



**Marcus Vinicius Pereira de Souza**

**Uma Abordagem Bayesiana para o Cálculo  
dos Custos Operacionais Eficientes das  
Distribuidoras de Energia Elétrica**

**Tese de Doutorado**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro

Maio de 2008



**Marcus Vinicius Pereira de Souza**

**Uma Abordagem Bayesiana para o Cálculo  
dos Custos Operacionais Eficientes das  
Distribuidoras de Energia Elétrica**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Reinaldo Castro Souza**  
**Orientador**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. Madiagne Diallo**  
**Co-orientador**

Departamento de Engenharia Industrial

**Prof. José Francisco Moreira Pessanha**  
CEPEL

**Prof. Tufi Machado Soares**  
Universidade Federal de Juiz de Fora

**Profa. Mariane Branco Alves**  
UERJ

**Prof. Marcos Azevedo da Silveira**  
Departamento de Engenharia Elétrica – PUC-Rio

**Prof. José Eugenio Leal**  
Coordenador Setorial do Centro  
Técnico Científico

Rio de Janeiro, 13 de maio de 2008

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Marcus Vinicius Pereira de Souza**

Nascido em Andrelândia-MG, 1972. Graduou-se em Engenharia Elétrica (1999) pela Universidade Federal de Juiz de Fora (UFJF). Mestre em Engenharia Aeronáutica e Mecânica, Área de Produção (2002) pelo Instituto Tecnológico de Aeronáutica (ITA). Suas pesquisas de interesse incluem as áreas de redes neurais artificiais, inteligência artificial, estatística aplicada e pesquisa operacional.

#### Ficha Catalográfica

Souza, Marcus Vinicius Pereira de

Uma abordagem bayesiana para o cálculo dos custos operacionais eficientes das distribuidoras de energia elétrica / Marcus Vinicius Pereira de Souza ; orientador: Reinaldo Castro Souza. – 2008.

153 f. ; 30 cm

Tese (Doutorado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Análise envoltória de dados. 3. Fronteiras estocásticas. 4. Rede neural de Kohonen. 5. Regulação econômica. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

*Esta Tese é dedicada à memória do meu  
amado Pai, Sebastião Teixeira de Souza, que  
certamente está num plano maior  
intercedendo por mim; e também para minha  
querida Mãe Elizabeth Meer Pereira de  
Souza, essencial em todos os momentos da  
minha vida.*

## Agradecimentos

A Deus, à Nossa Senhora Aparecida, aos Mestres da Grande Fraternidade Branca e a todos os Gloriosos Santos por mais esta conquista.

Consigno meus sinceros agradecimentos e respeito ao professor Dr. Reinaldo Castro Souza pela oportunidade; orientando-me com paciência, dedicação e incentivo em prol desta tese.

Também quero externar o meu apreço sincero ao professor Dr. José Francisco Moreira Pessanha que aconselhou-me e encorajou-me durante toda essa trajetória.

Ao grande amigo e professor Dr. Madiagne Diallo, pelo apoio moral e constante estímulo na concretização desta importante etapa da minha vida.

Quero ainda expressar minha gratidão à Helaine Aparecida da Silva pelos conselhos, incentivos e sua valiosa ajuda na efetivação desta tese.

Muitas outras pessoas contribuíram para o sucesso deste trabalho, em especial: Airam Regina de Aquino Martins, Alexandra Mello Schmidt, Alexandre Bessa dos Santos, Alexandre Zanini, Álvaro Veiga, Ana Paiva, André Luis Marques Marcato, Arildo de Jesus Batista, Armando Zeferino Milioni, Cairo Lúcio Nascimento Jr., Carlos Galeno Ladeira Pereira, Célia Maria de Souza Pereira, Cristiano Augusto C. Fernandes, Eduardo Eliphaz Pereira, Ernesto Fagundes Neto, Eurides Moura, Fátima Mendes de Carvalho, Felipe Fortes Carvalho Silva, Fernando César Coelli, Fernando Neves Breseguello, Gisele Teixeira Rocha, Hélio Francisco da Silva, Isnarde Antônio Ernesto, Jacqueline Furtado Vital, José Luiz Bellini Leite, José Marcelino de Jesus, Juliana Garcia Cespedes, Lúcia Helena Lima Miranda e Silva, Lúcia Margareth Pereira, Luiz Tarciso de Andrade, Marcia Brum Guerra Gomes, Márcia Doná, Márcio Vicente Rizzo, Maria Alcina Portes, Maria Aparecida Gonçalves, Mariane Antunes Maia, Marley M. B. Rebuzzi Vellasco, Mischel Carmen Neyra Belderrain, Mônica Barros, Nair de Carvalho Pereira, Neimar Fernandes, Paula Adriana Soares, Pedro Gomes Barbosa, Ricardo Tanscheit, Sebastião Marcos M. P. Vasconcelos, Tatiane Cristina da Silva, Vander Menengoy da Costa, Wilma Galgane.

