



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA

Tese

Diagnóstico de perdas comerciais de energia elétrica
na distribuição usando redes Bayesianas

Paulo Roberto F. de Moura Bastos

Campina Grande

Março de 2011

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA

Diagnóstico de perdas comerciais de energia elétrica
na distribuição usando redes Bayesianas

Tese apresentada à Coordenação do Programa de Pós-graduação em Engenharia Elétrica da UFCG, em cumprimento às exigências do Programa de Doutorado em Ciências no domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Energia.

Paulo Roberto F. de Moura Bastos

Orientadores: Benemar A. de Souza, D. Sc.

Niraldo Roberto Ferreira, D. Sc.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

B327d Bastos, Paulo Roberto F. de Moura
Diagnóstico de perdas comerciais de energia elétrica na distribuição usando rede Bayesiana / Paulo Roberto F. de Moura Bastos. — Campina Grande, 2011.
125 f.: il. col.

Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Orientadores: Prof. Dr. Benemar A. de Souza e Prof. Dr. Nivaldo Roberto Ferreira.

Referências.

1. Perda de energia elétrica. 2. Redes Bayesianas. 3. Perda não técnica de energia. 4. Perda comercial. 5. Distribuição de energia. 6. Redução das perdas I. Título.

CDU 621.3.095.2(043)

UFCG - BIBLIOTECA - CAMPUS I	
2578	05-05-011

**DIAGNÓSTICOS DE PERDAS COMERCIAIS DE ENERGIA ELÉTRICA NA
DISTRIBUIÇÃO USANDO REDES BAYESIANAS**

PAULO ROBERTO FERREIRA DE MOURA BASTOS

Tese Aprovada em 18.03.2011

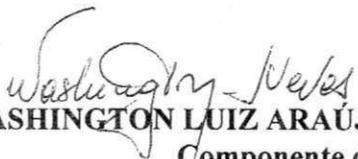

BENEMAR ALENCAR DE SOUZA, D.Sc., UFCG
Orientador


NIRALDO ROBERTO FERREIRA, Dr., UFBA
Orientador


CHRISTIANO LYRA FILHO, Dr., UNICAMP
Componente da Banca


ANTONIO PADILHA FELTRIN, Dr., UNESP
Componente da Banca


WELLINGTON SANTOS MOTA, Ph.D., UFCG
Componente da Banca


WASHINGTON LUIZ ARAÚJO NEVES, Ph.D., UFCG
Componente da Banca

**CAMPINA GRANDE – PB
MARÇO - 2011**

DEDICATÓRIA

Aos meus pais, Luiz e Edith. E lembro: “quem honra o próprio pai alcança o perdão dos pecados, e quem respeita sua mãe é como quem ajunta um tesouro” (Eclesiástico, 3).

Àquelas que muito me ensinaram: avó Belinha, tias Mercês e Juju, Claudinha e D. Maria.

A estes com os quais compartilho momentos mais intensos: Téo, Fábio, Mateus, Cíntia e Eliane.

Aos sobrinhos Lucas e Sofia, e ao neto Eduardo: vocês terão pela frente grandes desafios, mas tenho certeza que vencerão, percorrerão bons caminhos.

AGRADECIMENTOS

Agradeço a Deus e ao Menino Jesus de Praga pelas bênçãos recebidas.

A minha esposa Têu e ao nosso filho Mateus pelo amor e apoio.

Ao Prof. Benemar, pela paciência e incentivos, muito contribuindo com sua experiência e sabedoria.

Aos professores e funcionários da UFCG com os quais convivi, pela dedicação e amizade, especialmente os Profs. Benedito, Núbia e Edson, além de Ângela, Chico, Pedro, Suênia e Selma.

Aos colegas da UFBA, Jês, Fernando, Caiuby, Niraldo, Amaury, Jurandyr, Núbia e Jailton.

Aos irmãos e amigos Luiz, Clarice, Marta, Sito, Bel e Neto, pelas palavras e carinho recebidos ao longo desta jornada.

Àqueles que pagam seus impostos e, portanto nos financiam.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa recebida durante dois destes anos.

RESUMO

Este trabalho propõe uma metodologia para diagnóstico das perdas não técnicas ou comerciais, usando redes Bayesianas e a investigação de conformidade. Tal metodologia permite estimar as mais significativas perdas comerciais, quantificando-as regionalmente e conforme as causas de ordem interna ou externa. O diagnóstico torna possível avaliar relações custo-benefício para ações visando minimizar as perdas não técnicas, e conseqüentemente a definição de um eficiente plano de redução. Um estudo de caso exemplifica cada etapa do diagnóstico, descreve as ações a implantar ordenando-as segundo a análise econômica. Isto permite a escolha de um programa eficiente de redução das perdas comerciais. O Brasil, como vários outros países em desenvolvimento, apresenta elevados índices de perdas comerciais de energia, especialmente as empresas distribuidoras. É proposto ainda um indicador relacionando as perdas comerciais e as técnicas para servir como ferramenta comparativa entre empresas distribuidoras de energia elétrica, e o mesmo pode ser também um indicador do controle do processo. Neste trabalho, uma aplicação deste indicador mostra a situação das perdas de energia dentre as empresas distribuidoras brasileiras.

Palavras-chave:

Perda de energia elétrica. Perda comercial. Perda não técnica de energia. Redes Bayesianas. Redução das perdas. Distribuição de energia

ABSTRACT

The following work introduces a methodology for diagnosing non-technical losses by using Bayesian networks and research relating to compliance assessment. This proposed methodology allows estimating the most significant non-technical losses and their spatial distribution, identifying their external or internal causes. The diagnosis makes it possible to evaluate cost-benefit ratios for actions aiming to minimize non-technical losses and, consequently, define an efficient reduction plan. A case study exemplifies each stage of the diagnosis, describing actions to be taken and arranging them according to their economic impact. It also chooses an efficient program for non-technical loss reduction. Brazil, as well as in several other developing countries, presents high non-technical energy loss rates, particularly the distribution companies. An indicator relating non-technical and technical losses is also proposed as a comparative tool for electrical power distribution companies. So the indicator can be useful for controlling the process. It is applied, in this work, to describe the situation of energy losses in Brazilian distribution companies.

Keywords:

Electrical energy loss. Non-technical loss. Non-technical energy loss. Bayesian networks. Losses reduction. Energy distribution.

Diagnóstico de perdas comerciais de energia elétrica na distribuição usando redes Bayesianas

Sumário:

Capítulo 1 – INTRODUÇÃO	01
1.1 - Considerações iniciais	01
1.2 - Motivações do trabalho	03
1.3 - Objetivo e contribuição da tese	05
1.4 - Desenvolvimento e estruturação do trabalho	06
Capítulo 2 – REVISÃO BIBLIOGRÁFICA	08
2.1 – Considerações relativas a perdas técnicas	09
2.2 – Revisão bibliográfica referente a perdas comerciais	14
2.3 – Breve revisão referente a Redes Bayesianas	19
Capítulo 3 – NOÇÕES DE REDES BAYESIANAS	23
3.1 - Alguns conceitos de probabilidade e o teorema de Bayes	23
3.1.1 - Probabilidade condicional, TPC e fator de marginalização	23
3.1.2 - Teorema de Bayes	27
3.1.3 - Constante de normalização, regras fundamentais e independência entre variáveis	28
3.2 - Introdução às redes Bayesianas	33
3.2.1 – Definição das redes Bayesianas, estruturação e regra de cadeia	33
3.2.2 – Inferência, entropia em redes Bayesianas e explosão combinatória	35
3.3 – Exemplos de RB	41
Capítulo 4 – DIAGNÓSTICO DAS PERDAS COMERCIAIS	45

4.1- Justificativa para o diagnóstico e proposta de um índice comparativo	45
4.2 – Metodologia e modelo de rede para o diagnóstico das perdas Comerciais	52
4.2.1- Redes e estados das variáveis	55
4.2.2- Amostras e teste de validação	59
4.2.3- Aplicação da metodologia na estimativa das perdas comerciais dos clientes do grupo B de uma concessionária	63
Capítulo 5 – DIAGNÓSTICO DAS PERDAS COMERCIAIS: ESTUDO DE CASO COMPLETO	67
5.1 - Dados de mercado e cálculo da perda comercial	67
5.2 – Diagnóstico sob hipótese inicial	69
5.2.1 – Hipótese inicial e definição das amostras	69
5.2.2 – Conformidade e estimativa das perdas comerciais para os grupos de clientes	73
5.2.3 – Conformidade e estimativa das perdas comerciais para os grupos de não clientes	83
5.2.4 – Validação da hipótese inicial	87
5.3 – Diagnóstico sob nova hipótese	88
5.3.1 – Nova hipótese, amostras e validação	88
5.3.2 – Verificação da conformidade e estimativa das PC para os grupos de clientes	89
5.3.3 – Verificação da conformidade e estimativa das PC para os grupos de não clientes	91
5.2.4 – Estimativa das perdas e inferências	92
5.4 – Análise econômica das ações e definição de um programa eficiente para redução das perdas comerciais	95
Capítulo 6 – CONCLUSÕES E TRABALHOS FUTUROS	102

6.1 - Conclusões	102
6.2 - Trabalhos Futuros	104
6.3 - Publicações	105
Anexo A - Aplicação da metodologia na estimativa das perdas comerciais dos clientes do grupo B de uma concessionária.	107
Anexo B – Tabelas referentes às inspeções nos diversos grupos usadas no diagnóstico sob a nova hipótese.	114
REFERÊNCIAS BIBLIOGRÁFICAS	118

LISTA DE ABREVIATURAS

- ANEEL – Agência Nacional de Energia Elétrica
- ABRADE – Associação Brasileira de Distribuidores de Energia Elétrica
- ABNT – Associação Brasileira de Normas Técnicas
- AT – Alta tensão (no Brasil, distribuição acima de 1 kV, embora muitos tratem a faixa entre 1 kV e 69 kV como MT, ou média tensão)
- BT – Baixa tensão, em geral circuitos ou atendimento a unidades em tensão inferior a 1 kV.
- CODI – Comitê de Distribuição
- EDF – Electricité de France
- ELETROBRÁS – Centrais Elétricas Brasileiras S. A.
- Fp – Fator de perdas
- GRUPOS A e B – Fornecimento de energia em alta e baixa tensão, respectivamente
- IA – Inteligência artificial
- INMETRO – Instituto Nacional de Metrologia, Normatização e Qualidade Industrial
- IP – Iluminação pública
- I_{stp} – Índice de situação de perdas
- LC – Ligação clandestina
- MME – Ministério de Minas e Energia
- NT – Nota técnica
- PC – Perdas comerciais de energia elétrica (aqui usada com o mesmo significado de PNT)
- PNT – Perdas não técnicas de energia elétrica
- PT – Perdas técnicas de energia elétrica
- P&D – Pesquisa e desenvolvimento
- RB – Redes Bayesianas (em inglês usa-se BN, Bayesian network)
- RP – Rede primária de distribuição
- RS - Rede secundária de distribuição
- SNPTEE – Seminário Nacional de Produção e Transmissão de Energia Elétrica
- TPC – Tabela de probabilidade condicional
- UC – Unidade (s) consumidora (s)

LISTA DE FIGURAS

Figura 3.1 – Estrutura genérica de RB aplicada ao conjunto de n variáveis (POURRET et al, 2008, p.13).

Figura 3.2 - Exemplo de RB - (a) Conexão convergente e (b) Conexão divergente (adaptadas de RUSSELL e NORVIG, 2004, e de JENSEN e NIELSEN, 2007, respectivamente).

Figura 3.3 - (a) Estrutura da rede Grama Molhada, e (b) Mesma RB no software Netica® (NORSYS, 2008).

Figura 3.4 - RB da Grama Molhada, (a) após inseridos dados, (b) instanciado o estado *Sim*.

Figura 3.5 - Três redes ilustrando as possíveis situações de d-separação.

Figura 3.6 – RB relativa a doenças e sintomas; cada variável tem apenas dois estados, sendo mostradas as probabilidades de um deles (original em NEAPOLITAN, 2004, p.125).

Figura 3.7 – RB relativa a doenças e sintomas - (a) sem dados de TPC, e (b) depois de processada.

Figura 3.8 – Instanciado o estado um em Bronquite - (a) reprodução do NEAPOLITAN, 2004, p.137, (b) depois de simulado no Netica® (NORSYS, 2008).

Figura 3.9 – Modelos de RB para o sorteio de cartas, (a) engano – variáveis dependentes, e (b) RB correta.

Figura 3.10 – RB para o sorteio de cartas com dados teóricos, (a) modelo com engano, e (b) RB correta.

Figura 3.11 - Sorteio de cartas, instanciação - (a) RB com variáveis dependentes, e b) RB correta.

Figura 3.12 - Sorteio de cartas, instanciado *Dois_Mesmo_Grupo* - (a) dados teóricos, e (b) dados pesquisa.

Figura 4.1 – Perda comercial versus perda técnica em distribuidoras brasileiras (fonte ARAÚJO, 2007).

Figura 4.2 – Fluxograma para determinação das parcelas de perdas comerciais por grupo de origem e causa.

Figura 4.3 – Rede Bayesiana adotada para os clientes ativos e os não clientes.

Figura 4.4 – Identificação das PC, escolha dos estados do nó *Inspeção* para os grupos A e B.

Figura 4.5 – RB sugerida, (a) para os clientes do grupo A, (b) iluminação pública, e (c) ex-clientes Desligados.

Figura 4.6 – Rede Bayesiana para diagnóstico da PC do grupo B, Salvador (dados de DANTAS, 2006).

Figura 4.7 – Rede Bayesiana do Grupo B, simulação no Netica® (NORSYS, 2008).

Figura 5.1 – Repartição da energia anual comprada (2.712.397 MWh).

Figura 5.2 – Energia perdida (469.066 MWh/ano ou 100%) distribuída segundo hipótese inicial.

Figura 5.3 – RB do grupo A, (a) não conformidade 2,76% e (b) instanciação do estado *Fraude*.

Figura 5.4 – RB relativa ao grupo B, (a) não conformidade 7,17%, e (b) instanciação simultânea.

Figura 5.5 – RB do grupo IP, (a) não conformidade 2,50%, (b) instanciado o estado *Pontos da Região BB*.

Figura 5.6 - RB relativa aos Desligados, (a) não conformidade 8,16%, e (b) estado *Região BB* instanciado.

Figura 5.7 – RB do grupo A, (a) não conformidade 3,84%, e (b) estado *Região AA* instanciado.

Figura 5.8 – – RB do grupo B, (a) não conformidade 8,14%, e (b) instanciação do estado *Não Conforme*.

Figura 5.9 - RB do grupo Desligados, (a) não conformidade 8,98%, e (b) instanciação do estado *Região BB*.

Figura 5.10 – Distribuição da perda global da empresa que é de 17,3%.

Figura 5.11 – Perdas técnica e comercial por região.

Figura 5.12 – Perda comercial conforme as causas.

Figura 5.13 – Modelo de RB para investigação em uma faixa de consumo específica das UC residenciais.

Figura 5.14 – Modelo de RB para diagnóstico das PC na Região AA, grupo Desligados

Figura A.1 – RB utilizada, instanciado o estado *Fraude* do nó *Classe_Grupo_B*.

Figura A.2 – RB apresentando a instanciação simultânea dos estados *Fraude* (do nó *Inspeção_Agrupada*) e *Residencial* (do nó *Classe_Grupo B*).

LISTA DE TABELAS

Tabela 2.1 – Principais referências relativas a perdas comerciais.

Tabela 2.2 - Síntese das referências bibliográficas sobre redes Bayesianas.

Tabela 3.1 – Resultado da observação dos sorteios de cartas de ouros e paus.

Tabela 3.2 – Probabilidades conjuntas, sorteios de cartas de ouros e paus.

Tabela 3.3 – Probabilidades condicionais entre cartas de ouros e paus em sorteio.

Tabela 4.1 – Perdas comerciais na distribuição: localização, origem e possíveis soluções.

Tabela 4.2 – Análise das distribuidoras brasileiras quanto ao índice de situação de perdas.

Tabela 4.3 – Quantidades mínimas da amostra função do total de unidades.

Tabela 4.4 – Tipos de erro no teste de hipótese das perdas comerciais.

Tabela 4.5 – Correspondência entre os estados das variáveis *Inspeções* e *Inspeção Agrupada*.

Tabela 5.1 - Dados de mercado da empresa usada no estudo de caso (adaptado de CODI, 1996).

Tabela 5.2 – Quantidades de unidades dos grupos por classe e região.

Tabela 5.3 - Valores calculados das perdas técnicas por segmento e estimativa da perda comercial

Tabela 5.4 – Perda técnica e estimativa de PC segundo hipótese inicial.

Tabela 5.5 – Quantidades das amostras a inspecionar por Classe e Região, grupos A e Desligados.

Tabela 5.6 - Mínimo de pontos de IP e de UC residencial nas zonas das amostras dos grupos IP e LC.

Tabela 5.7 – Resultados da amostra inspecionada no grupo A.

Tabela 5.8 – Grupo A, probabilidades condicionais do nó *Inspeção*.

Tabela 5.9 – Estimativas de unidades por grupo e região função dos estados do nó *Inspeção*.

Tabela 5.10 – Estimativa (já arredondada) de unidades por classe e média anual de consumo.

Tabela 5.11 – Resultados da amostra inspecionada do grupo B.

Tabela 5.12 – Pontos e energia anual da amostra de IP, além da não conformidade por região.

Tabela 5.13 - Resultados das inspeções do grupo Desligados.

Tabela 5.14 – Dados das inspeções das amostras e estimativa de LC por região.

Tabela 5.15 – Não conformidade e perda por grupo.

Tabela 5.16 - Perdas Comerciais (MWh/ano): avaliação por região aplicado o diagnóstico.

Tabela 5.17 – Relações custo/benefício e outras estimativas para algumas ações de inspeção.

Tabela 6.1- Artigos publicados em seminários e congressos.

Tabela A.1- Consumidores por classe, probabilidade e dados da amostra inspecionada (DANTAS, 2006).

Tabela A.2 – Probabilidades conjuntas entre estados das variáveis *Inspeções* e *Classe_Grupo_B*.

Tabela A.3 – Probabilidades condicionais da variável *Inspeções*, conforme fornecido ao software Netica®.

Tabela A.4 – Estimativa de unidades por estado da inspeção e classes, para o Grupo B, Salvador – (dados de DANTAS, 2006).

Tabela B.1 - Resultados da nova amostra inspecionada no grupo A.

Tabela B.2 – Resultados da nova amostra inspecionada no grupo B.

Tabela B.3 - Resultados das novas quadrículas inspecionadas no grupo IP (pontos e energia).

Tabela B.4 – Resultados da nova amostra inspecionada no grupo Desligados.

Tabela B.5 - Resultados das novas quadrículas inspecionadas no grupo LC e estimativas por região.

1. INTRODUÇÃO

1.1- Considerações iniciais

As perdas totais de energia elétrica em uma concessionária de distribuição são definidas como a diferença entre o somatório da energia requerida pelo sistema elétrico e a energia realmente faturada aos diversos clientes. As perdas são normalmente tratadas nos aspectos perdas técnicas e comerciais (ou não técnicas).

As perdas técnicas são aquelas inerentes ao sistema devido à passagem da corrente elétrica nos meios físicos e nos materiais utilizados. Existem muitos estudos, pesquisas e publicações sobre o assunto, além de diversas metodologias de cálculo. As perdas técnicas são geralmente calculadas por nível de tensão na transmissão, subtransmissão e distribuição. Na distribuição são calculadas por segmento, por exemplo, na tensão secundária de distribuição determinam-se as perdas na rede, nos ramais de ligação, nos medidores e as perdas outras como nas conexões.

As perdas comerciais são causadas por problemas internos e externos à empresa distribuidora. Dentre os motivos internos há os defeitos apresentados no sistema de medição das unidades consumidoras, os erros de leitura ou no banco de dados para faturamento das mesmas, e diferenças originadas no faturamento sem medição como é o caso da iluminação pública. Já com respeito às causas externas existem as fraudes efetuadas por consumidores, as auto-religações de unidades desligadas e as ligações clandestinas realizadas diretamente na rede de distribuição.

Em geral as perdas comerciais não são prospectadas ou calculadas, sendo definido o montante total pela diferença entre o total de perda de energia elétrica existente na empresa e o total das perdas técnicas calculadas.

No Brasil, como em vários outros países em desenvolvimento, as perdas não técnicas de energia elétrica são historicamente altas, causadas por fatores sócio-econômicos e culturais. Dentre os primeiros estão o desemprego, o baixo nível de renda da população, a ausência do estado com políticas habitacionais e provendo infra-estrutura, os preços relativamente elevados da eletricidade, associados à impunidade no que diz respeito à fraude de energia. Há ainda a cultura

de “levar vantagem” e a contaminação pela vizinhança que levam à disseminação das irregularidades. A propagação em maior escala da fraude e roubo desmotiva a fiscalização tanto da empresa como dos órgãos oficiais.

Em 2008, as perdas totais de energia elétrica no Brasil atingiram 77.081GWh ou 18,0% da energia consumida conforme dados do Balanço Energético Nacional (Ministério de Minas e Energia do Brasil- MME, 2009), aí incluídas perdas na geração e na rede básica de transmissão. Tomando apenas a perda nas distribuidoras como 14,5% e sendo a comercial metade disto, significa que as perdas comerciais têm o valor de R\$4,1 bilhões ou R\$7,9 bilhões anuais, conforme se considere o custo da compra de energia pelas concessionárias ou o preço médio de fornecimento de energia praticado pelas distribuidoras (ANEEL – Agência Nacional de Energia Elétrica, 2010; Tarifas médias no Brasil referidas a janeiro de 2009, valor de R\$254,71/MWh).

Em algumas concessionárias de distribuição as perdas totais até ultrapassam os 30% sendo a parcela de perda não técnica bastante significativa (ARANHA NETO et al, 2008). A ANEEL até o momento não regulamentou as perdas, tendo divulgado uma Nota Técnica (NT) recomendando uma metodologia unificada para o cálculo das perdas técnicas (NT-035/SRD/ANEEL/2007), a qual serviu de base para o módulo sete dos “Procedimentos de distribuição” da própria agência (ANEEL - PRODIST, 2008). Quanto a perdas comerciais, essa agência convidou especialistas e promoveu um workshop sobre o tema (ANEEL, acesso em 07.05.07); tem acompanhado os quantitativos das empresas, e por ocasião do segundo ciclo das revisões tarifárias das distribuidoras propôs uma metodologia para o tratamento das perdas não técnicas (NT-348/SRE/ANEEL/2007).

A extinção, ou num primeiro horizonte a redução da perda não técnica, significa importante ganho para a empresa e a sociedade. Entretanto, algumas concessionárias procuram reduzir suas perdas comerciais muito focadas em inspeções, buscando identificar fraudes ou enganos e defeitos no sistema de medição. Um diagnóstico procurando prospectar tais perdas de modo mais amplo inclusive examinando os não clientes é bastante importante, pois pode trazer mais eficiência ao processo.

O principal objetivo deste trabalho é elaborar um diagnóstico em nível macro para prospecção das perdas comerciais objetivando quantificar suas parcelas segundo as principais causas e regiões de incidência dentro da área de concessão da distribuidora

Redes Bayesianas (RB) são modelos gráficos probabilísticos adequados para trabalhar com incertezas. Em tais modelos o problema é colocado como um conjunto de variáveis e suas relações de probabilidades. As RB têm sido aplicadas com sucesso na área médica, em investigações geológicas e criminais, dentre outras. Na área médica, por exemplo, são empregadas visando facilitar o diagnóstico de doenças através das relações entre os sintomas e exames, caracterizando estas relações através de probabilidades condicionais. Aplicações em diversas áreas podem ser encontradas em (POURRET et al, 2008) e em artigos recentes.

1.2 – Motivações do trabalho

Mesmo com todo o desenvolvimento tecnológico as perdas de energia elétrica têm crescido mundialmente nos últimos anos. Em muitos países em desenvolvimento as perdas comerciais são elevadas e significam valores altos economicamente. Tais fatos aliados à constatação de que as perdas comerciais têm sido menos estudadas e pesquisadas no meio acadêmico que as perdas técnicas, constituem fortes motivações para este trabalho.

As perdas comerciais de energia elétrica, de modo semelhante a outras perdas comerciais como as de água, ou as dos serviços de transporte, são em geral significativas nos países em desenvolvimento. Para reduzi-las devem existir ações inibidoras e coercitivas das concessionárias, da polícia, além das ações punitivas da justiça.

Como consequência, no Brasil, estas perdas são maiores nas empresas que atuam em estados mais pobres como Piauí, Alagoas e Maranhão, ou naqueles onde as ações de combate ao crime têm sido mais difíceis a exemplo do Rio de Janeiro. Até em empresas estruturadas como a Companhia Energética de Minas Gerais, as perdas comerciais cresceram desde 0,59% até 1,48% no período entre 2000 e 2003 (ALVES, 2003).

Mesmo a perda técnica sendo por vezes menor que a comercial, tradicionalmente é bastante mais estudada, pois a perda comercial vinha sendo tratada apenas como um problema gerencial. Para a redução das perdas técnicas são construídas novas subestações, novas redes, são usados bancos de capacitores, e tais estudos por resultarem em investimentos envolvem sempre análises técnicas, econômica e financeira.

O tema perdas ou sua redução tem sido objeto de projetos de pesquisa e desenvolvimento em várias concessionárias no Brasil podendo-se citar CIPOLI et al (2001), VALENTE et al (2005), COMETTI et al (2007), PENIN (2008) e FERREIRA (2008). Também tem sido objeto de estudo em universidades (ARANHA NETO et al, 2009), (ARAÚJO, 2007), (BRAZ, 2003), (DANTAS, 2006), (ELLER, 2003), (FERNANDES, 2003), (MÉFEE, 2001), (QUEIROZ, 2005) e (VIEIRALVES, 2005).

