dentro deste campo: a de interesse político privado, no qual grupos de interesse competem entre si para capturar o regulador; e a de interesse econômico privado, que utiliza análise custo benefício, de modo que a regulação reflete os interesses dos grupos politicamente mais fortes [39].

As teorias institucionalistas são geralmente vistas como um balanceamento entre as teorias do interesse público e as teorias do interesse privado. Dentro delas, existem o tripartismo, a teoria do espaço regulatório e a teoria dos sistemas. O tripartismo parte do conceito de que terceiros, que não o regulador e o regulado, também têm interesse na regulação. Dessa forma, esses entes se organizariam de forma a interferir no resultado da regulação.

A teoria do espaço regulatório afirma que a regulação é o resultado de uma barganha entre diversos atores. Grandes organizações teriam, então, força política e poderiam afetar a regulação, não sendo apenas objetos da regulação.

A teoria dos sistemas tem seu foco nas relações entre os sistemas econômico e político, de modo que estes não devem ser analisados separadamente, pois um influencia diretamente no outro.

É possível observar, portanto, que há diversas teorias para a existência das agências reguladoras, de forma que um determinado regulado, ao se sentir prejudicado, pode argumentar que a regulação não está sendo efetuada de modo a seguir o propósito original, de melhoria do bem-estar social. Desse modo, para justificar suas ações perante a sociedade, as ações tomadas por uma agência reguladora devem ser baseadas em transparência e fortemente motivadas pelo princípio do interesse público, de modo a possibilitar críticas por parte do setor regulado e da sociedade em geral. Tal comportamento, inclusive, minimiza as chances de corrupção, a qual implica má utilização de recursos e perda na eficiência. Cabe notar que os problemas associados à corrupção vem sendo uma causa de preocupação inclusive na Europa, graças ao aumento do número de casos observados [45].

Não obstante, é importante que o próprio processo regulatório possua uma base conceitual que justifique as ações tomadas, e que essa base seja disposta à consulta pública, bem como seja aberta a modificações devidamente fundamentadas. De modo a minimizar insurgência dos regulados, é necessário comprovar que o processo utilizado é justo quando observado por variados pontos de vista.

Nesse sentido, quando são verificadas potenciais falhas no processo, é necessário saná-las, não apenas para a tranquilidade do regulador, mas para a comprovação, perante a sociedade, de que o interesse público é o foco da agência reguladora.

2.1.3 Regulação Tarifária

O setor de distribuição de energia elétrica brasileiro é, majoritariamente, composto por empresas privadas. Considerando que experiências empíricas internacionais demonstram que só foi possível alcançar eficiência na privatização da distribuição elétrica quando esta for acompanhada da regulação [41], e, ainda, que as distribuidoras de energia são vistas como monopólios naturais, torna-se necessária a regulação do setor, realizada hoje por meio da ANEEL. Como estratégia de regulação, a ANEEL impõe tarifas que devem ser cumpridas pelas empresas, de modo a simular o comportamento de mercados perfeitamente competitivos em um ambiente real, onde não existe concorrência.

Para simular o preço correto a ser cobrado, existe dois modelos clássicos, o modelo cost-plus e o modelo price-cap. Além desses, modelos mais brandos têm sido testados,

como monitoramento de preços, negociação direta e regras de consulta.

O modelo cost-plus, também conhecido como Tarifação pelo Custo do Serviço ou Regulação da Taxa Interna de Retorno, parte do princípio de que o preço cobrado deve remunerar os custos totais e conter uma margem que gere uma taxa interna de retorno atrativa ao investidor [44]. Procura-se então, um preço que equilibre a modicidade tarifária com a atratividade de investidores. Uma vantagem desse método é que ele evita desincentivos de investimento por parte dos investidores. Entretanto, requer do regulador constante monitoramento dos custos das empresas reguladas. Além disso, a taxa de retorno é calculada através de negociação com a prestadora de serviço ou através de metodologias pré-estabelecidas. Para tanto, são necessárias diversas informações, que, a priori, apenas o regulado possui. Isso gera um grande problema de assimetria de informações, de forma que é difícil estabelecer um custo eficiente para as atividades da empresa. Isso geralmente se traduz em baixo poder do método em incentivar eficiência nas empresas reguladas.

O mecanismo de tarifação conhecido como price-cap constitui-se na definição de um preço-teto para os preços médios da firma, corrigido de acordo com a evolução de um índice de preços ao consumidor, o Retail Price Index (RPI), menos um percentual equivalente a um fator X de produtividade, para um período fixo de tempo, normalmente, 5 anos [44]. A fórmula geral do modelo price-cap, para os propósitos deste trabalho, é dada por (2.1).

$$P_t = P_{t-1} + \pi \pm X \pm Q \quad (2.1)$$

Onde:

- P_t é o preço no período t ;
- P_{t-1} é o preço no período $t-1$;
- π é a inflação no período;
- X é o fator relacionado à produtividade da empresa.
- Q é o fator relacionado à qualidade dos serviços prestados.

A função do fator X é de replicar os ganhos de produtividade sobre os preços, que seria o que ocorreria em ambientes competitivos, ou elevar os preços no caso de perda de produtividade, que pode ser advinda de diversos fatores.

Um dos desafios do regulador é o de incluir incentivos relacionados a qualidade em conflito com os incentivos à eficiência de custo [12]. Como há pressão por redução dos custos, pode-se reduzir a qualidade para reduzir os custos, o que acaba por prejudicar o consumidor final na maioria dos casos. Como este efeito é indesejável, há a inclusão de um fator relacionado à qualidade no modelo do price-cap, o fator Q. No Brasil, os fatores relacionados à qualidade podem incluir pressão popular e penalidades financeiras [12].

O modelo price-cap permite que a empresa retenha parte dos lucros obtidos entre os períodos de revisão tarifária. Isso serve de incentivo para que os investidores busquem a eficiência, pois eles se apropriarão das economias advindas dos incrementos de eficiência até a revisão tarifária seguinte. Isso, segundo Nepal, [42], tem um efeito menos desejável de reduzir o incentivo a economias conforme o avanço do período regulatório. Uma estratégia para reduzir esse efeito, utilizada pela Austrália, é a de utilizar um modelos de blocos, conforme demonstrado por um relatório do ente Regulador Australiano [1]. Além disso, segundo Mountain [40], estabelecer tetos de receita ou tarifas por 5 anos, período normal da revisão tarifária, gera fortes incentivos às empresas de descobrir necessidades e convencer o ente regulador dessas necessidades. Vale notar que o regulador deve avaliar se essas necessidades expostas são realmente relevantes para o setor, haja vista que, em geral, acarretarão no aumento de tarifas cobradas do consumidor.

Um dos grandes desafios do modelo price-cap está na adequada calibração do fator X, de modo a inibir lucros extraordinários do provedor de serviços, enquanto não prejudica a rentabilidade do empreendimento de forma a repelir investidores ou inibir investimentos, prejudicando a qualidade dos serviços.

Por se tratar de um elemento que afeta diretamente a rentabilidade dos regulados, é necessário que o modo de cálculo do fator X seja embasado em uma robusta explicação de cada passo envolvido, de forma a convencer tanto a sociedade quanto as empresas reguladas de que o processo utilizado é justo com todas as partes. Inconsistências podem servir de base para discussões, gerando argumentos fortes o suficiente para ensejar a modificação do método até então utilizado.

2.2 Análise Envoltória de Dados(DEA)

2.2.1 Conceituação

Quando a avaliação de eficiência é um fator de importância para algum projeto ou estudo, como para o cálculo do fator X, pode ser interessante utilizar a análise envoltória de dados. Também chamada de DEA (do inglês Data Envelopment Analysis) [8], essa abordagem utiliza problemas de programação linear de modo a comparar diversas unidades, denominadas de DMUs, a fim de avaliar sua eficiência.

Compreende-se por DMU (do inglês Decision Making Unit) qualquer unidade que realize algum processo de modo que possibilite a comparação com outras unidades que realizem processos similares. Desse modo, no contexto deste trabalho, as distribuidoras de energia são compreendidas como DMUs, no sentido de que realizam o processo de distribuição de energia, o qual pode ser comparado em termos de entradas e saídas para todas as distribuidoras.

Uma das principais vantagens de DEA é a capacidade de se analisar processos que possuam mais de uma entrada ou saída. Isso é particularmente útil quando é desejável comparar unidades que efetuem trabalhos relativamente complexos, como é o caso das distribuidoras de energia elétrica. Por esse motivo, a utilização de modelos DEA vem sendo implementada em nível internacional, como, por exemplo, em um estudo [2] com três modelos DEA executado na Espanha, para avaliação de eficiência de distribuidoras de energia, ou um estudo [43] que avalia modos de transporte no Japão. O artigo de Ali Emrouznejad e Guo-liang Yang [19] apresenta outras aplicações de DEA e a evolução no uso dessa abordagem.

A Análise Envoltória de Dados é uma ferramenta matemática para a medida de eficiência de unidades produtivas. Portanto, para que possamos estudá-la, torna-se necessário, primeiro, que saibamos o que é eficiência [15].

A eficiência compara os resultados atingidos por um processo com os resultados atingidos por processos semelhantes, ao mesmo tempo em que compara a utilização de recursos dos mesmos. É importante, portanto, que as entradas e saídas do processo a ser analisado possam ser quantificadas.

A eficiência de uma DMU é definida como a razão da soma ponderada de seus produtos pela soma ponderada dos insumos necessários para gerá-los[8]. Isso significa que é possível definir matematicamente a eficiência de uma DMU pela equação 2.2.

$$Eficiencia = \frac{\sum_{j=1}^s u_j y_j}{\sum_{i=1}^r v_i x_i} \quad (2.2)$$

Onde:

- u são os pesos associados às saídas do processo;
- y são as saídas do processo;
- j é o índice indicador da saída do processo;
- v são os pesos associados às entradas do processo;
- x são as entradas do processo;
- i é o índice indicador da entrada do processo;
- s é o número de saídas;
- r é o número de entradas;

Dentro desse conceito de eficiência, determinada DMU pode alterar os pesos de seus inputs e outputs. Entretanto, para haver coerência, o valor de nenhuma DMU, ao utilizar esses mesmos pesos, deve ser maior que 1. Uma DMU é considerada ineficiente caso sua eficiência seja menor que 1 [8]. Dessa forma, pode-se dizer que DEA forma um espaço de possibilidades de produção [8]. A problemática de DEA se resume, portanto, em grande parte dos casos, a encontrar os pesos que maximizem a eficiência de uma DMU ao mesmo tempo em que não violem a condição de máxima eficiência para nenhuma DMU.

Cabe ressaltar que, em geral, modelos DEA não levam em conta erros de medida ao definir os valores das eficiências [24]. Dessa forma, espera-se que os valores fornecidos contenham o mínimo de erros possível. Dessa forma, o presente trabalho parte da premissa de que não existem erros de medição nos dados colhidos, ou que esse erros são suficientemente pequenos, de forma a não impactar a análise.

2.2.2 Eficiência de Escala

Há dois modelos clássicos em DEA: o modelo CCR [8] e o modelo BCC [5]. Ambas nomenclaturas são acrônimos formados a partir dos nomes dos autores dos respectivos

modelos. A principal diferença entre esses modelos reside no conceito de eficiência de escala. A eficiência de escala é o fator que mede se o tamanho do processo de uma DMU afeta sua eficiência. Pensando em um processo de uma fazenda, por exemplo, um questionamento possível é se seria justo que um produtor com ampla extensão de terra tivesse que produzir proporcionalmente o mesmo que um pequeno agricultor para alcançar a mesma eficiência. Considerando que as técnicas empregadas e os custos de mão de obra e transporte, entre outros, provavelmente não seriam proporcionais, nem sempre é justo assumir que existe proporcionalidade. Entretanto, outros processos podem ser tomados como proporcionais sem prejuízo da análise proposta. A Figura 2.1 demonstra o caso de eficiência de escala, comparando a abordagem do modelo CCR e BCC.

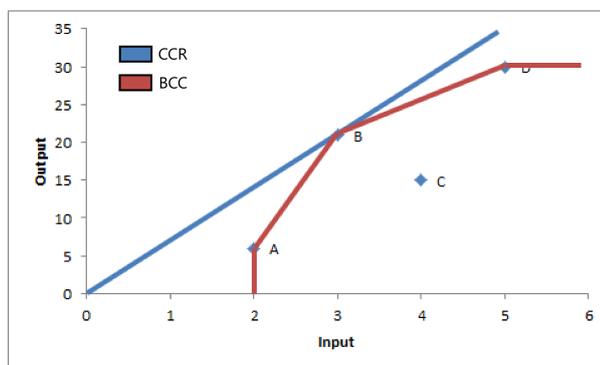


Figura 2.1: Eficiência segundo os modelos CCR e BCC (Adimensional)

Na Figura 2.1 são apresentadas 4 DMUs : A, B, C, D. No eixo vertical é dado o valor da saída (output) do processo para cada DMU. No eixo horizontal é dado o valor da entrada (input) do processo para cada DMU. Cabe notar que este exemplo é apenas conceitual, sendo as unidades adimensionais.

Desconsiderando-se os efeitos de eficiência de escala, a DMU que possui maior eficiência seria a DMU B, que possui a maior relação output/inputs. Considerando-se também que nenhuma outra DMU possui valor igual dessa relação, essa DMU seria a única eficiente. Cria-se, então, uma envoltória de eficiência, a qual, teoricamente, nenhuma outra DMU conseguiria ultrapassar, que se torna o objetivo das demais DMUs alcançar. Essa envoltória é dada pela reta que sai da origem e passa pelo ponto B, e representa a barreira de eficiência do modelo CCR.

Considerando-se os efeitos da eficiência de escala, o conceito de eficiência sofre uma transformação, e essa envoltória é transformada. Passa-se a considerar eficiente também a DMU que consome menos recursos, a DMU que produz maior output e as DMUs que ultrapassem a barreira teórica de eficiência gerada pelas retas entre as DMUs eficientes.

Essa é a envoltória de eficiência do modelo BCC, demonstrada na Figura (1) pela curva que passa pelos pontos A, B, D.

Observe-se que na Figura 2.1 há uma área onde aumentos no input causam aumentos ainda maiores no output. Essa região é denominada de região de retornos crescentes de escala, e é a região na qual a DMU A se encontra. Além dela, há a região onde aumentos no input levam a aumentos menores no output, como no caso da DMU D. a essa região dá-se o nome de região de retornos decrescentes de escala.

Seria ainda possível uma terceira região, onde um aumento do input acarretaria aumento proporcional no output. Nessa região, denominada região de retornos constantes, a eficiência BCC é equivalente à eficiência CCR. Observe-se que o modelo CCR é composto apenas por retornos constantes, motivo pelo qual também é chamado de modelo CRS (do inglês Constant Returns to Scale). Analogamente, o modelo BCC, que possui retornos crescentes e decrescentes, é chamado de modelo VRS (do inglês Variable Returns to Scale).

A eficiência de escala indica, então, em qual das três regiões possíveis a DMU se encontra, e seu módulo é dado pela divisão entre a eficiência BCC e a eficiência CCR.