À **PUC-Rio**, pelo suporte financeiro do projeto.

Finalmente, mas não menos importante, desejo registrar a minha admiração a todos que me honraram apresentando críticas e sugestões concernentes ao meu trabalho, em particular: Elizabeth Meer Pereira de Souza, Dr. Marcos Azevedo da Silveira, Dra. Mariane Branco Alves e Dr. Tufi Machado Soares.

## Resumo

Souza, Marcus Vinicius Pereira de; Souza, Reinaldo Castro (Orientador). **Uma Abordagem Bayesiana para o Cálculo dos Custos Operacionais Eficientes das Distribuidoras de Energia Elétrica**. Rio de Janeiro, 2008. 153p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro

Esta tese apresenta os principais resultados de medidas de eficiência dos custos operacionais de 60 distribuidoras brasileiras de energia elétrica. Baseado no esquema *yardstick competition*, foi utilizado uma Rede Neural de Kohonen (KNN) para identificar grupos de empresas similares. Os resultados obtidos pela KNN não são determinísticos, visto que os pesos sinápticos da rede são inicializados aleatoriamente. Então, é realizada uma simulação de Monte Carlo para encontrar os *clusters* mais frequentes. As medidas foram obtidas por modelos DEA (*input oriented*, com e sem restrições aos pesos) e modelos Bayesianos e frequencistas de fronteira estocástica (utilizando as funções Cobb-Douglas e Translog). Em todos os modelos, DEA e SFA, a única variável *input* refere-se ao custo operacional (OPEX). Os índices de eficiência destes modelos representam a potencial redução destes custos de acordo com cada concessionária avaliada. Os *outputs* são os *cost drivers* da variável OPEX: número de unidades consumidoras (uma *proxy* da quantidade de serviço), montante de energia distribuída (uma *proxy* do produto total) e a extensão da rede de distribuição (uma *proxy* da dispersão dos consumidores na área de concessão). Finalmente, vale registrar que estas técnicas podem mitigar a assimetria de informação e aprimorar a habilidade do agente regulador em comparar os desempenhos das distribuidoras em ambientes de regulação incentivada.

## Palavras-chave

Análise Envoltória de Dados; Fronteiras Estocásticas, Rede Neural de Kohonen, Regulação Econômica.

## Abstract

Souza, Marcus Vinicius Pereira de; Souza, Reinaldo Castro (Advisor). **A Bayesian Approach to Estimate the Efficient Operational Costs of Electrical Energy Utilities**. Rio de Janeiro, 2008. 153p. Doctorate Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro

This thesis presents the main results of the cost efficiency scores of 60 Brazilian electricity distribution utilities. Based on yardstick competition scheme, it was applied a Kohonen Neural Networks (KNN) to identify and to group the similar utilities. The KNN results are not deterministic, since the estimated weights are randomly initialized. Thus, a Monte Carlo simulation was used in order to find the most frequent clusters. Therefore was examined the use of the DEA methodology (input oriented, with and without weight constraints) and Bayesian and non-Bayesian Stochastic Frontier Analysis (centered on a Cobb-Douglas and Translog cost functions) to evaluate the cost efficiency scores of electricity distribution utilities. In both models the only input variable is operational cost (OPEX). The efficiency measures from these models reflect the potential of the reduction of operational costs of each utility. The outputs are the cost-drivers of the OPEX: the number of customers (a proxy for the amount of service), the total electric power supplied (a proxy for the amount of product delivered) and the distribution network size (a proxy of the customers scattering in the operating territory of each distribution utility). Finally, it is important to mention that these techniques can reduce the information asymmetry to improve the regulator's skill to compare the performance of the utilities in incentive regulation environments.

## Keywords

Data Envelopment Analysis; Stochastic Frontier Analysis, Kohonen Neural Networks, Economic Regulation.