Muitos dos trabalhos se concentram em um aspecto ou apresentam apenas os resultados de ações de redução de perdas, e raramente há um enfoque mais abrangente, especialmente voltado para a prospecção e diagnóstico das perdas comerciais, como forma de facilitar o planejamento das ações e estabelecimento de metas de combate às perdas.

Embora tenha havido grande desenvolvimento tecnológico nos últimos anos e aumento da eficiência energética em vários setores, a perda de energia elétrica cresceu mundialmente de 11,7% em 1980 para 16,2% em 2000 (SMITH, 2004), tomada uma amostra de 102 países. Conforme SIQUEIRA e ARAÚJO (2006) *“os países com índices de perdas superiores a 30% caracterizam-se por baixo nível de desenvolvimento econômico e social”*. Tendo em vista a dimensão do problema das perdas comerciais, é relevante a busca de uma metodologia adequada ao diagnóstico das mesmas.

Corroborando com esta idéia constatou-se que com respeito a perdas comerciais de energia é sempre difícil calcular ou estimar as parcelas devido à fraude, aos erros de medição, engano no banco de dados do sistema de faturamento, ou aos popularmente chamados “gatos”, incluídos consumidores auto-religados e aqueles não cadastrados conectados através das ligações clandestinas.

Como dito, as perdas comerciais na distribuição podem significar cerca de 7,9 bilhões de reais/ano, e se as ações de combate e redução destas perdas forem mais fundamentadas certamente haverá mais eficiência no processo.

Isto significa que conhecendo as origens e onde estão concentradas as perdas não técnicas, as distribuidoras devem reduzi-las mais rapidamente e com menor custo, pois poderão realizar análises econômicas e então melhor direcionar os recursos quer seja no dimensionamento de suas equipes de inspeção em campo, na programação de treinamento interno visando aprimorar a

execução das ligações e leituras, quer na aquisição e substituição dos equipamentos de medição, ou mesmo nas atualizações cadastrais.

Espera-se enfim facilitar a recuperação das perdas contribuindo para maior eficiência deste processo nas empresas distribuidoras.

1.3 – Objetivo e contribuição da tese

Considerando a importância das perdas não técnicas, consolidou-se a motivação para abordar o tema, e verificando-se que vários dos trabalhos na área são relativos ao uso de técnicas de inteligência artificial objetivando identificar clientes com perdas nas empresas, optou-se por trabalhar com algo anterior à decisão de inspecionar as unidades consumidoras, ou seja, com um diagnóstico completo, em toda a área de concessão de uma distribuidora, envolvendo seus clientes e não clientes, pois estes podem ser consumidores de energia elétrica.

Optou-se por prospectar as perdas comerciais através da concepção de um diagnóstico capaz de estimar as perdas comerciais conforme as causas e regiões, possibilitando realizar uma análise econômica das ações a propor para redução de perdas, e assim obter um plano contemplando ações mais eficientes. Deparou-se com incertezas e probabilidades, verificando-se que as redes Bayesianas poderiam ser utilizadas, bem como o exame de conformidade.

Deste modo o principal objetivo deste trabalho é a proposição de uma metodologia para diagnóstico das perdas comerciais na distribuição, usando RB e investigando a conformidade.

Posteriormente, verificou-se que em várias empresas do país as perdas comerciais superavam as técnicas, e que um índice relacionando tais perdas possibilitaria comparar distribuidoras. Se usado pelo órgão regulador poderá ajudar na avaliação do controle do processo pelas distribuidoras, e quanto aos planos e ações a exigir das mesmas. Assim, poderia facilitar a identificação de um grupo de empresas que necessitam de um diagnóstico das perdas comerciais além de um acompanhamento mais efetivo dos seus programas de redução.

Dentre as contribuições do trabalho destacam-se:

- i. A proposta de um diagnóstico que permite obter estimativas de perdas comerciais conforme as causas e a distribuição regional dentre as áreas da distribuidora. Isto

- possibilita uma análise econômica das ações, comparando-as e então viabilizando metas e ações mais consistentes num programa de redução;
- ii. O modelo de rede Bayesiana usado mostrou-se flexível a alterações nas empresas;
 - iii. As relações de probabilidade condicional *a posteriori* necessárias para suprir a RB podem ser obtidas de inspeções em campo realizadas em amostras aleatórias, calcadas na investigação da conformidade, tanto para clientes como para não clientes;
 - iv. O indicador sugerido é de fácil apuração e permite comparar empresas distribuidoras avaliando o processo de perdas.

Tendo em vista a necessidade do estudo das perdas técnicas, objeto de um dos “Projeto e Pesquisa” e de algumas publicações, há ainda a seguinte contribuição: na quantificação das perdas técnicas existem aspectos importantes que em geral não estão sendo abordados, como por exemplo: o envelhecimento dos medidores, a inadequação do intervalo de integração da demanda tomado como quinze minutos, e o provável aumento das perdas nos transformadores de distribuição após recuperação dos mesmos.

1.4 – Desenvolvimento e estruturação do trabalho

Propõe-se uma metodologia de diagnóstico de perdas comerciais que as quantifica indicando as principais causas e regiões de incidência. São usadas as redes Bayesianas e a prospecção está fundamentada em conceitos de conformidade.

Para os consumidores ativos no banco de dados da distribuidora, as unidades encontradas com perda, sejam por fraude, defeito de medição, engano de cadastro e mesmo a determinação de lâmpadas de iluminação pública acesas durante o dia são irregularidade ou não conformidade. Para os consumidores não clientes da distribuidora se constituem não conformidade a identificação da unidade auto-religada dentre aqueles ex-clientes mantidos no banco de dados, e a identificação de redes ou conexões ilegais ligadas diretamente à rede da distribuidora.

Usa-se um modelo de rede Bayesiana aplicado a três grupos de consumidores ativos, aqueles dos atendidos em alta tensão (grupo A), em baixa (grupo B), e a iluminação pública, e dois grupos de não clientes, quais sejam as unidades desligadas e ligações clandestinas. São estabelecidas hipóteses iniciais referentes a não conformidade nestes grupos, as quais em seguida são validadas

estatisticamente. Pesquisas amostrais investigando a conformidade em campo por grupo são usadas para obtenção das relações de probabilidade condicional exigidas pela RB. Em função de hipóteses iniciais relativas a perdas comerciais nos diversos grupos e do número de unidades, são determinadas as amostras, dentro dos erros e intervalos de confiança desejados.

Com respeito à estruturação, no Capítulo Dois são feitas algumas considerações sobre as perdas técnicas, uma revisão bibliográfica relativa a perdas não técnicas, e uma breve revisão referente às redes Bayesianas.

O Capítulo Três apresenta os principais conceitos relativos à probabilidade condicional, algumas regras e definições referentes às redes Bayesianas, bem como as formas usuais de modelagem e estruturação destas redes.

Os capítulos Quatro e Cinco são os mais importantes deste trabalho. O Quarto Capítulo se inicia com a justificativa para o diagnóstico, segue-se a proposição de um índice de perdas comparativo entre empresas e a descrição da metodologia para o diagnóstico das perdas comerciais. São apresentados o fluxograma, os modelos de redes Bayesianas adotados com descrição dos estados das variáveis, a definição das amostras e a validação.

No Capítulo Cinco é feita uma completa aplicação do diagnóstico usando-se os dados da empresa exemplificada em relatório do extinto Comitê de Distribuição (CODI, 1996), isto porque até o momento não se dispõe de dados atuais completos de uma distribuidora. Tornando esta aplicação mais didática a hipótese inicial referente à repartição das perdas comerciais por grupo é rejeitada necessitando-se de uma nova hipótese. Com a validação desta segue-se com inferências a respeito da origem e regionalização das perdas, definição de ações de redução, análise de viabilidade econômica destas ações e proposição de um programa de redução das perdas comerciais.

O Capítulo Seis apresenta as conclusões e propostas de trabalhos futuros, além das publicações realizadas no período. Seguem-se dois anexos e as referências bibliográficas. O Anexo A traz a aplicação da metodologia aos clientes de baixa tensão de uma concessionária de distribuição, podendo-se constatar a flexibilidade dos modelos de rede Bayesianas utilizados, e o Anexo B contém alguns dados das inspeções amostrais do diagnóstico.

2. REVISÃO BIBLIOGRÁFICA

As perdas de energia elétrica acontecem em todo o sistema elétrico, portanto devem ser tratadas nas áreas de geração, transmissão e distribuição. Aqui é tratada a perda de energia na distribuição. Como já foi dito as perdas de energia são resultado da diferença entre o total de energia colocado no sistema e a energia global faturada, sendo normalmente classificada em perda técnica e não técnica ou comercial.

Os trabalhos da ANEEL, enquanto órgão regulador, são importantes. Dentre os Procedimentos de Distribuição (PRODIST/ANEEL, 2007) destacam-se os módulos um e sete que tratam respectivamente da introdução e das perdas técnicas. Ainda da Agência Reguladora há as notas técnicas 035/SRD/2007, que trata das perdas técnicas, e 348/SRE/2007, que apresenta o tratamento regulatório das perdas não técnicas. Preferivelmente será seguida a nomenclatura adotada pela ANEEL como estas definições referentes a perdas (PRODIST, módulo 1, 2007):

Perdas globais (PG): diferença entre a energia requerida (ou injetada - E_{in}) e a energia fornecida pela distribuidora, expressas em megawatt-hora(MWh).

Perdas técnicas (PT): montante de energia elétrica, expresso em megawatt-hora (MWh), dissipada no sistema de distribuição, decorrente das leis físicas relativas aos processos de transporte, transformação de tensão e medição. Corresponde à soma de três parcelas: joule, corona e magnética.

Perdas não técnicas (ou comerciais - PC): apurada pela diferença entre as Perdas Globais e as Perdas Técnicas, considerando portanto todas as demais perdas associadas à distribuição de energia elétrica, tais como furtos de energia, erros de medição, etc.

Devido à sazonalidade de diversas unidades consumidoras e aos ciclos de leitura das concessionárias o período de apuração das perdas normalmente é de um ano, portanto em geral a perda global bem como a técnica vem expressa em MWh/ano. Assim, pode-se definir o índice percentual de perdas (I_p) da empresa distribuidora como

$$I_p = 100 \cdot (PG/E_{in}) \quad (2.1)$$

Também E_{in} , o total de energia injetada na empresa distribuidora, é em MWh/ano.

Deste modo as perdas comerciais (PC) são

$$PC = PG - PT \quad (2.2)$$

Nesta classificação das perdas globais há pesquisadores que tratam a energia consumida pelas próprias empresas de geração, transmissão e distribuição como “perdas administrativas” (ORILLAZA et al, 2006). Aqui esta parcela não é tratada como perda visto que no Brasil geralmente é faturada e identificada no mercado da empresa como consumo próprio.

Neste capítulo inicialmente são apresentadas algumas considerações relativas a perdas técnicas, em seguida uma revisão bibliográfica referente a perdas comerciais e por fim uma breve revisão com respeito às redes Bayesianas.

2.1 – Considerações relativas a perdas técnicas

Os sistemas de distribuição são caracterizados por linhas e redes primárias em diferentes níveis de tensão, grande quantidade de transformadores de distribuição e suas redes secundárias, visando atender as cargas que se apresentam nas mais diversas formas.

Em distribuição, as perdas técnicas são geralmente calculadas em função dos parâmetros, tipologia da rede e dos dados das cargas. Na verdade as cargas são funções aleatórias do tempo sobre as quais não se tem controle, embora se possa inferir expectativas quanto às variações ao longo dos dias e meses do ano. Em geral as perdas são calculadas por nível de tensão segregadas nos diversos segmentos: linhas de sub-transmissão, subestações abaixadoras, rede primária, transformadores de distribuição, rede secundária, ramais de serviço, medidores e outros (capacitores, reguladores, etc).

Em relação a linhas de subtransmissão e subestações as PT podem ser medidas embora nestes dois segmentos e na rede primária sejam quantificadas usando-se programas computacionais de fluxo de carga. O mais comum é o cálculo das perdas de demanda em condições de carga máxima, associada ao respectivo fator de potência, e através do fator de perdas (GONEN, 1987) se estimarem as perdas médias de demanda e as perdas de energia.

Da mesma forma que o fator de carga relaciona a demanda média e a demanda máxima, o fator de perda (F_p) é a relação entre as perdas de demanda média e aquela determinada na condição máxima, sendo muitas vezes estimado como uma função quadrática do fator de carga (BULLER,

1928, citado por GUSTAFSON, 1983), (GONEN, 1987). Há estudos relativos ao fator de perdas no Brasil (CODI, 1996), (ANEEL, NT-035, 2007).

Em geral a perda de energia é determinada pelo produto da perda de demanda média no período por T que é o período analisado, sendo na maioria das vezes um ano ou 8.760 horas. Logo, se pode escrever que a perda técnica (em MWh) anual é

$$PT = 8.760 \cdot F_p \cdot P_{Dm\acute{a}x} \quad (2.3)$$

F_p é o fator de perdas;

$P_{Dm\acute{a}x}$ é a perda de demanda na situação de carga máxima, em MW.

A partir da década de 1980 apareceram mais estudos os quais têm se aprofundado em quatro vertentes:

- Determinação das perdas técnicas totais das concessionárias e sua quantificação por segmento;
- Utilização de modelos mais exatos através da caracterização das cargas e/ou sua tipificação, e da representação mais apurada das redes;
- Minimização das perdas técnicas;
- Determinação das perdas técnicas com fins regulatórios.

Relativos à primeira vertente se destacam: SUN et al (1980) que estudaram perdas em alimentadores e transformadores de distribuição desenvolvendo um programa que modela a carga como potência constante, impedância constante ou um misto, bem como analisaram configuração desbalanceada e sensibilidade à variação de tensão; FLATEN (1988) apresentando equações para o cálculo das perdas em vários segmentos em função da carga percentual, e que corrige o valor da resistência do elemento analisado conforme condições específicas; e, GUSTAFSON e BAYLOR (1989) que tratam as perdas na transmissão, distribuição e outras, separando as que não dependem da carga (magnéticas, como aquelas que ocorrem nos núcleos dos transformadores e nos medidores), das perdas que dependem das cargas e, portanto das correntes, correlacionando para estas o valor das perdas com o consumo global.

No Brasil surgiram vários trabalhos na década de 1990 e dentre outros, há o Relatório do Comitê de Distribuição- CODI, *Método para determinação, análise e otimização das perdas técnicas em*

sistemas de distribuição (CODI, 1996), que apresenta métodos específicos para o cálculo das perdas nos diversos segmentos, tendo se tornado uma referência nacional.

Os estudos buscando modelos mais exatos além de melhor representarem a rede, modelam a carga, por exemplo, segregando diversas curvas de carga conforme atividade principal típica, usando valores de demanda medidos com intervalos de quinze minutos, etc. Vários destes estudos apresentam desenvolvimento de softwares específicos e acompanharam a evolução dos computadores. Destacam-se RAJIÈIÉ e TALESKI (1996) que além de curvas típicas de consumidores residenciais, comerciais e industriais consideram tensões nos vários nós diferentes da nominal de acordo com o fluxo de carga e empregam o método da soma das potências, e POVEDA (1999) com um tratamento diferenciado para a carga, sugerindo o acompanhamento da variação da demanda registrada a cada 15 minutos, o que conduz a simulação de 96 fluxos de carga diários.

No Brasil destacam-se CIPOLI et al (2001) que apresentam um método que considera várias destas curvas típicas por classe de consumidores, e MÉFFE (2001) que além disto faz uma representação completa das redes primária e secundária.

Por outro lado tais programas computacionais caso necessitem de base de dados muito grande, dispersa por diversos órgãos na empresa, ou requeiram execução demorada, tendem a não ser empregados nas concessionárias devido à própria dificuldade de implementação e atualização.

Mais recentemente têm sido usadas outras ferramentas como em KANG et al (2006) que empregaram redes neurais artificiais no cálculo das perdas em alimentadores, e outros preferiram utilizar métodos estatísticos, como CALMEYER e FURIE (2004) que calculam perdas técnicas em oito segmentos da distribuição apresentando um intervalo esperado. MIKIÉ (2007) considera a componente devida à carga média e à sua variância, fazendo ainda considerações a respeito da caracterização das cargas e das medições variando-se os intervalos de integração.

Dentre os que empregam a metodologia probabilística: as variações da carga são abordadas por BERASATEGUI et al (2001), através de intervalos de previsões probabilísticas para a curva de carga; STRAUCH (2002) trata das perdas em transformadores de distribuição, rede secundária, ramais e medidores; MARQUES et al (2005) com tratamento estatístico das cargas apresentam a perda global dentro de uma faixa esperada além de uma alocação das perdas por consumidor, mais no enfoque tarifário.

Para minimização exclusiva das perdas técnicas diversos métodos de otimização têm sido empregados, podendo-se citar BRAZ (2003) e QUEIROZ (2005) como exemplos no Brasil.

No enfoque regulatório os métodos de cálculo devem ser simples o suficiente para permitir o uso em todas as empresas, validando a comparação entre as mesmas. Do contrário, se o método for mais sofisticado, dependerá de dados em quantidade e, sobretudo, em nível de detalhamento que poderá não existir em todas elas. Por outro lado, os métodos devem fornecer resultados de boa precisão e de fácil auditoria. Após a reforma do setor elétrico, pode-se dizer que nos últimos quinze anos muitos são os trabalhos nesta direção, tais como NADIRA e DORTOLINA (2005), LASSO et al (2006) e QUEIROZ et al (2009).

Outras referências que merecem destaque são RAO e DEEKSHIT (2006) e ORILLAZA et al (2006) pela revisão bibliográfica que fizeram, além do recente trabalho de MÉFFE e OLIVEIRA (2009) no qual estes propõem o uso de fatores de correção além da consideração das PC no cálculo das PT.

Os métodos probabilísticos são mais adequados ao cálculo das perdas técnicas face à variação dos dados, podendo-se usar a média e o desvio padrão das estimativas de perdas ou seus intervalos mais prováveis.

No início deste trabalho de tese a perda técnica foi objeto de um “Projeto e Pesquisa”. Neste, verificou-se que há aspectos que necessitam de maior aprofundamento visto que, como atualmente considerados nas metodologias, podem conduzir a imprecisão ou introduzir erros (BASTOS et al, 2008, 2009). Quanto a isto destacam-se os seguintes pontos:

- A curva de carga tomada a intervalos de quinze minutos;
- A não consideração do envelhecimento dos medidores eletromecânicos ou seu real estado em campo;
- A corrente de partida dos medidores que é relativamente elevada face às cargas cada vez menores dos usuários;
- A não consideração de perdas mais elevadas em transformadores reformados.

Vários métodos caracterizam as cargas usando curvas típicas levantadas mediante campanhas de medição nas quais as demandas são tomadas com intervalos de integralização de 15 minutos, como é usual. Entretanto, há muitas cargas como chuveiro elétrico e forno de micro-ondas, que têm potência elevada e são usadas em intervalos inferiores a 15 minutos. Analiticamente, conclui-

se que sempre que o eletrodoméstico for ligado por um período t_l inferior ao intervalo de integração (T , igual a 15 minutos), estará havendo erro significativo com relação ao cálculo da perda técnica de demanda na proporção do quadrado da relação t_l/T , e para as perdas de energia o erro é diretamente proporcional a t_l/T (BASTOS et al, 2008).

O envelhecimento dos medidores eletromecânicos aumenta os atritos nas engrenagens e mancais. Portanto, em função do desgaste sofrido durante a vida útil, tendem a apresentar erro médio negativo. RIOS e HENRIQUES (2006) avaliaram mais que três mil e trezentos medidores que estavam em funcionamento havia mais de vinte anos, dos quais 218 (6,5% do total) foram reprovados no ensaio de dielétrico, representando risco de acidente e perdas por corrente de fuga. Os demais medidores foram calibrados e apresentaram os seguintes resultados: *“erro médio do total de 3.139 medidores calibrados é -6,63% o que é muito expressivo e inspira atenção (valor influenciado pelos medidores com erro igual a -100% em pelo menos um elemento motor ou condição de carga),... 73,8% dos medidores apresentaram erro médio negativo com uma expressiva concentração de medidores na faixa de erro entre 0 e -1,5%, e 82,8% dos medidores apresentaram erro médio dentro dos limites permitidos, que são $\pm 3,0\%$ ”*. Em 2005, quando estas pesquisas foram efetuadas ainda não havia sido publicada a Portaria 088 do órgão regulador da metrologia (INMETRO, 2006) que ampliou a faixa aceitável dos testes em campo para $\pm 4,0\%$.

Os transformadores reformados apresentam problema análogo aos motores reconicionados, ou seja, mesmo sendo exigidos e acompanhados os ensaios, por vezes acabam tendo perdas maiores que as originais. Isto se verificou quando em 2005 a Universidade Federal da Bahia reformou três transformadores seus: em dois deles as perdas no núcleo e no cobre apresentaram nos ensaios pós-reforma valores bem acima dos máximos das normas (as perdas totais foram 14% acima em um e 72% superior no outro).

Uma metodologia simplificada para o cálculo das perdas técnicas foi desenvolvida, tendo sido denominada de SIMPLES (BASTOS et al, 2008). Recentemente a mesma incorporou três alterações:

- Para os alimentadores primários: como a subestação abaixadora normalmente está próxima ao centro de carga há maior concentração de carga no trecho inicial. Então, tomou-se a distribuição da carga ao longo do alimentador sendo um quarto do total nos

primeiros um oitavo do comprimento, e os restantes três quartos da carga como uniformemente distribuída.

- Para perdas nos ramais de serviço: período médio de uso de equipamentos que ficam ligados por curto intervalo como 7,5 minutos, frente ao intervalo de integração de medidores de 15 minutos, e que isto ocorre em 50% das cargas residenciais.
- Para a perda nos medidores: por envelhecimento considerou-se erro médio negativo de 1% aparecendo em 50% das unidades, e a hipótese de que em 10% dos clientes durante 6 horas/dia pode haver carga inferior a 90% da corrente de partida.

Concluindo, as pesquisas buscando metodologias precisas e simplificadas que permitam associar as responsabilidades dos consumidores nas perdas estão ganhando espaço face ao mercado mais aberto e competitivo. Também os modelos probabilísticos e a consideração das incertezas têm se mostrado mais apropriados na quantificação das perdas técnicas devido aos muitos aspectos desconhecidos e ao aumento da heterogeneidade dos parâmetros à medida que a avaliação da perda se dá nos segmentos próximos à carga.

2.2 – Revisão bibliográfica referente a perdas comerciais

Quanto às perdas não técnicas, estas sempre foram vistas como um problema gerencial e nem se constituíam em objeto de pesquisa até dez ou quinze anos atrás. Com a perspectiva de reforma do setor elétrico e discussão da privatização das empresas, o empresário vislumbrou um atrativo nas empresas com altas perdas visto que atuando para reduzi-las teria maiores ganhos.

As perdas não técnicas em geral não existem nos sistemas de geração e transmissão. No varejo, que é a distribuição, estão associadas a causas de origem interna ou externa à empresa (EDF, No.1, 1998). Dentre as causas internas há, por exemplo, defeito na medição, erro na constante de uma unidade consumidora, engano no cadastro da iluminação pública que é faturada por estimativa, e as causas externas são os vários tipos de fraude realizados por clientes, a auto-religação de uma unidade de ex-consumidor ou uma ligação clandestina direta na rede.

Conforme NT-348/ANEEL (2007) a perda não técnica está classificada em:

- “*Perdas comerciais*”, relacionadas com unidades consumidoras regulares e as que já o foram;

- “*Perdas por consumo de ligações clandestinas não acessáveis*”, que são as ligações irregulares sem relação contratual mesmo que de adesão, em locais de impedimento legal ou de fato para concessionária atuar, como áreas de litígio fundiário, invasões, etc. Inserindo um comentário: ao colocar “não acessáveis” entende-se que a NT quer se referir a ligações em locais com impedimento de acesso legal;
- “*Outras perdas não técnicas*”, correspondendo à perda técnica causada pela própria perda não técnica.

Vê-se que a ANEEL não inclui as perdas por inadimplência entre as perdas não técnicas. Aqui é usada a perda comercial com o mesmo significado da não técnica, sendo dispersa em diversos segmentos de clientes (grupos A, B e IP) e não clientes (unidades desligadas ou ex-clientes e as ligações clandestinas). Já a perda técnica causada pela não técnica deve compor o bloco “outras” visto que o órgão regulador não a aceita; óbvio que não existindo perdas não técnicas este excesso de perda técnica não existiria.

Até dez anos atrás eram raros os trabalhos e pesquisas dedicados às perdas não técnicas. Mais recentemente tem havido pesquisas sobre o tema, muitas direcionadas a dois sub-temas: fraude, e busca dos fatores que levam ao furto de energia. Com relação à prospecção das perdas, isto é, a quantificação das suas parcelas dentre os vários motivos e áreas, são escassos os trabalhos: há a dissertação de DANTAS (2006) e como a mesma fez parte de um projeto de pesquisa e desenvolvimento está descrita em VALENTE et al (2005) e em SILVEIRA et al (2006).

Em termos internacionais, dentre os trabalhos dedicados a fraude destaca-se JIANG et al (2002) que propõe uma ferramenta automática para identificar fraudadores combinando múltiplos classificadores e a técnica Wavelet. Dentre os que visam identificar os fatores que conduzem ao furto de energia, SMITH (2004) apresenta uma análise comparativa referente ao roubo de energia, sendo uma pesquisa importante por ter analisado relatórios e trabalhos de aproximadamente cem países correlacionando as perdas comerciais não apenas à ineficiência das empresas, mas também a fatores sociais e culturais associados à governança.