Cabe notar que, no caso em pauta, a ANEEL utiliza para o cálculo do Fator X um modelo particular, que considera a envoltória como uma junção entre as envoltórias do modelo CCR e do modelo BCC. Esta envoltória começa como a do modelo BCC, porém, ao alcançar a DMU B, passa a ser a envoltória do modelo CCR. Esse tipo de modelo é denominado de Retornos de Escala Não Decrescentes, pois se inicia com retornos de escala crescentes e, posteriormente, aceita apenas retornos constantes de escala..

2.2.3 Modelo CCR

Matematicamente, o modelo DEA CCR maximiza o quociente entre a combinação linear dos outputs e a combinação linear dos inputs [13]. Partindo do conceito de eficiência apresentado na equação 2.2, é possível montar um modelo para encontrar as eficiências, considerando as restrições de não negatividade e de máxima eficiência. Para cada DMU observada, doravante denominada DMU_o, seria necessário resolver o sistema dado pelas formulas 2.3 a 2.5, de forma a encontrar os pesos que maximizem a eficiência.

$$Max \quad Eff_o = \frac{\sum_{j=1}^s u_j y_{jo}}{\sum_{i=1}^r v_i x_{io}} \quad (2.3)$$

S.a.

$$\frac{\sum_{j=1}^s u_j y_{jk}}{\sum_{i=1}^r v_i x_{ik}} \leq 1, \forall k \quad (2.4)$$

$$u_j, v_i \geq 0, \forall j, i. \quad (2.5)$$

Onde:

- u são os pesos associados às saídas do processo;
- y são as saídas do processo;
- j é o índice indicador da saída do processo;
- v são os pesos associados às entradas do processo;
- x são as entradas do processo;
- i é o índice indicador da entrada do processo;
- s é o número de saídas;
- r é o número de entradas;
- Eff_o é a eficiência da DMU_o;
- k é o identificador da DMU;
- O subscrito o significa observado e indica que o valor descrito é o da DMU para a qual o modelo está sendo executado naquele momento.

Percebe-se que modelo não está linearizado e é não limitado, o que impossibilita sua utilização como um problema de programação linear. Dessa forma, torna-se necessário linearizá-lo de forma a obter um índice de eficiência. Um modo de linearização consiste em assumir que o denominador da eficiência do modelo é igual a 1, o que é possível graças à multiplicidade de pesos. Para eliminar as divisões nos termos referentes às demais DMUs, multiplica-se os dois lados da inequação pelo denominador e rearranja-se a inequação. Dessa maneira, obtém-se o modelo descrito pelas fórmulas 2.6 a 2.9.

$$Max \quad Eff_o = \sum_{j=1}^s u_j y_{jo} \quad (2.6)$$

S.a.

$$\sum_{i=1}^r v_i x_{io} = 1 \quad (2.7)$$

$$\sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^r v_i x_{ik} \leq 0, \forall k \quad (2.8)$$

$$u_j, v_i, \geq, 0 \forall j, i. \quad (2.9)$$

Onde:

- u são os pesos associados às saídas do processo;
- y são as saídas do processo;
- j é o índice indicador da saída do processo;
- v são os pesos associados às entradas do processo;
- x são as entradas do processo;
- i é o índice indicador da entrada do processo;
- s é o número de saídas;
- r é o número de entradas;
- Eff_o é a eficiência da DMUo;
- k é o identificador da DMU;
- O subscrito o significa observado e indica que o valor descrito é o da DMU para a qual o modelo está sendo executado naquele momento.

A função objetivo da formulação 2.6 significa que o objetivo desse modelo é maximizar o valor dos outputs, através da variação dos pesos. Como resultado da linearização, a equação 2.7 mantém o denominador da eficiência da DMUo igual a um. A equação 2.8 impede eficiências maiores que um para qualquer DMU, enquanto a última equação 2.9 garante que os pesos sejam maiores ou iguais a zero.

2.2.4 Modelo BCC

O modelo BCC foi criado por Banker, Charnes e Cooper [5]. Como supõe retornos variáveis de escala, outra denominação possível para esse modelo é VRS (do inglês Variable Returns to Scale). Em contrapartida, o modelo CCR recebe a denominação de CRS (do inglês Constant Returns to Scale), por supor retornos constantes de escala.

Partindo do modelo CCR, e considerando variação nos retornos de escala, admite-se a inserção de uma nova variável, η , que indica em qual tipo de região se está trabalhando: retornos crescentes, constantes ou decrescentes. Dessa forma, as equações 2.10 a 2.13 apresentam o modelo BCC.

$$Max \quad Eff_o = \sum_{j=1}^s u_j y_{jo} + \eta \quad (2.10)$$

S.a.

$$\sum_{i=1}^r v_i x_{io} = 1 \quad (2.11)$$

$$\sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^r v_i x_{ik} + \eta, \leq; 0, \forall k \quad (2.12)$$

$$u_j, v_i, \geq, 0 \forall j, i. \quad (2.13)$$

$$\eta \in \mathbb{R} \quad (2.14)$$

Onde:

- u são os pesos associados às saídas do processo;
- y são as saídas do processo;
- j é o índice indicador da saída do processo;
- v são os pesos associados às entradas do processo;
- x são as entradas do processo;
- i é o índice indicador da entrada do processo;
- s é o número de saídas;
- r é o número de entradas;
- Eff_o é a eficiência da DMU_o;

- k é o identificador da DMU;
- O subscrito o significa observado e indica que o valor descrito é o da DMU para a qual o modelo está sendo executado naquele momento.

A variável η é chamada de fator de escala. Sua interpretação geométrica é a intercessão com o eixo x das retas-suportes das faces da fronteira de eficiência. No modelo apresentado, indicam retornos crescentes de escala se forem positivos; retornos decrescentes de escala se forem negativos ou retornos constantes de escala, caso sejam nulos.

2.2.5 Restrição aos Pesos

Pode ser observado, tanto no modelo CCR, quanto no modelo BCC, que existem variáveis associadas aos pesos dados às entradas e saídas do processo. Cabe notar que os pesos, que partem da definição de eficiência apresentada na equação 2.2, podem ser entendidos como a relevância dada pela DMU à variável em questão. Entretanto, cabe notar que os pesos encontrados na resolução dos modelos apresentados não são necessariamente únicos, sendo possível que diversas combinações de pesos, quando colocados na fórmula de eficiência, alcancem a mesma eficiência. Tal efeito é conhecido como multiplicidade de pesos. Cabe notar que o valor dos pesos em si não é necessariamente significativo, pois depende da escala das variáveis dadas como entrada. Algumas soluções que minimizam esse problema consistem em normalização das variáveis do modelo.

Não obstante, após algumas análises e observando a opinião de pessoas intimamente relacionadas com o processo, pode ser possível e razoável gerar relações entre os pesos das variáveis utilizadas. Isso pode ser utilizado para garantir que será dada importância adequada a determinada variável quando comparada a outra, pois reduz a flexibilidade dada aos valores que um determinado peso pode assumir [31] Tal relação é denominada Restrição aos Pesos. Um modo bastante comum de impor restrição aos pesos em determinado modelo DEA é a manutenção de um valor mínimo (ou máximo) quando da divisão do valor de um peso atribuído a uma variável pelo valor do peso atribuído a outra variável.

2.2.6 Modelos Network DEA

Um problema identificado nos modelos clássicos em DEA é o fato de não haver esclarecimento sobre o que ocorre dentro do processo [9]. Como se trabalha apenas com inputs e outputs, não se observa o modo como ocorre essa transformação. Porém, em

alguns casos, não é possível caracterizar uma variável nem como input nem como output. No caso do cálculo da eficiência para fins de definição do fator X em distribuidoras de energia, por exemplo, isso acontece com a extensão da rede. A extensão da rede não pode ser vista como input, pois é necessário que haja OPEX para gerar ou manter a rede de distribuição. Entretanto, também não é possível observá-la puramente como output, já que ela é um recurso de entrada para a geração de consumo e aumento do número de consumidores.

Para tentar minimizar esse tipo de problema, foram criados os modelos Network DEA. Esses modelos dividem o processo em partes. Considerando-se uma divisão do processo em apenas duas partes, para exemplificar, teríamos a primeira etapa voltada para transformar as entradas do processo em variáveis intermediárias, que serão utilizadas no processo, e a segunda etapa voltada para transformar essas variáveis intermediárias em saídas do processo. Isso pode ser visualizado na Figura 2.2. No caso em pauta, isso se traduz em uma operação voltada a transformar o OPEX em extensão de rede, e uma etapa voltada a transformar a extensão de rede em consumo e número de consumidores.

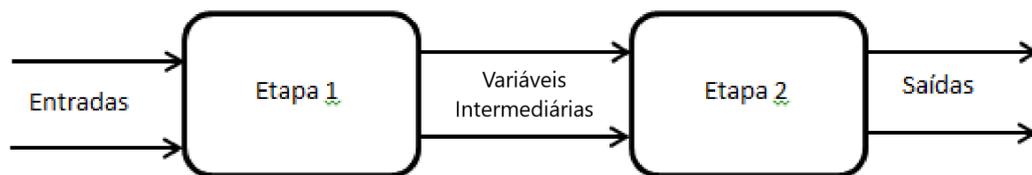


Figura 2.2: Esquematização do processo de network DEA, adaptado [25]

A Figura 2.2 demonstra o funcionamento geral de um modelo Network DEA. Uma das vantagens ao se utilizar esse tipo de modelo é a possibilidade de se calcular uma eficiência para a etapa 1 e outra eficiência para a etapa 2. A partir de uma operação com as eficiências das duas etapas é possível obter a eficiência do processo como um todo. Isso possibilita saber se, em um conjunto de DMUs, dada DMU é mais eficiente gerando variáveis intermediárias ou transformando essas variáveis intermediárias em outputs, além de permitir a comparação com as demais DMUs.

2.2.7 Modelo Network DEA Aditivo

O modelo aditivo foi proposto em 2010 por Wade D. Cook et al. [11]. Este modelo foi pensado considerando-se retornos variáveis de escala em conjunto com as particularidades do modelo network. Esse modelo calcula a eficiência geral de uma DMU como uma soma

ponderada das eficiências de cada etapa do processo, e é explicitado matematicamente nas formulas 2.15 a 2.19.

$$Max \quad Eff_o = \sum_{j=1}^s u_j y_{jo} + \sum_{t=1}^T w_t z_{to} + \eta_1 + \eta_2 \quad (2.15)$$

S.a.

$$\sum_{t=1}^T w_t z_{rk} + \sum_{i=1}^r v_i x_{ik} = 1 \quad (2.16)$$

$$\sum_{j=1}^s u_j y_{jk} - \sum_{t=1}^T w_t z_{tk} + \eta_2, \leq; 0, \forall k \quad (2.17)$$

$$\sum_{t=1}^T w_t z_{tk} - \sum_{i=1}^r v_i x_{ik} + \eta_1, \leq; 0, \forall k \quad (2.18)$$

$$u_j, v_i, w_t, \geq, 0 \forall j, i, t. \quad (2.19)$$

$$\eta_1, \eta_2 \in \mathbb{R} \quad (2.20)$$

Onde:

- u são os pesos associados às saídas do processo;
- y são as saídas do processo;
- j é o índice indicador da saída do processo;
- v são os pesos associados às entradas do processo;
- x são as entradas do processo;
- z são as variáveis intermediárias;
- w são os pesos associados as variáveis intermediárias;
- i é o índice indicador da entrada do processo;
- t é o índice indicador da variável intermediária;
- s é o número de saídas;
- T é o número de variáveis intermediárias;
- r é o número de entradas;
- Eff_o é a eficiência da DMUo;

- k é o identificador da DMU;
- O subscrito o significa observado e indica que o valor descrito é o da DMU para a qual o modelo está sendo executado naquele momento.
- η_1, η_2 são as variáveis livres correspondentes a retornos de escala variáveis.

No modelo descrito pelas equações 2.15 a 2.19, ocorre o aparecimento de uma nova variável, z , e um novo peso, w . Essa nova variável corresponde à quantidade do produto intermediário, e o novo peso corresponde ao peso relativo a essa variável. Vale notar que esse peso se mantém, o que gera diferença entre utilizar essa abordagem de cálculo da eficiência ou calcular as eficiências separadamente.

Resolvido o modelo, é possível encontrar as eficiências de cada etapa a partir dos pesos encontrados para cada DMU. Desse modo, basta que se aplique a fórmula de eficiência para cada etapa.

2.2.8 Modelo Network DEA Aditivo Não Decrescente

Na distribuição de energia elétrica, a ANEEL considera que os retornos de escala devem ser não decrescentes, pois o custo não é diluído com o tamanho da empresa [17]. Tal conceituação foi mantida neste trabalho. No entanto, cabe notar que existem estudos [3] que questionam se o modelo BCC é válido para descrever o setor elétrico no país, argumentando se, na verdade, grandes companhias sofrem retornos decrescentes de escala.

Para se adequar ao modelo de retornos não decrescentes, utilizado pela ANEEL, e permanecer utilizando modelos network, é necessário modificar o modelo aditivo, de modo a considerar retornos não decrescentes de escala. Para tanto, modificou-se a restrição às variáveis livres, permitindo-as assumir apenas valores positivos ou valor nulo. O modelo então tornou-se o modelo descrito pelas equações 2.21 a 2.25.

$$Max \quad Eff_o = \sum_{j=1}^s u_j y_{jo} + \sum_{t=1}^T w_t z_{to} + \eta_1 + \eta_2 \quad (2.21)$$

S.a.

$$\sum_{t=1}^T w_t z_{rk} + \sum_{i=1}^r v_i x_{ik} = 1 \quad (2.22)$$

$$\sum_{j=1}^s u_j y_{jk} - \sum_{t=1}^T w_t z_{tk} + \eta_2, \leq; 0, \forall k \quad (2.23)$$

$$\sum_{t=1}^T w_t z_{tk} - \sum_{i=1}^r v_i x_{ik} + \eta_1, \leq; 0, \forall k \quad (2.24)$$

$$u_j, v_i, w_t, \eta_1, \eta_2, \geq, 0 \forall j, i, t. \quad (2.25)$$

Onde:

- u são os pesos associados aos outputs;
- y são os outputs;
- j é o índice indicador do output;
- v são os pesos associados aos inputs;
- x são os inputs;
- z são as variáveis intermediárias;
- w são os pesos associados as variáveis intermediárias;
- i é o índice indicador do input;
- t é o índice indicador da variável intermediária;
- s é o número de outputs;
- T é o número de variáveis intermediárias;
- r é o número de inputs;
- Eff_o é a eficiência da DMUo;
- k é o identificador da DMU;
- O subscrito o significa observado e indica que o valor descrito é o da DMU para a qual o modelo está sendo executado naquele momento.
- η_1, η_2 são as variáveis correspondentes a retornos de escala variáveis.

2.3 Redes Neurais

2.3.1 Conceituação

É interessante tentar simular o funcionamento do cérebro humano, pois o cérebro rotineiramente executa tarefas de percepção de padrões (como por exemplo reconhecer uma face familiar em uma cena desconhecida) em aproximadamente 100-200 ms, enquanto tarefas de complexidade muito menor podem levar dias para serem executadas em um computador convencional [21].

Isso demonstra a grande capacidade da mente humana de processar informações. Tarefas incrivelmente complexas, que podem ser extremamente importantes, podem ser feitas de modo quase instantâneo. Portanto, não é de se estranhar que várias abordagens tentem emular o modo de funcionamento do cérebro. Uma dessas abordagens são as redes neurais[21].