## Sumário

1. Introdução	18
1.1. Considerações gerais	18
1.2. Revisão bibliográfica	23
1.3. Objetivos	24
1.4. Estrutura da tese	26
2. Redes Neurais Auto-Organizáveis	29
2.1. Introdução	29
2.2. Componentes de uma tarefa de <i>clustering</i>	29
2.3. Aprendizado por competição	31
2.4. Rede Neural de Kohonen (RNK)	32
2.5. Exemplo ilustrativo da Rede Neural de Kohonen	34
2.6. Rede Neural de Kohonen via simulação de Monte Carlo (MCRNK)	38
3. Análise Envoltória de Dados	40
3.1. Introdução	40
3.2. Conjunto de Possibilidades de Produção	42
3.3. Conceito de eficiência	44
3.4. <i>Benchmarks</i>	46
3.5. Retornos de escala	47
3.6. Modelos DEA	48
3.6.1. Modelo DEA CCR <i>input-oriented</i>	49
3.6.2. Modelo DEA BCC <i>input-oriented</i>	53
3.6.3. Outros modelos DEA	56
3.6.4. Seleção de variáveis <i>inputs</i> e <i>outputs</i>	57
3.6.5. Técnicas baseadas em Restrições aos Pesos	58
3.7. <i>Adjusted contingent restrictions on weights</i>	60
4. Conceitos Básicos de Estatística Bayesiana e Simulação Estocástica	62
4.1. Elementos de inferência Bayesiana	62
4.2. Distribuição a priori	63
4.3. Distribuição amostral (Função de Verossimilhança)	64
4.4. Distribuição a posteriori	65
4.5. Exemplo	65
4.6. Obtenção de resumos de interesse através de simulação	68
4.6.1. Integração via Monte Carlo	69
4.7. Monte Carlo via Cadeias de Markov (MCMC)	72
4.7.1. Algoritmo de Metropolis-Hastings	74
4.7.2. Amostrador de Gibbs	77
4.7.3. Monitoração da convergência	82
5. Modelos Econométricos	84
5.1. Fronteira Estocástica de Produção	84
5.2. Fronteira Estocástica de Custos	89
5.3. Fronteira Estocástica sob o paradigma Bayesiano	91
5.3.1. Modelos Bayesianos para Fronteira Estocástica	92



5.3.2. <i>Varying Efficiency Distribution (VED) models</i>	93
5.4. Inferência Bayesiana usando Amostrador de Gibbs	95
6. <i>Clustering</i> : Definindo Grupos Estratégicos	98
6.1. Introdução	98
6.2. Análise de dados	99
6.3. Resultados obtidos via <i>cluster analysis</i>	99
7. Mensuração da Eficiência Relativa utilizando Modelos DEA	105
7.1. Introdução	105
7.2. Resultados DEA e discussões	105
8. Estimação da Eficiência Técnica usando Fronteiras Estocásticas	110
8.1. Introdução	110
8.2. Resultados dos modelos econométricos com enfoque clássico	110
8.3. Resultados dos modelos econométricos com enfoque Bayesiano	115
9. Análise Multivariada de Dados	123
9.1. Introdução	123
9.2. Matriz de correlação de Pearson e Spearman	123
9.3. Análise Fatorial	127
10. Conclusões e Perspectivas	133
10.1 Conclusões	133
10.2 Perspectivas	135
Referências bibliográficas	138
Anexo 1 - Conjunto de Dados	149
Anexo 2 - Modelo Bayesiano	150

## Lista de figuras

Figura 1.1: Processo de revisão tarifária [4]	19
Figura 2.1: Etapas de um processo de <i>clustering</i>	30
Figura 2.2: Rede de Kohonen com as unidades de saída organizadas como grade bi-dimensional	33
Figura 2.3: Disposição bi-dimensional das entradas	35
Figura 2.4: Disposição das unidades no espaço das entradas	36
Figura 2.5: Resultado do processo de identificação dos <i>clusters</i>	38
Figura 3.1: Comparação de DEA e regressão [18]	41
Figura 3.2: Eficiências Técnica, Alocativa e Econômica	45
Figura 3.3: Envoltória determinada pelo modelo CCR <i>input-oriented</i> [18]	52
Figura 3.4: Envoltória determinada pelos modelos CCR e BCC ( <i>input-oriented</i> ) [18]	55
Figura 4.1: Densidades a priori do parâmetro $\theta$	66
Figura 4.2: Função de Verossimilhança para $\theta$	67
Figura 4.3: Densidades a posteriori do parâmetro $\theta$	68
Figura 4.4: <i>Scatterplot</i> dos valores simulados	71
Figura 4.5: Valores simulados através do algoritmo Metropolis-Hastings	76
Figura 4.6: Algoritmo Metropolis-Hastings na geração da distribuição de Cauchy	77
Figura 4.7: Dados simulados para $\lambda$ e $\beta$ utilizando o algoritmo de Gibbs	80
Figura 4.8: Resultado das médias amostrais para o parâmetro $\lambda$	80
Figura 4.9: Distribuição marginal estimada de $p(x)$	81
Figura 4.10: Gráficos de autocorrelação amostral dos parâmetros $\lambda$ e $\beta$	83
Figura 5.1: Densidades de probabilidade da combinação de $v_j + u_j$	91
Figura 6.1: Resultado de um processo de <i>clustering</i> em uma simulação	100
Figura 6.2: Histograma dos <i>clusters</i>	101
Figura 6.3: Histograma dos 13 <i>clusters</i> mais frequentes	103
Figura 7.1: <i>Scatterplot Matrix</i> dos índices de eficiência (Modelos DEA)	108
Figura 8.1: <i>Boxplot</i> dos índices de eficiência (Modelos SFA Clássica)	114
Figura 8.2: Representação gráfica dos índices de eficiência (Modelos Bayesianos)	119