Há outros que destacam ações ou experiências bem sucedidas no combate ao furto e fraude: GIMENEZ (2009) que reporta a utilização de medidores totalizadores e o balanço de energia em determinada área do Paraguai, e VALENZUELA e MONTAÑA (2005) que citam medidas técnicas como o uso da rede antifurto (cujos cabos construtivamente dificultam as conexões para

furto), medidas punitivas a exemplo da retirada das ligações clandestinas e divulgação nos meios de comunicação, e medidas comerciais como postos móveis de cobrança, facilidades de pagamento, além do programa de regularização de usuários empregados no Chile.

No Brasil, com referência às pesquisas direcionadas à detecção de fraude alguns trabalhos estão voltados à aplicação de inteligência artificial ou computacional especialmente procurando aumentar o índice de acerto na descoberta de unidade com perda por ocasião da inspeção em campo. Em DURÃES et al (2006) há uma boa descrição destes modelos. ELLER (2003) desenvolve um sistema de gerenciamento das perdas comerciais, e VIEIRALVES (2005) propõe uma metodologia de avaliação das mesmas focando efetivamente as perdas por fraude e metodologias para detectar este tipo de contravenção.

Nesta linha, ROCHA et al (2006) apresenta uma metodologia para identificar divergências e classificar clientes baseada em algoritmo Fuzzy-C-Means analisando as curvas de consumo, e os fatores de carga e demanda. COMETTI et al (2007) descrevem um sistema baseado em conhecimento capaz de melhorar o índice de acerto na detecção de irregularidades quando das inspeções. ALMEIDA (2006) aplica técnicas de mineração de dados por atividade visando maior eficiência nas inspeções, e traz estudo de caso para algumas atividades da classe comercial.

Há autores que apenas descrevem o processo de gestão de perdas não-técnicas, as estratégias e ações para redução (SILVEIRA e MORAES, 2006), com enfoque nos resultados (ALVES, 2003), e no combate à fraude (ARANHA NETO et al, 2008). Assim, em quase todos os trabalhos não é abordada a perda nos não clientes, o que é claro em FERREIRA (2008, p.7) ao colocar: *“não é escopo deste trabalho detectar em áreas em que o fraudador não é cliente da empresa, isto é, quando ele não tem cadastro nem medidor como, por exemplo, favelas e invasões”*.

Neste caminho de detecção e combate às perdas comerciais PENIN (2008) vai um pouco além ao abordar medidas preventivas e incluir a análise econômico-financeira ao buscar um nível de perdas ótimo para a empresa.

Há aqueles que procuram identificar os fatores que levam ao furto de energia. Destacam-se FRANCISCO e FAGUNDES (2006), URANI et al (2008) e ARAÚJO (2007). Os primeiros usam um modelo de regressão multivariada para analisar áreas de propensão às perdas considerando as variáveis renda, concentração de favelas, e de clientes cortados por falta de pagamento. URANI et al (2008) determinam um indicador de perdas globais em função de

parâmetros como o percentual de domicílios com mais de duas pessoas por dormitório, óbitos devido a homicídio por cem mil habitantes, percentual de domicílios com ar condicionado, e outros, associados às dimensões desenvolvimento, densidade, favelização, violência, ambiente institucional, inadimplência e consumo regional diferenciado.

ARAÚJO (2007, p.14) busca resposta a quatro questões, duas das quais extremamente importantes: *“que fatores devem ser considerados para a análise do problema de perdas e inadimplência? Os fatores que contribuem para os problemas de perdas e inadimplência são iguais para todo o país, ou existem diferenças entre as regiões?”*. A análise é conjunta, perdas e inadimplência, pois o autor acredita que estes dois problemas andam lado a lado e que não é apenas o motivo renda que está na origem dos dois. Examina a correlação referente a mais que cinquenta parâmetros e propõe um modelo composto só por três parâmetros setoriais (tarifa da concessionária, percentual de consumo residencial e a posse de ar condicionado) e um fator relativo ao desenvolvimento (o percentual de crianças entre 7 e 14 anos no ensino fundamental). Este modelo explica bem as perdas nas regiões Norte, Nordeste e Centro-Oeste enquanto que para as regiões Sudeste e Sul também incluiu os parâmetros índice de posse de bens e taxa de óbitos por agressão. Para a inadimplência são outros os aspectos a considerar e as conclusões.

Da maior importância com respeito às perdas não técnicas é a NT/348/SRE/ANEEL (2007) na qual é apresentado o modelo proposto para o tratamento regulatório das perdas comerciais considerado quando do segundo ciclo de revisão tarifária. Uma questão relevante colocada é quais fatores considerar nas diferentes áreas do país. Segue-se um estudo objetivando *“elaborar um índice que permite diferenciar as áreas de concessão quanto a determinadas características sócio-econômicas”* (p.11), no qual há uma parcela atribuída à ineficiência gerencial da empresa e outra relacionada às características sócio-econômicas da área de concessão; nesta parcela são examinadas algumas variáveis ligadas às dimensões infraestrutura, educação, violência, e judiciário. Ao final posiciona as empresas brasileiras quanto à tendência de maior ou menor presença de perdas comerciais. Importante é reconhecer que o órgão regulador pode determinar o nível máximo ou uma trajetória para as perdas, visando sinalizar a eficiência. A NT/348 cita a importância de um *“diagnóstico de perdas”*, e de *“um plano de combate às perdas”* cujas ações possam ser avaliadas economicamente (pp.25-26).

A Tabela 2.1 apresenta as principais referências relativas às perdas comerciais.

Tabela 2.1 - Principais referências relativas a perdas comerciais.

Autor, ano	Título	Descrição
Smith, 2004	Electricity Theft: a comparative analysis	Analisa as perdas por furto e falta de pagamento em cerca de cem países. Mostra crescimento das perdas e correlação destas com a governança.
Vieiralves, 2005	Proposta de uma Metodologia para Avaliação das Perdas Comerciais dos Sistemas Elétricos: O Caso Manaus	Analisa PC na região Norte (caso Manaus) procurando identificar causas e efeitos para a sociedade e empresários. Muito voltado à busca de perdas por fraude.
Dantas, 2006	Avaliação de perdas de energia elétrica não técnicas, metodologia aplicada ao município de Salvador	Descreve a metodologia e o desenvolvimento de um software para prospecção das PC no que denominou “matriz de perdas”.
Almeida, 2006	Metodologia de identificação de perdas não-técnicas em unidades consumidoras por atividade de consumo	Metodologia para identificação de PC por atividade de consumo visando melhor identificar UC com maior probabilidade de irregularidade.
Durães et al, 2006	Inteligência Computacional nas Distribuidoras de Energia Elétrica: evolução tecnológica, aplicações e impactos na redução das perdas não-técnicas.	Apresenta as principais aplicações de IC e seus impactos na redução das PNT, no que tange a seleção de UC para inspeção visando melhor índice de acerto, e destaca técnicas mais promissoras.
Francisco e Fagundes, 2006	Estudo geostatístico entre variáveis sócio-econômicas e perda de energia	Modelo de regressão multivariada através da avaliação espacial. Associa renda, clientes cortados e localização de favelas. Propõe um indicador de propensão.
NT/348/SRE, ANEEL, 2007	Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica	Apresenta os critérios e metodologia adotados pela ANEEL quanto ao nível máximo de PNT para o segundo ciclo de revisão tarifária.
Cometti, et al 2007	Uma abordagem baseada no conhecimento para identificação das perdas elétricas	Propõe uma abordagem capaz de automatizar o processo de seleção das UC candidatas a inspeção visando detectar perdas e fraudes.
Ferreira, 2008	Uso de ferramentas de aprendizado de máquina para prospecção de perdas comerciais em distribuição de energia elétrica	Identificação de fontes das PC focando fraude, medidores defeituosos. Analisa ferramentas de aprendizado e compara quatro estratégias de classificação.
Aranha Neto et al, 2008.	Combate às perdas não técnicas no Brasil	Discussão sobre fraudes e PNT. Destaque aos tipos e ao enquadramento jurídico.

2.3 – Breve revisão referente a redes Bayesianas

As redes Bayesianas (RB) são grafos acíclicos orientados cujos “nós” representam as variáveis e aos “ramos” que interligam estes nós estão associadas relações de dependência entre tais variáveis em um modelo probabilístico. As RB têm se mostrado uma ferramenta poderosa e adequada na aplicação em sistemas inteligentes.

Os sistemas inteligentes fornecem as práticas, as regras do raciocínio e do conhecimento humanos às máquinas visando resolver problemas, e “*o campo da inteligência artificial tenta não apenas compreender, mas também construir entidades inteligentes*” (RUSSEL e NORVIG, p.3, 2004). Há muitas definições para inteligência artificial, algumas destacando os processos de pensamento e raciocínio, outras que as máquinas usem a linguagem mais natural possível, e ainda aquelas que destacam a atuação ou comportamento os mais racionais possíveis.

São muitas as dificuldades quando se deseja traduzir em regras e conhecimento as informações obtidas. A ciência usa modelos e teorias para representar a realidade, e face à almejada universalização das explicações científicas a pesquisa segue uma padronização metodológica. Tal padronização envolve a apresentação do problema, a colocação das hipóteses, os modelos e algoritmos usados na solução, o desenvolvimento bem como testes e validação. Quando se tem um problema é importante saber quais os objetivos e até onde se deseja chegar, além dos métodos mais indicados e aproximações desejadas. A solução certamente virá através do raciocínio, das representações e heurísticas.

O estudo das perdas, embora parcialmente observável, é estocástico devido às ações dos agentes externos envolvidos e trabalha com muitas incertezas. Não ter certeza sobre um ambiente não quer dizer que não se deve agir ou tomar decisões neste ambiente. RUSSEL e NORVIG (2004, p.450) acrescentam: “*diagnóstico é uma tarefa que quase sempre envolve incerteza*”. As redes Bayesianas são aplicadas em sistemas inteligentes visando diagnóstico. A falta de conhecimento pode ter origem na impossibilidade ou na ignorância teórica que é não dispor de todo o domínio do problema. A ignorância e impossibilidade estão presentes nos problemas de perdas, pois obter todas as informações de modo confiável é extremamente trabalhoso e oneroso.

Assim, em ambientes com incertezas as decisões podem ser tomadas usando-se raciocínios apropriados e dentre estes está o probabilístico. As RB trabalham com probabilidades e se

fundamentam nas probabilidades condicionais. O termo homenageia a Thomas Bayes que no século XVIII as aplicou para obter estimativas (ou *crenças*) relativas à probabilidade *a priori* conhecidas as probabilidades *a posteriori*.

Em probabilidade há três tipos de abordagem:

- A clássica que está calcada na simetria do experimento,
- O enfoque axiomático estabelecido por Kolmogorov no início do século XX e que une a Teoria dos Conjuntos com a Teoria da Medida (ALENCAR, 2008),
- A abordagem por frequência relativa (também denominada de frequentista).

Segundo NEAPOLITAN (2004, p.17) foi Richard Von Mises que “em 1919 desenvolveu a abordagem de frequência relativa aplicada à probabilidade” a qual envolve a repetibilidade de experimentos idênticos. O importante é que a interpretação da probabilidade permite associá-la a uma medida da incerteza.

A aplicação prática da probabilidade enquanto *chance* está associada a uma avaliação subjetiva no sentido de *crença*, bastante voltada ao tratamento das incertezas. Isto pode ser ilustrado com perguntas como: Qual a probabilidade da safra brasileira de grãos ultrapassar 140 milhões de toneladas, em 2009? Ou, fará sol amanhã pela tarde? As respostas podem variar conforme a sensibilidade do entrevistado ou função das informações disponíveis pelo especialista.

Como curiosidade a safra de grãos no país foi de 144 milhões de toneladas em 2008. Ao final de 2008 (06.11.08) o Instituto Brasileiro de Geografia e Estatística, que congrega especialistas, estimou a safra para 2009 em 140,8 milhões (IBGE, 2009), e colocou que “o cenário atual não é definitivo em face das incertezas geradas pela crise mundial... o que poderá redundar em menor utilização de tecnologia e uma safra 3,3% menor”. Mais adiante, em 09.3.09 já se falava em uma safra de 135,3 milhões (CANALRURAL, 2009), e ao final de 2009, chegando-se ao final da safra, o Conselho Nacional de Abastecimento – CONAB divulga que “este foi o segundo melhor resultado brasileiro, tendo sido produzidos 134,8 milhões de toneladas de grãos” (AGROSOFT, 2009). Vê-se que mesmo entre especialistas as estimativas são distintas inclusive variando no tempo em função de novas informações.

JUDEA PEARL (1991) estudou as regras de propagação das RB, usou circuitos lógicos para demonstrar o processo de crença e introduziu um algoritmo de formação de agrupamentos no final dos anos 80 que permitiu o desenvolvimento de ferramentas computacionais adequadas para

tratar incertezas. Isto possibilitou a aplicação da teoria de Bayes a outras áreas do conhecimento inclusive o desenvolvimento de sistemas especialistas.

Assim é que as RB têm sido aplicadas com sucesso tanto na área médica visando obter conclusões a respeito do diagnóstico de doenças através da análise dos sintomas, como em exploração mineral (CHIEN et al, 2002). Também tem sido largamente usada em genética, investigações criminais ou dos seus fatores, gerenciamento de risco terrorista, problemas de inferência na ciência forense e outros (POURRET et al, 2008).

Aplicações em sistemas de potência têm sido freqüentes na área de confiabilidade: YAN et al (2008), YIN et al (2009) e RAMOS et al (2007), dentre outros. Também tem sido empregada em diagnóstico de falta (YONGLI et al, 2006), em localização de faltas (CHIEN et al, 2002) e na análise de risco de exceder o limite térmico em linhas de transmissão (GALLUS et al, 2002). No Brasil, as RB também têm sido utilizadas como classificadores a exemplo de MATOS (2004) e algumas aplicações na área médica.

Não foram encontradas pesquisas aplicando redes Bayesianas na prospecção das perdas comerciais de energia.

Os conhecimentos teóricos relativos às RB são apresentados em alguns livros que juntamente com os artigos mais importantes estão apontados na Tabela 2.2.

Tabela 2.2 – Síntese das referências bibliográficas sobre redes Bayesianas.

Autor, ano	Título	Origem	Importância
Pearl, J. 1991	Probabilistic Reasoning in intelligent systems: networks of plausible inference	Morgan Kaufmann Publishers, Inc., San Francisco.	Fundamental no estudo das RB, agrupamento e algoritmos.
Neapolitan, R. E., 2004.	Learning Bayesian networks.	Pearson Education, Inc., New Jersey, EUA.	Básico em inferência e aprendizado.
Jensen and Nielsen, 2007.	Bayesian networks and decision graphs.	Springer Science, New York, USA.	Completo, ajuda na confecção da rede.
Madsen A.L., and Kjaerulff U.B, 2008.	Bayesian Networks and Influence Diagrams: a guide to construction and analysis.	Science+Business Media, New York, EUA.	Apresenta bem os fundamentos teóricos.
Pourret O.; et al, 2008.	Bayesian networks: a practical guide to applications.	John Wiley & Sons Ltd. England.	Contem aplicações a diversas áreas.
Pearl, J., 2009.	Causality: models, reasoning, and inference.	Cambridge University	Apresenta modelos e diagramas causais.

Autor, ano	Título	Origem	Importância
Yan et al, 2008.	Bayesian network based time-sequence simulation for power system reliability assessment	7th Mexican International Conference on Artificial Intelligence, Atizapan de Zaragoza.	Confiabilidade. Usa algoritmo de inferência com RB e seqüência de amostras.
Yin et al, 2009.	Systems reliability analysis and fault diagnosis based on Bayesian networks	International Workshop on Intelligent System and Application, Wuhan, China	Modela sistemas série e paralelo e traz aplicação à indústria.
Chien, Chen and Lin, 2002.	Using Bayesian network for fault location on distribution feeder	Transactions on Power Delivery, Vol.17, No.3, july 2002	Aplicação a sistemas de potência, localização de faltas.
Ramos L. et al, 2007.	Análisis de confiabilidad de sistemas industriales aplicando RB, considerando aspectos de PQ e seguridad – Caso de Estudio Sistema IEEE493	IEEE Latin America Tansactions Vol.5, No. 8, December, 2007	Uma das aplicações das RB em confiabilidade. Caso de estudo o sistema IEEE.

3. NOÇÕES DE REDES BAYESIANAS

Com respeito a probabilidade já foi colocado no segundo capítulo que há três tipos de tratamento: o clássico, o axiomático e o de frequência relativa (ou *frequencista*). A estrutura de rede Bayesiana empregada no diagnóstico das perdas usa os conceitos de probabilidade, está baseada na observação de eventos ao determinar probabilidades condicionais de algumas variáveis, e utiliza a probabilidade com o sentido de crença ou *probabilidade subjetiva* (subjective probability) como nomeia NEAPOLITAN (2004, p.22). No modelo usado para o diagnóstico das perdas, o levantamento das probabilidades condicionais especialmente para a variável *Inspeção* é frequencista, enquanto a análise da conformidade tem o sentido de crença.

As redes Bayesianas se constituem em uma boa estratégia para “*trabalhar com incertezas especialmente quando não se tem todas ou muitas informações*” (RUSSEL e NORVIG, 2004, p.451), e aqui sempre estará exprimindo uma probabilidade no sentido de crença.

Este capítulo está desenvolvido em três itens: o primeiro trata de alguns conceitos de probabilidade especialmente o teorema de Bayes, o segundo traz uma introdução às redes Bayesianas (RB), e o último apresenta exemplos de modelos de RB. Outras técnicas e conceitos referentes às RB podem ser encontrados nas referências bibliográficas.

3.1 – Alguns conceitos de probabilidade e o Teorema de Bayes

São apresentadas noções de probabilidade, os teoremas da probabilidade condicional e de Bayes, montadas tabelas de probabilidade condicional (TPC), e enfatizados os conceitos de marginalização e normalização. Destacam-se a importância da verificação da independência entre variáveis e a definição clara dos seus estados que devem ser mutuamente exclusivos. Os exemplos visam facilitar a compreensão.

3.1.1 – Probabilidade condicional, TPC e fator de marginalização

Na probabilidade tradicional o conjunto de todas as possibilidades é tratado como universo, mas sob o aspecto frequencista se coloca tal conjunto como espaço amostral ou simplesmente

amostra. NEAPOLITAN (2004) destaca que os cientistas sociais preferem mesmo falar de população.

Qualquer subconjunto é um evento. Assim, por exemplo, jogar um dado comum cujas faces têm os algarismos de um a seis é um evento; tratando-se de um dado não viciado, a probabilidade de após a jogada a face apresentada ser três (terno) é $1/6$. Matematicamente se coloca que o universo é o conjunto E

$$E = \{1, 2, 3, 4, 5, 6\} \quad (3.1)$$

Neste caso a probabilidade de sair cada um dos elementos do espaço amostral é a mesma

$$P(x_i) = P(3) = 1/6 \quad (3.2)$$

Dentre as regras seguidas pela Teoria da Probabilidade é importante esta que trata da probabilidade de cada elemento

$$0 \leq P(x_i) \leq 1 \quad (3.3)$$

Logo a probabilidade de um evento ser igual a 1,0 é a certeza, e sendo zero corresponde à impossibilidade.

Chamando o espaço amostral ou universo de Ω , terminologia usual, e X_i um evento do subconjunto deste espaço, os três axiomas estabelecidos por Andrei Kolmogorov são

$$P(\Omega) = 1 \quad (3.4)$$

$$P(X_i) \geq 0 \quad (3.5)$$

$$P\left(\bigcup_i^N X_i\right) = \sum_i^N P(X_i) \quad (3.6)$$

Estas noções de probabilidade estão colocadas com eventos discretos, entretanto também se aplicam a variáveis contínuas; neste caso usa-se a função densidade de probabilidade.

Sejam dois eventos “a” e “b”, sendo

$$P(b) > 0 \quad (3.7)$$

a probabilidade condicional é expressa como

$$P(a|b) = \frac{P(a \cap b)}{P(b)} \quad (3.8)$$

Onde:

$P(a|b)$ é a probabilidade condicional, acontecer “a” tal que tenha acontecido “b”;

$P(a \cap b)$ é a probabilidade de acontecer “a” e “b” simultaneamente, da teoria dos conjuntos “a” inter “b”, ou o “e” lógico $P(a \wedge b)$.

A equação (3.8) na versão frequencista se apresenta assim

$$P(a|b) = \frac{P(a \cap b)}{P(b)} = \frac{n_{ab}/n}{n_b/n} = \frac{n_{ab}}{n_b} \quad (3.9)$$

Onde:

n é o total da amostra ou sorteio (ou a população);

n_b é o número de vezes que se registrou o evento b;

n_{ab} é o número de vezes que aconteceu a simultaneidade a e b.

Introduz-se um exemplo com cartas usando todo um baralho (52 cartas), tratando os conjuntos *ímpar*, *par* e *figurada*, e os naipes de forma completa: ouros, espadas, copas e paus. Uma questão de probabilidade condicional pode ser colocada como: qual a probabilidade de ter sido tirada uma carta de copas tal que tenha sido figurada? Isto significa que quem retirou a carta olha-a e diz que a mesma é figurada. Existem 12 cartas figuradas no universo de 52, e figuradas de copas são três. Logo a probabilidade é

$$P(\text{copas}|\text{figurada}) = \frac{3/52}{12/52} = \frac{3}{12} = 0,25$$

As redes Bayesianas tratam de variáveis e seus estados mutuamente exclusivos e assim neste exemplo das cartas a variável naipe teria quatro estados.

A expressão da probabilidade condicional quando aplicada a variáveis X e Y em geral é apresentada com esta notação (PEARL, 1991, p.31)

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} \quad (3.10)$$

A variável X podendo assumir n estados e a variável Y outros m estados, $P(X,Y)$ são as probabilidades dos $n \times m$ possíveis estados conjuntos.

Aplicando às cartas, X pode assumir os *estados* seguintes: par, ímpar e figurado. Para a variável Y, os naipes, são quatro *estados* logo existe uma matriz ou *tabela de probabilidade conjunta* neste caso com dimensão 3x4:

		<i>ouros</i>	<i>paus</i>	<i>copas</i>	<i>espadas</i>
$P(X, Y) =$	<i>par</i>	$5/52$	$5/52$	$5/52$	$5/52$
	<i>ímpar</i>	$5/52$	$5/52$	$5/52$	$5/52$
	<i>figurada</i>	$3/52$	$3/52$	$3/52$	$3/52$

Tabelas de probabilidade conjunta mostram os valores das probabilidades relativos às combinações dos possíveis estados das variáveis, devendo a soma total ser um (ou 100%). Alerta-se que com tabelas de probabilidade condicional são as linhas que somam um.

Atendendo o conceito de estados mutuamente exclusivos (sem intersecção entre si) são n as opções de estados X_i para a variável X . Neste exemplo das cartas n é 3 correspondendo a *par*, *ímpar* e *figurada*. Portanto obtém-se a probabilidade da “carta ser par” pelo somatório da primeira linha da matriz anterior com a probabilidade conjunta e neste caso é $20/52$. A probabilidade da carta ser figurada é o somatório da terceira linha ou $12/52$. Generalizando, este é o *fator de marginalização*

$$P(X) = \sum_Y P(X, Y) \tag{3.11}$$

Ou em termos dos x_i estados da variável X e y_j estados da variável Y

$$P(x_i) = \sum_{j=1}^m P(x_i, y_j) \tag{3.12}$$

Assim, aplicado ao exemplo

$$P(X) = P(x_1, x_2, x_3) = P(\textit{par}, \textit{ímpar}, \textit{figurado}) = (20/52, 20/52, 12/52)$$

Conforme RUSSEL e NORVIG (2004, p.462) era uma “*prática comum entre as companhias de seguros de escrever as somas das freqüências observadas nas margens de tabelas de seguros*”, daí a denominação *marginalização*. MADSEN e KJAERULFF, (2008, pp.41-42) denominam o fator de marginalização de *regra da probabilidade total* apresentada de um modo como eles próprios dizem “*bem menos compacto*”, aqui com $m=3$ fica

$$P(X) = (x_1, x_2, x_3) = (\sum_{j=1}^3 P(x_1, y_j), \sum_{j=1}^3 P(x_2, y_j), \sum_{j=1}^3 P(x_3, y_j)) \tag{3.13}$$

Ainda do exemplo anterior o fator de marginalização também se aplica às colunas, por exemplo, a probabilidade sair uma carta de *ouros* é igual ao somatório da primeira coluna, ou seja, $13/52$.

3.1.2 - Teorema de Bayes

A equação (3.8) mostra a probabilidade condicional de a tal que tenha acontecido b . Tal equação na forma recíproca, ou seja, a condicional de b tal que tenha acontecido a é

$$P(b|a) = \frac{P(b \cap a)}{P(a)} \quad (3.14)$$

No século XVIII Thomas Bayes mostrou que a intersecção dos conjuntos é comutativa, ou

$$P(a \cap b) = P(b \cap a) \quad (3.15)$$

Logo, igualando a probabilidade destas interseções e re-arrumando as equações (3.8) e (3.14) resulta na expressão conhecida como *Regra ou Teorema de Bayes*

$$P(a|b) = \frac{P(b|a).P(a)}{P(b)} \quad (3.16)$$

O importante é que Bayes apresentou um método de estimar a crença relativa a um evento “a” conhecidas informações sobre um evento “b”. Normalmente $P(a)$ é uma probabilidade *a priori*, $P(a|b)$ é chamada de probabilidade *a posteriori* e $P(b|a)$ é denominada de esperança ou crença de que aconteça “b” tal que tenha acontecido “a” (JENSEN e NIELSEN, 2007, p.5).

Há os que preferem apresentar a regra de Bayes como uma relação de causa e efeito, ou

$$P(causa|efeito) = \frac{P(efeito ,causa)}{P(efeito)} = \frac{P(efeito |causa)P(causa)}{P(efeito)} \quad (3.17)$$

Bayes também mostrou que a expressão (3.16) é válida para variáveis. Daí

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)} \quad (3.18)$$

Combinando esta com a equação (3.11), resulta

$$P(Y|X) = \frac{P(X|Y).P(Y)}{\sum_Y P(X,Y)} = \frac{P(X|Y).P(Y)}{\sum_Y P(Y).P(X|Y)} \quad (3.19)$$

A equação (3.19) e a regra de Bayes como em (3.17) são muito usadas nos modelos de RB aplicados na área médica, relacionando doenças e sintomas. Estas expressões também são importantes no desenvolvimento dos algoritmos de inferência nas redes Bayesianas.