O cérebro humano organiza e processa suas informações por meio de neurônios, que são entidades capazes de processar informações, os quais, a partir de um conjunto de experiências, sofrem um processo de aprendizagem, modificando o modo como trabalham as informações recebidas.

Redes Neurais podem ser entendidas como uma maneira matemática de tentar simular o funcionamento físico do cérebro. Para tanto, elas são compostas de unidades computacionais, chamadas de neurônios, capazes de executar tarefas computacionais simples individualmente, mas complexas quando estabelecidos em uma rede [18].

É comum haver diversos modelos de redes neurais para efetuar tarefas diferentes, cada modelo tentando se aproximar do modo como o cérebro interpretaria aquela tarefa. Inclusive, uma das principais vantagens da utilização de redes neurais é a capacidade de generalização, ou seja, operar sobre dados não vistos anteriormente [18]. Alguns modelos mais conhecidos são o de backpropagation e perceptron.

De modo a obter utilidade prática para uma rede neural, é necessário que ela possa se aprimorar, podendo obter resultados mais próximos do desejável, modificando o comportamento de seus neurônios. Para tanto, ela é submetida a um processo de aprendizagem, no qual dados de entrada conhecidos são dados a ela e, normalmente, comparados com os dados de saídas desejáveis, aprimorando os valores das relações utilizadas pelos neurônios. Normalmente, no processo de aprendizagem é deixada uma porcentagem de dados para conferir a confiabilidade da rede.

Existem algumas vantagens ao se utilizar redes neurais, sendo elas: não-linearidade, mapeamento input-output, adaptatividade, resposta evidente, informação contextual, tolerância a falhas, implementabilidade VLSI, uniformidade de análise e projeto, analogia neurobiológica [21].

A não-linearidade pode estar presente ou não em uma rede neural, e é especialmente útil quando o problema a ser tratado não é linear. A não linearidade geralmente se apresenta quando se trata de uma rede que possui feedback ou que tenha mais de uma camada de neurónios.

2.3.2 Mapas auto organizáveis de Kohonen

Mapas auto organizáveis de Kohonen são redes neurais cuja principal finalidade é encontrar similaridades em um grupo de elementos. Nesse modelo, os dados de entrada são distribuídos a todos os neurónios, porém apenas um neurónio é atribuído a cada elemento. Segundo (Borkowska et Al, 2014) [7], um mapa auto-organizável se diferencia de outros métodos de clusterização por se organizar de modo em que unidades próximas sejam similares entre si. Ainda segundo (Borkowska et Al, 2014) [7], mapas auto-organizáveis são de fácil compreensão e geralmente são melhores que análises estatísticas tradicionais. Uma representação gráfica dessa rede neural é dada na Figura 2.3.

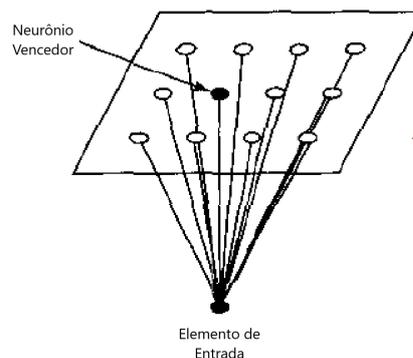


Figura 2.3: Representação de um mapa auto organizável de Kohonen. Adaptado de (Haykin,1994) [21]

Na Figura, pode-se observar que o input é distribuído a todos os neurónios, porém apenas um é ativado. A ideia predominante é de que a conexão mais forte será o neurónio a ser ativado. Essa estratégia é normalmente chamada de “o ganhador leva tudo”. Conforme os neurónios são ativados, eles são modificados, de forma a se aproximarem do input que os ativou. Isso gera um “mapa” formado pelos neurónios conforme se ajustam. Por isso é

chamado de mapa auto organizável.

O algoritmo de um SOM (Self Organizing Map, mapa auto organizável, em inglês) é composto pelas seguintes etapas [21] :

1. Inicialização
2. Competição
3. Cooperação
4. Adaptação Sináptica.

Vale notar que esse processo deve ser repetido algumas vezes até que a rede se torne estável, ou seja, não sejam notadas mudanças significativas na rede quando houver uma iteração seguinte. Para tanto, após o passo 4, é quase sempre necessário retornar ao passo 2, onde um novo elemento de entrada do grupo a ser analisado é submetido à rede, até sua estabilização.

Cada elemento a ser analisado pela rede possui um vetor contendo suas características. Por sua vez, cada neurônio possui um vetor, de mesma dimensão do vetor dos elementos de entrada, que indica sua posição topológica, e, conseqüentemente, sua similaridade com o elemento analisado. O processo de inicialização pode ser feito atribuindo valores aleatórios pequenos para os vetores dos neurônios.

O processo de competição é voltado para encontrar qual neurônio possui maior similaridade com o input analisado no momento. Isso pode ser representado pela equação 2.26.

$$i(X) = \operatorname{argmin}_j \operatorname{dist}(X, W_j) \quad (2.26)$$

Onde:

- $i(X)$ é o índice referente ao neurônio ganhador;
- X é o vetor do input analisado;
- W_j é o vetor associado ao neurônio j ;
- dist é a função distância utilizada que pode ser a distância euclidiana.

A fórmula 2.26 indica que é buscado o índice do neurônio que possua a menor distância entre o vetor de entrada X e o vetor W do neurônio. Esse será o neurônio vencedor.

O processo de cooperação é voltado a encontrar a vizinhança do neurônio, e calcular uma função para esta, de modo que no passo seguinte os neurônios próximos possam ser afetados pela mudança no neurônio vencedor. Uma função normalmente definida para a vizinhança é dada pela fórmula 2.27 (adaptado de [21]).

$$h_{j,i(x)}(n) = \exp \frac{-d(i,j)}{2(\sigma_0 \exp \frac{n}{t1})^2} \quad (2.27)$$

Onde:

- $h_{j,i(x)}(n)$ é a função vizinhança;
- $i(X)$ é o índice referente ao neurônio ganhador;
- X é o vetor do input analisado;
- $d(i, j)$ é a distância entre os neurônios i e j ;
- n é o número de iterações;
- σ_0 é uma constante relacionada à distribuição gaussiana de vizinhança;
- $t1$ é uma constante temporal.

A fórmula 2.27 indica que a vizinhança decresce conforme o número de iterações e a distância entre os neurônios. Isso significa que conforme o número de iterações cresce, a rede se estabiliza e os neurônios representantes de cada grupo de elementos são movimentados cada vez menos.

O processo de adaptação é voltado a corrigir os vetores dos neurônios. Kohonen [28] define a função de adaptação na forma da equação 2.28 (adaptado), de modo que apenas o neurônio ganhador, ou seja, aquele cujo valor da distância ao elemento considerado seja mínimo, seja atualizado.

$$W_j(n+1) = W_j(n) + \alpha(n)[X - W_j(n)] \quad (2.28)$$

Onde:

- W é o vetor de pesos associados ao neurônio j ;
- α é uma função que varia de acordo com a distância e o número de iterações;
- n é o número de iterações;
- X é o elemento considerado;

Como é necessário que o processo se estabilize após determinado número de iterações, a função α é escolhida de modo a propiciar esse ajuste. Dessa forma, essa função decresce exponencialmente com o número de iterações. Além disso, pelos preceitos do SOM, a atualização é proporcional a distância entre o elemento considerado e o neurônio. Dessa forma, a função de vizinhança é incorporada a função α . Com essas alterações, a equação 2.28 pode ser reescrita na forma da equação 2.29 [21].

$$W_j(n+1) = W_j(n) + \eta_0 \exp \frac{-n}{t_2} h_{j,i(x)}(n) [X - W_j(n)] \quad (2.29)$$

Onde:

- W é o vetor de pesos associados ao neurônio j ;
- $h_{j,i(x)}(n)$ é a função vizinhança;
- n é o número de iterações;
- X é o elemento considerado;
- η_0 é o parâmetro inicial de aprendizagem;
- t_2 é uma segunda constante temporal.

A fórmula 2.29 indica que a vizinhança de um neurônio também será atingida pela adaptação ao elemento de entrada. Entretanto, como a função vizinhança decresce com o tempo, a tendência é de que essas movimentações se tornem cada vez menores, estabilizando a rede. Também é possível notar que, para assegurar a estabilização da rede, o parâmetro de aprendizagem é reduzido com o tempo de maneira exponencial.

Ao final do processo, é possível observar um mapa topológico da rede, que mostra, a partir das distâncias entre os neurônios, as similaridades entre os dados que poderiam ser difíceis de se identificar [37].

2.3.3 Definição do número de clusters

Cabe notar que, ao se utilizar um mapa auto organizável, há a necessidade de se estabelecer previamente o número de agrupamentos (neurônios, clusters) a ser criados. Caso o problema em estudo não dê uma razão forte o suficiente para definir quantos agrupamentos serão utilizados, é necessário uma estratégia para definir tal valor.

Uma estratégia possível para tanto é a de criar diversas simulações, de diferentes tamanhos de rede, e, posteriormente, analisar qual delas apresenta o melhor resultado. Entretanto, isso requer uma previsão de um intervalo de valores sobre quantos agrupamentos devem ser formados. Ainda, dependendo do tamanho do problema, podem ser necessários inúmeros valores nesse intervalo, o que, considerando que cada simulação necessita de milhares de iterações, em geral, acarreta elevado esforço computacional e de análise, tornando-se inviável na prática.

Outra estratégia, que requer menos esforço, consiste em analisar quantos agrupamentos são formados naturalmente pelos elementos de entrada [14]. Para tanto, roda-se uma rede de dimensões consideravelmente maiores que o necessário e analisam-se graficamente as distâncias entre as posições de seus neurônios e as distâncias entre os vetores de pesos. O gráfico gerado pela distância entre as posições dos neurônios é chamado de Matriz U. A matriz de distância unificada pode ser interpretada como uma imagem através da coloração dos pixels de acordo com a intensidade de cada componente da matriz. Valores altos correspondem a neurônios vizinhos dissimilares e valores baixos correspondem a neurônios vizinhos similares. Regiões com baixos valores do gradiente correspondem a vales que agrupam neurônios especializados em padrões similares. Regiões com valores altos correspondem a fronteiras entre agrupamentos [14]. Tal abordagem foi utilizada neste trabalho, e um exemplo da matriz U é dado na Figura 5.2, enquanto a matriz de pesos resultante da análise neste trabalho é mostrada na Figura 5.3.

2.3.4 Utilização com variáveis ambientais

A análise das variáveis ambientais é de grande importância devido às grandes variações entre as regiões brasileiras [50]. Além disso, é mostrado em um estudo anterior [30] que as variáveis ambientais podem influenciar a eficiência de um distribuidor de maneira linear ou não linear. A estratégia utilizada neste trabalho é de aproveitar as variáveis ambientais analisadas pela Aneel, de modo a utilizá-las como entrada para um SOM. Isso é interessante, pois permite agrupar as operadoras conforme as similaridades ambientais.

Ao se utilizar como entrada as variáveis ambientais, é possível observar áreas de similaridade no ambiente de atuação de cada distribuidora de energia elétrica, permitindo observar as dificuldades impostas a cada ambiente. Isso pode permitir comparar distribuidoras que atuem em ambientes similares e o modo como reagem às características desses mesmos ambientes.

2.4 Análise Multivariada

2.4.1 Conceituação

Análise multivariada refere-se a todos os métodos estatísticos que simultaneamente analisam múltiplas medidas sobre cada indivíduo ou objeto sob investigação. Qualquer análise simultânea de mais de duas variáveis de certo modo pode ser considerado análise multivariada [26]. Dessa forma, a principal diferença da análise multivariada com relação a análise univariada é que, enquanto na univariada utiliza-se uma variável para explicar um resultado obtido, na análise multivariada é possível inserir diversas variáveis para explicar o resultado observado, tornando a análise mais complexa, porém, normalmente, mais próxima da realidade.

2.4.2 Regressão Múltipla

A análise de regressão múltipla é uma técnica estatística que pode ser usada para analisar a relação entre uma única variável dependente (critério) e várias variáveis independentes (preditoras) [26]. Assim como na regressão simples, a análise de regressão múltipla busca encontrar uma relação entre variáveis. Entretanto, enquanto a regressão simples utiliza apenas uma variável independente e uma dependente, a regressão múltipla utiliza várias variáveis independentes, de modo que a variação dessas variáveis acarrete a variação da variável dependente. Dessa forma, a análise de regressão propicia pesos a cada uma das variáveis independentes, que são referentes a participação delas na formação da variável dependente.

A regressão múltipla é um dos modelos estatísticos mais utilizados em pesquisas [48]. Sua principal aplicação, após encontrar a relação matemática, é produzir valores para a variável dependente quando se têm as variáveis independentes [47]. Ou seja, caso se obtenha os dados das variáveis dependentes, como no caso do modelo utilizado pela ANEEL, as variáveis ambientais, torna-se possível estimar a variável dependente, como, no caso

do modelo de cálculo do fator X pela ANEEL, a eficiência. Entretanto, outra aplicação importante para a regressão múltipla é saber quais variáveis são realmente importantes para definir a variável dependente. Ou seja, a partir de uma análise de regressão múltipla entre as variáveis ambientais e a eficiência é possível descobrir quais variáveis ambientais realmente influenciam na eficiência.

É importante notar que, para uma análise ser válida, deve haver alguma correlação entre as variáveis independentes e a variável dependente, algum sentido lógico implícito entre as variáveis. É importante também que as variáveis independentes não sejam inter-relacionadas entre si, ou seja, que seja mínima, ou, se possível, inexistente, a multicolinearidade, que é a correlação entre três ou mais variáveis independentes.

A equação geral da regressão múltipla é dada pela equação 2.30 [26].

$$Y = b_0 + \sum_{i=1}^k b_i X_i + e \quad (2.30)$$

Onde:

- Y é a variável dependente;
- b_0 é o viés da equação, ou seja, o valor que a variável dependente assumiria se não houvesse erros na análise e todas as variáveis independentes assumissem valor nulo;
- i é o índice indicador da variável independente;
- k é o número de variáveis independentes;
- b_i é a variação na variável dependente em relação à variável independente X_i ;
- X_i é a variável independente i;
- e é o erro associado, também chamado de resíduo.

Os resíduos são obtidos a partir da diferença entre o valor observado da variável de resposta e o valor predito pela regressão [16].

Existem algumas pressuposições necessárias para a utilização da análise de regressão múltipla [22]. São elas:

1. A variável dependente é função linear das variáveis explanatórias;

2. os valores das variáveis explanatórias são fixos;
3. $E(e_j) = 0$;
4. Os erros são homocedásticos, isto é, $E(e^2) = \sigma^2$
5. Os erros são não-correlacionados entre si, isto é, $E(e_j e_h) = 0$ para j diferente de h;
6. Os erros têm distribuição normal.

As pressuposições 1, 2 e 3 são necessárias para demonstrar que os estimadores de mínimos quadrados são não-tendenciosos e as cinco primeiras pressuposições permitem demonstrar que tais estimadores são estimadores lineares não-tendenciosos de variância mínima (teorema de Gauss-Markov). A pressuposição VI é necessária para realizar testes de hipótese e para construir intervalos de confiança para os parâmetros [22]. No âmbito deste trabalho, a preposição 6 tem maior importância, dada que são necessários testes de hipótese para verificar a importância de uma determinada variável ambiental na influência sobre a eficiência de determinada distribuidora.