Figura 8.3: Estimativa pontual dos coeficientes de regressão para diferentes valores de $r^*$	122
Figura 9.1: Coeficientes de correlação de Pearson e Spearman	125
Figura 9.2: Coeficientes de correlação de Pearson e Spearman	126
Figura 9.3: Coeficientes de correlação de Pearson e Spearman	127
Figura 9.4: Gráfico qui-quadrado (Q-Q <i>plot</i> )	128
Figura 9.5: Gráfico de dispersão dos <i>loadings</i> (sem rotação ortogonal).	130
Figura 9.6: Gráfico de dispersão dos <i>loadings</i> (com rotação ortogonal).	130
Figura 9.7: Gráfico de dispersão de objetos segundo os <i>factor scores</i>	131

## Lista de tabelas

Tabela 2.1 - Centro dos <i>clusters</i> com seus respectivos pontos de treinamento	34
Tabela 3.1 - Representação algébrica de retornos de escala	47
Tabela 3.2 - Relação entre retornos de escala e elasticidade total da produção	48
Tabela 6.2 - <i>Clusters</i> obtidos pela Rede de Kohonen em uma simulação	101
Tabela 6.3 - Sequência ordenada dos <i>clusters</i> mais frequentes	102
Tabela 6.4 - Agrupamentos finais	104
Tabela 7.1 - Variáveis consideradas	105
Tabela 7.2 - Índices de eficiência $\theta_j^{DEA}$	107
Tabela 7.3 - Estatística descritiva dos modelos DEA	108
Tabela 8.1 - Índices de eficiência $\theta_j^{SFA-C / HN}$	112
Tabela 8.2 - Índices de eficiência $\theta_j^{SFA-C / NT}$	113
Tabela 8.3 - Índices de eficiência $r_j^{SFA-B}$	117
Tabela 8.4 - Índices de eficiência $r_j^{SFA-B}$	118
Tabela 8.5 - Estimativa dos coeficientes de regressão e parâmetros do modelo de fronteira estocástica (enfoque Bayesiano)	120
Tabela 8.6 - Índices de eficiência $r_j^{SFA-B / NT-TL}$ ; ( $r^* = 0,95$ )	121
Tabela 9.1 - Coeficientes de Correlação de Pearson	124
Tabela 9.2 - Coeficientes de Correlação de Spearman	124
Tabela 9.3 - Comparação entre os <i>factor loadings</i> não rotacionados e rotacionados	129
Tabela 9.4 - <i>Ranking</i> das concessionárias	132

## Símbolos, Acrônimos e Abreviaturas

**d.p.** - desvio-padrão;

$e$  - vetor de 1's;

$e$  - Elasticidade Total;

**exp** - função exponencial;

$\text{Exp}(\cdot)$  - distribuição exponencial com parâmetro  $(\cdot)$ ;

$E(X)$  - valor esperado da variável aleatória  $X$ ;

$f(x|\theta)$  - densidade de  $X$ , condicional ao parâmetro  $\theta$ ;

$h(\cdot)$  - função Translog (TL) ou Cobb-Douglas (CD);

$i$  - indexador de *inputs*,  $(i=1, \dots, m)$ ;

*i.i.d* - independentes e identicamente distribuídos;

$j$  - indexador de DMU's,  $(j=1, \dots, n)$ ;

$j_0$  - indexador da DMU que está sendo analisada;

$jt$  - indexador de DMU's eficientes;

$l$  - função de verossimilhança;

**ln** - logaritmo natural;

$L$  - logaritmo natural da função de verossimilhança;

$m$  - número de *inputs*;

**max**( $\cdot$ ) - máximo de  $(\cdot)$ ;

**min**( $\cdot$ ) - mínimo de  $(\cdot)$ ;

$n$  - número de DMU's;