3.1.3 – Constante de normalização, regras fundamentais e independência entre variáveis

As variáveis nas redes Bayesianas assumem valores contínuos ou discretos; quando discretos os valores de probabilidade ou crença são associados aos diversos *estados* da variável. Por vezes para obter as tabelas de probabilidades condicionais (TPC) são usadas as probabilidades conjuntas de duas ou mais variáveis. No caso de valores contínuos são usadas funções densidade de probabilidade, ou os valores são discretizados e se trabalha com intervalos.

Já a *inferência probabilística* trata com hipóteses: em geral se escolhe um estado e o mesmo é evidenciado, ou seja, é tratado como se tivesse 100% de chance de acontecer. É como no exemplo das cartas, ao tirar uma delas passa a ser conhecido o estado do naipe; caso se anuncie ser de espadas, as probabilidades de que seja par, ímpar ou figurada são respectivamente 5/52, 5/52 e 3/52. A equação (3.19) equivale a:

$$P(Y|X) = \frac{P(X \cap Y)}{\sum_Y P(X, Y)} \quad (3.20)$$

Portanto a partir das probabilidades conjuntas pode ser conhecida a probabilidade da carta ser par dado que o naipe seja espadas, por exemplo, visto que é uma inferência probabilística em “naipe de espadas”, e a crença é obtida aplicando-se a expressão anterior:

$$P(\text{par}|\text{naipe} = \text{espadas}) = 100 \left(\frac{5/52}{5/52 + 5/52 + 3/52} \right) = 38,46\%$$

Isto posto pode-se dizer que quando se busca a inferência probabilística se usa uma evidência (ε) de uma variável, isto é, se fixa um estado desta variável. Assim, existindo outra variável X define-se a *constante de normalização* alfa (α) por:

$$\alpha = P(\varepsilon) = \sum_X P(X, \varepsilon) \quad (3.21)$$

Neste caso diz-se que a constante de normalização pode ser determinada “pelo somatório sobre X das probabilidades conjuntas da evidência (ε) e da variável” (MADSEN e KJAERULFF, 2008, p.109), ou seja, somatório de $P(X, \varepsilon)$ em X .

Essencial é verificar a relação de dependência entre as variáveis (HADI et al, 1998, pp.75-76), e certificar-se que os estados escolhidos para as mesmas são mutuamente exclusivos. É importante

destacar que a soma de todos os eventos, isto é, todos os elementos da tabela de probabilidade conjunta é 1,0. É análogo à expressão (3.6) aplicada a variáveis; matematicamente tem-se:

$$P(X, Y) = \sum_{i=1}^n \sum_{j=1}^m P(X = x_i, Y = y_j) = 1 \quad (3.22)$$

Com relação à análise da dependência entre as variáveis, a variável X é dita independente da variável Y quando com respeito às suas probabilidades se pode escrever:

$$P(X|Y) = P(X) \quad (3.23)$$

Para a independência vale a comutatividade então neste caso também Y é independente de X.

No caso de variáveis independentes, a probabilidade conjunta corresponde ao produto de cada uma das probabilidades independentes. Das expressões (3.18) e (3.23), vem:

$$P(X, Y) = P(X|Y).P(Y) \quad \Rightarrow \quad P(X, Y) = P(X).P(Y) \quad (3.24)$$

Generalizando a equação (3.24) para n variáveis aleatórias independentes conduz a:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i) \quad (3.25)$$

A expressão (3.25) é semelhante àquela normalmente usada no cálculo de probabilidades em eventos independentes, e é definida em HADI et al (1998, p.75) como *independência de um conjunto de variáveis*. Também está associada à *regra de cadeia*, importante propriedade das RB a ser abordada no item seguinte.

A seguir faz-se outro exemplo para mostrar a importância da observação dos dados e da relação de dependência entre os eventos. Neste exemplo a abordagem é frequencista. Destaca-se a comparação das probabilidades calculadas conforme uma pesquisa e o resultado teórico esperado, sendo testada a significância da amostra. Ao trabalhar como se os eventos fossem dependentes, é preparada a tabela de probabilidade conjunta oriunda de dados da pesquisa mostrando-se como através do fator de marginalização se chega à TPC. Posteriormente usa-se este caso na montagem de uma rede Bayesiana.

Foram usadas as cartas de ouros (O) e paus (P). Fez-se o sorteio de duas, uma de cada naipe, sendo classificadas em par, ímpar e figura. Como eram poucas cartas, para melhor embaralhá-las foi usado um baralho cujo verso das cartas de ouros era verde, e vermelho para as cartas de paus. Foi feito assim para que não houvesse a chance de no mesmo sorteio saírem duas cartas de ouros

ou duas de paus. Depois as cartas foram novamente baralhadas e fez-se novo sorteio. Foram 301 eventos cujos dados estão mostrados na Tabela 3.1.

Observa-se que houve predominância da carta par, especialmente dentre aquelas de ouros. Visto que teoricamente (usado o índice t) para ambos os naipes, paus ou ouros, as probabilidades de sair cada uma das cartas conforme classificadas em par, ímpar e figura compõem uma *hipótese inicial* (H_0), e neste exemplo é:

$$P_t(Ouros_i) = \left(\frac{5}{13}, \frac{5}{13}, \frac{3}{13} \right) = (0,3846 \quad 0,3846 \quad 0,2308)$$

Tabela 3.1 - Resultado da observação dos sorteios de cartas de ouros e paus.

		Paus			
		Par	Ímpar	Figura	Total
Ouros	Par	48	50	28	126
	Ímpar	40	39	30	109
	Figura	28	22	16	66
Total		116	111	74	301

Inicialmente considerem-se independentes os eventos dentro de cada sorteio. A probabilidade de se tirar uma carta par do naipe ouros é de 126 dividido por 301 resultando em 0,4186 diferente portanto do valor teórico anteriormente estabelecido. Se tirar uma figura de paus é $74/301$, ou seja, 0,2458, também distinto do teórico.

Trabalhando-se com eventos independentes, pode-se perguntar: qual a probabilidade de sair uma carta par dentre as do naipe ouros e ímpar dentre aquelas de paus?

A resposta teórica para esta questão, não considerada a pesquisa amostral seria:

$$P_t(O = par, P = ímpar) = (0,3846) \cdot (0,3846) = 0,1479$$

O valor esperado em função da amostra pesquisada, considerando eventos independentes, é determinado empregando-se o índice EI para caracterizar a probabilidade calculada como eventos independentes, ou seja,

$$P_{EI}(O = par, P = ímpar) = \left(\frac{126}{301} \right) \cdot \left(\frac{111}{301} \right) = 0,1544$$

Trabalhando como se fossem eventos dependentes (o que não é correto neste exemplo), todos os dados da Tabela 3.1 seriam divididos pelo total da amostra (301) para se obter a tabela de

probabilidade conjunta com estas relações de frequência. É usado o índice D para caracterizar que se está considerando eventos dependentes. Assim, a probabilidade de sair uma carta par dentre as do naipe ouros e ímpar de paus seria

$$P_D(O = \text{par} \cap P = \text{ímpar}) = 50/301 = 0,1661$$

A Tabela 3.2 apresenta as probabilidades conjuntas.

Tabela 3.2 - Probabilidades conjuntas, sorteios de cartas de ouros e paus.

		Paus			
		Par	Ímpar	Figura	Total
Ouros	Par	0,1595	0,1661	0,0930	0,4186
	Ímpar	0,1329	0,1296	0,0997	0,3621
	Figura	0,0930	0,0731	0,0532	0,2193
Total		0,3854	0,3688	0,2458	1,0000

Estão indicados os totais por linha e coluna, pois permitem verificar as crenças com respeito a sair cartas de ouros segundo seus estados e também as crenças referentes às cartas de paus. Entretanto os valores são distintos daqueles teoricamente esperados. Por exemplo, a expectativa de que saia uma carta ímpar de ouros corresponde à probabilidade teórica de 0,3846, mas com os dados da amostra sorteada a crença em tal evento é de 0,3621. Enfim, a crença quanto à saída das cartas de ouros é o somatório de cada linha, conforme o fator de marginalização; aplicando a expressão (3.12) tem-se

$$P(\text{ouros}_i) = \sum_{j=1}^m P(\text{ouros}_i, \text{paus}_j) = (0,4186 \quad 0,3621 \quad 0,2193)$$

Para as probabilidades acima indicadas não foi usado qualquer índice visto que, para a distribuição entre naipes, considerar como eventos dependentes ou independentes conduz ao mesmo resultado. Conforme visto, para avaliar a probabilidade de sair determinada carta de ouros e outra de paus é que há influência quanto à dependência dos eventos.

Com estas probabilidades totais mostradas na Tabela 3.2 pode-se fazer um teste de significância em relação à *hipótese inicial* (H_0) pela qual as probabilidades teoricamente esperadas de sair carta par, ímpar (38,46% cada) ou figura (23,08%), de modo a validar a amostra. Usando para teste a

distribuição qui-quadrado (VARDEMAN, 1998), com grau de liberdade quatro (cada variável tem três estados). O valor do qui-quadrado amostral resulta 0,652, valor bem abaixo do limite crítico (9,49). Conclui-se que a amostra é válida.

Prossegue-se com a consideração dos sorteios das cartas como eventos dependentes apenas para mostrar como a partir da tabela de probabilidade conjunta se chega à tabela de probabilidade condicional (TPC). São usados os fatores de marginalização conforme equação (3.11), que são os valores totais por linha na Tabela 3.2. Efetuados os cálculos para todos os elementos, a Tabela 3.3 mostra os valores da TPC que compõem uma matriz de dimensão $n \times m$.

Tabela 3.3 - Probabilidades condicionais entre cartas de ouros e paus em sorteio.

		Paus		
		Par	Ímpar	Figura
Ouros	Par	38,0952	39,6825	22,2222
	Ímpar	36,6973	35,7798	27,5229
	Figura	42,4242	33,3333	24,2424

Para finalizar este exemplo, ressalta-se que ter os resultados da pesquisa amostral não é suficiente para estimar uma probabilidade correta, para fazer inferências, sendo importante a consideração quanto às variáveis. Estas sendo dependentes, se poderia perguntar ainda: qual a probabilidade de ter saído uma carta par do naipe ouros tal que a de paus foi ímpar?

Respondendo à questão considerando como variáveis dependentes são usadas as probabilidades condicionais. Assim, o termo que representa a probabilidade percentual da carta sorteada de ouros ser par tal que a de paus tenha sido ímpar é:

$$P_D(Ouros = par|Paus = ímpar) = 0,1661/0,3688 = 0,4505$$

Tomando os eventos como independentes o resultado seria como definido por (3.23), ou

$$P_{EI}(Ouros = par|Paus = ímpar) = P(Ouros = par) = 0,4186$$

Já que a análise deste problema com cartas é frequencista, o resultado (41,86%) não necessariamente deve ser igual ao teórico (38,46%). Por outro lado a consideração incorreta de que as variáveis são dependentes conduz a um resultado numérico errado (45,05%). Daí, ressalta-se a importância do exame da relação de dependência entre variáveis.

3.2 – Introdução às redes Bayesianas

São introduzidos neste item: a definição de RB, algo a respeito de conceitos como regra de cadeia, do modo de estruturação das redes, além de observações quanto aos modelos, inferências, entropia nas RB, e comentários sobre a explosão combinatória. Isto está dividido em dois subitens.

3.2.1 – Definição das redes Bayesianas, estruturação e regra de cadeia

“Qualquer distribuição de probabilidade conjunta pode ser representada por uma rede Bayesianas” (POURRET et al, 2008, p.12) e estas apresentam uma estrutura genérica sujeita às relações de probabilidade conjunta aplicada a n variáveis aleatórias como mostrada na Figura 3.1. Estes autores advertem que qualquer problema determinístico pode ser visto como um caso particular de um modelo probabilístico, podendo quaisquer destes ser representados pelas RB.

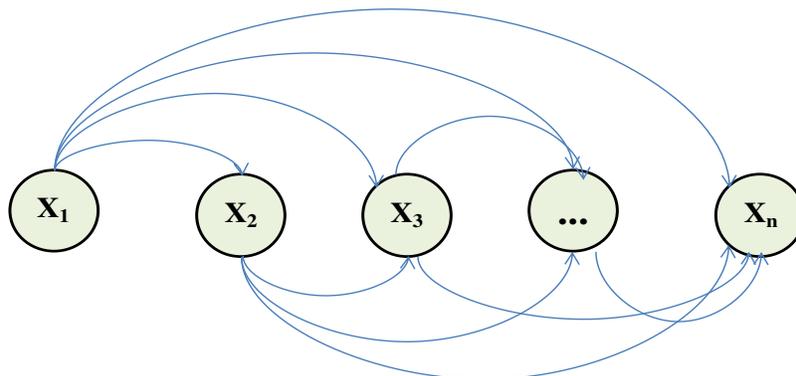


Figura 3.1 – Estrutura genérica de RB aplicada ao conjunto de n variáveis (POURRET et al, 2008, p.13).

As redes Bayesianas são grafos acíclicos orientados cujos “nós” representam as variáveis e cujos “ramos” associam as relações de dependência entre as mesmas em um modelo probabilístico. Diz-se também que os ramos ou arestas direcionadas representam as ligações de causa-e-efeito entre variáveis. São denominadas variáveis *Pais* aquelas causais (ou a montante), e de nós *Filhos* aquelas “sob efeito de” (ou colocadas a jusante). A Figura 3.2 traz exemplos de RB com conexão convergente (a), e divergente (b). Na Figura 3.2 (a) a variável *Chove* é pai de *Grama_Molhada*. Em 3.2 (b) a variável *A* é pai de *B*, *C*, ... e *E*.

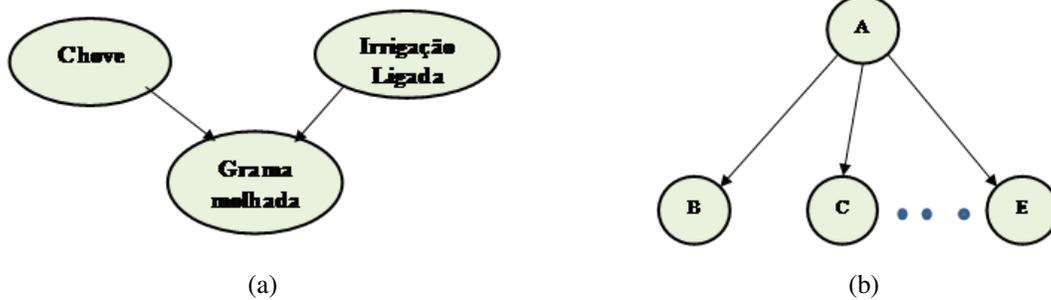


Figura 3.2- Exemplos de RB - (a) Conexão convergente, e (b) Conexão divergente (adaptadas de RUSSELL e NORVIG,2004, e de JENSEN e NIELSEN, 2007, respectivamente).

Uma rede bayesiana por definição consiste em (JENSEN e NIELSEN, 2007, pp.32-35):

- Um conjunto de variáveis $U = \{A_1, A_2, \dots, A_n\}$ e de linhas direcionadas entre variáveis (um grafo orientado).
- Cada variável tem um conjunto finito de estados mutuamente exclusivos.
- Devem ser grafos direcionados acíclicos (em inglês, DAG ou *directed acyclic graph*), ou seja, não devem conter ciclos ou malhas fechadas.
- Para cada variável pai A estão associadas probabilidades através de uma tabela de probabilidades condicionais em relação aos descendentes ou filhos.
- Variáveis de entrada como A não têm pais, são *causas-raiz* (RUSSEL e NORVIG, 2004), e a estas deve ser associada uma tabela de $P(A)$.

A construção da rede deve fornecer a descrição completa do domínio, sendo importante o conhecimento das relações entre as variáveis, suas dependências condicionais, e o uso de uma linguagem formal e unificada. HADI et al (1998, pp.115-175) traz bom conteúdo a respeito da teoria dos grafos, o que é fundamental. Nas referências bibliográficas há outros conceitos como o estabelecimento de cenários de evidência, a propagação em árvore de junção, e classificações das RB.

Muitos modelos de RB são elaborados recorrendo-se ao conhecimento dos especialistas na área envolvida, mas outros utilizam o aprendizado automático conforme dados e parâmetros numéricos. O exame das relações de dependência, categorização e linguagem com respeito às variáveis é essencial na própria escolha do modelo, podendo-se em geral trabalhar com cinco tipos: síntese, medidas, causa-efeito, reconciliação e indução (NEIL et al, 2000 citado por MADSEN e KJAERULFF, 2008, p.160).

Diagramas causais, estrutura contra-factual, e outros conceitos relativos à causalidade estão em PEARL (2009). Vários autores alertam mostrando modelos causais não bem estruturados: o modelo para avaliação da *trigliceris elevada* (POURRET et al, 2088, p.23), e o teste da *inseminação positiva na vaca* (JENSEN e NIELSEN, 2007, pp.55-56).

Ainda com respeito a aspectos teóricos, um dos teoremas mais importantes é o da *regra da cadeia* (*chain rule*), que trata da probabilidade da junção para qualquer distribuição de probabilidade (JENSEN e NIELSEN, 2007, pp.35-36), que é uma aplicação do teorema de Bayes visto no item anterior. Para duas variáveis (A_1 e A_2) é:

$$P(A_1, A_2) = P(A_2|A_1).P(A_1) \quad (3.26)$$

Como generalização, seja o universo U contendo n variáveis, diz-se que a probabilidade conjunta das variáveis de uma RB é o produto de todas as probabilidades condicionais especificadas na rede (PEARL, 2009, p.69), e é expressa por:

$$P(U) = \prod_{i=1}^n P(A_i|pais(A_i)) \quad (3.27)$$

Onde $pais(A_i)$ denotam os valores específicos das variáveis pais.

3.2.2 – Inferência, entropia em redes Bayesianas e explosão combinatória

A partir do Teorema de Bayes e da regra de cadeia (3.26), PEARL (2009, p.5) destaca que a crença em uma hipótese H obtida através de uma evidência e pode ser função do produto da crença inicial em H , isto é, $P(H)$, pela probabilidade $P(e|H)$, que é a probabilidade desta evidência se concretizar tal que H , acrescentando que “*o coração para a inferência Bayesiana*” é:

$$P(H|e) = \frac{P(e|H).P(H)}{P(e)} \quad (3.28)$$

Onde:

$P(H|e)$ é a probabilidade *a posteriori*,

$P(H)$ é uma probabilidade *a priori*, e

$P(e)$ é a probabilidade da evidência.

Em geral $P(e)$ é normalizada, não com o fator de marginalização como na expressão (3.11), mas usando a hipótese e sua negação ($\neg H$). Deste modo, $P(e)$ é dado por:

$$P(e) = P(e|H).P(H) + P(e|\neg H).P(\neg H) \quad (3.29)$$

A RB da Figura 3.2 (a) relativa ao problema da grama molhada é modelada com nova rede na Figura 3.3 (a): o nó *grama molhada* tem como pais as variáveis *chove* e *irrigação*, sendo esta filha do nó *controlador*; o qual tem como pais *chove* e *reservatório*. Estas duas últimas variáveis são de entrada e para ambas devem ser fornecidas suas tabelas de probabilidade independente. Para as demais devem ser obtidas as TPC. A Figura 3.3 (b) é a tela do programa Netica® (NORSYS, 2008), mostrando os nós e seus estados ainda sem os dados de probabilidade (nesta situação os estados aparecem com distribuição uniforme de probabilidades entre os mesmos).

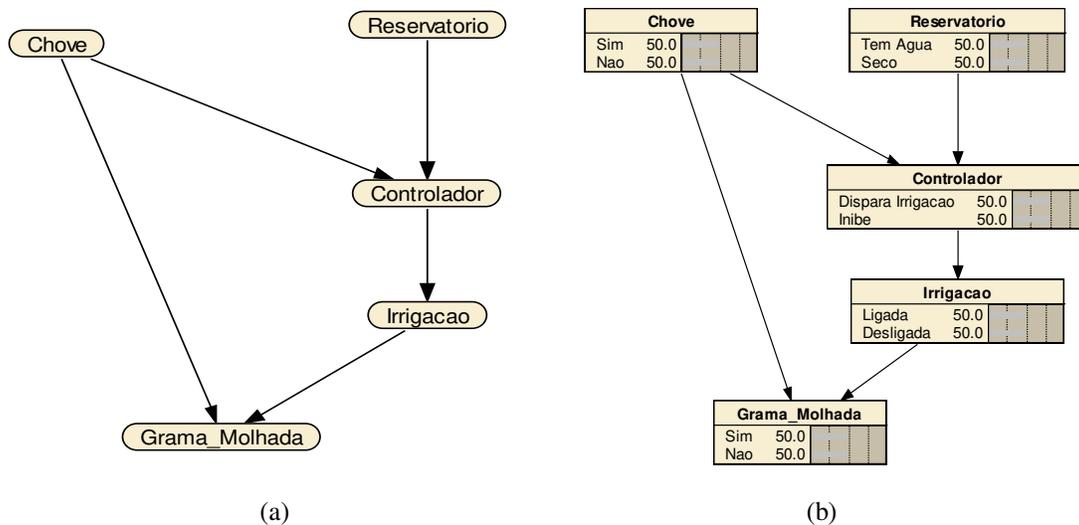


Figura 3.3- (a) Estrutura da rede *Grama Molhada*, e (b) Mesma RB no software Netica® (NORSYS, 2008).

Depois de inseridas as tabelas de probabilidade, a simulação da rede é vista na Figura 3.4 (a). Conforme dados fornecidos, vê-se que a probabilidade de chuva é 10% (estado *Sim*), e de que há água no reservatório é 60%. As TPC não são aqui mostradas. A grama tem 64% de chance de estar molhada (estado *Sim* do nó *Grama Molhada*), ou se diz que esta crença é de 64%.

Instanciar um estado é assumir para o mesmo a certeza ou a evidência (100%). É o que foi feito para o estado *Sim* da variável *Grama Molhada*, na Figura 3.4 (b). A conclusão é que estando a grama molhada há 15,6% de probabilidade (ou crença) de que chova, e uma chance de 93,8% do reservatório ter água, ou que esteja cheio.

Com base nas evidências disponíveis, a atualização das probabilidades *a posteriori* das variáveis de uma RB é denominada *inferência* e para tal existem algoritmos específicos.

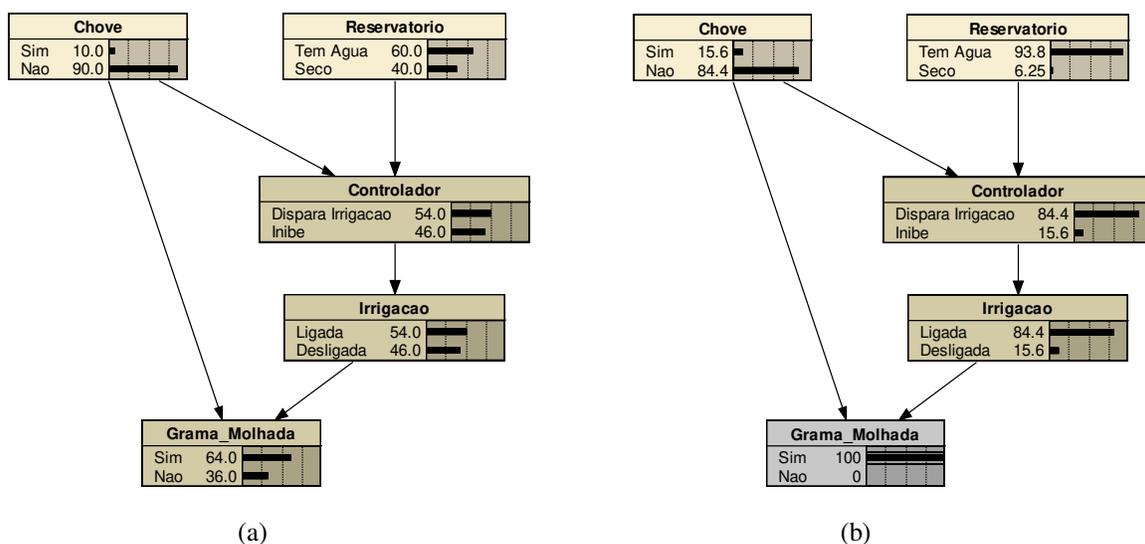


Figura 3.4- RB da *Grama Molhada*, (a) após inseridos dados, (b) instanciado o estado *Sim*.

Para a operação de *inferência* e *instanciação* um ponto importante é a definição de *d-separação* (*d-separation*) associada a situações nas quais uma evidência bloqueia um caminho ou, em outras palavras, as variáveis tornam-se “*separadas*” devido à evidência. Estas situações são três (JENSEN e NIELSEN, 2007, p.30) ilustradas na Figura 3.5.

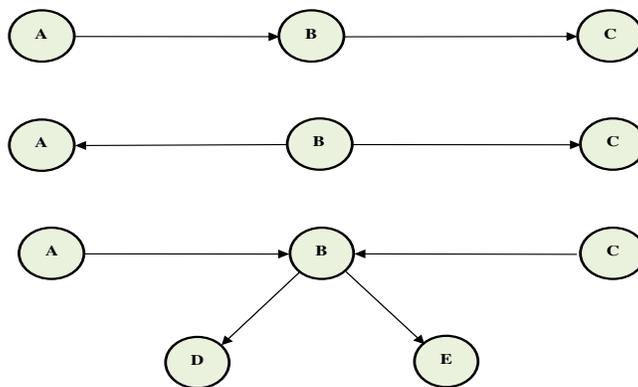


Figura 3.5 – Três redes ilustrando as possíveis situações de d-separação.