A análise de regressão múltipla, assim como a regressão simples, faz uso do método dos mínimos quadrados para determinar os valores ideais dos coeficientes ao tentar minimizar o valor do resíduo, e é possível verificar o grau de explicação fornecido pela regressão pelo coeficiente de determinação, R^2 , que pode ser determinado pela equação 2.31.

$$R^2 = \frac{\text{Soma de quadrados da regressão}}{\text{Soma total de quadrados}} \quad (2.31)$$

Onde a soma de quadrados da regressão é dada pela fórmula 2.32, e a soma total de quadrados é dada pela fórmula 2.33.

$$\text{Soma de quadrados da regressão} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (2.32)$$

$$\text{Soma total de quadrados} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.33)$$

Onde:

- y_i é o valor medido da observação i;
- \bar{y} é a média de todas as observações;

- \hat{y}_i é o valor previsto da observação i através do método de mínimos quadrados.

A soma de quadrados da regressão corresponde à parte da soma total de quadrados que é explicada pela regressão. Há também a soma de quadrados de erros, que corresponde à parte não explicada pelo modelo. A soma das duas, como esperado, é igual à soma total de quadrados.

2.4.3 Análise de Fatores

Um conceito possível para a análise de fatores é de que esta é uma classe de métodos estatísticos multivariados cujo propósito principal é definir a estrutura subjacente em uma matriz de dados [26].

Outro conceito diz que a análise fatorial objetiva explicar as correlações existentes entre um conjunto grande de variáveis em termos de um conjunto de poucas variáveis aleatórias não observáveis, denominadas fatores [6].

Isso significa que a análise de fatores atua em uma matriz de dados, de forma a, analisando as relações entre as variáveis, encontrar certos itens não explícitos nos dados, mas que, de alguma maneira, geram variações nas características como um todo. A esses itens, dá-se o nome de fatores.

Uma das vantagens da análise de fatores, e o principal motivo da utilização desta técnica no trabalho é a possibilidade de reduzir o número de variáveis, eliminando a multicolinearidade, mantendo, pelo menos em boa parte, o poder explicativo dessas variáveis.

Como o foco da análise de fatores é a correlação entre as variáveis, não se utiliza o conceito de variáveis dependentes ou independentes. Todas as variáveis são, em teoria, interdependentes entre si, em maior ou menor grau.

Para se utilizar a análise de fatores é desejável que a amostra apresente pelo menos um pouco de multicolinearidade, para que seja possível definir um fator que explique com maior grau de confiança a variação apresentada pelas variáveis representada por esse fator.

A técnica da análise fatorial consiste em encontrar os valores dos coeficientes a_{ij} tais que reproduzam com o maior grau de confiança as variáveis a partir dos fatores, de forma a se adequarem à equação 2.34

$$Y_i = \sum_j a_{ij}F_j + a_iU_i \quad (2.34)$$

Onde:

- Y é a variável de entrada;
- F é o fator;
- i é o indicador na variável no elemento de entrada;
- j é o indicador do fator;
- U é o fator específico associado.
- a é a carga do fator.

Os coeficientes a também são conhecidos como carga do fator e, contanto que os fatores sejam não correlacionados, expressam a correlação entre a variável e um dado fator. As cargas servem como pesos, mostrando como cada y_i depende individualmente dos F [46]. A soma do quadrado desses coeficientes é chamado de comunalidade e representa o quanto da variância da variável é explicada pelos fatores considerados. O coeficiente U corresponde ao fator específico associado àquela variável, ou seja, corresponde à parte da variável que não pode ser explicada pelos fatores que aquele elemento possui em comum com os demais elementos analisados.

Dentro do escopo de análise fatorial são considerados três tipos de variância: a variância comum, a variância específica e a variância do erro. A variância comum corresponde àquela parte da variância que é partilhada pelas demais variáveis na análise. A variância específica é associada apenas a variável observada. A variância do erro, por sua vez, é devida a erros de amostra, imprecisões do processo de agrupamento ou uma componente aleatória na variável observada.

A análise de componentes principais, um dos dois modelos básicos utilizados para obter soluções fatoriais, procura fatores que possuam pouca variância única ou do erro, buscando o menor número de fatores que expliquem a maior parte da variância do grupo de variáveis iniciais. O outro modelo utilizado, a análise de fatores comuns, faz uso das comunalidades, que são estimativas da variância comum entre as variáveis.

Um problema referente à análise de fatores é a respeito do número de fatores a serem extraídos. Uma das soluções para determinar esse valor, a qual foi utilizada para a análise nesse trabalho, é o critério da raiz latente [26], que diz que um fator só é considerado relevante se consegue representar a variância de pelo menos uma variável. Logo, apenas

os fatores que apresentem autovalor maior do que 1 são considerados relevantes para a análise, sendo os demais considerados insignificantes e descartados. Outros critérios possíveis são: definir a princípio o número de fatores, calcular com base em uma porcentagem da variância total representada, realizar o teste scree que define o ponto onde a variância única passa a dominar a variância comum.

Ao se calcular os fatores, é interessante notar que o primeiro fator é o que representa a maior parte da variância dos dados, e explica melhor a variância do que qualquer outra combinação linear das variáveis. O segundo fator, que, por definição, é ortogonal ao primeiro fator, é o que melhor explica a variância das variáveis após a eliminação do primeiro fator, e assim sucessivamente. Existem limitações para a interpretação dos resultados da análise fatorial, pois há indeterminação dos fatores [36]. Isso pode significar que um fator esteja demonstrando mais de um traço ou variável, embora tal ocorrência não seja usual.

Para verificar a importância de cada variável dentro de cada fator, é válido recorrer ao conceito de cargas fatoriais, que indicam o grau de correspondência entre a variável e o fator. Como a análise inicial dos fatores não costumam indicar um padrão significativo das cargas fatoriais, é comum utilizar a rotação de fatores, que geralmente melhora a interpretação. Ao se utilizar a rotação de fatores, geralmente as cargas são mais bem definidas, de modo a facilitar a percepção de qual fator melhor explica qual variável.

Existem três abordagens principais de rotação de fatores: Quartimax, Varimax e Equimax. A abordagem Quartimax procura maximizar a carga de uma variável em um fator e minimizar a carga dessa variável em todos os demais fatores. Essa abordagem geralmente apresenta o problema de colocar quase todas as variáveis em um único fator, independente da estrutura subjacente a essas variáveis, o que geralmente torna essa abordagem imprópria para a análise fatorial.

A abordagem Varimax procura definir as cargas como o mais próximas de 0 ou 1 em cada fator, ao invés de tentar sempre maximizá-los. O método varimax procura tornar as cargas ou grandes ou pequenas, de modo a facilitar a interpretação [46]. Essa abordagem é analiticamente mais complexa do que a Quartimax, porém tem sido considerado bem-sucedida para a obtenção de uma rotação ortogonal. Portanto, é o método utilizado neste trabalho.

O método Equimax procura ser um meio termo entre os dois métodos anteriores [26]. Entretanto, geralmente não é bem aceito ou utilizado.

Não obstante, para se executar a análise fatorial, é esperado que boa parte da variância dos dados seja comum a todas as variáveis. De modo a verificar essa hipótese, indicando a adequação ou não do uso da análise fatorial, pode ser executado o teste KMO (Kaiser–Meyer–Olkin), descrito pela equação 2.35.

$$KMO = \frac{\sum_{j \neq k} \sum R_{jk}^2}{\sum \sum_{j \neq k} R_{jk}^2 + \sum \sum_{j \neq k} P_{jk}^2} \quad (2.35)$$

Onde:

- KMO é o resultado do teste;
- R_{jk}^2 é o coeficiente de correlação simples entre as variáveis X_j e X_k ;
- P_{jk}^2 é o coeficiente de correlação parcial entre as variáveis X_j e X_k , dadas as demais variáveis;

Dessa forma, em geral, assume-se que a análise fatorial é adequada quando o resultado do teste KMO for, no mínimo, igual a 0,5 [26].

Capítulo 3

Revisão da Literatura

Para melhor compreensão do problema enfrentado, pode ser interessante observar os trabalhos já publicados sobre o assunto abordado. Dessa forma, esse capítulo aborda outros artigos encontrados que tratam do tema em discussão. Foram encontrados alguns artigos que abordam o tema de avaliação de eficiência em distribuidoras elétricas no Brasil, os artigos de Andrade et al. [20], Tschaffon e Angulo-Meza [49], Moreno et al. [38], Machado et al. [33] e Mardani et al. [34], que serão discutidos a seguir.

O artigo de Andrade et al.(2014) [20] trata de uma avaliação das distribuidoras de energia elétrica utilizando mapas auto organizáveis de Kohonen (SOM) e DEA.

É proposta pelos autores uma nova modelagem para avaliação das distribuidoras de energia com relação ao modelo utilizado pela ANEEL. Segundo o autor, uma solução possível para encontrar as eficiências seria agrupar as distribuidoras por fatores ambientais em clusters para, posteriormente, executar o modelo DEA tradicional. Para o modelo DEA, o autor utiliza o próprio modelo RCM utilizado pela ANEEL.

Para o agrupamento em clusters, foi utilizado um SOM (Self Organizing Map), com as seguintes variáveis ambientais como entrada: Índice de precipitação; Salário médio; Complexidade média; Unidades consumidoras/Área; Consumo médio.

O foco do artigo é propor uma alternativa à modelagem em dois estágios utilizada pela ANEEL, substituindo-a por uma modelagem com SOM, de forma a simplificar a análise sem implicar necessariamente perda na confiabilidade. Os autores expõem como conclusão que as eficiências encontradas pelo modelo utilizado pela ANEEL podem estar subestimadas.

O artigo, ao contrário deste trabalho, não considera a rede como variável intermediária, não utilizando modelos network DEA. Isso é compreensível tendo em mente que a proposta

do autor era a simplificação do modelo de análise. Outro ponto de interesse consiste no fato do presente trabalho realizar uma análise com os dados históricos disponíveis, de forma a verificar se dada variável ambiental é realmente importante para a avaliação de eficiência, algo que não foi demonstrado no artigo analisado.

O 2º artigo analisado é o de Tschaffon e Angulo-Meza [49], que estuda a utilização de outputs (saídas) indesejáveis em DEA com aplicação no setor de energia elétrica.

O artigo aborda quatro maneiras distintas de se tratar outputs indesejáveis, utilizando a fronteira inversa para aumentar a discriminação entre as alternativas. A primeira abordagem utilizada é a *Additive Inverse*, onde os outputs indesejáveis são trocados pelos seus simétricos. Como segunda abordagem, é utilizada a *Incorporating undesirable Outputs as Inputs*, que considera os outputs indesejáveis como inputs (entradas). A terceira abordagem é a *Multiplicative Inverse*, que utiliza o inverso do output indesejável na elaboração do PPL. A última abordagem utilizada é a *Translation*, que soma à primeira abordagem um valor suficientemente grande para que os valores se tornem positivos.

Como entrada, foi utilizado o custo operacional (OPEX). Como saídas, foram utilizados: número de consumidores, consumo em TWh, extensão da rede, Duração Equivalente de Interrupção por Unidade Consumidora (DEC), Frequência Equivalente de Interrupção por Unidade Consumidora (FEC) e Índice Aneel de Satisfação do Consumidor. DEC e FEC são tratados como outputs (saídas) indesejáveis, tendo em vista que são produtos do processo, porém quanto menores, melhor para a eficiência da DMU.

O artigo aponta como DMUs eficientes as distribuidoras: Eletro Paulo, AES Sul, Cemar, Cemig, Coelba, Coelce, CPFL Paulista, CPFL Piratininga, Elektro e RGE.

Vale notar que, embora algumas análises da ANEEL utilizem as variáveis DEC e FEC, optou-se por não utilizá-las no presente trabalho por alguns motivos. O primeiro é baseado na análise dessas variáveis como outputs indesejáveis. Como outputs indesejáveis, percebe-se pelo artigo em pauta que não há um único modelo a ser utilizado como a melhor alternativa para tratamento desse tipo de variável. A utilização dessas variáveis abriria margem para questionamentos por parte das distribuidoras que eventualmente fossem prejudicadas pelo modo de avaliação dessas variáveis. Além disso, há o fato da interpretação delas como variáveis do processo. Pode haver discussão sobre o fato de considerar tais variáveis como produtos da extensão da rede, pois pode não ser percebido como uma relação muito lógica. Por último, optou-se por manter a coerência com o modelo em análise da ANEEL, o modelo RCM, que não utiliza essas variáveis.

O artigo de Moreno et al. [38] estuda a aplicação de um modelo network DEA com inputs compartilhados para análise da eficiência em distribuidoras de energia brasileiras. Por inputs compartilhados, entende-se que as entradas do processo são utilizadas não apenas para a geração das variáveis intermediárias, mas também para as variáveis finais do processo.

Foram utilizados os dados do ano de 2010 das seguintes distribuidoras: Eletropaulo, AES Sul, Ampla, Bandeirante, CEEE, Celesc, CELG, Celpa, Celpe, Cemar, Cemig, Coelba, Coelce, Copel, CPFL Paulista, CPFL Piratininga, Elektro, Escelsa, Light, RGE.

São utilizados dois modelos network DEA na análise: o modelo aditivo, já explicado neste trabalho, e o modelo relacional. O modelo relacional é um modelo network DEA que considera retornos constantes de escala, e cuja eficiência do processo é obtida pela multiplicação das eficiências das etapas. Cabe notar que o próprio artigo aponta como possível a utilização do modelo aditivo, tendo em vista a grande diferença de tamanho entre as DMUs. Entretanto, o artigo também aponta que uma vantagem da utilização do modelo relacional é o alto grau de discriminação nas eficiências entre distribuidoras alcançado neste modelo.

As variáveis para o modelo seguem a mesma lógica utilizada neste trabalho. Dessa forma, como entrada do processo é utilizado o OPEX. A rede é considerada como variável intermediária, enquanto o consumo e número de consumidores são tratados como saídas do processo.

Quando do uso do modelo relacional, o artigo aponta como menos eficiente as distribuidoras CELG e CEMIG, ao passo em que as mais eficientes foram Coelba, RGE e CPFL Piratininga.

Quando utilizado o modelo aditivo, o artigo encontra como menos eficiente a distribuidora CELG. Neste modelo foram encontradas 4 DMUs eficientes: COELBA, RGE, CEMAR e CEMIG.

A proposta de utilizar entradas compartilhadas é interessante, e pode representar melhor o processo analisado. Tal proposta é incluída como sugestão de trabalhos futuros, ao final deste trabalho.

O artigo de Machado et al. [33] utiliza teoria dos jogos aplicado a DEA e posteriormente utiliza clusterização de modo a avaliar as ditribuidoras de energia brasileiras no ano de 2009. São consideradas 61 distribuidoras. Com entrada para o modelo DEA é utilizado o OPEX. Como saídas são utilizados o consumo e o número de consumidores. Uma

vantagem da utilização de teoria dos jogos com DEA é o dato de a solução encontrada ser um equilíbrio de Nash, e portanto, estável.