$\text{Poi}(\cdot)$  - distribuição de Poisson com parâmetro  $(\cdot)$ ;

$r$  - indexador de *outputs*,  $(r=1, \dots, s)$ ;

$\mathbf{R}_*^+$  - Conjunto dos números Reais inteiros positivos;

$\mathbf{R}_+^s$  - Conjunto dos números Reais inteiros positivos de dimensão  $s$ ;

$\mathbf{R}_+^m$  - Conjunto dos números Reais inteiros positivos de dimensão  $m$ ;

$s$  - número de *outputs* (DEA);

**T** - Conjunto de Possibilidades de Produção;

$u_r$  - peso atribuído ao  $r$ -ésimo *output*, modelos (DEA) *input-oriented*;

$v_i$  - peso atribuído ao  $i$ -ésimo *input*, modelos (DEA) *input-oriented*;  
 $\mathbf{W}_i^k$  - vetor de pesos sinápticos do neurônio ( $i$ ) no instante ( $k$ );  
 $X_j$  = vetor de variáveis aleatórias (*inputs*) da DMU  $j$  com dimensão  $1 \times m$ ;  
 $\mathbf{X}$  - matriz de *inputs* observados de dimensão  $n \times m$ ;  
 $\mathbf{x}_j$  = vetor de *inputs* observados da DMU  $j$  com dimensão  $1 \times m$ ;  
 $\mathbf{x}$  = vetor de realizações;  
 $Y_j$  = vetor de variáveis aleatórias (*outputs*) da DMU  $j$  com dimensão  $1 \times s$ ;  
 $\mathbf{Y}$  - matriz de *outputs* observados de dimensão  $n \times s$ ;  
 $\mathbf{y}_j$  = vetor de *outputs* observados da DMU  $j$  com dimensão  $1 \times s$ ;  
 $\mathbf{y}$  = vetor de realizações;  
 $\Gamma(a, b)$  ou Gama( $a, b$ ) - distribuição Gama com parâmetros  $a$  e  $b$ ;  
 $Unif(a, b)$  - distribuição Uniforme com parâmetros  $a$  e  $b$ ;  
 $\pi(\cdot)$  - distribuição a posteriori;  
 $[\cdot]^T$  - transposto de  $[\cdot]$ ;  
 $\beta$  - coeficientes da fronteira estocástica;  
 $\theta$  - índice de eficiência no modelo DEA CCR e vetor de parâmetros (na abordagem Bayesiana) dependendo do contexto;  
 $\theta^*$  - valor ótimo do índice de eficiência no modelo CCR *input-oriented*;  
 $\phi(\cdot)$  - função de densidade da Normal-padrão;  
 $\Phi(\cdot)$  - função de distribuição acumulada da Normal-padrão;  
 $\varepsilon$  - infinitésimo positivo ou erro composto dependendo do contexto;  
 $\prod$  - produto;

**ANEEL** - Agência Nacional de Energia Elétrica;  
**BCC** - Banker, Charnes e Cooper;  
**CB** - Cobb-Douglas;  
**CCR** - Charnes, Cooper e Rhodes;  
**CED** - *Common Efficiency Distribution*;  
**DEA** - *Data Envelopment Analysis*;  
**DMU** - *Decision Making Unit*;  
**EA** - Eficiência Alocativa;

**EE** - Eficiência Econômica;  
**EEF** - Equilíbrio Econômico Financeiro;  
**ET** - Eficiência Técnica;  
**HN** - Distribuição *Half-Normal*;  
**IRT** - Índice de Reajuste Tarifário;  
**LINDO** - *Linear Interactive and Discrete Optimizer*;  
**MCMC** – Monte Carlo via Cadeias de Markov;  
**MCRNK** – Redes Neurais de Kohonen via simulação de Monte Carlo;  
**NT** - Distribuição Normal-Truncada;  
**PL** - Programação Linear;  
**PPL** - Problema de Programação Linear;  
**RCE** - Retorno Constante de Escala;  
**RNA** - Redes Neurais Artificiais;  
**RVE** - Retorno Variável de Escala;  
**SEB** - Sistema Elétrico Brasileiro;  
**SFA** – *Stochastic Frontier Analysis*;  
**TL** - Translog;  
**VED** - *Varying Efficiency Distribution*;  
**WinBUGS** - *Bayesian inference Using Gibbs Sampling for Windows*.

"Bem-aventurado o homem que acha Sabedoria, e o homem que adquire  
Conhecimento."  
(Provérbios 3-13)



“Ninguém ignora tudo, ninguém sabe tudo. Por isso, aprendemos sempre”.

(Paulo Freire)