As situações caracterizadas na Figura 3.5 são:

- A rede superior é serial A – B – C, ou seja, representa uma condição na qual A influencia B que influencia C; porém se B tem um estado evidenciado (ou conhecido), o canal entre A e C torna-se bloqueado;
- A rede intermediária também A - B - C sendo divergente em B que está evidenciado e, portanto há um bloqueio entre A e C;

- A terceira, rede inferior da figura, é convergente em B vinda de A e C, porém nem B nem qualquer dos seus descendentes (D e E) estão evidenciados.

Ainda com relação à inferência há a *inferência abduativa (abductive inference)*, denominada “*explicação mais provável*” (NEAPOLITAN, 2004, pp. 216-218). Tem aplicação típica em problemas como a identificação da falha mais provável no funcionamento de um veículo ou da mais provável causa de uma doença. JENSEN e NIELSEN (2007, p.98) destacam que muitas RB têm características especiais para o cálculo da explicação mais provável.

Uma das características das RB é a facilidade para obter inferências e a literatura traz muitos algoritmos específicos para isto.

Conforme se apresente o grafo da rede, há rotinas para inferência das redes conectadas por um só ramo, das redes multiconectadas, das redes em árvores, etc. NEAPOLITAN (2004, pp. 125-173) traz vários algoritmos de inferência em redes com variáveis discretas. Em lugar de repetir algum destes algoritmos se prefere analisar inferências através do exemplo, até para mostrar como algumas redes são representadas, já com suas TPC junto às variáveis.

Na Figura 3.6 é mostrada uma rede com cinco variáveis relativas a doenças (bronquite, B, e câncer de pulmão, L) e situações ou sintomas (histórico de fumante é H, fadiga é F, e raio X de tórax, C). Esta rede é usada por NEAPOLITAN (2004, pp. 125-144) para explicar o algoritmo *Message-Passing* de Pearl. A Figura 3.6 é a RB com suas probabilidades.

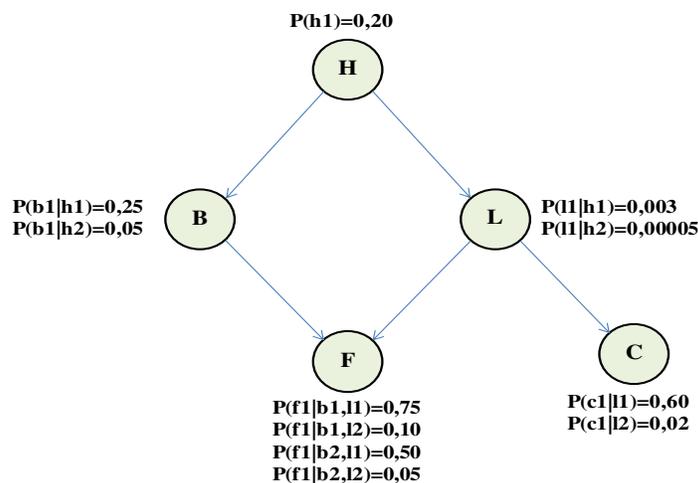


Figura 3.6 – RB relativa a doenças e sintomas; cada variável tem apenas dois estados, sendo mostradas as probabilidades de um deles (original em NEAPOLITAN, 2004, p.125).

Para este modelo de RB fez-se a simulação usando um programa específico, o que é visto na Figura 3.7 (a) antes de colocados os dados das probabilidades condicionais. Para facilitar foram complementados os nomes das variáveis. Todas têm dois estados, e os valores de probabilidade estão em percentual. Após inserir as probabilidades de H e as TPC das demais variáveis vê-se que a crença de que haja bronquite (*sim*) é 9,00% e de câncer de pulmão é 0,064%, Figura 3.7 (b).

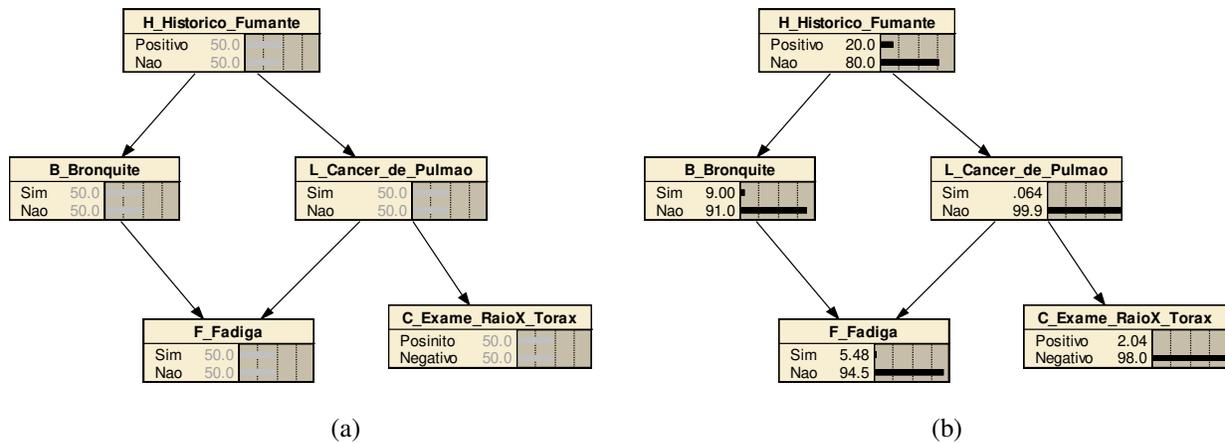


Figura 3.7 – RB relativa a doenças e sintomas - (a) sem dados de TPC, e (b) depois de processada.

A Figura 3.8 (a) reproduz a original que NEAPOLITAN (2004) usa na explicação do mencionado algoritmo fazendo *instanciação* do estado *um* de B, enquanto a Figura 3.8 (b) mostra a simulação desta *inferência*: comparativamente à Figura 3.7 (b) crescem as chances para os estados *sim* ou *positivo* das demais variáveis.

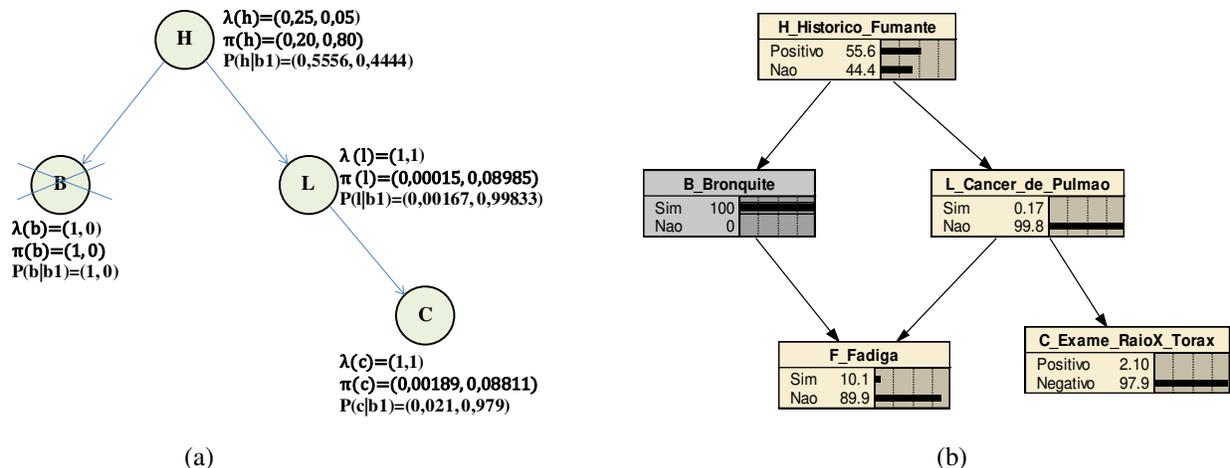


Figura 3.8 – Instanciado o estado um em Bronquite- (a) reprodução do NEAPOLITAN (2004, p.137), (b) depois de simulado no Netica® (NORSYS, 2008).

Na Figura 3.8 (a) as funções λ e π são rotinas do algoritmo de inferência em árvores apresentado em NEAPOLITAN (2004, p.129), vendo-se que as probabilidades, por exemplo, $P(c|b1)$ coincidem nas duas figuras.

Os conceitos de entropia, informação mútua e entropia condicional são aplicados também às RB sendo úteis em estudos de correlação e na análise das influências entre variáveis.

A entropia mede o grau de organização e, como na teoria da informação, quanto maior a desordem maior a entropia, e maior a quantidade de informação. Para uma função contínua a entropia máxima corresponderia a uma distribuição uniforme. Assim, seja X uma variável discreta randômica com n estados x_1, x_2, \dots, x_n , e com distribuição de probabilidade $P(X)$. A entropia de X é $H(X)$ sendo dada por (PEARL, 1991, p.321)

$$H(X) = - \sum_x P(x) \cdot \log[P(x)] \quad (3.30)$$

Se também Y é uma variável randômica, a informação mútua $I(X,Y)$ é a medida da informação compartilhada entre X e Y . A informação mútua (ou entropia cruzada) é definida por (MADSEN and KJAERULFF, 2008, p.293)

$$I(X,Y) = H(X) - H(X|Y) \quad (3.31)$$

Onde $H(X|Y)$ é a entropia em X dada uma observação em Y , ou entropia condicional.

Voltando à rede da Figura 3.7 (b) a análise da entropia mútua mostra, por exemplo, como mais fortemente correlacionados com a variável $B_Bronquite$ o nó pai $H_Histórico_Fumante$ ($45,10 \times 10^{-3}$) e o nó filho F_Fadiga ($2,45 \times 10^{-3}$).

Analisadas as correlações do nó $L_Câncer_de_Pulmão$, as variáveis que apresentam maior entropia mútua são $C_Exame_RaioX_Tórax$ ($1,55 \times 10^{-3}$) e $H_Histórico_Fumante$ ($1,19 \times 10^{-3}$).

Para finalizar, apenas uma introdução à questão da explosão combinatória nas redes Bayesianas. Quando se olha uma RB como a da Figura 3.3 (b) referente ao problema da *Grana_Molhada* no qual cada uma das variáveis tem dois estados, verifica-se que se todas elas se relacionassem entre si as tabelas de probabilidade conjunta já envolveriam um espaço de trinta e duas possibilidades, ou seja, o número de combinações possíveis é proporcional a 2^n , onde n é o número de variáveis. Num problema com quarenta variáveis no qual cada uma tenha seis estados o espaço de possibilidades já seria $1,34 \times 10^{31}$ e caso as variáveis tenham mais estados ainda acontece a

explosão combinatória visto que em verdade para p estados e n variáveis a dimensão seria p^n . Na prática, felizmente, cada variável é influenciada e influencia umas poucas outras e assim a complexidade espacial do problema é bastante reduzida.

Nas RB usuais cada uma das variáveis apenas se relaciona com aquelas que são seus *pais e filhos*, geralmente outras seis no máximo, e também os estados, na maioria das vezes, limitam-se a quatro ou cinco, o que resulta em probabilidades condicionais envolvendo quando muito a dimensão de 15 mil. Como destaca MATOS (2004) o problema de inferência em RB “*é em geral NP-Difícil (NP-HARD), entretanto, quando o grafo que o compõe é uma poliárvore existem algoritmos de complexidade polinomial*”.

O que acontece na prática nas RB reduzindo a complexidade polinomial é algo semelhante às redes dos sistemas de potência: nestes, cada barra se relaciona apenas com umas poucas adjacentes, e nas RB cada variável se relaciona somente com seus próprios pais e filhos, que como dito, não são muitos.

3.3 – Exemplos de RB

Os exemplos apresentados têm a finalidade de ressaltar a importância de bem definir as variáveis examinando-se a questão da dependência entre elas, o que é essencial no modo de interligar os nós, bem como destacar que os estados das mesmas devem ter uma nomenclatura tal que ressalte com clareza sua exclusividade mútua.

Parte-se das probabilidades conjuntas originárias de amostragem e introduzindo-se propositalmente engano nas relações de dependência, são mostradas as diferenças construtivas das RB e os erros que aparecem nos resultados. Os dados são os mesmos do problema visto no subitem 3.1.3 referente ao sorteio de cartas de ouros e paus, amostra com 301 sorteios.

Como o verso das cartas do naipe *ouros* é verde, e o das cartas de *paus* é vermelho, há a garantia de ao se retirar uma de verso verde e outra vermelha se está sorteando uma do naipe *ouros* e outra do *paus*. Como após cada sorteio há sempre a reposição das cartas, os eventos são variáveis independentes. Com dúvida ou havendo engano a rede poderia ter sido modelada considerando similaridade com a do problema anterior, isto é, a variável *evento_DOIS_paus* sendo dependente da primeira *evento_UM_ouros*.

Com os dados da amostra, as duas RB considerando estes eventos como dependentes e independentes são vistas respectivamente nas Figuras 3.9 (a) e 3.9 (b).

Cada uma das duas variáveis de entrada tem três estados e assim o nó que representa o acompanhamento da pesquisa amostral é denominado *Pesquisa_Dois_Sorteios* e tem nove estados. Cria-se outra variável com os quatro estados desejados, que é o nó *Agrupamento_Final*.

Para este nó pensou-se nos estados *Dois_Mesmo_Grupo*, *Sem_Figura*, *Sem_Par* e *Sem_Impar*, entretanto viu-se que não haveria tanta clareza em ressaltar a exclusividade destes estados. Isto porque os estados *Par_Par* e *Impar_Impar* do nó pai tanto poderiam ser agrupados no estado *Dois_Mesmo_Grupo* como em *Sem_Figura* da variável *Agrupamento_Final*. Então, do modo definido, por exemplo, o estado *Grupos_Distintos_Sem_Figura* do nó *Agrupamento_Final*, nele só podem ser agrupados os estados *Par_Impar* e *Impar_Par* da variável pai.

Retomando a questão da independência, comparando-se a Figuras 3.9 (a) e (b) cujas RB têm os nós de entrada dependentes e independentes, respectivamente, observa-se que caso tenha havido engano na relação de dependência entre variáveis, os resultados se modificam não apenas na variável *Pesquisa_Dois_Sorteios* como em *Agrupamento_Final*. Verifica-se erro a menor principalmente no estado *Dois_Mesmo_Grupo*, no qual a crença cai de 34,9% para 34,2%.

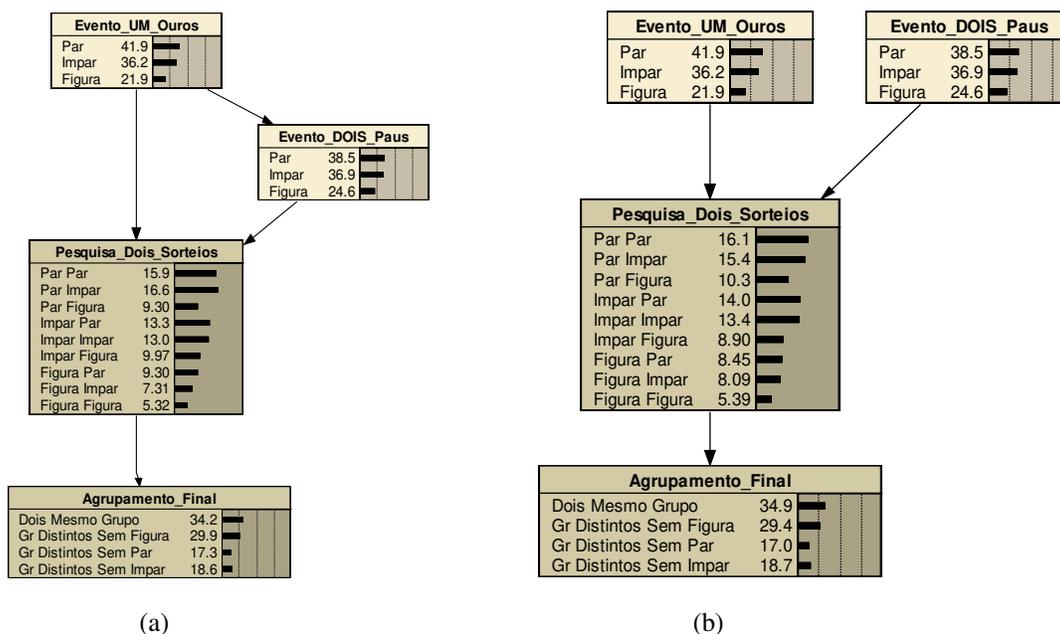


Figura 3.9 – Modelos de RB para o sorteio de cartas, (a) engano – variáveis dependentes, e (b) RB correta.

Individualmente, dentre os estados da variável *Pesquisa_Dois_Sorteios*, o maior desvio absoluto está no estado *Par_Ímpar* (16,6% contra 15,4%), e percentualmente no estado *Figura_Par*.

Para finalizar este exemplo são comparadas as duas redes usando as estimativas de probabilidade teóricas para os estados das cartas, e não os resultados da amostra cuja tendência era sair mais carta *par* de ambos os naipes. Os resultados são iguais para os nós de saída tanto no modelo de RB correto, Figura 3.10 (b), como no que considera os nós dependentes, Figura 3.10 (a).

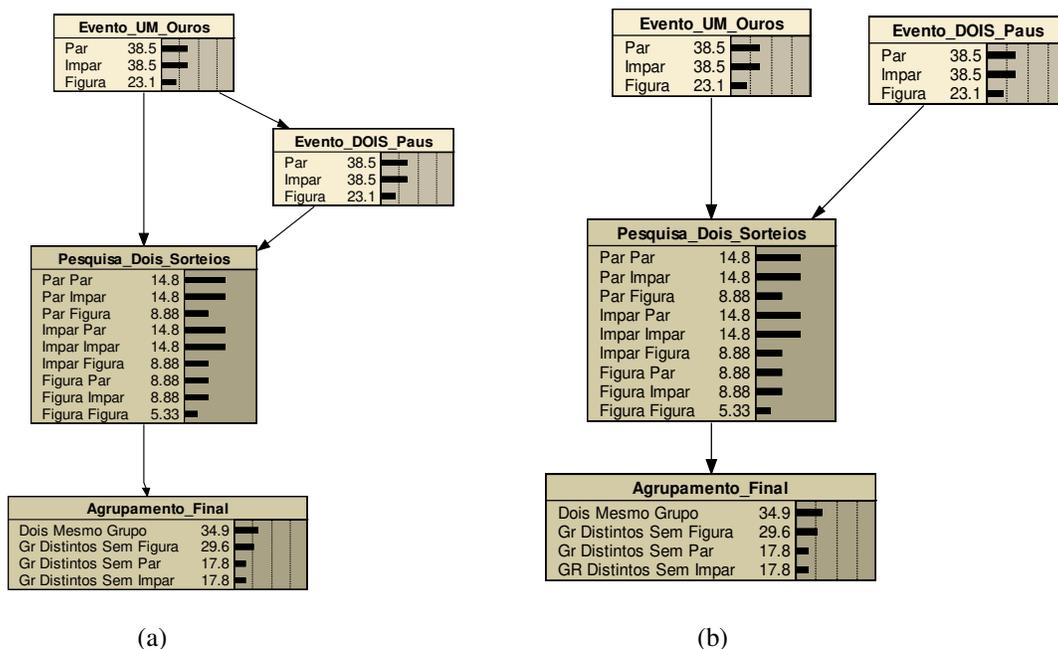


Figura 3.10 – RB para o sorteio de cartas com dados teóricos, (a) modelo com engano, e (b) RB correta.

Quando este exemplo foi proposto, no item 3.1.3 se perguntou qual a probabilidade da carta de *ouros* sorteada ter sido par tal que a de *paus* tenha sido ímpar. Nas RB busca-se a resposta instanciando-se o estado ímpar do nó *Evento_DOIS_Paus*, e a rede da Figura 3.11 (a) cujos nós de entrada são dependentes mostra uma crença de 45,0% para o estado *par naipe de ouros*, enquanto a RB cujos nós são independentes mostra o valor de 41,9%, Figura 3.11 (b).

Como neste problema são conhecidas as probabilidades teóricas, para estas duas redes são comparados ainda os resultados da instanciação do estado *Dois_Mesmo_Grupo* da variável *Agrupamento_Final*, o que está mostrado na Figura 3.12. Vê-se que teoricamente seria esperada a mesma probabilidade para saírem duplas de mesmo grupo sejam par ou ímpar, Figura 3.12 (a), mas novamente a tendência é que saiam mais duplas pares como visto na Figura 3.12 (b).

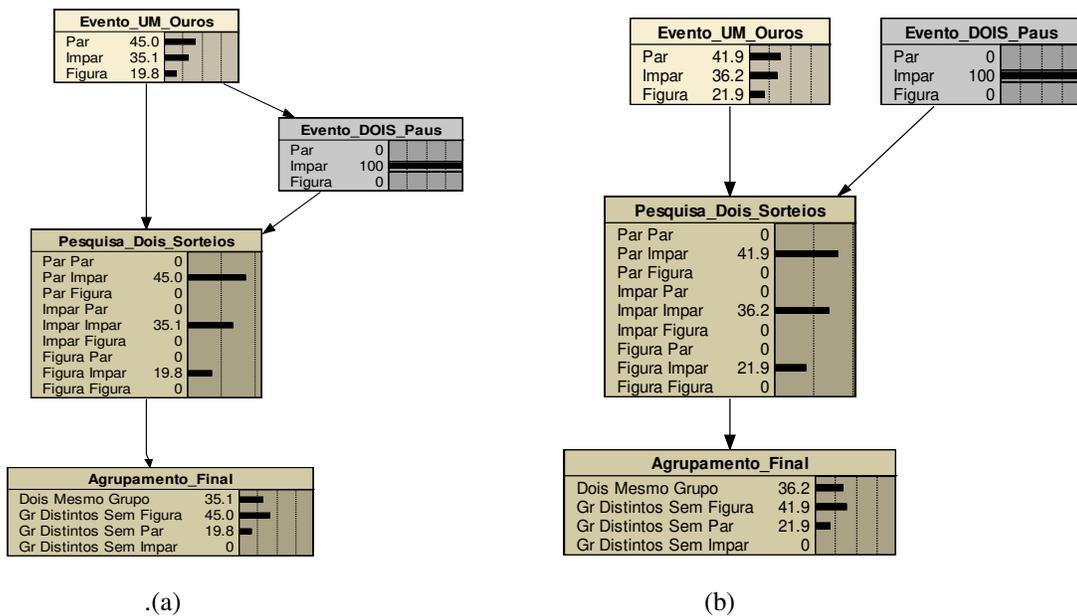


Figura 3.11 - Sorteio de cartas, instanciação. - (a) RB com variáveis dependentes, e (b) RB correta.

Observa-se na Figura 3.12 (b) que não houve diferenciação entre carta par de ouros ou paus, isto porque o algoritmo de instanciação desenvolve-se de baixo para cima e a probabilidade *par_par* em *Pesquisa_Dois_Sorteio* tem uma crença de 46,3%.

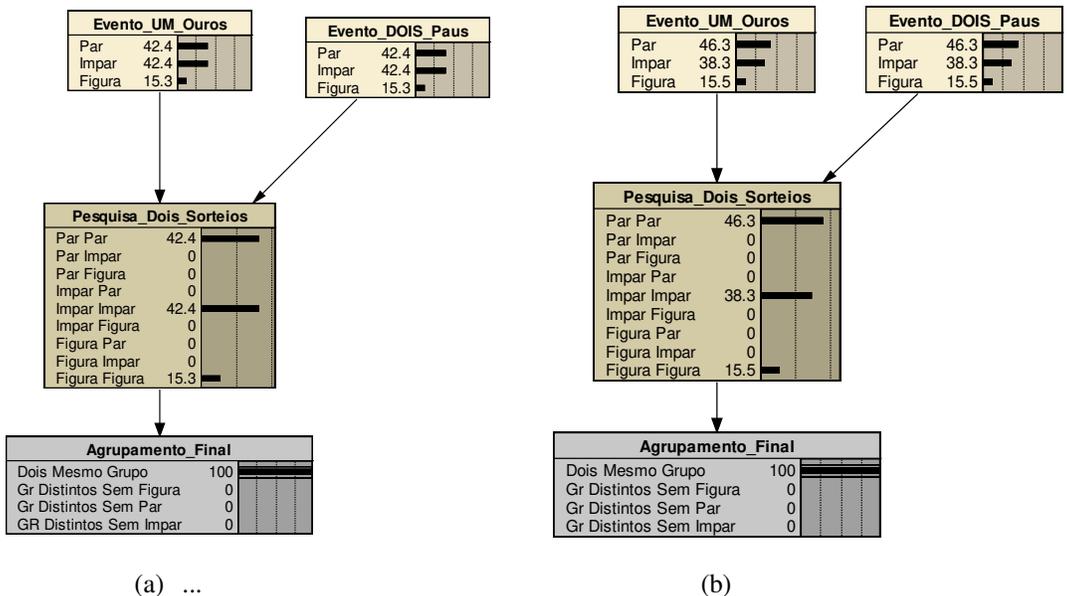


Figura 3.12 - Sorteio de cartas, instanciado *Dois_Mesmo_Grupos*- (a) dados teóricos, e (b) dados pesquisa.

Outros exemplos de redes Bayesianas são encontrados na literatura existente sobre o tema, e as RB aplicadas a perdas são vistas nos capítulos seguintes.

4. DIAGNÓSTICO DAS PERDAS COMERCIAIS

Neste capítulo inicialmente são apresentadas justificativas para o diagnóstico das perdas comerciais (PC) e é proposto um índice que permite agrupar e comparar empresas distribuidoras. O diagnóstico é um trabalho em nível macro de prospecção das PC objetivando identificar suas principais causas e regiões de incidência dentro da área de concessão da distribuidora. Em seguida discorre-se sobre a metodologia e a estrutura do modelo de RB sugerida, e finalmente faz-se a aplicação aos clientes atendidos em baixa tensão (grupo B) de uma concessionária.