As 61 distribuidoras utilizadas são: AES SUL, AME, AMPLA, BANDEIRANTE, BOA VISTA, CAIUA, CEAL, CEB, CEEE, CELESC, CELG, CELPA, CELPE, CELTINS, CEMAR, CEMAT, CEMIG, CEPISA, CERON, CFLO, CHESP, CJE, CLFM, CLFSC, CNEE, COCEL, COELBA, COELCE, COOPERALIANÇA, COPEL, COSERN, CPEE, CPFL PAULISTA, CPFL PIRATININGA, CSPE, DEMEI, DME-PC, EBO, EDEV, EEB, EFLJC, EFLUL, ELEKTRO, ELETROACRE, ELETROCAR, ELETROPAULO, ELFSM, EMG, ENERSUL, ENF, EPB, ESCELSA, ESE, FORCEL, HIDROPLAN, IENERGIA, LIGHT, MUX-ENERGIA, RGE, SULGIPE, UHENPAL.

A clusterização efetuada, ao contrário deste trabalho, não é realizada com mapas auto-organizáveis, mas com uma técnica de clusterização hierárquica denominada método de Ward. Além disso, no artigo analisado a clusterização é realizada utilizando os coeficientes de correlação entre as colunas da avaliação cruzada em DEA. Dessa forma, a clusterização proposta no artigo não busca definir o ambiente no qual as distribuidoras de energia se situam, mas sim definir o modo como realizam seus processos.

Como definição para a clusterização, foram utilizados três clusters, sendo que no cluster 1 se encontram as principais distribuidoras do Sul e do Sudeste do Brasil (ELETROPAULO, COPEL, CEMIG e LIGHT); no cluster 2 se encontram as principais distribuidoras do Norte e Nordeste (COELBA, CELPE, CELPA) e distribuidoras menores do Sul e do Sudeste; e no cluster 3 se encontram distribuidoras do Centro-Oeste, Norte e Nordeste.

Como alvos de eficiência foram definidas as distribuidoras: CPFL Piratininga, para o cluster 1; COELBA, para o cluster 2; e COELCE, para o cluster 3.

A possibilidade de agrupar distribuidoras baseado na maneira de execução do processo pode permitir avaliar, em estudos futuros, se distribuidoras similares em fatores ambientais realizam seus processos de forma semelhante e, caso contrário, a definir a melhor maneira de utilização dos recursos para cada tipo de ambiente enfrentado pelas distribuidoras de energia, possibilitando grandes ganhos de produtividade. Entretanto, tal análise foge ao escopo deste trabalho.

O artigo de Mardani et al. [34] revisa as principais aplicações de DEA no âmbito energético. Para tanto, foi utilizada uma análise PRISMA, contendo as etapas de busca de literatura, seleção do material encontrado, extração de dados e sumarização. Foram

analisadas bases online contendo arquivos desde 1983, data de publicação do primeiro artigo relacionado a eficiência energética. Ao todo foram analisados 3654 artigos, sendo que chegaram ao último estágio da análise 144 artigos.

O material encontrado foi dividido nas áreas: eficiência ambiental, Economia e eco-eficiência, Problemas de eficiência energética, Energia renovável e sustentável, eficiência aquática, performance energética, economia energética, eficiência energética integrada, e outras áreas de aplicação. O periódico com maior número de artigos adequados à última etapa da análise foi *Energy Policy*, sendo a maior parte dos artigos originários de autores chineses.

Os resultados do artigo mostram que DEA demonstra grande potencial para análises futuras sobre o tema de eficiência energética, onde é extremamente difícil de se obter uma função de produtividade, como no caso deste trabalho.

Capítulo 4

Metodologia

De forma a comparar as eficiências das distribuidoras, é utilizado um modelo DEA não decrescente, no sentido de que os retornos de escala não podem ser decrescentes. As variáveis utilizadas para a análise são o custo operacional, OPEX, a extensão da rede, o número de consumidores e o consumo. Em geral, o raciocínio é de que é mais eficiente a distribuidora que, com menor custo operacional, possa gerar elevada extensão da rede, alcançando um grande número de consumidores e proporcionando alto consumo.

Há um problema conceitual na definição da rede como uma variável do modelo DEA. Enquanto a rede é utilizada para aumentar o número de consumidores e o consumo de energia elétrica, ela também é fruto do investimento operacional da empresa. Dessa forma, não é correto conceituar a rede puramente como entrada ou saída do processo. De forma a resolver esse problema, utilizou-se o modelo network DEA, de modo que a rede foi considerada como uma variável intermediária, atuando como saída do OPEX, mas entrada para alcançar o número de consumidores e o consumo.

Entretanto, segundo a nota técnica nº 192-2014 [17], da ANEEL, existem “outras variáveis externas à gestão da empresa que podem afetar a eficiência estimada. Tais variáveis são comumente classificadas como variáveis ambientais”. Cabe notar que o termo “variáveis ambientais” é corriqueiramente associado em DEA com quaisquer variáveis que possam influenciar a eficiência, mas não são inclusas no modelo. No caso em pauta, há divergência de significado desse termo, no sentido em que é utilizado apenas para variáveis externas à gestão da empresa, mas que possuem potencial de influenciar a eficiência da mesma, sendo, primariamente, relacionadas ao ambiente de atuação da distribuidora de energia elétrica.

Em um país com ampla extensão territorial, como é o caso do Brasil, tais variáveis

podem variar intensamente de uma região para a outra, de modo a influenciar muito a comparação entre as distribuidoras. Para corrigir as variações de eficiência decorrentes das variáveis ambientais, a ANEEL utiliza um método em dois estágios, representado na Figura 4.1, onde o primeiro estágio representa o modelo DEA e o segundo estágio é uma regressão linear, utilizando, como variável dependente, a eficiência encontrada previamente, e, como variáveis independentes, as variáveis ambientais [20].

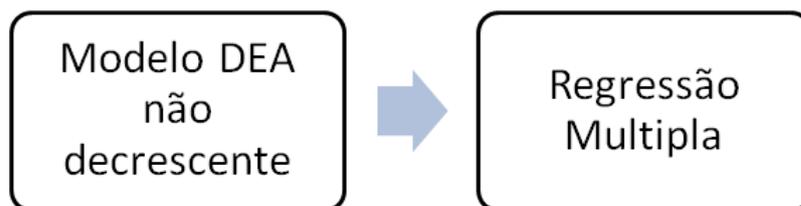


Figura 4.1: Esquematização do método utilizado atualmente pela ANEEL

Essa abordagem da utilização das variáveis ambientais, entretanto, pode não ser conceitualmente ideal. Isso porque, ao utilizar a regressão múltipla, gera-se a ideia de que as variáveis ambientais são, na realidade as responsáveis pela eficiência da distribuidora. Dessa forma, caso fosse possível obter dados das variáveis ambientais de uma nova distribuidora, não incluída previamente no modelo, e os dados referentes aos fatores intrínsecos de gestão dessa distribuidora estejam indisponíveis, ainda assim seria possível obter a eficiência esperada dessa distribuidora, o que, de certa forma, gera conflito com o próprio objetivo da medição da eficiência, que é a definição de quão bem os recursos são utilizados pela distribuidora. Portanto, esse é um ponto do modelo atual sujeito a mudanças.

As distribuidoras analisadas são as que possuíam dados disponíveis para o ano de referência (2012). Estas são: AES SUL, AMPLA, BANDEIRANTE, BRAGANTINA, CEB, CEEE, CELESC, CELG, CELPA, CELPE, CELTINS, CEMAR, CEMAT, CEMIG, COELBA, COELCE, COPEL, COSERN, CPFL PAULISTA, DME-Poços de Caldas, ELEKTRO, ELETROPAULO, ENE. PARAÍBA, ENE. SERGIPE, ENERSUL, ESCELSA, LIGHT, NACIONAL, PIRATININGA, RGE, SANTA MARIA, SULGIPE.

Pode ser interessante para posterior análise definir a área de atuação de cada distribuidora dentro do Brasil. Considerando as variações geográficas, culturais e econômicas entre os estados, que podem influenciar na eficiência das distribuidoras, segue a lista de distribuidoras analisadas, separadas por estado de atuação.

- BAHIA: COELBA;

- CEARÁ: COELCE;
- DISTRITO FEDERAL: CEB;
- ESPÍRITO SANTO: ESCELSA;
- GOIÁS: CELG;
- MARANHÃO: CEMAR;
- MATO GROSSO: CEMAT;
- MATO GROSSO DO SUL: ENERSUL; ELEKTRO;
- MINAS GERAIS: DME-Poços de Caldas; BRAGANTINA; CEMIG;
- PARÁ: CELPA;
- PARAÍBA: ENE. PARAÍBA;
- PARANÁ: CFLO; COPEL;
- PERNAMBUCO: CELPE;
- RIO DE JANEIRO: AMPLA; LIGHT;
- RIO GRANDE DO NORTE: COSERN;
- RIO GRANDE DO SUL: AES SUL; RGE; SANTA MARIA; CEEE
- SANTA CATARINA: CELESC;
- SÃO PAULO: BANDEIRANTE; BRAGANTINA; ELEKTRO; ELETROPAULO;
CPFL PAULISTA; PIRATININGA; NACIONAL;
- SERGIPE: SULGIPE; ENE. SERGIPE;
- TOCANTINS: CELTINS.

Cabe notar que a lista acima não contempla todos os estados brasileiros. Isso ocorre por causa da ausência de dados que possibilitem a análise das distribuidoras que operam nesses estados.

De modo a se obter uma eficiência preliminar para a análise, foi efetuada uma análise histórica das eficiências das distribuidoras com o modelo RCM (nomenclatura utilizada

pela ANEEL para denotar as saídas do processo em análise: Rede, Consumidores e Mercado ou consumo). Outros modelos DEA poderiam ser utilizados, mas optou-se pelo modelo RCM por ser atualmente utilizado pela ANEEL. A análise histórica foi efetuada com cada distribuidora para cada ano em que havia dados disponíveis, levando a um total de 411 DMUs. Cabe notar outras distribuidoras foram consideradas nessa análise e nas análises posteriores. Isso ocorre por diversos motivos, desde ausência de dados no ano de referência, 2012, até aquisição dessas distribuidoras por outras empresas. Essas empresas, que só possuem dados de anos anteriores, foram omitidas da listagem anterior a fim de dirimir eventuais confusões.

Uma nova metodologia de avaliação das eficiências das distribuidoras de energia elétrica para o cálculo do fator X que possa ser mais coerente com a realidade das operadoras, em sentido de considerar o ambiente de atuação de cada uma, sem deixar de considerar as ponderações atuais da ANEEL deve utilizar um modelo DEA não decrescente, bem como fazer uso das variáveis ambientais, de forma a corrigir a avaliação de acordo com a realidade da área de atuação de cada distribuidora. Um pensamento inicial seria o de incluir as variáveis ambientais no próprio modelo DEA. Entretanto, essa técnica seria inviável por diversos fatores. A primeira dificuldade seria definir quais variáveis ambientais seriam utilizadas como entrada ou saída do processo, especialmente quando não há relação direta dessas variáveis com as demais variáveis já definidas como importantes para o modelo DEA. Além disso, haveria uma quantidade de variáveis grande o suficiente para inviabilizar a análise. Regras empíricas ditam que é necessário utilizar, no mínimo, cerca de 3 DMUs para cada variável utilizada. Considerando o número de DMUs utilizadas para o ano de 2012, praticamente todas as DMUs alcançariam elevada eficiência, muito próximas de 1. Por último, seria conceitualmente errado colocar variáveis ambientais, externas às empresas, como responsáveis pela eficiência da empresa, que deve ser vista como um fator que é influenciado, primariamente, pela gestão das distribuidoras.

Percebe-se, então, que é necessária uma análise onde as variáveis ambientais sejam consideradas, porém sem entrar diretamente no modelo DEA. A solução proposta foi a de se separar as distribuidoras com base na semelhança entre suas áreas de atuação, normalizando as eficiências encontradas com base na mais eficiente do grupo formado por essas similaridades. Para formar esses grupos, a opção utilizada foi utilizar mapas auto organizáveis de Kohonen.

Entretanto, pode ser possível que nem todas as variáveis ambientais definidas previamente pela ANEEL influenciem realmente na eficiência das distribuidoras. Um novo

problema a ser resolvido implica saber quais das variáveis ambientais realmente possuem algum grau de influência sobre as eficiências que seja forte o suficiente para afetar de maneira significativa a eficiência de uma distribuidora.

Aproveitando-se da técnica já utilizada pela ANEEL de regressão múltipla, a solução encontrada foi utilizar a regressão múltipla, não como uma ferramenta para definir diretamente a eficiência, mas para avaliar a influência das variáveis ambientais na eficiência das operadoras. Para tanto, foi necessária uma análise primária das eficiências das operadoras, que foi executada pelo modelo RCM [17], o qual é um modelo não decrescente com restrição aos pesos, utilizado pela ANEEL. O modelo RCM pode ser descrito como um modelo BCC, com as restrições de que o valor de η ser positivo ou nulo e com restrições aos pesos, descritos nas inequações 4.1, 4.2 e 4.3.

$$\frac{v_{consumidores}}{u} \geq 30 \quad (4.1)$$

$$\frac{v_{rede}}{u} \geq 580 \quad (4.2)$$

$$\frac{v_{consumo}}{u} \geq 1 \quad (4.3)$$

$$(4.4)$$

As restrições aos pesos acima implicam, de maneira geral, que, para cada unidade de OPEX incrementada no modelo, deve haver, para manutenção da eficiência, no mínimo: elevação de 30 unidades de consumidores, 580 unidades de rede, uma unidade de consumo.

De maneira a melhorar os resultados da análise, a fim de evitar falsas interpretações com relação ao resultado da regressão múltipla, foram incluídas na regressão múltipla e, conseqüentemente, na análise prévia de eficiência, os dados históricos das distribuidoras, incluindo também as demais distribuidoras que possuíam dados de anos anteriores, mas sobre as quais não havia disponibilidade de dados sobre 2012. Dessa forma foi possível obter uma base de dados bem maior e mais confiável para a análise. Por sua complexidade, uma representação gráfica do processo pode ajudar a compreender todas as etapas realizadas no estudo. Essa representação pode ser visualizada na Figura 4.2.

A Figura 4.2 é dividida em níveis, no sentido de que todas as atividades que compõem determinado nível devem ser concluídas para avançar ao nível seguinte. As atividades (fases) desenvolvidas foram numeradas, e os fluxos de atividades foram identificados com letras maiúsculas.