4.1- Justificativa para o diagnóstico e proposta de um índice comparativo

O diagnóstico das perdas comerciais visa quantificar as suas parcelas de acordo com as diversas causas dentre os consumidores, sejam eles clientes ou não da concessionária, identificando as contribuições conforme as regiões. A justificativa para tal objetivo deve-se aos seguintes fatos:

- As perdas de um modo global têm crescido nos últimos anos, tanto mundialmente (SMITH, 2004), quanto no Brasil (MME, 2009), (ALVES, 2003), (SIQUEIRA e ARAÚJO, 2006);
- Várias empresas dos países em desenvolvimento apresentam elevados níveis de perdas de energia (SMITH, 2004), (NADIRA e DORTOLINA, 2005);
- As perdas comerciais por vezes são até maiores que as perdas técnicas (ARAÚJO, 2007), entretanto vinham sendo menos pesquisadas. Isto porque muitos as consideravam apenas um problema gerencial, porém nos últimos dez anos há maior número de trabalhos sobre este tema.

Nos países em desenvolvimento não apenas as perdas comerciais das empresas de eletricidade são altas, mas também em empresas de serviços de um modo geral como água, esgoto e transporte. No Brasil, em particular no setor elétrico, tais perdas médias estão em 8,75% enquanto as perdas técnicas médias no Brasil são da ordem de 8,11% (ARANHA NETO et al, 2008). Com isto se conclui que o estudo das perdas comerciais é da maior importância,

justificando ultimamente a existência de vários estudos e pesquisas dedicados a perdas não técnicas como VIEIRALVES (2005), DANTAS (2006), ALMEIDA (2006), COMETTI et al (2007), FRANCISCO E FAGUNDES (2006), ARAUJO (2007), ARANHA NETO et al (2008) e FERREIRA (2008).

Conforme ressalta FERREIRA (2008) e a NT-348/SRE/ANEEL (2007) reconhece, as perdas comerciais dependem de dois fatores: eficiência gerencial da empresa e a vulnerabilidade ou propensão da área às perdas. Há fatores sócio-econômicos e culturais que induzem as pessoas ao furto e ao não pagamento das faturas, e para reverter tal situação as empresas devem ser eficientes, e o Estado mostrar-se mais responsável e presente. Por outro lado toda a sociedade deverá ser mais cooperativa e participativa, colaborando com empresas e órgãos governamentais na fiscalização de abusos.

Como exemplo destes fatores podem-se citar: altas tarifas da energia, baixa renda média da população, custo elevado do padrão de ligação, condições precárias de infra-estrutura urbana (muitas sem abastecimento de água, sem rede de esgoto e localizadas em ruas sem calçamento), existência de residências não legalizadas ou invasões.

Interessante é observar que nesta questão de perdas comerciais o órgão regulador e as empresas concessionárias têm visão um pouco distintas. No Brasil, até o momento, a agência reguladora não especificou metas a serem alcançadas relativas a perdas técnicas nem comerciais, embora por ocasião do segundo ciclo de revisão tarifária tenha emitido a NT-348 divulgando a metodologia e os critérios a adotar no cálculo do nível máximo de perdas não técnicas a admitir por concessionária, o que se constitui um avanço significativo. Um ponto importante colocado nesta nota técnica é quais fatores considerar nas diferentes áreas de concessão.

Na referida NT há a descrição de um estudo cujo objetivo era “*elaborar um índice que permite diferenciar as áreas de concessão quanto a determinadas características sócio-econômicas, observando os dados existentes em institutos oficiais como IBGE, STF e Ministério da Saúde*” (ANEEL, NT-348, 2007, p.11). Foram pesquisados diversos indicadores como número de óbitos por agressão, concessionárias com maiores níveis de pobreza, de analfabetismo e cobertura de abastecimento de água. Escolhido o modelo, foram consideradas as variáveis pobreza, violência, educação, cobertura de abastecimento de água e região metropolitana, entretanto a tarifa não foi considerada. Nos resultados as empresas foram ordenadas em ordem decrescente tendo maiores

índices CEAL, CELPA, CERON, CELPE, ELETROACRE, CEMAR; dentre aquelas com menores índices estão CPFL-PAULISTA, VALE DO PARANAPANEMA, CAIUÁ, CNEE-NACIONAL e CJE-JAGUARI. Relevante nesta NT é o regulador admitir que a empresa tem grande capacidade de gerenciar as perdas e que ele, o regulador, pode determinar um nível máximo ou uma trajetória para as perdas, além de ressaltar a importância de um diagnóstico das perdas e de um plano de combate às perdas cujas ações possam ser avaliadas economicamente. Objetivando introduzir algumas sugestões de ações úteis no combate às PC na Tabela 4.1 são relacionadas as origens das perdas, onde as mesmas estariam acontecendo (se no grupo dos clientes ou não) e algumas possíveis soluções ou medidas mitigadoras.

Tabela 4.1 – Perdas comerciais na distribuição: localização, origem e possíveis soluções.

Grupo	Origem	Possíveis soluções
Clientes da empresa	Interna: erro de leitura, defeito ou obsolescência do medidor, erro na ligação, especificação inadequada da medição, engano no banco de dados.	Acompanhamento do processo de ligação e faturamento; atualização do banco de dados; treinamento dos empregados que especificam o sistema de medição e dos ligadores; inspeções de medidores (lote e individual), outras.
	Externa: fraudes como desvio embutido, desvio antes da medição, circuito de potencial interrompido, medidor avariado, ponte nos terminais, ponte no transformador de corrente, etc.	Inspeções; acompanhamento do histórico de consumo; instalação de medição totalizadora; desenvolvimento de ferramentas de IA que ajudem a identificar desvios de consumo mesmo que sazonais, melhorando a taxa de acerto das inspeções; lacres no sistema de medição; uso de rede antifurto e outras.
Não-clientes	Externa: ex-clientes desligados que se auto religam	Adequação tarifária; subsídios intra-setorial; facilidades de negociação de débitos; inspeção das unidades desligadas, lacres no sistema de medição e outras.
	Externa: conexões ilegais, como o vendedor na praça aos domingos ou as extensões de redes em invasões e prédios	Mapeamento das zonas; proposições de políticas sociais conjuntas com prefeituras, polícia e justiça; financiamento do padrão de ligação; rede antifurto, dentre outras.

Ressalta-se que os problemas com clientes cujas causas sejam de origem interna são de responsabilidade da distribuidora; legalmente só pode ser efetuada a cobrança da energia não faturada no mês anterior à identificação, caso contrário tais perdas são não recuperáveis. A empresa deve ter políticas específicas para eliminar tais problemas, com metas e prazos definidos.

As causas de origem externa, como fraude, podem ter cobrança retroativa da energia, entretanto enquadrar o fraudador conforme o código penal é difícil, exige perícias externas, enfim a cooperação de delegacias e da Justiça em dar andamento a tais processos. Isto é bem abordado em ARANHA NETO et al (2008).

As conexões ilegais em grande escala são impulsionadas pelos fatores mencionados anteriormente. O exame destes fatores em cada área de concessão e o zelo para que zonas específicas não se tornem propensas ou vulneráveis ao aparecimento de irregularidades, desde ligações clandestinas à inadimplência, devem ser de competência da empresa.

Do ponto de vista das empresas, estas devem se empenhar em diagnosticar efetivamente o problema das perdas não técnicas em sua área de concessão, sendo importante uma prospecção em nível macro. Elas dispõem do cadastro dos seus clientes, conhecem os débitos e características de pagamentos dos mesmos e inclusive dos ex-clientes. A empresa deve conhecer bem a dimensão do problema e sua distribuição espacial, logo deve ter melhores condições para dimensionar os recursos de combate às perdas comerciais, podendo ser mais eficiente.

Conhecida a fundo toda a situação das perdas comerciais deve a concessionária distinguir aquilo que ela própria pode resolver daquilo que exigirá a colaboração do Governo, da Polícia e Justiça, além da cooperação da sociedade.

Dependendo da profundidade do problema há estratégias e políticas que levam a ações diferenciadas: podem ser estabelecidas políticas conjuntas com delegacias e Justiça visando coibir as LC, retirar as existentes e punir os culpados, bem como podem ser propostas medidas conjuntas com os Governos Municipal e Estadual objetivando legalizar e urbanizar áreas invadidas ou degradadas. Concluindo, feito o diagnóstico deve a empresa elaborar seu plano de redução das PC, sem esquecer que sendo composta por pessoas estas devem estar empenhadas nos resultados.

Entretanto, na prática, o que tem acontecido em algumas empresas se distancia um pouco do que foi descrito visto que por vezes a maior preocupação nestas distribuidoras é buscar perdas apenas dentre os clientes, tendo como ação preponderante as inspeções. Infelizmente isto em nem sempre contribuiu significativamente para redução das PC, e ao menor custo.

Neste trabalho propõe-se um indicador que possibilite agrupar e comparar as empresas concessionárias, o que permitiria ao regulador exigir estudos e análises diferenciados por grupos de empresas. Tal indicador é denominado de “*índice de situação de perdas*” (I_{stp}), e pode ser definido como a relação percentual entre perda comercial e técnica, ou

$$I_{stp} = 100. \left(\frac{PC}{PT} \right) \quad (4.1)$$

Considerando a equação (2.2), este índice é

$$I_{stp} = 100. \left[\frac{PG}{PT} - 1 \right] \quad (4.2)$$

A perda global é facilmente acompanhada pelo órgão regulador e uma vez que este disponha de metodologia unificada e auditável para o cálculo das perdas técnicas, torna-se fácil estabelecer e acompanhar este *índice de situação de perdas*. Deve-se cuidar para que não haja super-avaliação das PT, pois isto distorceria tal índice.

Com o agrupamento as distribuidoras seriam caracterizadas com respeito ao controle sobre o processo de perdas, permitindo propor estratégias distintas às empresas. Como sugestão quanto a níveis pode-se adotar um valor inicial de referência de 20% e um segundo de 70%. Deste modo as concessionárias seriam classificadas e os estudos exigidos conforme se segue:

- *Concessionária com controle do processo* – toda distribuidora cujo I_{stp} seja inferior à referência inicial de 20%. Como esta empresa detém o controle do problema das PC poderia o órgão regulador ser flexível, concedendo autonomia para a mesma continuar definindo as suas políticas e ações de redução, sendo mantida a supervisão;
- *Concessionária com altas perdas comerciais* – aquelas para as quais o I_{stp} situa-se entre 20% e 70%, caracterizando que a empresa pode estar perdendo o controle do processo de perdas sendo, portanto, ineficiente em tratar as PC. Provavelmente há zonas nas quais a fraude e os roubos de energia acontecem com maior incidência. Para estas empresas o

regulador deveria exigir um diagnóstico quanto as PC e um plano de redução, acompanhando-os regularmente;

- *Concessionária com PC bastante alta* – empresas nas quais as PC já superam a 70% das PT. Nesta situação as PC já fogem totalmente ao controle da empresa e provavelmente há zonas nas quais o roubo e a fraude de energia estão consolidados. Além da ineficiência da distribuidora possivelmente existe descaso do poder constituído (judiciário que não pune e executivo omissivo), contribuindo para o delito ser incorporado na cultura. Para tais empresas o regulador exigiria o diagnóstico das perdas, o plano de combate incluindo provavelmente solicitação de apoio da Polícia e do Estado, e a tudo acompanharia.

Com base nos dados de PC e PT das distribuidoras por ocasião da primeira revisão tarifária contidos em ARAÚJO (2007, pp.15-16), fez-se uma aplicação deste indicador. Os resultados relativos a 59 distribuidoras estão na Figura 4.1, dividindo-se as áreas do gráfico de acordo com o agrupamento sugerido. Os dados originais correspondiam a 61 concessionárias, mas como em três delas as perdas técnicas eram exatamente iguais, considerou-se apenas uma destas.

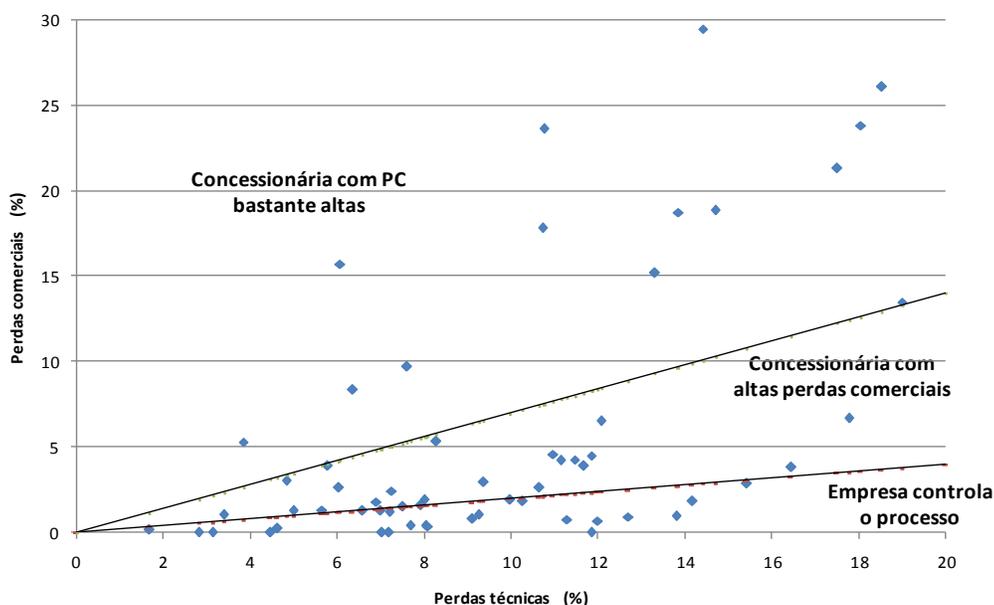


Figura 4.1 – Perda comercial versus perda técnica em distribuidoras brasileiras (fonte ARAÚJO, 2007).

Conforme estes dados 40,7% das concessionárias de distribuição do país estariam agrupadas como eficientes e para 59,3% delas o órgão regulador deveria exigir a elaboração do diagnóstico

das perdas comerciais. A Tabela 4.2 apresenta o resumo desta análise vendo-se que em 22,0% das empresas as PC são maiores que 70% das PT.

Tal indicador supera o limite sugerido de 70% em treze empresas dentre as quais estão LIGHT, MANAUS, CERON, BOA VISTA, CEPISA, CELB, ELETROACRE, ELETROPAULO, CEAL e CEMAR. Interessante observar que neste grupo há empresas cujas perdas comerciais superam 20% (MANAUS, CEPISA e CEMAR), enquanto outras têm PC menores que 9% (ELETROPAULO e CELB). Muitas destas também aparecem com elevado índice de perdas nos estudos da Nota Técnica 348 (ANEEL, 2007, pp.22-23). CNEE-NACIONAL, CAIUÁ e VALE DO PARANAPANEMA estão entre aquelas cujo indicador é inferior a 20%.

Tabela 4.2 – Análise das distribuidoras brasileiras quanto ao índice de situação de perdas.

Índice de situação de perdas	Quantidade de distribuidoras	% do total de empresas
Istp \leq 20%	24	40,7
20% < Istp \leq 70%	22	37,3
70% < Istp	13	22,0

SMITH (2004) estudou empresas de energia elétrica de todo o mundo, examinando a situação em 102 países no período entre 1980 e 2000 para demonstrar que as perdas globais têm aumentado mundialmente. Analisou as perdas na transmissão e distribuição classificando-as por faixa, constatando que em 33,3% dos países as perdas eram menores que 10% entretanto em 42,2% as mesmas superavam 16%. Dentre os países com elevado índice de perdas SMITH (2004) cita Bangladesh, Malásia, Paquistão, México, Tailândia e África do Sul. Outros pesquisadores ressaltam altas perdas comerciais de energia em países como Chile, Argentina, Peru, Turquia, Panamá, Brasil, Uganda e Venezuela (NADIRA e DORTOLINA, 2005), além da Índia (RAO E DEEKSHIT, 2006), Gana (EDF, Vol.15, 1998) e Filipinas (ORILLAZA et al, 2006).

Deste modo fica claro a importância das PC e que o nível elevado das mesmas é característico de vários países em desenvolvimento, entretanto no início da eletrificação em muitas destas nações algumas comunidades não pagavam pela energia elétrica a qual por outro lado era fornecida apenas em curtos períodos do dia.

Há ainda fatores sócio-econômicos e culturais que tornam os consumidores propensos ao furto e ao não pagamento das faturas de energia.

Em resumo, optou-se por propor um índice que relacione a situação das perdas comerciais frente às técnicas com o objetivo de agrupar empresas e permitir ao órgão regulador um fácil acompanhamento, por ser de fácil apuração, orientando quanto a possíveis exigências a serem feitas às distribuidoras, a exemplo da necessidade de estudos de diagnóstico de perdas, definição de estratégias, metas e programa de ações para redução das mesmas.

4.2 – Metodologia e modelo de rede para o diagnóstico das perdas comerciais

Estudar, analisar o problema das perdas comerciais em uma empresa de distribuição, e propor medidas de controle e redução destas perdas envolve etapas e a primeira deve ser a prospecção ou o diagnóstico destas perdas não técnicas.

O modelo proposto incorpora conceitos dos estudos de qualidade e emprega as redes Bayesianas nas quais os valores estimados têm o sentido de crença.

Há uma analogia com as aplicações de RB na área de saúde: nestas para se conseguir os dados das TPC necessárias à rede são ouvidos os pacientes, tratados os dados relativos aos exames, enquanto nas áreas de concessão das distribuidoras devem ser pesquisadas as perdas nas unidades consumidoras (UC) e as ligações ilegais feitas pelos não clientes. Isto pode ser feito através da inspeção das UC em campo, investigação quanto a áreas onde existam ligações clandestinas ou ex-clientes auto-religados, auditoria no processo de leitura e faturamento, verificação do banco de dados dos clientes especialmente o cadastro das informações para faturamento, e outras ações relevantes.

Já foi colocado que as perdas não técnicas são causadas por:

- Fraudes nas UC de clientes, particularmente nos grupos A e B;
- Unidades consumidoras ligadas que estão com engano no cadastro ou banco de dados do sistema de faturamento, bem como aquelas cujos equipamentos de medição apresentam defeito ou funcionamento incorreto;
- Incorreções no faturamento da iluminação pública, a exemplo de defeitos nos equipamentos e de situação cadastral irregular (expansão da rede pela Prefeitura sem que a concessionária tenha tal informação);

- Ligações clandestinas, realizadas por pessoas que não são clientes da empresa, e em geral através extensão ilegal de rede;
- Unidades desligadas que se auto-religaram (cortados que se ligaram à revelia).

As três primeiras causas acima descritas dizem respeito aos clientes da distribuidora, isto é, UC ativas no banco de dados de faturamento. As últimas são não clientes, ou seja, são consumidores irregulares: as ligações clandestinas (desconhecidas da empresa) e as unidades auto-religadas, as quais em algum momento já foram cadastradas na companhia.

Diagnosticar e quantificar as perdas comerciais relacionando-as às origens e à distribuição espacial tanto para clientes quanto para não clientes da empresa é um modo racional da distribuidora planejar ações mais eficientes para redução e minimização das perdas.

O faturamento dos clientes ativos pode ter engano causado por fraude ou motivos relacionados à concessionária. Como os clientes estão em três grupos (A, B e iluminação pública – IP), e a empresa conhece os dados dos mesmos, a idéia é identificar não conformidades através inspeção em amostras randômicas nas unidades destes grupos.

Relativamente aos não clientes, para as unidades desligadas se dispõe da localização podendo-se dirigir também uma investigação randômica para exame da normalidade, enquanto que para as ligações clandestinas um sorteio de zonas ou quadrículas a investigar torna-se o procedimento adequado. Ressalta-se que o uso de informações vindas de denúncias ou identificação de locais com LC, ainda que de entrevistas com leituristas e técnicos de campo, seria um procedimento não correto caracterizando estatisticamente a pesquisa como viciada. Poderiam, sim, ser usadas para checar se a investigação randômica em campo identificou estes problemas e/ou áreas.

Para definir o modelo de RB é necessário bem definir as variáveis, seus estados e as relações entre as variáveis. Foram consultados alguns especialistas em perdas, funcionários e ex-funcionários de distribuidoras, tentando-se chegar a relações causais entre as variáveis, com base em classificações e agrupamentos usuais das empresas. Optou-se pelo exame da conformidade segundo os grupos de clientes e não clientes porque são independentes.

O fluxograma da metodologia proposta está mostrado na Figura 4.2. Inicia-se com os dados de mercado e o cálculo das perdas. Segue-se com a definição de uma hipótese inicial – H_0 (estimativa a priori da não-conformidade) que significa estimar índices de não conformidade (p) por grupo, tendo como condição de contorno o total das perdas comerciais. Trabalha-se com

cinco grupos de consumidores: três grupos de clientes (grupo A, grupo B e Iluminação Pública - IP), e dois grupos de não clientes (as unidades Desligadas e as ligações clandestinas - LC).

O passo seguinte é o dimensionamento de amostras aleatórias por grupo, para inspeção da conformidade sob a ótica de perda. Para os grupos A, B e Desligados a empresa dispõe do banco de dados do sistema de faturamento logo o sorteio é simples.

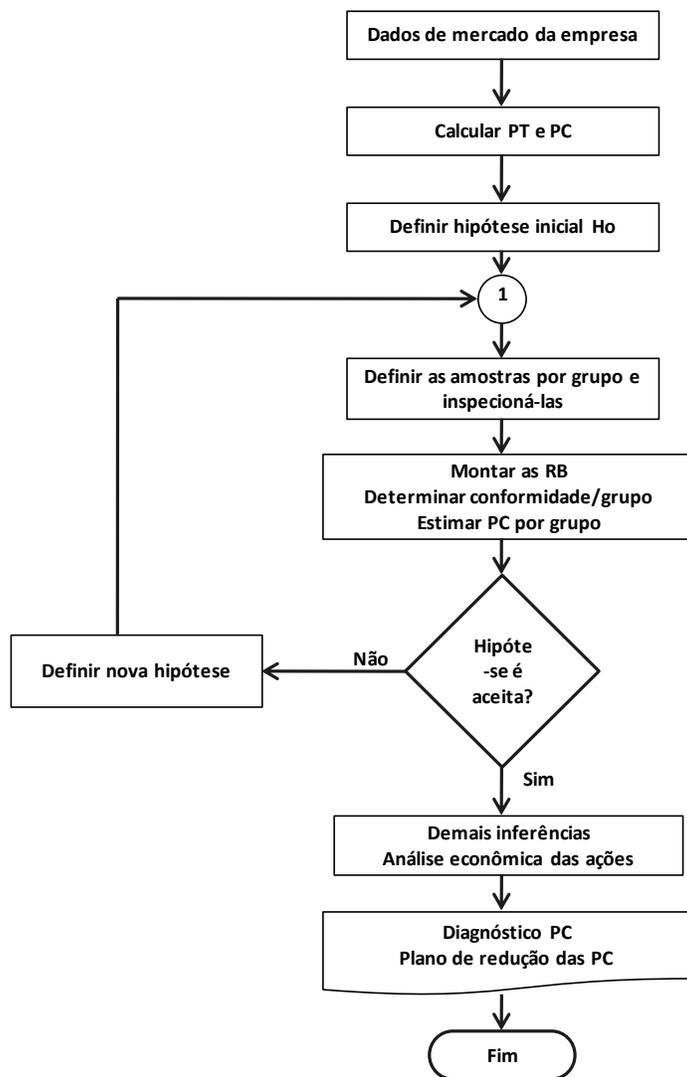


Figura 4.2 – Fluxograma para determinação das parcelas de perdas comerciais por grupo de origem e causa.

Para o grupo IP a amostra é uma quantidade do total de pontos de iluminação sorteando-se quadrículas ou bairros das cidades. No grupo de Ligações Clandestinas (LC) faz-se correlação com as unidades dos clientes residenciais em baixa tensão e são investigadas também quadrículas por região.

Os resultados das inspeções são as relações de probabilidade condicional da RB, o que permite obter os índices de conformidade dos vários grupos. Faz-se o teste de validação, caso se aceite H_0 então são quantificadas as perdas comerciais segundo regiões e causas.

Assumindo consumos médios por grupos ou classes são estimadas as parcelas de PC de energia relativas a cada grupo, sendo associadas às causas. Como consumo médio pode-se trabalhar com a própria média do respectivo grupo ou com a média daqueles encontrados na situação de não conformidade. Preferiu-se não utilizar qualquer acréscimo aos valores médios, seja com justificativa de aumento de perda técnica ou por consumo em excesso, porque não se possui respaldo através de pesquisa.

São feitas algumas inferências objetivando orientar-se quanto às estratégias e ações. Segundo a análise econômica define-se um eficiente plano de redução das PC, e se encerra o diagnóstico.

Não sendo aceita H_0 , faz-se nova hipótese e retorna-se ao passo 1 para dimensionamento das amostras para inspeção.

4.2.1 – Redes e estados das variáveis

Considerando que para os grupos de clientes e para os ex-clientes desligados as quantidades por classe de consumo são conhecidas, e que praticamente a totalidade das empresas usa divisão territorial, foram escolhidas as variáveis *Grupo* cujos estados são as classes de consumo, e *Região*, na qual os estados são as regiões existentes na distribuidora.

Nas sub-redes de iluminação pública e ligação clandestina não há classes, a variável de entrada é territorial.

Conforme o modelo de RB adotado, mostrado na Figura 4.3, os nós de entrada de cada uma das sub-redes dos grupos *A*, *B* e *Desligados* são as classes de consumo, cujos quantitativos a distribuidora conhece na sua base de dados, por exemplo, *Residencial*, *Comercial*, *Industrial*, *Serviços Públicos*, *Poderes Públicos*, *Rural* e *Outros* (inclui consumo próprio).

A não conformidade pesquisada é binomial (há ou não há perdas?). Os estados da variável *Inspeção* também são mutuamente exclusivos.

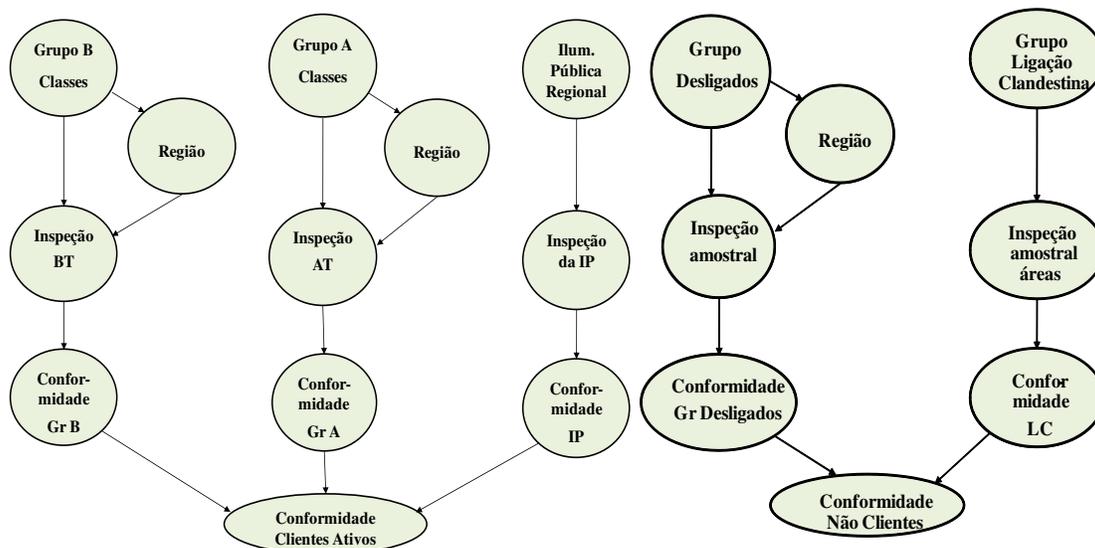


Figura 4.3 – Rede Bayesiana adotada para os clientes ativos e os não clientes.

Para os grupos A e B se procura distinguir os estados de não conformidade, ou seja, aqueles onde acontecem perdas de acordo com as causas externas e internas às empresas:

- “Perda: origem externa, fraude” – corresponde àquela unidade encontrada com desvio embutido, circuito de potencial interrompido, medidor avariado, e outras iniciativas do cliente com intenção de fraudar a empresa, portanto levando a perda comercial;
- “Perda: origem interna” – a causa não é de responsabilidade do cliente, mas diz respeito aos problemas internos da empresa. Pode-se separar os defeitos ou problemas nos equipamentos de medição, dos erros na base de dados quais sejam os enganos de relação, de constante ou leitura, e outros.

Os estados das inspeções nos grupos A e B que correspondem à situação de conformidade são *Normal* que corresponde à UC encontrada absolutamente normal, e *Irregular sem perda* que significa aquela unidade sem selo, com a caixa avariada ou com outra anormalidade, mas que não implica em haver perdas de energia.

No diagrama da Figura 4.4 procura-se identificar as perdas comerciais nos Grupos e Classes segundo os tipos de irregularidades ou não conformidades encontradas no processo de inspeção das unidades, e assim melhor escolher os estados da variável *Inspeção* para os grupos A e B.

Para o grupo “*UC Desligadas*” a conformidade é examinada ainda por classe, mas são outros os estados da variável *Inspeção* e, por conseguinte para *Conformidade*, por exemplo, *unidade*

demolida, desabitada, já regularizada e realmente desligada que são situações conformes, e a caracterização da não conformidade que é a unidade *auto-religada* ou ligada à revelia.

Nos demais grupos quais sejam *Iluminação Pública* e *Ligações Clandestinas*, não cabe a segmentação por classe de consumo então a sub-rede é mais simples, ver Figura 4.3. Nestes dois grupos os nós de entrada assumem uma regionalização, por número de pontos de iluminação no primeiro grupo, e por correlação com as unidades residenciais em baixa tensão no grupo de LC. As amostras podem ser por quadrículas ou por zonas dentro dos municípios.

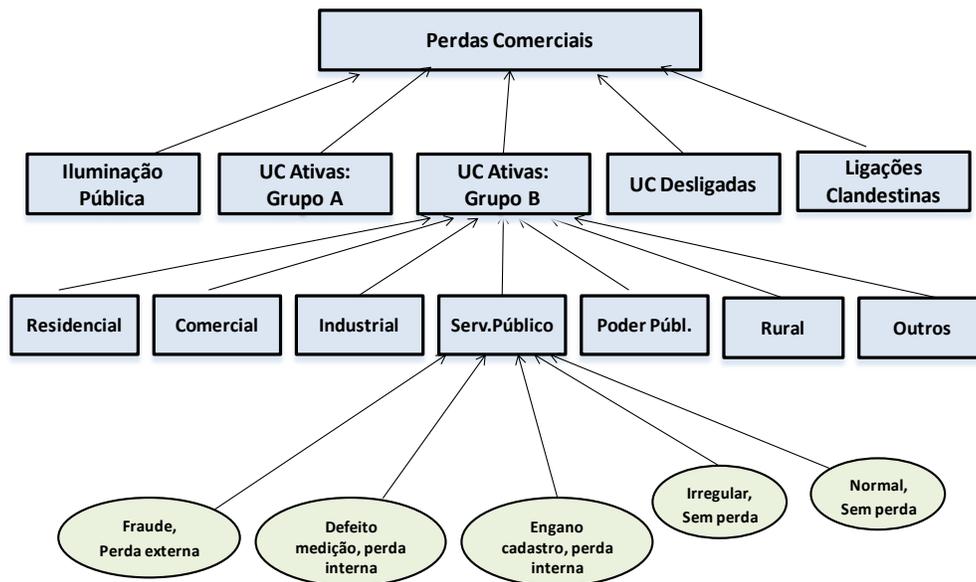


Figura 4.4 – Identificação das PC, escolha dos estados do nó *Inspeção* para os grupos A e B.

Em termos de não conformidade no grupo IP a inspeção deve investigar condições próprias da operação irregular das lâmpadas, que são as lâmpadas acesas durante o período diurno (perda positiva) e aquelas apagadas à noite (perda negativa) para ser feito um balanço, bem como se verificar a atualização do cadastro (banco de dados) de IP.

Para o grupo das LC o fato de encontrar as próprias ligações ilegais caracterizaria a não conformidade. Para estabelecer a amostra considerando a hipótese inicial quanto à perda comercial fez-se a correlação com as unidades residenciais do grupo B, pois a maioria das LC têm uso apenas residencial e umas poucas apresentam características comerciais.

As probabilidades dos nós de entrada são independentes e para os grupos A e B, por exemplo, são provenientes dos dados de mercado existentes no banco de dados. Logo, a probabilidade de ser comercial e do grupo A é

$$P(\text{comercial, grupoA}) = \frac{\text{número unidades comerciais do grupo A}}{\text{total das unidades do grupo A}} \quad (4.3)$$

Para o nó *Região* as probabilidades são condicionais e sendo esta variável filha do nó *Grupo_A* a TPC é determinada com os dados de mercado da empresa. Exemplificando, a probabilidade de uma unidade estar na *Região BB* desde que seja comercial é

$$P(\text{regiãoBB|comercial}) = \frac{\text{número unidades comerciais Gr.A da região BB}}{\text{total das unidades comerciais do Grupo A}} \quad (4.4)$$

As relações de probabilidade condicional do nó *Inspeção* são obtidas dos dados das inspeções naquelas unidades sorteadas para amostragem, o que caracteriza a obtenção das probabilidades *a posteriori*. O elemento, por exemplo, que indica a probabilidade de que há perda em unidades comerciais do grupo A é determinado aplicando-se aos dados da amostra:

$$P(\text{há perda|comercial}) = \frac{P(\text{comercial |há perda}) \cdot P(\text{há perda})}{P(\text{comercial})} \quad (4.5)$$

Pode-se também encontrar esta crença de que há perda recorrendo-se à expressão da probabilidade condicional tal como em (3.8).

Ao aplicar a metodologia em uma empresa os estados da variável *Grupo*, devem ser definidos em comum acordo, aqui sugeridos as sete classes dos grupos A, B e Desligados. Também há flexibilidade para os estados do nó *Inspeção*, e estes devem ser adequadamente classificados conforme os estados do nó *Conformidade*. A Figura 4.5 (a) mostra a rede sugerida para os clientes do grupo A, a 4.5 (b) a RB do grupo iluminação pública, e em 4.5 (c) aquela do grupo dos desligados.

Após a pesquisa são determinadas as quantidades de unidades não conformes dentro das diversas classes e regiões, o que é feito aplicando-se a regra de cadeia, expressão (3.27).

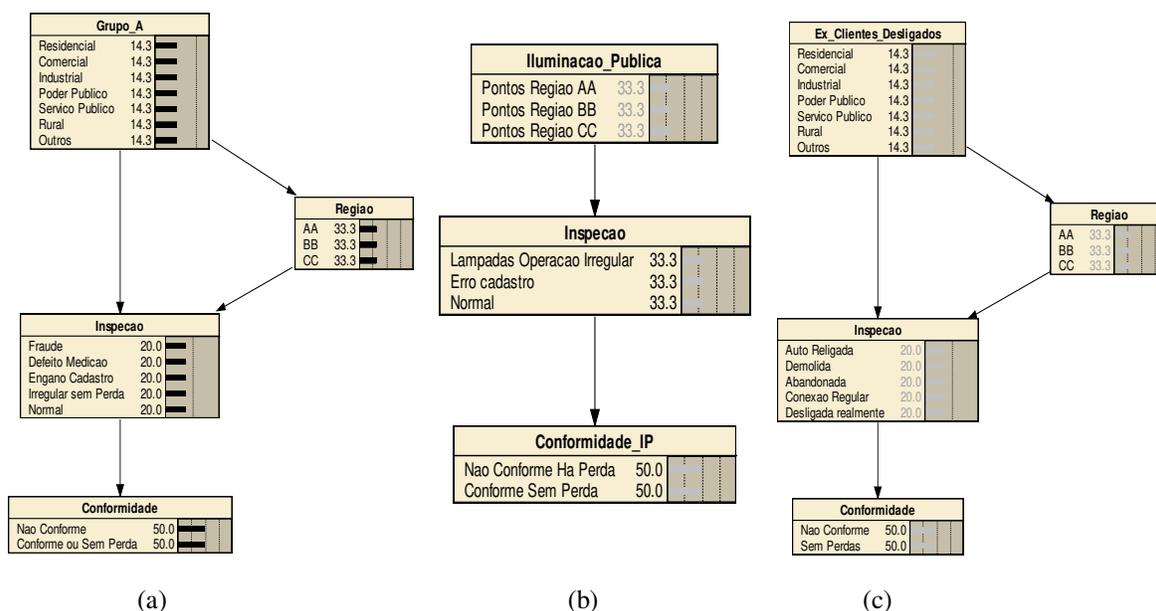


Figura 4.5 – RB sugerida, (a) para os clientes do grupo A, (b) iluminação pública, e (c) ex-clientes Desligados.

4.2.2 – Amostras e teste de validação

De modo semelhante a aplicações estatísticas na área de qualidade, se deseja examinar a conformidade das unidades, portanto se trata de uma distribuição binomial, isto é, se não há perdas ou com perdas (esta a situação não conforme, sendo que p é o índice de não conformidade).

Isto justifica a necessidade de hipóteses *a priori* para estimar a quantidade de unidades não conformes por grupo (VARDEMAN, 1994, p.87). Tais hipóteses iniciais (H_0) serão testadas e validadas posteriormente. Os *Grupos*, segmentos independentes nos quais acontecem as perdas comerciais são cinco: Grupo A, Grupo B, Ligação Clandestina (LC), Iluminação Pública (IP) e Desligados (unidades consumidoras já desconectadas há mais tempo). A composição da hipótese inicial é a estimativa dos índices de não conformidade p para estes cinco grupos tendo como condição de contorno que o total de PC de energia seja aquele encontrado subtraindo-se da perda global a perda técnica calculada.

Fazendo analogia com as aplicações de RB na área médica que usam os resultados de exames para suprir as tabelas de probabilidades condicionais (TPC), no modelo para as perdas comerciais os resultados quanto à conformidade provenientes das inspeções das amostras nos diversos *Grupos* são os dados das TPC.

O trabalho de campo no exame da conformidade tem base estatística, sendo válido o teorema do limite central. Denominando-se p a fração de não conformidade, para universos maiores que 10 mil unidades a variância é (VARDEMAN, 1994, p.326)

$$\sigma^2 = \frac{p \cdot (1-p)}{n} \quad (4.6)$$

onde,

σ^2 é a variância;

n , tamanho da amostra;

p , índice de não conformidade das unidades (com perdas), que de início é estimado.

O erro (E) em por unidade é dado por (VARDEMAN, 1994, p.279)

$$E = z \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \quad (4.7)$$

Onde z é o parâmetro da curva normal segundo o intervalo de confiança desejado. Desenvolvendo a expressão anterior, a amostra pode ser dimensionada pela equação seguinte cujos parâmetros já foram definidos:

$$n = p \cdot (1 - p) \cdot \frac{z^2}{E^2} \quad (4.8)$$

O valor máximo da amostra máxima em função da não conformidade ocorre para p igual a 0,5.

Define-se então o intervalo de confiança e o erro desejado e dimensiona-se a amostra. Por exemplo, para um índice de não conformidade elevado, p de 25%, intervalo de confiança de 95% e erro de 5% a amostra é de 289 unidades; se o erro for de 3% a amostra é de 801 unidades.

Para grandes números a amostra independe do universo, assim mantidas as demais condições e com erro de 2%, a amostra é de 1.801 unidades; nesta situação caso o p atinja 0,5 a amostra máxima é de 2.401 unidades, próximo ao valor usado em pesquisas eleitorais.

Enfim, com base nos quantitativos de clientes da empresa, nos dados atuais de mercado, e nas hipóteses a priori quanto a PC é que são dimensionadas as amostras a inspecionar.

Há empresas nas quais o universo (N) de consumidores é menor que 10 mil em algum dos grupos, como o Grupo A; nestes casos a amostra é estatisticamente expressa por

$$n = \frac{z^2 p (1-p) N}{E^2 (N-1) + z^2 p (1-p)} \quad (4.9)$$

Onde todos os parâmetros já foram anteriormente definidos.

Verifica-se que considerando o universo N muito grande em (4.9), chega-se à expressão (4.8). Por outro lado, caso haja muita não conformidade no grupo A as PC serão significativas, logo é recomendável se utilizar as duas expressões e tomar a maior das amostras, afinal o ideal neste grupo seria inspecionar todos os clientes no máximo a cada dois ou três anos.

Um ponto relevante é que em vários grupos, ao distribuir a amostra proporcionalmente entre as diversas classes há algumas que resultam pouco representativas, ou seja, que ficariam com quantidades não expressivas. É o caso, por exemplo, da classe residencial dentro do grupo A, bem como das classes industrial e rural nos grupos B e Desligados. Assim, se preferiu estabelecer um critério que garanta uma quantidade mínima da amostra na *classe* em função do número de unidades, como indicado na Tabela 4.3, sabendo que isto conduz a um acréscimo nas quantidades das amostras.

Portanto, para obter o número de unidades a inspecionar em cada *classe* aplicam-se os dois critérios, a divisão proporcional e o da quantidade mínima; escolhe-se aquele que representa maior valor e aplica-se uma folga de 10%.

Tabela 4.3 – Quantidades mínimas da amostra função do total de unidades.

Total de Unidades	Mínimo da amostra
Até 30	8,0%, mais 0,2, ou mínimo de 2
31 a 100	5% + 1,1
101 a 500	1,80% + 4,3
501 a 1.000	1,00% + 8,3
1.001 a 10.000	0,20% + 16,3
10.001 a 50.000	0,05% + 31,3
Maior que 50.000	0,025% + 43,8

Outra observação quanto às amostras diz respeito aos grupos de iluminação pública e de ligações clandestinas. No banco de dados de IP nas empresas geralmente se dispõe do número de pontos de iluminação por região, daí a amostra mínima poder ser definida pela expressão 4.8, mas como a proposta é a realização de inspeções por quadrículas o que se faz é sortear uma primeira quadrícula e verificar quantos pontos há. Certamente é menor que o número da amostra, então se sorteia outra e mais outras, etc, até que se atinja o valor de pontos da amostra mínima.

Para as LC, as inspeções são feitas em quadrículas representando zonas nas quais se procura este tipo de irregularidade. A definição da amostra considera a correlação com as UC residenciais em

baixa tensão logo, faz-se o sorteio consecutivo de quadrículas até que o número de unidades residenciais do grupo B existentes nas mesmas supere o número da amostra mínima.

Entretanto, considerando que as vistorias para identificação de LC e de não conformidade em IP são mais simples e baratas que as inspeções de consumidores, ampliaram-se as amostras e em lugar de considerar apenas a amostra mínima determinada pela citada expressão, tomou-se inspecionar em IP quadrículas cujo número de pontos sejam 0,4% do total existente caso se trabalhe com erro de 5%, 1,1% do total quando se use erro de 3%, e 1,5% dos pontos de IP ao tomar erro de 2%. Raciocínio semelhante é feito para a pesquisa em campo das ligações clandestinas. Quando das aplicações nos diversos grupos se destacará a influência destes critérios no aumento das amostras.

Definidas as amostras nos vários grupos as inspeções são realizadas e verifica-se a conformidade possibilitando obter os dados das TPC da rede Bayesiana, e na seqüência inferir o número de unidades com perdas em cada grupo.

Ao final, tendo-se a não conformidade quantitativa nos grupos é feito o teste de validação, aqui empregado o qui-quadrado, sendo recomendado o nível de significância de 5%. São duas as situações de rejeição da hipótese inicial (H_0) e suas conseqüências conforme o tipo de erro.

A hipótese inicial é que o processo de faturamento não funciona bem ou que existe PC elevada na empresa. Numa analogia com a medicina a hipótese inicial é que o paciente (a empresa) está doente (tem PC elevada), e parte-se para o diagnóstico. Pode-se resumir o quadro como na Tabela 4.4, adaptada de ZAR (1974, pp.44-47, erros estatísticos e teste de hipótese).

Tabela 4.4 – Tipos de erro no teste de hipótese das perdas comerciais.

Diagnóstico	Existe PC elevada (H_0)	
	Verdadeiro	Falso
Rejeita-se H_0	Erro tipo I	Sem erro, OK
Aceita-se H_0	OK	Erro tipo II

Rejeitar H_0 sendo tal hipótese falsa significa que a empresa não tem perda comercial alta e que o diagnóstico realmente não a detectou, logo o processo estaria sem erro. O erro tipo I acontece quando H_0 é verdadeiro, mas as inspeções das amostras ou o diagnóstico não identifica tais perdas elevadas na empresa (conclusão do falso-negativo). É o caso de se ampliar o espaço pesquisado refazendo-se as amostras, se trabalhar com erro menor ou mesmo redefinir H_0 .

Quando H_0 é verdadeira e esta hipótese inicial é aceita significa que as conformidades nos grupos estão dentro das margens de erro, pode-se prosseguir com a quantificação das perdas comerciais por causa, região e classe de modo a se aplicar as ações ou remédios necessários. Já a aceitação de H_0 quando a mesma é falsa (erro tipo II) significa que há algo errado nos cálculos das perdas técnicas, e nas medições daí a perda global não estaria correta, ou que as irregularidades encontradas são recentes visto que se trabalha com valores médios de perdas anuais. De qualquer modo para este tipo de erro deveriam ser providenciadas ações corretivas.

Cabe destacar que mesmo havendo o erro tipo I isto não produziria gastos excessivos na distribuidora visto que não desencadearia nenhum grande projeto de redução de perdas, afinal as mesmas existem, entretanto não foram devidamente localizadas. Iniciar-se qualquer ação de redução significa não chegar ao objetivo de redução das PC. Como o processo é dinâmico tal risco é pequeno.

Todos os cálculos deste diagnóstico podem ser feitos em uma planilha ou se pode usar um dos softwares específicos que trabalham com RB como o NETICA®, ELVIRA, HUGIN® ou o ToolBox do MATLAB®. O uso apenas destes programas não é suficiente, pois as estimativas de energia são feitas em separado empregando as médias de consumo. Foram experimentadas duas versões demonstrativas dos programas comerciais, e aqui as figuras que são imagens das telas de software foram originadas do Netica® (NORSYS, 2008).

Destaca-se a comodidade dos softwares específicos para RB ao fazer inferências, facilitando a busca de informações por região ou tipo de ocorrência. Por exemplo, a instanciação simultânea do estado *fraude* da variável *Inspeção*, e uma das classes, *comercial*, dentro de um nó *Grupo*.

4.2.3 – Aplicação da metodologia na estimativa das perdas comerciais dos clientes do grupo B de uma concessionária

Não tendo sido encontrados dados públicos para verificar comparativamente a metodologia e considerando que apenas se conseguiu a aplicação parcial em uma distribuidora, optou-se por usar os dados de UC do grupo B da Região Metropolitana de Salvador (DANTAS, 2006). Entretanto não foi possível seguir exatamente a rede Bayesiana proposta porque faltavam dados das UC por região (quantidades e resultado das inspeções), e de consumo médio por faixa.

Adaptou-se o modelo de RB ficando como mostra a Figura 4.6. Esta aplicação completa compõe o Anexo A, e aqui é feito apenas um resumo desta adaptação.

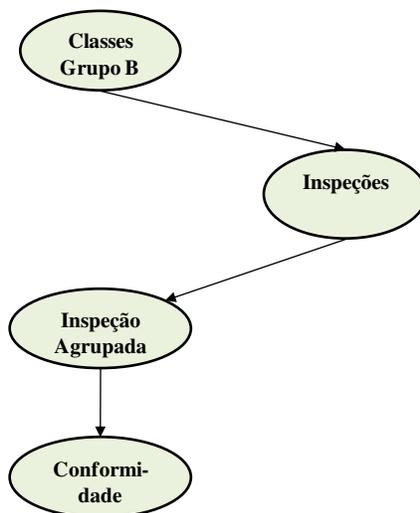


Figura 4.6 – Rede Bayesiana para diagnóstico da PC do grupo B, Salvador (dados de DANTAS, 2006).

Considerou-se a variável *Classes_Grupo_B* com os estados anteriormente citados, mas o nó *Inspeções* assumiu dez estados sendo necessário criar o nó *Inspeção Agrupada* com quatro estados para tratar as PC. Não foram usados aqueles cinco estados desejados na metodologia e mostrados na Figura 4.4; aqui o estado *Defeito_Perda_Empresa* engloba as ocorrências de defeito na medição e de engano no cadastro. A Tabela 4.5 correlaciona os estados destes nós.

Tabela 4.5 - Correspondência entre os estados das variáveis *Inspeções* e *Inspeção Agrupada*.

Inspeções	Inspeção Agrupada
Ponte no bloco de terminais	Fraude
Ligação direta	Fraude
Interrupção bobina de potencial	Fraude
Desvio antes do medidor	Fraude
Desvio embutido na parede	Fraude
Medidor avariado e selo violado	Fraude
Medidor defeituoso	Defeito Perda Empresa
Medidor danificado	Defeito Perda Empresa
Irregular Sem Perdas	Irregular Sem Perda
Normal	Normal

No grupo B existiam na Região Metropolitana de Salvador por ocasião da realização da pesquisa 825.698 consumidores e foram feitas 374 inspeções, cujos resultados já agrupados estão no Anexo A.

Inserindo-se todos os dados na RB adaptada fez-se a simulação verificando-se, conforme Figura 4.7, que a não conformidade é de 14,2% (UC com perda) dentre as unidades do Grupo B, sendo que a parcela atribuída a fraude (8,64%) excede a defeito (5,53%).

Após os cálculos são obtidas as estimativas quanto a não conformidade e a crença é que há 117.015 UC com perdas, das quais 111.934 na classe residencial. Estes valores correspondem ao que foi determinado usando o software específico descrito por DANTAS (2006).

Com base no consumo médio anual de cada uma das classes é feita a estimativa da perda anual de energia que corresponde a um total de 185.393 MWh (ou 11,14% da energia vendida no Grupo B), concentrando-se mais nas classes residencial (165.886 MWh) e comercial (16.992 MWh). O trabalho original estima a PC neste grupo em 152.068 MWh, mas cita o emprego de estratificação por faixas de consumo e uso das médias encontradas nas unidades não conformes. Como há mais irregularidades nas faixas de menor consumo isto contribuiu para reduzir a estimativa das PC, entretanto tais médias não estão apresentadas na referência.

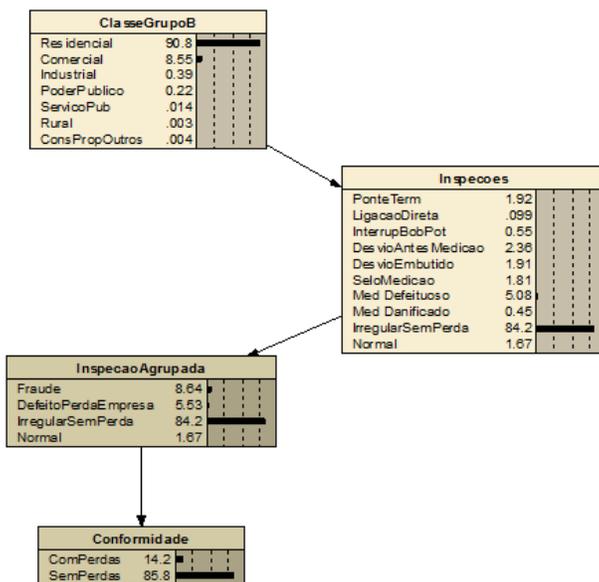


Figura 4.7 – Rede Bayesiana do Grupo B, simulação no Netica® (NORSYS, 2008).