Utilizando a Figura 4.2 como base, a fase 1 é a utilização de um modelo RCM para

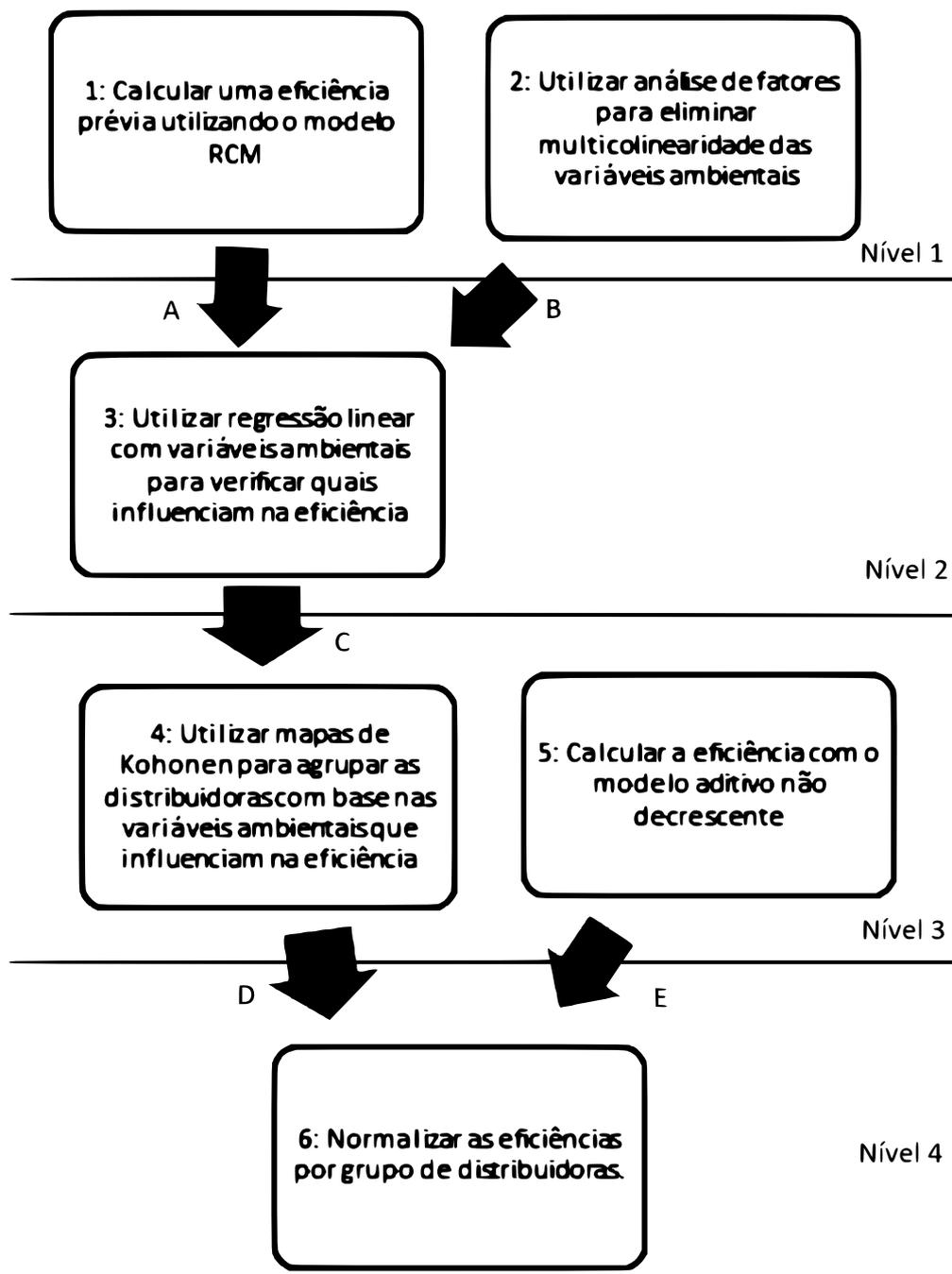


Figura 4.2: Esquematização do modelo proposto

uma verificação rápida de eficiência das operadoras. Como o resultado das eficiências será utilizado para definir quais variáveis ambientais realmente influenciam a eficiência na fase 3, foram utilizados dados históricos de todas as distribuidoras que possuíam dados disponíveis a partir do ano de 2003, ano em que começam a aparecer dados para consulta. Dessa forma, foi possível obter 411 eficiências para comparação posterior na regressão múltipla. Para esse modelo RCM, foi utilizada a opção da ANEEL de utilizar a rede como output para o processo. Dessa forma, as entradas e saídas do processo utilizadas foram:

- Custo operacional, OPEX, como entrada, ao término do ano de atuação, segundo fonte da ANEEL (Audiência Pública 023/2014);
- Tamanho da rede de transmissão, Rede, como saída, ao término do ano de atuação, segundo fonte da ANEEL (Audiência Pública 023/2014);
- Consumo, CONS, como saída, ao término do ano de atuação, segundo fonte da ABRADÉE;
- Número de consumidores, NCONS, como saída, ao término do ano de atuação, segundo fonte da ANEEL (Audiência Pública 023/2014).

Para determinar as eficiências usando o modelo RCM, foi utilizado o software MATLAB [35].

Com relação a fase 2, foi notado que algumas variáveis ambientais possuem certo grau de correlação entre si. Isso implica que essas variáveis, na realidade, fazem parte de uma condição mais abrangente, e que esse fator é o verdadeiro responsável pelas mudanças apresentadas por essas variáveis. Caso essas variáveis fossem utilizadas sem serem transformadas em um fator elas poderiam influenciar bem mais a formação dos clusters na fase 4 da Figura 4.2, além de causar distorções importantes na regressão múltipla. Dessa forma, utilizando os dados ambientais das mesmas DMUs utilizadas na fase 1, foi executado, no software SPSS [23], uma análise de fatores, utilizando rotação Varimax e extraindo os fatores com autovalores maior que 1. Essa atividade compõe a fase 2 da Figura 4.2. As variáveis ambientais foram obtidas da Audiência Pública 023/2014 da ANEEL, e correspondem a:

- Densidade de consumidores, entendido como o número de consumidores por área;
- Densidade de rede entendido como a extensão da rede por área;

- Complexidade, um índice de complexidade socioeconômica utilizado para definição dos níveis regulatórios de Perdas Não Técnicas, tais como: Violência acumulada, óbitos por agressão, cobertura de coleta de lixo, cobertura de disponibilidade de água, precariedade, número de pessoas em domicílios subnormais;
- Violência, entendido como o número de óbitos por habitante;
- Subnormal, entendido como a percentagem de pessoas vivendo em domicílios subnormais;
- Precipitação pluviométrica média anual (mm);
- Descargas, um indicador de descargas atmosféricas médias em km^2 por área de atuação;
- Área de atuação da distribuidora, em km^2 ;
- Vegetação, um indicador de incidência elevada de vegetação natural na Área de Atuação por concessionária que implique maiores custos operacionais em km^2 ;
- Vegetação baixa, em km^2 ;
- Vegetação média, em km^2 ;
- Vegetação alta, em km^2 ;
- Declividade, um indicador de declividade média do relevo da área de atuação de cada concessionária, em pontos percentuais;
- Pavimentação, um indicador da extensão das estradas pavimentadas existentes.

Dessa forma, as atividades 1 e 2 compõem o nível 1, sendo o término de ambas necessário para o início do nível 2. Da atividade 1 é levado para a atividade 3, conforme disposto pela seta A, os valores de eficiências históricas das 411 DMUs. Da atividade 2 são levados para a atividade 3 os valores dos fatores associados às variáveis ambientais relacionadas às mesmas DMUS utilizadas na atividade 1, representado pela seta B.

Em seguida, na fase 3, utilizaram-se os fatores encontrados para executar uma regressão múltipla, tendo como variável dependente a eficiência do modelo CCR encontrada na fase 1 da Figura 4.2 e como variáveis independentes os valores dos fatores encontrados na fase 2. Com base nessa regressão, executada no SPSS [23], alguns fatores mostraram-se

correlacionados com a eficiência encontrada. A atividade 3 conclui o nível 2, e seu resultado é a identificação de quais variáveis ambientais influenciam no cálculo da eficiência. A identificação dessas variáveis é passada, pela seta C, à atividade 4.

Para o ano de 2012, os fatores que influenciaram a eficiência na fase 3 da Figura 4.2 foram selecionados para inclusão em um SOM, sendo esta a atividade 4. No SOM foram incluídos apenas os dados referentes a 2012, ano da análise. Foi criada a matriz U para determinar o tamanho ideal do mapa a ser gerado. Com o resultado obtido, foi criada uma matriz 2x2, de modo que as distribuidoras foram agrupadas em 4 clusters. Essa análise foi executada no MATLAB [35].

Para encontrar a eficiência não corrigida das distribuidoras, atividade 5 do método proposto, foi executado um modelo network DEA aditivo não-decrescente com os dados de 2012, que foi executado no MATLAB [35] por código próprio criado para essa finalidade. A Figura 4.3 apresenta um diagrama do modelo DEA utilizado nesta atividade.

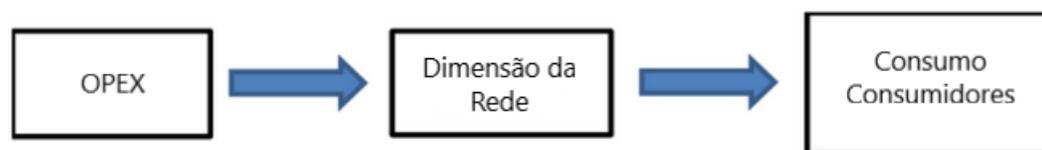


Figura 4.3: Esquematização do modelo DEA proposto

As atividades 4 e 5 finalizam o nível 3. O resultado da etapa 4 é a formação de grupos com base em similaridades ambientais, permitindo observar qual distribuidora se encontra em qual grupo. Essa mensagem é enviada à atividade 6 por meio da seta D. O resultado da etapa 5 é o valor não normalizado das eficiências utilizando o modelo network DEA RCM aditivo com restrição aos pesos. Os valores dessas eficiências são a mensagem transmitida pela seta E.

Por fim, as eficiências foram normalizadas por cluster, sendo esta a etapa 6, que finaliza o algoritmo, gerando a eficiência para o cálculo do fator X.

Capítulo 5

Resultados

De pronto, vale ressaltar que nem todas as distribuidoras analisadas possuíam dados para análise em todos os anos de análise. De forma a possibilitar a avaliação das eficiências, as DMUs que não possuíam dados para análise em um determinado ano foram excluídas da análise. Isso implica que há certas distribuidoras que possuem suas eficiências explicitamente demonstradas para alguns anos, mas não para outros.

Cabe notar que, como foram analisadas 411 DMUs, os dados colhidos de entradas e saídas não foram incluídos no presente trabalho por questões de espaço e relevância, enquanto as eficiências do modelo RCM também foram omitidas pelas mesmas razões. Ressalta-se que, para o presente trabalho, a relevância desses dados é primariamente para análise estatística, de modo que não se percebe necessidade de incorporá-los ao trabalho, pois geraria desgaste desnecessário ao leitor que não busca fazer nova análise sobre esses dados. Desse modo, o resultado da atividade 1, enviado pela seta A à atividade 3 da figura 4.2 não é apresentado de forma explícita neste trabalho.

Com relação à atividade 2 da figura 4.2, foi necessária uma intervenção, relacionada ao valor do teste KMO. Para que o valor do teste KMO fosse superior a 0.5 e a análise fosse considerada viável, foram eliminadas da análise duas variáveis ambientais: Área e Vegetação baixa. Dessa maneira, obtiveram-se quatro fatores na análise. De forma a melhor visualizar a importância de cada variável em cada fator, a figura 5.1 mostra a participação das variáveis nos fatores, obtida a partir da normalização da matriz de coeficiente de escore de componente.

A partir da análise da figura 5.1, pode-se notar que:

- O fator 1 é composto primariamente pelas variáveis: densidade de consumidores, densidade de rede e pavimentação;

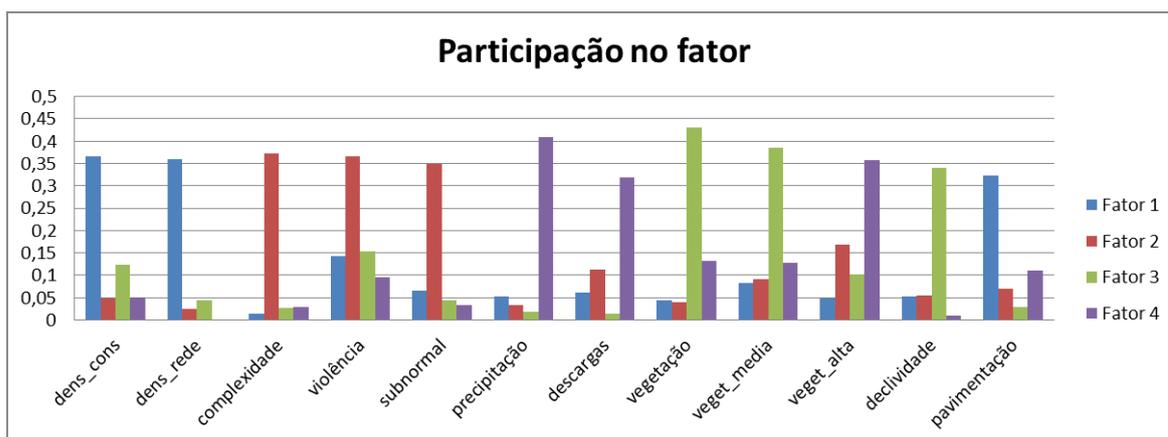


Figura 5.1: Participação nos fatores

- O fator 2 é composto primariamente pelas variáveis: complexidade, violência e subnormal;
- O fator 3 é composto primariamente pelas variáveis: vegetação, vegetação média e declividade;
- O fator 4 é composto primariamente pelas variáveis: precipitação, descargas e vegetação alta.

Similarmente aos resultados da atividade 1, considerando-se que a análise de fatores é efetuada para todas as 411 DMUs, o resultado da atividade 2, enviado pela seta B à atividade 3 da figura 4.2 não é apresentado de maneira explícita neste trabalho.

É válido notar que as variáveis dentro de cada fator possuem, pela própria conceituação de análise de fatores, alta correlação entre si. Além disso, como os fatores basicamente possuem variáveis bem determinadas, a significância desses fatores na regressão múltipla se refletirá na significância das próprias variáveis.

Na regressão linear, atividade 3 da figura 4.2, executada pelo SPSS, obteve-se valor de R^2 igual a 0,356. Isso significa que existe alguma correlação entre os fatores e a eficiência RCM, mas essa relação não é suficiente para que os fatores expliquem, por si só, a eficiência das distribuidoras. Esse fato reforça a tese de que pode não ser a estratégia mais adequada a utilização de regressão múltipla como ajuste final das eficiências, pois pode passar a interpretação de que as variáveis ambientais, demonstradas nos fatores, são as principais responsáveis pela eficiência, sendo os fatores intrínsecos à administração da empresa relegados a segundo plano, ou seja, caso se soubessem os valores das variáveis ambientais de uma dada distribuidora, e não se soubesse nada sobre as entradas e saídas

dessa mesma distribuidora, seria, ainda assim, possível determinar a eficiência dessa distribuidora, com base no ambiente em que estaria localizada. Entretanto, isso contradiz diretamente o resultado da correlação encontrada. Dessa maneira, pode ser mais adequado a utilização de clusterização, que assumiria igualdade de condições ambientais, porém o fator principal de eficiência permaneceria sendo as variáveis relacionadas à administração, que influenciam diretamente a eficiência, ou seja, as entradas e saídas do modelo DEA.

Para saber quais fatores afetam a eficiência encontrada, é necessário observar o nível de significância dos mesmos. Serão considerados significantes aqueles que obtiverem nível de significância inferior a 0,05 [26]. Os dados referentes aos níveis de significância foram obtidos pelo SPSS [23] e se encontram na tabela 5. Nessa tabela a primeira coluna refere-se à variável independente da regressão. A segunda coluna indica o coeficiente B de cada fator na regressão. A terceira coluna indica o valor de t, utilizado para o cálculo do nível de significância da quarta coluna. A quinta e sexta colunas indicam os limites inferiores e superiores dos intervalos de confiança de t para a significância, respectivamente.

	B	t	Sig.	Limite Inferior	Limite Superior
(Constante)	0,282	65,279	0,000	0,274	0,291
Fator 1	-0,05	-11,593	0,000	-0,059	-0,042
Fator 2	0,01	2,375	0,018	0,002	0,019
Fator 3	0,022	4,992	0,000	0,013	0,03
Fator 4	-0,027	-6,124	0,000	-0,035	-0,018

Tabela 5.1: Coeficientes da regressão múltipla

Pela tabela 5 nota-se que todos os fatores foram considerados importantes para a análise. Dessa forma, todas as variáveis ambientais contidas nesses fatores devem ser consideradas para a clusterização. Esse é o resultado da atividade 3, enviado pela seta C à atividade 4 da figura 4.2.

Para a clusterização não são mais necessários os dados de anos anteriores, pois não há interesse em descobrir a eficiência desses anos, apenas do ano de 2012, e a entrada dos dados anteriores não traria benefícios a essa etapa no quesito de confiabilidade. Ao contrário, poderia prejudicar a interpretação da clusterização, pois provavelmente cada distribuidora seria colocada no mesmo cluster que ela própria em anos anteriores, a menos que haja modificação ambiental entre esses anos. Como o propósito da clusterização é descobrir quais distribuidoras possuem similaridades ambientais, utilizar dados de anos anteriores prejudicaria a análise, pois provavelmente o objetivo seria distorcido. Dessa forma, optou-se por utilizar apenas os dados referentes a 2012 quando da realização da atividade 4 da figura 4.2.

A clusterização foi executada no MATLAB, partindo da utilização da matriz U e da

matriz de pesos para descobrir o número natural de neurônios. Para tanto, foi gerada uma rede 20x20 com 7000 iterações, valor escolhido pela estabilização da rede, e observados os resultados, presentes nas figuras 5.2 e 5.3. Na análise da Matriz U e da matriz de pesos, deve-se observar se há regiões claramente separadas do restante do mapa, etapa realizada visualmente através da coloração observada no mapa formado.

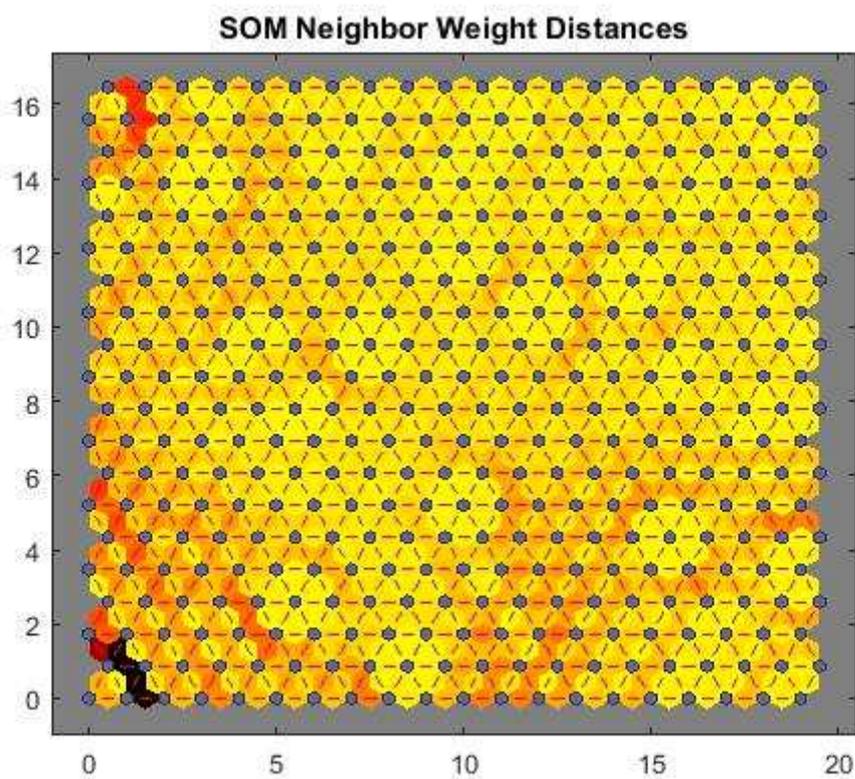


Figura 5.2: Matriz U

Observando a figura 5.2, percebe-se que não é possível extrair um número determinado de clusters formados naturalmente pela rede, ou seja, não existe diferenciação clara dos clusters formados, pois a distribuição não apresenta regiões bem definidas. Para melhor análise, observam-se os pesos, expressos na figura 5.3. É possível perceber um cluster na parte inferior esquerda do mapa, demonstrado pelos pesos 1, 2 e 12, e ao outro ponto também na parte inferior direita, demonstrado pelos pesos 4, 8 e 9. Também é visível um cluster no meio da parte superior da rede ao analisar os pesos 3, 4, 7, 8 e 9. Além disso, pelos pesos 3, 5, 6, e 10 é possível observar um cluster na parte superior esquerda. Portanto, uma sugestão natural de número de clusters para se montar a rede seria utilizar quatro clusters. Tal sugestão foi utilizada e o resultado da clusterização, enviado pela seta D à atividade 5 da figura 4.2, é apresentado na tabela 5.2.

Distribuidora	Cluster	Região
CELPE	1	PE
CEMIG	1	MG
COELBA	1	BA
COELCE	1	CE
COSERN	1	RN
ENE. PARAÍBA	1	PB
ENE. SERGIPE	1	SE
ESCELSA	1	ES
SANTA MARIA	1	RS
SULGIPE	1	SE
CEB	2	DF
PIRATININGA	2	SP
AMPLA	3	RJ
BRAGANTINA	3	SP
CEEE	3	RS
CELG	3	GO
CEMAR	3	MA
CPFL PAULISTA	3	SP
ELEKTRO	3	MS
ENERSUL	3	MS
NACIONAL	3	SP
AES SUL	4	RS
BANDEIRANTE	4	SP
CELESC	4	SC
CELPA	4	PA
CELTINS	4	TO
CEMAT	4	MT
COPEL	4	PR
DME-PC	4	MG
ELETROPAULO	4	SP
LIGHT	4	RJ
RGE	4	RS

Tabela 5.2: Resultado da clusterização

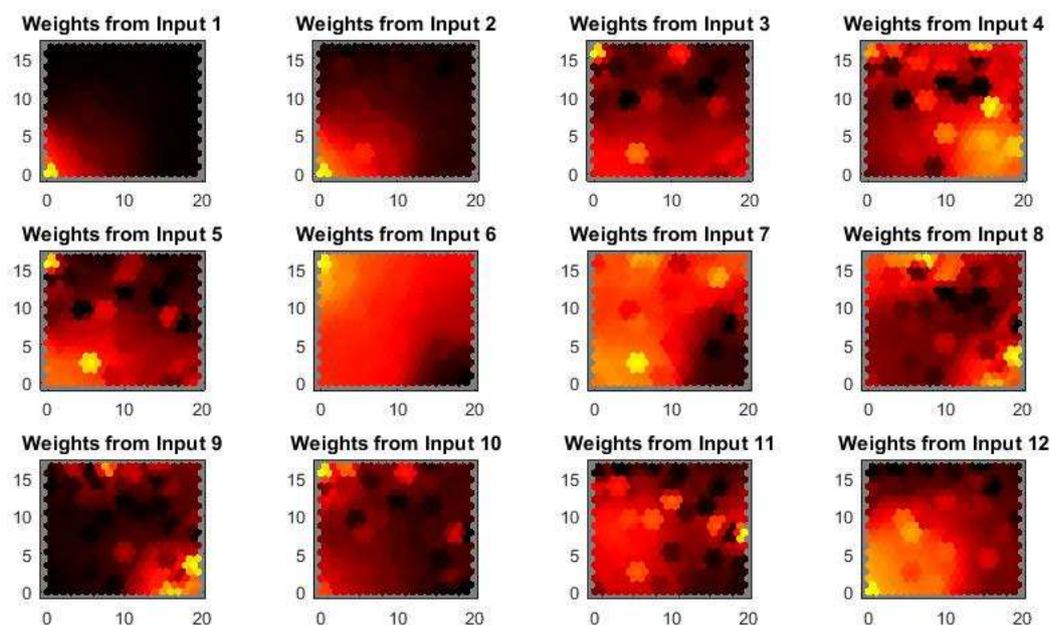


Figura 5.3: Matriz de pesos

Pela tabela 5.2, é possível observar que o cluster 1 ficou com 8 distribuidoras, o cluster 2 ficou com 7 distribuidoras, o cluster 3 englobou 4 distribuidoras e o cluster 4 conta com 9 distribuidoras. Para melhor visualização dos resultados da clusterização, os clusters foram representados em um mapa do Brasil na figura 5.4.

Na figura 5.4 percebe-se que há alguns estados sem representação de cluster. Isso ocorre devido à falta de informação sobre as distribuidoras atuantes nesses estados. Além disso, por não conhecer completamente a distribuição territorial das distribuidoras em cada estado, os estados com participação em mais de um cluster estão representados com as cores de todos os clusters em que participam. Cabe notar que a representação da figura pode não corresponder à realidade geográfica, em especial com relação a estados distribuídos entre 2 ou mais clusters. A divisão apresentada na figura 5.4 foi efetuada de modo a ser representativa da divisão dos estados por cluster de maneira puramente estética.

Pela análise da figura 5.4, é possível observar que o cluster 1 abrange principalmente as regiões do nordeste e parte do sudeste, enquanto o cluster 3 abrange parte do sudeste e do centro oeste, além de MA e parte do RS. O cluster 2 abrange parte de SP e DF. O cluster 4 engloba as regiões sul, parte do norte e centro oeste e parte de MG e RJ.

Para o cálculo das eficiências no modelo RCM aditivo não decrescente, atividade 5 da

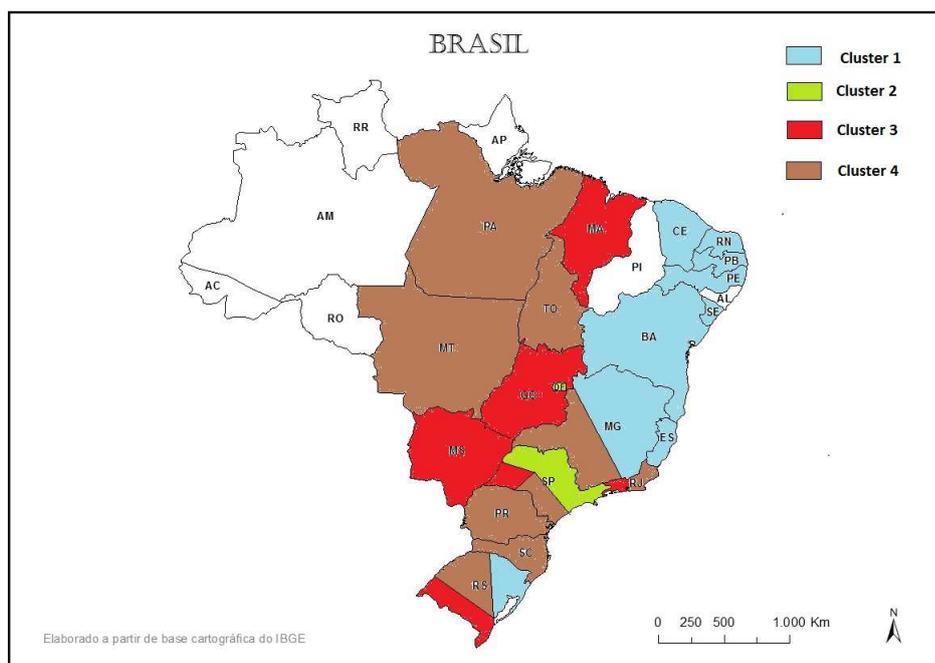


Figura 5.4: Divisão regional dos clusters

figura 4.2, foi utilizado o MATLAB, foram utilizadas todas as distribuidoras que possuíam dados em cada ano de análise, o que rendeu 411 eficiências. Tais eficiências são os dados enviados pela seta E da figura 4.2.

Na atividade 6 da figura 4.2, as eficiências do ano de 2012 foram separadas dos dados recebidos pela seta E da figura 4.2 e normalizadas por cluster, segundo a divisão recebida pela seta D da figura 4.2. Tal resultado e a comparação com o modelo da ANEEL se encontram na tabela 5.3.

Percebe-se pela tabela 5.3 que não há uma relação direta do resultado do modelo com o modelo da ANEEL. De fato, a correlação entre os modelos é de vinte e quatro por cento, com R quadrado de apenas 0,06. Tal divergência se dá por causa das grandes mudanças efetuadas, principalmente com relação ao modelo network DEA. A comparação entre o resultado dos modelo apresentado e o modelo da ANEEL pode ser visualizada graficamente na figura 5.5.

Pela figura 5.5, percebe-se que existe ganho de eficiência de algumas distribuidoras, enquanto há perda de eficiência por outras. Essas diferenças, se adotadas para o cálculo do fator X, refletem-se diretamente no valor tarifário permitido à cada distribuidora, afetando diretamente suas receitas e, conseqüentemente, seus lucros. Deve-se ressaltar que o valor tarifário é repassado diretamente ao consumidor, que arcará com eventuais aumentos ou será beneficiado pelas reduções tarifárias. Vale notar, entretanto, que existem fatores

Distribuidora	EFF Modelo	Eficiência Normalizada	Eficiência Aneel	Diferença
AES SUL	0,36	0,82	0,8	-0,02
AMPLA	0,23	0,54	0,7	0,16
BANDEIRANTE	0,19	0,44	0,77	0,33
BRAGANTINA	0,12	0,3	0,74	0,44
CEB	0,2	0,7	0,49	-0,21
CEEE	0,27	0,65	0,39	-0,26
CELESC	0,38	0,89	0,59	-0,3
CELG	0,29	0,69	0,66	-0,03
CELPA	0,39	0,89	0,5	-0,39
CELPE	0,49	1	0,79	-0,21
CELTINS	0,43	1	1	0
CEMAR	0,41	1	0,82	-0,18
CEMAT	0,34	0,79	0,71	-0,08
CEMIG	0,21	0,43	0,64	0,21
COELBA	0,44	0,9	0,89	-0,01
COELCE	0,43	0,87	0,95	0,08
COPEL	0,27	0,62	0,6	-0,02
COSERN	0,4	0,82	0,89	0,07
CPFL PAULISTA	0,19	0,46	0,84	0,38
DME-PC	0,25	0,58	0,47	-0,11
ELEKTRO	0,11	0,26	0,76	0,5
ELETROPAULO	0,28	0,64	0,67	0,03
ENE. PARAÍBA	0,34	0,7	0,75	0,05
ENE. SERGIPE	0,1	0,2	0,56	0,36
ENERSUL	0,35	0,85	0,69	-0,16
ESCELSA	0,38	0,77	0,78	0,01
LIGHT	0,41	0,94	0,69	-0,25
NACIONAL	0,33	0,79	0,65	-0,14
PIRATININGA	0,28	1	0,88	-0,12
RGE	0,19	0,43	1	0,57
SANTA MARIA	0,44	0,88	0,9	0,02
SULGIPE	0,33	0,68	0,66	-0,02

Tabela 5.3: Resultado das eficiências e comparação com o modelo da ANEEL

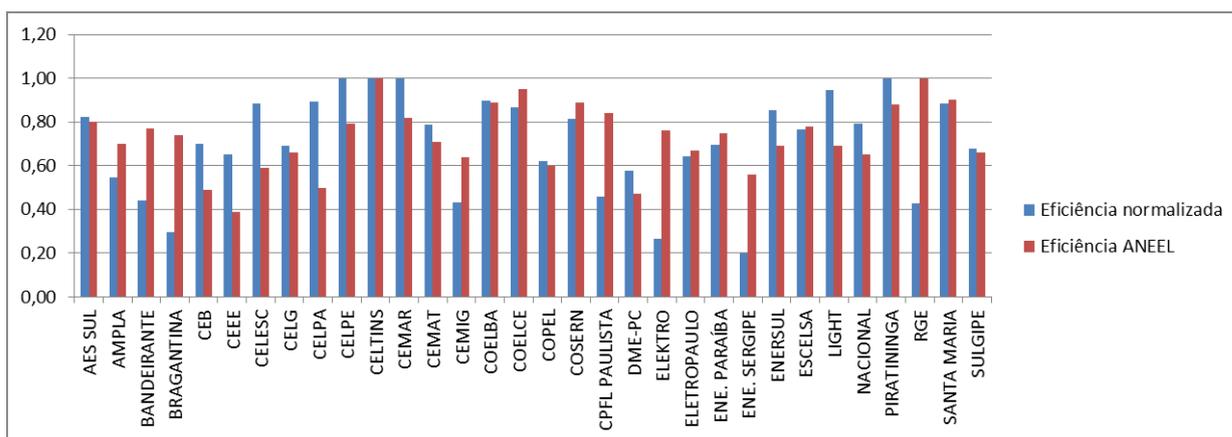


Figura 5.5: Comparação entre o resultado do modelo e da ANEEL

limitantes na análise efetuada, como o fato de não ser considerada a inflação e o OPEX não ser considerado também como entrada do consumo e do número de consumidores.

Capítulo 6

Conclusão

Foi proposto um modelo diferenciado para a avaliação de eficiência das distribuidoras de energia brasileiras, a fim de permitir os cálculos envolvendo o fator X, referente à produtividade no modelo price-cap.

De pronto, ressalta-se que a dúvida relacionada à utilização da rede como entrada ou saída do processo não existe no método de cálculo apresentado, tendo em vista que, como variável intermediária, é tanto saída do OPEX quanto entrada do consumo e do número de consumidores.

Há de se considerar, que, para o cálculo do fator X, houve a divisão das distribuidoras em alguns clusters, e pode ser interessante realizar a comparação dentro de cada cluster no cálculo final do fator X. Apesar dessa divisão, ainda é possível a comparação entre as distribuidoras, de forma que seja possível certa proporcionalidade nas correções das tarifas energéticas definidas que deverão ser cobradas, o que é um ponto positivo do método apresentado.

O método apresentado procura refletir aspectos do ambiente de atuação das distribuidoras, entretanto é razoável assumir que podem ocorrer distorções por causa da ausência de correção dos efeitos inflacionários.

O método de avaliação proposto apresentou diferenças de resultado no cálculo da eficiência com relação ao método atualmente utilizado pela ANEEL, com algumas distribuidoras tendo sua eficiência elevada, enquanto outras tiveram sua eficiência reduzida. Conclui-se que não há correlação direta entre os métodos.

Este trabalho pode ser utilizado em países de grande extensão territorial que possuam algum nível de influência do setor privado na distribuição de energia, ou em outros setores regulados, desde que se efetuem os devidos ajustes relacionados às variáveis a serem

utilizadas.

O método apresentado permite avaliação objetiva da eficiência, o que é relevante quando a discussão envolve direitos econômicos, como é o caso das distribuidoras brasileiras de energia. Além disso, a eficiência encontrada considera aspectos ambientais relacionados à produtividade das distribuidoras, mas não é definida por esses aspectos.

Capítulo 7

Trabalhos Futuros

Existem alguns pontos a serem considerados na construção de próximos trabalhos baseados neste. Um primeiro ponto de interesse é a inclusão do efeito da inflação no cálculo da eficiência da etapa 1 da figura 4.2. Essa inclusão pode afetar as eficiências encontradas, de modo que a análise histórica de eficiência seja mais condizente com a realidade. Isso implica na possibilidade de mudanças nas variáveis ambientais tidas como relevantes para a eficiência.

Além disso, pode ser interessante incluir o OPEX como entrada para as demais saídas do processo, utilizando o modelo de inputs compartilhados. Isso é importante, visto que essa variável também possui relação direta com as variáveis de saída consumo e número de consumidores, e não apenas com a rede. Tal modificação acarretaria mudanças diretas no modelo network DEA, o que pode alterar a eficiência final das distribuidoras.

Não obstante, o método de clusterização utilizado pode não ser necessariamente o mais eficiente. É interessante verificar o resultado da análise com número diferente de agrupamentos. Ainda, pode ser relevante a comparação do resultado obtido através de mapas de Kohonen com outros métodos de clusterização alternativos, tal como k-means.

Referências

- [1] AER, A. E. R. Annual report 2012 - 2013, inaugural report. Relatório Técnico, AER, Australia, 2012.
- [2] ARCOS-VARGAS, A., NÚÑEZ-HERNÁNDEZ, F., VILLA-CARO, G. A dea analysis of electricity distribution in spain: An industrial policy recommendation. *Energy Policy* 102 (2017), 583–592.
- [3] AZEVEDO-COSTA, M., MIRANDA-LOPES, A., DE PINHO MATOS, G. B. B. Statistical evaluation of data envelopment analysis versus cols cobb-douglas benchmarking models for the 2011 brazilian tariff revision. *Socio-Econ. Plan. Sci.* 49 (2015), 47–60.
- [4] BALDWIN, R., CAVE, M., LODGE, M. *Understanding Regulation*, 2 ed. Oxford University Press, 2012.
- [5] BANKER, R. D., CHARNES, A., COOPER, W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30 (1984), 1078–1092.
- [6] BAPTISTELLA, M., STEINER, M. T. A., NETO, A. C. O uso de redes neurais e regressão linear múltipla na engenharia de avaliações: determinação dos valores venais de imóveis urbanos. Em *XXXVIII Simpósio Brasileiro de Pesquisa Operacional* (2006), p. –.
- [7] BORKOWSKA, E. M., KRUK, A., JEDRZEJCZYK, A., ROZNIECKI, M., JABLONOWSKI, Z., TRACZYK, M., CONSTANTINOU, M., BANASZKIEWICZ, M., PIETRUSINSKI, M., SOSNOWSKI, M., HAMDY, F. C., PETER, S., CATTO, J. W., KALUZIEWSKI, B. Molecular subtyping of bladder cancer using kohonen self organizing maps. *Cancer Medicine* 3 (2014), 1225–1234.
- [8] CHARNES, A., COOPER, W., RHODES, E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2 (1978), 429–444.
- [9] CHEN, C., YAN, H. Network dea model for supply chain performance evaluation. *European Journal of Operational Research* 213 (2011), 147–155.
- [10] COLLEGE, O. *Principles of Economics*, 1 ed. Textbook Equity, 2014.
- [11] COOK, W. D., ZHU, J., BI, G., YANG, F. Network dea: Additive efficiency decomposition. *European Journal of Operational Research* 207 (2010), 1122–1129.
- [12] CORTON, M. L., ZIMMERMANN, A., PHILLIPS, M. A. The low cost of quality improvements in the electricity distribution sector of brazil. *Energy Policy* 97 (2016), 485–493.

- [13] CUNHA, T., ANGULO-MEZA, L. A avaliação cruzada: uma revisão bibliográfica e implementação computacional. Em *XXXVII Simpósio brasileiro de pesquisa operacional -SBPO* (2006), p. –.
- [14] DA SILVA, M. A. S. Mapas auto-organizáveis na análise exploratória de dados geoespaciais multivariados. Dissertação de Mestrado, INPE, São José dos Campos, 2005.
- [15] DE MELLO, J. S., ANGULO-MEZA, L., NETO, L. B. Curso de análise de envoltória de dados. Em *XXXVII Simpósio Brasileiro de Pesquisa Operacional (SBPO)* (2005), p. 1–28.
- [16] DE OLIVEIRA SILVA, R., DA SILVA CHRISTO, E., COSTA, K. A. Analysis of residual autocorrelation in forecasting energy consumption through a java program. *Advanced Materials Research 962* (2014), 1753–1756.
- [17] DE REGULAÇÃO ECONÔMICA SRE, S. Análise de eficiência dos custos operacionais das distribuidoras de energia elétrica. Relatório Técnico, ANEEL, Brasília, 2014.
- [18] DI TOLLO, G., TANEV, S., DAVIDE, D. M., MA, Z. Neural networks to model the innovativeness perception of co-creative firms. *Expert Systems with Applications 39* (2012), 12719–12726.
- [19] EMROUZNEJAD, A., LIANG YANG, G. A survey and analysis of the first 40 years of scholarly literature in dea: 1978 - 2016. *Socio-Economic Planning Sciences 61* (2018), 4–8.
- [20] ET AL., G. N. A. Evaluating electricity distributors efficiency using self-organizing map and data envelopment analysis. *IEEE LATIN AMERICA TRANSACTIONS 12* (2014), 1456–1464.
- [21] HAYKIN, S. *Neural networks: a comprehensive foundation*, 1 ed. Prentice Hall PTR, 1994.
- [22] HOFFMANN, R. *Análise de regressão : uma introdução à econometria*, 1 ed. O autor, 2005.
- [23] IBM. Spss.
- [24] JEBALI, E., ESSID, H., KHRAIEF, N. The analysis of energy efficiency of the mediterranean countries: A two-stage double bootstrap dea approach. *Energy 134* (2017), 991–1000.
- [25] JÚNIOR, S. F. G., BELTRÁN, P. M., DE MELLO, J. C. C. B. S., ANGULO-MEZA, L. Utilização de modelo network dea na avaliação de cursos de pós-graduação stricto sensu em engenharia. Em *XVII Simpósio de Pesquisa Operacional e Logística da Marinha* (2014), p. –.
- [26] JOSEPH F. HAIR, J., ANDERSON, R. E., TATHAM, R. L., BLACK, W. C. *Análise multivariada de dados*, 5 ed. Bookman, 2005.
- [27] KAHN, A. E. *The Economics of Regulation: Institutional Issues*, 1 ed. Wiley, New York, 1971.

- [28] KOHONEN, T. The self-organizing map. Em *Proceedings of the Institute of Electrical and Electronics Engineers* (1990), p. 1464–1480.
- [29] KRUNGMAN, P., WELLS, R. *Introdução à Economia*, 1 ed. Elsevier, 2007.
- [30] LEME, R. C., PAIVA, A. P., SANTOS, P. E. S., BALESTRASSI, P. P., DE LIMA GALVÃO, L. Design of experiments applied to environmental variables analysis in electricity utilities efficiency: The brazilian case. *Energy Economics* 45 (2014), 111–119.
- [31] LINS, M. E., DA SILVA, A. M., LOVELL, C. Avoiding infeasibility in dea models with weight restrictions. *European Journal of Operational Research* 181 (2006), 956–966.
- [32] LOPES, D. G. Custos de transação e direitos de propriedade: Arranjos institucionais da comercialização de energia elétrica no ambiente de contratação livre brasileiro. Dissertação de Mestrado, UFSC, Florianópolis, 2015.
- [33] MACHADO, L. G., DE MELLO, J. C. C. B. S., ROBOREDO, M. C. Efficiency evaluation of brazilian electrical distributors using data envelopment analysis game and cluster analysis. *IEEE Latin America Transactions* 14 (2016), 4499–4505.
- [34] MARDANI, A., ZAVADSKAS, E. K., STREIMIKIENE, D., JUSOH, A., KHOSHNOUDI, M. A comprehensive review of data envelopment analysis (dea) approach in energy efficiency. *Renewable and Sustainable Energy Reviews* 70 (2017), 1298–1322.
- [35] MATHWORKS. Matlab.
- [36] MCFARLAND, D. J. Simulating the effects of common and specific abilities on test performance: An evaluation of factor analysis. *Journal of Speech, Language and Hearing Research* 57 (2014), 1919–1928.
- [37] M.LÓPEZ, VALERO, S., SENABRE, C., APARICIO, J., GABALDON, A. Application of some neural networks to short-term load forecasting: The spanish electricity market case study. *Electric Power Research* 91 (2012), 18–27.
- [38] MORENO, P., ANDRADE, G., ANGULO-MEZA, L., DE MELLO, J. C. B. S. Evaluation of brazilian electricity distributors using a network dea model with shared inputs. *IEEE Latin America Transactions* 13 (2015), 2209–2216.
- [39] MORGAN, B., YEUNG, K. *Introduction to Law and Regulation: Text and Materials*, 1 ed. Cambridge University Press, 2007.
- [40] MOUNTAIN, B. Independent regulation of government-owned monopolies: An oxymoron? the case of electricity distribution in australia. *Utilities Policy* 31 (2014), 188–196.
- [41] NEPAL, R., FOSTER, J. Electricity networks privatization in australia: An overview of the debate. *Economic Analysis and Policy* 48 (2015), 12–24.
- [42] NEPAL, R., MENEZES, F., JAMASB, T. Network regulation and regulatory institutional reform: Revisiting the case of australia. *Energy Policy* 73 (2014), 259–268.

-
- [43] OUM, T. H., PATHOMSIRI, S., YOSHIDA, Y. Limitations of dea-based approach and alternative methods in the measurement and comparison of social efficiency across firms in different transport modes: An empirical study in japan. *Transportation Research E* 57 (2013), 16–26.
- [44] PIRES, J. C. L., PICCININI, M. S. Modelos de regulação tarifária do setor elétrico. *Revista do BNDES* 5 (1998), 147–168.
- [45] RANDRIANARISOA, L. M., BOLDUC, D., CHOO, Y. Y., OUM, T. H., YAN, J. Effects of corruption on efficiency of the european airports. *Transportation Research Part A* 79 (2015), 65–83.
- [46] RENCHER, A. C. *Methods of Multivariate Analysis*, 2 ed. Wiley-Interscience, 1934.
- [47] SASSI, C. P., PEREZ, F. G., MYAZATO, L., YE, X., FERREIRA-SILVA, P. H., LOUZADA, F. Modelos de regressão linear múltipla utilizando os softwares e estatística: uma aplicação a dados de conservação de frutas. *ICMC – USP CP668* (2011), –.
- [48] SKIDMORE, S. T., THOMPSON, B. Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement* 70 (2010), 777–795.
- [49] TSCHAFFON, P. B., ANGULO-MEZA, L. Um estudo de outputs indesejáveis em dea com aplicação no setor de distribuição de energia elétrica. Em *XLIII Simpósio brasileiro de pesquisa operacional -SBPO* (2011), p. –.
- [50] XAVIER, S., LIMA, J. W. M., LIMA, L. M. M., LOPES, A. L. M. How efficient are the brazilian electricity distribution companies? *Journal of Control, Automation and Electrical Systems* 26 (2015), 283–296.