Esta aplicação demonstrou que o modelo de RB usado é coerente e valida a metodologia do diagnóstico calcada na estimativa de unidades não conformes. Constatou-se ainda a facilidade de adaptação da estrutura de RB aplicada, inclusive no que tange aos estados das variáveis de acordo com práticas usuais nas empresas.

CAPÍTULO 5

5. DIAGNÓSTICO DAS PERDAS COMERCIAIS: ESTUDO DE CASO COMPLETO

Aqui são usados os dados contidos em CODI (1996), sendo feitas estimativas quanto à energia vendida por classe e ao número de consumidores por região já que era conhecido apenas o total (585.555). Os quantitativos de inspeção colocados mais adiante são fictícios, tendo sido extrapolados de uma distribuidora que fez o diagnóstico apenas parcialmente. A energia anual comprada ou injetada na rede da empresa é 2.712.393 MWh, esta fornece aos clientes 2.243.327 MWh, portanto há 17,3% de perdas globais, ou 469.066MWh/ano.

5.1 – Dados de mercado e cálculo da perda comercial

Tal empresa teria dados de mercado conforme indicado na Tabela 5.1.

No total são 1.718 consumidores no grupo A cujo consumo é de 493.566 MWh/ano, e existem 583.760 UC no grupo B consumindo anualmente 1.664.339 MWh. Para a iluminação pública, são 77 unidades com consumo de 85.426 MWh/ano, que corresponde a 3,81% da energia vendida ficando próximo da média nacional que é de 3,44% (FRÓES DA SILVA, 2006).

Tabela 5.1 - Dados de mercado da empresa usada no estudo de caso (adaptado de CODI, 1996).

Classe	No. de Unidades	Classe	Energia (MWh/ano)
Residencial	532.895	Residencial	759.573
Comercial	36.928	Comercial	855.736
Industrial	2.158	Industrial	474.357
Poder Público	5.030	Rural	21.811
Serviço Público	642	Outras	46.424
Rural	7.703	Iluminação Pública	85.426
Cons.Próprio e Outros	122	Energia Vendida	2.243.327
Ilum. Pública (IP)	77	Energia comprada	2.712.393
Total	585.555	Perda global	469.066

Outros dados também foram estimados a exemplo das quantidades de consumidores segundo as três regiões (AA, BB e CC) para os grupos A, B e para os Desligados, estando apresentados na Tabela 5.2. Em geral tais dados estão disponíveis no sistema de faturamento da empresa.

Tabela 5.2 - Quantidades de unidades dos grupos por classe e região.

	Grupo B			Grupo A			Desligados		
	AA	BB	CC	AA	BB	CC	AA	BB	CC
Residencial	181.630	178.033	173.190	23	11	8	34.373	35.500	34.009
Comercial	17.154	11.495	7.796	218	170	95	3.374	3.059	2.664
Industrial	726	542	294	295	210	91	246	227	192
Poder Público	1.846	1.644	1.383	81	45	31	396	364	372
Serv. Público	237	204	186	10	3	2	66	82	71
Rural	2.569	2.002	2.711	58	203	160	589	621	478
Outros	49	41	28	2	2	0	39	27	18
Total	204.211	193.961	185.588	687	644	387	39.083	39.880	37.804

A perda técnica média calculada é 205.298 MWh/ano ou 7,6%, tendo sido determinada pela metodologia SIMPLES (BASTOS et al, 2008), incluídas as três recentes alterações comentadas no item 2.1. Tal metodologia indica valores máximos e mínimos para as PT nos diversos segmentos, como mostrado na Tabela 5.3. Neste caso as PT são mais significativas na sub-transmissão e subestações abaixadoras (2,30%), e nos transformadores de distribuição.

Tabela 5.3 – Valores calculados das perdas técnicas por segmento e estimativa da perda comercial.

Segmentos	Mínimo (MWh/ano)	Máximo (MWh/ano)	Média (MWh/ano)	Média (%)
Subtransmissão e SE	54.248	70.522	62.385	2,30
Rede primária	32.190	33.653	32.921	1,21
Transformadores	39.300	40.538	39.919	1,47
Rede secundária	38.099	39.831	38.965	1,44
Ramais	8.792	13.188	10.990	0,41
Medidores	7.633	18.992	13.312	0,49
Diversas	6.300	7.310	6.805	0,25

Segmentos	Mínimo (MWh/ano)	Máximo (MWh/ano)	Média (MWh/ano)	Média (%)
Perda técnica total	186.562	224.034	205.298	7,6
Perda comercial total	245.032	282.504	263.768	9,7

Usando (2.2) a perda comercial média é 263.768 MWh/ano ou 9,7%. Para esta empresa o indicador I_{stp} é 128%, caracterizando descontrolo do processo de gestão das PC; neste caso, deveria a agência reguladora exigir da distribuidora o diagnóstico e um plano de redução das PC. A energia anual comprada (2.712.397 MWh) se divide entre venda aos consumidores e perdas como na Figura 5.1. Vê-se que o total de perdas é quase igual à energia vendida ao grupo A.

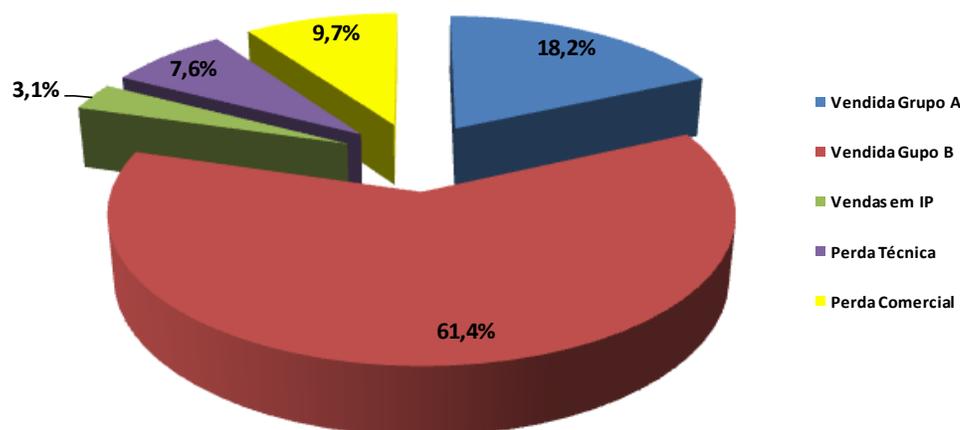


Figura 5.1 – Repartição da energia anual comprada (2.712.397 MWh).

As PC deverão ser estimadas dentro dos clientes (grupo A, grupo B e IP), e não clientes nos grupos ligações clandestinas e unidades desligadas (neste, as perdas são devido ao auto-religamento, ou seja, religação à revelia), sendo a soma aproximadamente o total das PC, ou seja, 9,7%.

5.2 – Diagnóstico sob hipótese inicial

5.2.1- Hipótese inicial e definição das amostras

Os índices de não conformidade inicial (p) em cada grupo devem ser estimados compondo a hipótese inicial ou *a priori*, pois são parâmetros importantes nas definições das amostras aleatórias para as inspeções da conformidade em campo.

Estabeleceu-se esta hipótese relativa à repartição das PC de modo semelhante ao que em geral fazem as distribuidoras, isto é, admitiu-se que quase toda a perda está nos clientes dos grupos A e B, e que a PC é pequena entre os Desligados e as ligações clandestinas. Assim, haveria 8,0% de não conformidade (p) no grupo A, 3,0% em IP, e para os não-clientes, 3,8% entre os desligados e 3,8% para as LC (correlação com os consumidores residenciais em BT).

Com os valores médios de consumo estimado para as unidades irregulares em cada grupo se calcula a perda de energia. Foram considerados para as unidades não conformes os mesmos valores dos respectivos consumidores normais da empresa. Para as LC o consumo médio esperado por unidade supôs-se igual ao do cliente residencial em BT.

A condição de contorno é que o somatório das perdas nos cinco grupos deve coincidir com o total das PC, daí imputou-se a não conformidade no grupo B de 10,8%. Deste modo, *a priori* estariam sendo perdidos 180.223 MWh/ano no grupo B e 28.847 MWh com as ligações clandestinas. A Tabela 5.4 apresenta a perda técnica e as estimativas de perda comercial conforme a hipótese inicial formulada para os vários grupos.

Tabela 5.4 – Perda técnica e estimativa de PC segundo hipótese inicial.

		Não conformidade		Perda (MWh/ano)
		(p)	Unidades	
Perda por Grupo	Técnica	xxx	xxx	205.298
	PC no Grupo A	8,0	137	39.485
	PC no Grupo B	10,8	63.212	180.223
	PC em IP	3,0	xxx	2.563
	PC no Grupo Desligados	3,8	4.437	12.650
	PC no Grupo LC	3,8	20.248	28.847
	Total das perdas	xxx	88.034	469.066

Com estas estimativas *a priori* das parcelas de PC pode-se construir o gráfico da Figura 5.2 que mostra como a perda total de energia (100%) se reparte em PT e perda comercial nos diversos grupos. Vê-se que apenas as PC nos grupos A e B superariam toda a perda técnica.

Esta é a hipótese inicial a ser validada através do diagnóstico das PC que usa o modelo Bayesiano e as inspeções das amostras para exame de conformidade.

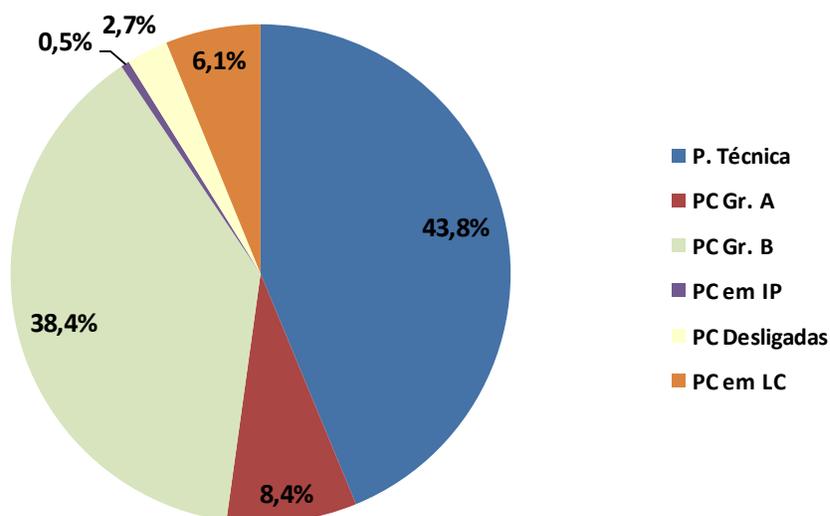


Figura 5.2 – Energia perdida (469.066 MWh/ano ou 100%) distribuída segundo hipótese inicial.

Na definição das amostras usou-se intervalo de confiança de 95% e recomenda-se usar o mesmo erro em todas as amostras dos grupos, máximo da ordem de 5% preferivelmente. Adotou-se este valor.

Assim, usando a expressão 4.8 as amostras em cada uma das três regiões seriam 114 unidades, 149 e 57 unidades, respectivamente para os grupos A, B e Desligados. Entretanto foi definido um número mínimo de unidades da amostra em cada *classe*, pois caso isto não seja feito, ao distribuir proporcionalmente a amostra de 149 unidades na região AA do grupo B, pelos estados correspondentes às classes, a amostra ficaria concentrada nas classes residencial e comercial e haveria o risco de se ter um número muito pequeno de inspeções nas demais. Por exemplo, são 204.211 unidades na região AA, grupo B, dos quais 2.569 rurais (Tabela 5.2); a distribuição proporcional conduziria a inspecionar apenas dois destes, entretanto aplicando o critério citado no item 4.2.2, Tabela 4.3, serão 22 unidades rurais a inspecionar.

Portanto, mesmo conduzindo a um acréscimo na amostra é adotado o critério mencionado anteriormente, e as amostras passam a ser de 363 unidades no grupo A, 354 no Desligados e 946 no grupo B, este inclusive ficando com valores um pouco diferentes por região.

Para o grupo B, como será visto adiante, as classes residencial e comercial foram sub-divididas por faixa de consumo: quatro na classe residencial e duas na comercial. Não se detalhou, mas das

17.154 UC comerciais da região AA, 9.221 têm consumo médio até 1.000kWh/mês e outras 7.933 têm média superior a esta.

Os valores distribuídos por *Classe* e *Região* para os grupos A e Desligados são mostrados na Tabela 5.5. Observa-se que no estado *outros* estão incluídas as unidades que correspondem ao consumo próprio.

Tabela 5.5 - Quantidades das amostras a inspecionar por Classe e Região, grupos A e Desligados.

	Grupo A				Desligados			
	AA	BB	CC	Total	AA	BB	CC	Total
Residencial	4	3	4	11	49	51	57	157
Comercial	37	31	28	96	24	23	22	69
Industrial	49	38	27	114	9	9	8	26
Poder Público	14	8	10	32	12	11	11	34
Serviço Público	5	3	2	10	5	6	5	16
Rural	10	36	50	96	15	15	13	43
Outros	2	2	0	4	4	3	2	9
Total	121	121	121	363	118	118	118	354

Para o grupo de IP são sorteadas quadrículas e há no banco de dados da empresa o número e localização dos pontos de iluminação por região. Considerando o que foi colocado no item 4.2.2 como se está usando erro de 5% o número mínimo de pontos que deve existir nas quadrículas sorteadas deve corresponder a 0,4% do total.

Para as quadrículas a investigar referentes a ligações clandestinas, estas devem ter também um mínimo de 0,4% das quantidades de unidades residenciais BT das regiões. Deste modo, são realizados os sorteios e as quadrículas com áreas a investigar para IP e LC devem ter os quantitativos conforme apresentado na Tabela 5.6.

Tabela 5.6 - Mínimo de pontos de IP e de UC residencial nas zonas das amostras dos grupos IP e LC.

Grupo Iluminação Pública				
	AA	BB	CC	Total
Consumo de IP (kWh/ano)	35.027	27.244	23.155	85.426
Total de pontos de IP	59.486	47.484	41.083	148.053
Mínimo de pontos da amostra	238	190	165	593
Quantidade de pontos da amostra	292	214	174	680

Grupo Ligação Clandestina				
	AA	BB	CC	Total
UC residenciais da região	181.630	178.033	173.190	532.853
Mínimo de UC residencial na amostra	727	713	693	2.133
Quantidade de UC residencial da amostra	739	727	712	2.178

5.2.2 – Conformidade e estimativa das PC para os grupos de clientes

Feitas as inspeções das amostras são determinados os valores percentuais da não conformidade em cada grupo pesquisado. Neste item há os resultados e estimativas para os grupos A, B e IP.

5.2.2.1 - Clientes do Grupo A

No grupo A os resultados da amostra inspecionada estão mostrados na Tabela 5.7.

Tabela 5.7 - Resultados da amostra inspecionada no grupo A.

	Fraude	Defeito medição	Engano cadastro	Irregular s/ perda	Normal	Total
Residencial Reg.AA	0	0	0	2	2	4
Residencial Reg.BB	0	0	0	1	2	3
Residencial Reg.CC	0	1	0	0	3	4
Comercial Reg.AA	1	0	1	10	25	37
Comercial Reg.BB	0	0	1	8	22	31
Comercial Reg.CC	1	0	0	11	16	28
Industrial Reg.AA	1	0	0	14	34	49
Industrial Reg.BB	0	1	0	14	23	38
Industrial Reg.CC	1	0	0	6	20	27
Poder Público Reg.AA	0	0	0	3	11	14
Poder Público Reg.BB	0	0	0	0	8	8
Poder Público Reg.CC	0	0	0	2	8	10
Serviço Púb. Reg.AA	0	0	0	1	4	5
Serviço Púb. Reg.BB	0	0	0	1	2	3
Serviço Púb. Reg.CC	0	1	0	0	1	2
Rural Reg.AA	0	0	0	3	7	10
Rural Reg.BB	1	0	0	11	24	36
Rural Reg.CC	0	1	0	8	41	50
Outros Reg.AA	0	0	0	0	2	2
Outros Reg.BB	0	0	0	1	1	2
Outros Reg.CC	0	0	0	0	0	0
Totais	5	4	2	96	256	363

Depois de analisados os dados das inspeções em campo (Tabela 5.7), o passo seguinte é a montagem da rede Bayesiana referente ao grupo A, cuja estrutura é vista na Figura 4.5(a). As probabilidades dos estados do nó de entrada *GrupoA* não são condicionais e são determinadas conforme os dados de mercado da empresa (Tabela 5.2). Assim, segundo equação 4.3 a probabilidade da unidade ser comercial é

$$P(\text{GrupoA} = \text{comercial}) = \frac{(218 + 170 + 95)}{(687 + 644 + 387)} = 0,281141$$

Como está sendo tratado o grupo A, para simplificar, pode-se usar letra minúscula referindo-se à probabilidade de um estado e letra maiúscula para probabilidade da variável. Assim, tem-se a probabilidade da classe ser rural como

$$P(\text{rural}) = \frac{(58 + 203 + 160)}{(687 + 644 + 387)} = 0,245052$$

E para a variável ou nó *GrupoA*, tem-se

$$P(\text{GrupoA}) = \begin{bmatrix} 0,024447 \\ 0,281141 \\ 0,346915 \\ 0,091385 \\ 0,008731 \\ 0,245052 \\ 0,002328 \end{bmatrix}$$

Para o nó região, as probabilidades são condicionais e são determinadas também a partir dos dados de mercado da Tabela 5.2. Por exemplo, a probabilidade de uma unidade pertencer à região AA desde que seja comercial é

$$P(\text{Região} = \text{AA} | \text{GrupoA} = \text{comercial}) = \frac{218}{(218 + 170 + 95)} = 0,451346$$

Daí a tabela de probabilidade condicional para o nó *Região* (com três estados) e que é filho de *GrupoA*, que tem sete estados, é uma matriz 7x3 na qual as linhas se referem às classes do nó *GrupoA* e as colunas às três regiões AA, BB e CC, cujos valores são

$$P(\text{Região}|\text{GrupoA}) = \begin{bmatrix} 0,547619 & 0,261905 & 0,190476 \\ 0,451346 & 0,351967 & 0,196687 \\ 0,494966 & 0,352349 & 0,152685 \\ 0,515924 & 0,286624 & 0,197452 \\ 0,666667 & 0,200000 & 0,133333 \\ 0,137767 & 0,482185 & 0,380048 \\ 0,500000 & 0,500000 & 0,000000 \end{bmatrix}$$

Como o nó inspeção é filho de *GrupoA* e *Região*, a TPC é uma matriz 21x5 na qual as linhas representam a combinação dos estados dos pais e as colunas são os estados de inspeção, nesta ordem: *fraude*, *defeito na medição*, *engano cadastro*, *irregular sem perda* e *normal*. Esta matriz é determinada com os resultados das inspeções.

A probabilidade condicional de uma unidade ser encontrada com fraude tal que seja da região AA e comercial é

$$P(\text{Inspeção} = \text{fraude} | \text{GrupoA} = \text{comercial}, \text{Região} = \text{AA}) = \\ = P(\text{fraude} | \text{comercial}, \text{regiãoAA}) = \frac{1}{(1 + 0 + 1 + 10 + 25)} = 0,027027$$

E este é o termo (4,1) da matriz da TPC. A Tabela 5.8 apresenta todas estas probabilidades.

Tabela 5.8 – Grupo A, probabilidades condicionais do nó *Inspeção*.

	Fraude	Defeito medição	Engano cadastro	Irregular s/ perda	Normal
Residencial Reg.AA	0,000000	0,000000	0,000000	0,500000	0,500000
Residencial Reg.BB	0,000000	0,000000	0,000000	0,333333	0,666667
Residencial Reg.CC	0,000000	0,250000	0,000000	0,000000	0,750000
Comercial Reg.AA	0,027027	0,000000	0,027027	0,270270	0,675676
Comercial Reg.BB	0,000000	0,000000	0,032258	0,258065	0,709677
Comercial Reg.CC	0,035714	0,000000	0,000000	0,392857	0,571429
Industrial Reg.AA	0,020408	0,000000	0,000000	0,285714	0,693878
Industrial Reg.BB	0,000000	0,026316	0,000000	0,368421	0,605263
Industrial Reg.CC	0,037037	0,000000	0,000000	0,222222	0,740741
Poder Púb. Reg.AA	0,000000	0,000000	0,000000	0,214286	0,785714
Poder Púb. Reg.BB	0,000000	0,000000	0,000000	0,000000	1,000000
Poder Púb. Reg.CC	0,000000	0,000000	0,000000	0,200000	0,800000
Serviço Púb. Reg.AA	0,000000	0,000000	0,000000	0,200000	0,800000
Serviço Púb. Reg.BB	0,000000	0,000000	0,000000	0,333333	0,666667
Serviço Púb. Reg.CC	0,000000	0,500000	0,000000	0,000000	0,500000
Rural Reg.AA	0,000000	0,000000	0,000000	0,300000	0,700000
Rural Reg.BB	0,027778	0,000000	0,000000	0,305556	0,666667
Rural Reg.CC	0,000000	0,020000	0,000000	0,160000	0,820000

	Fraude	Defeito medição	Engano cadastro	Irregular s/ perda	Normal
Outros Reg.AA	0,000000	0,000000	0,000000	0,000000	1,000000
Outros Reg.BB	0,000000	0,000000	0,000000	0,500000	0,500000
Outros Reg.CC	0,000000	0,000000	0,000000	0,000000	0,000000

A variável *Conformidade* é filha de *Inspeção*, e as relações de probabilidade condicional são a certeza do estado conforme (sem perda) quando os estados do nó *Inspeção* são *Normal* ou *Irregular sem Perdas*, e certeza de não conformidade (há perdas) para os estados *Fraude*, *Defeito medição* e *Engano cadastro*.

Calcular a probabilidade de unidades do grupo A com fraude desde que estejam na classe comercial e região AA, é determinar a probabilidade conjunta. Para isto se aplica a regra de cadeia para o estado *Fraude*, ou seja,

$$P(\text{fraude, comercial, regiãoAA}) = P(\text{fraude}|\text{comercial, RegiãoAA}) \cdot P(\text{RegiãoAA}|\text{comercial}) \cdot P(\text{comercial}) = 0,027027 \cdot 0,451346 \cdot 0,281141 = 0,003430$$

A quantidade esperada de unidades com fraude na classe comercial e região AA é o produto desta probabilidade pelo total de unidades do grupo A (1.718) resultando em seis unidades.

Do mesmo modo o valor esperado de unidades consumidoras normais na mencionada classe e região é o produto da probabilidade conjunta para o estado *Normal* na classe comercial e região AA pelo número total de unidades do grupo A, isto é,

$$UC_{\text{Normais}} = 1718 \cdot P(\text{normal, comercial, regiãoAA}) = 1718 \cdot 0,675676 \cdot 0,451346 \cdot 0,281141 = 147,3 \sim 147 \text{ unidades}$$

Os valores estimados de unidades por grupo e região ainda sem arredondamento, em função dos estados da inspeção estão indicados na Tabela 5.9.

Tabela 5.9 – Estimativas de unidades por grupo e região função dos estados do nó *Inspeção*.

Grupo A	Região	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total
Residencial	AA	0,0	0,0	0,0	11,5	11,5	23
Residencial	BB	0,0	0,0	0,0	3,7	7,3	11
Residencial	CC	0,0	2,0	0,0	0,0	6,0	8
Comercial	AA	5,9	0,0	5,9	58,9	147,3	218
Comercial	BB	0,0	0,0	5,5	43,9	120,6	170

Grupo A	Região	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total
Comercial	CC	3,4	0,0	0,0	37,3	54,3	95
Industrial	AA	6,0	0,0	0,0	84,3	204,7	295
Industrial	BB	0,0	5,5	0,0	77,4	127,1	210
Industrial	CC	3,4	0,0	0,0	20,2	67,4	91
Poder Público	AA	0,0	0,0	0,0	17,4	63,6	81
Poder Público	BB	0,0	0,0	0,0	0,0	45,0	45
Poder Público	CC	0,0	0,0	0,0	6,2	24,8	31
Serviço Público	AA	0,0	0,0	0,0	2,0	8,0	10
Serviço Público	BB	0,0	0,0	0,0	1,0	2,0	3
Serviço Público	CC	0,0	1,0	0,0	0,0	1,0	2
Rural	AA	0,0	0,0	0,0	17,4	40,6	58
Rural	BB	5,6	0,0	0,0	62,0	135,3	203
Rural	CC	0,0	3,2	0,0	25,6	131,2	160
Outros	AA	0,0	0,0	0,0	0,0	2,0	2
Outros	BB	0,0	0,0	0,0	1,0	1,0	2
Outros	CC	0,0	0,0	0,0	0,0	0,0	0
Total		24,3	11,7	11,4	469,7	1.200,8	1.718

Deste modo a não conformidade esperada é

$$\text{Não conformidade} = 100 \cdot \frac{(24,3 + 11,7 + 11,4)}{1718} = 2,76\%$$

Instanciar é ter uma forte evidência. Pode-se instanciar o estado *Fraude*, por exemplo, para verificar a sua distribuição por região ou classe. Então, dentre os clientes com fraude a expectativa de que se encontrem na região AA é de

$$P(\text{regiãoAA}|fraude) = \frac{(5,9 + 6,0)}{24,3} = 0,489927$$

Em percentual esta expectativa é 49,0%. Na Figura 5.3 (a) está apresentada a saída da RB do grupo A vendo-se que a não conformidade é 2,76%, enquanto a Figura 5.3 (b) mostra a instanciação do estado *Fraude*; observa-se que esta irregularidade concentra-se mais na região AA e nas classes industrial (38,6%) e comercial (38,2%).

A Tabela 5.10 mostra a estimativa da distribuição das unidades do grupo A por classe e segundo os estados do nó *Inspeção*, sendo que na última coluna estão as médias anuais de consumo dos

clientes normais da empresa, valores estes considerados para o cálculo da perda comercial esperada de energia.

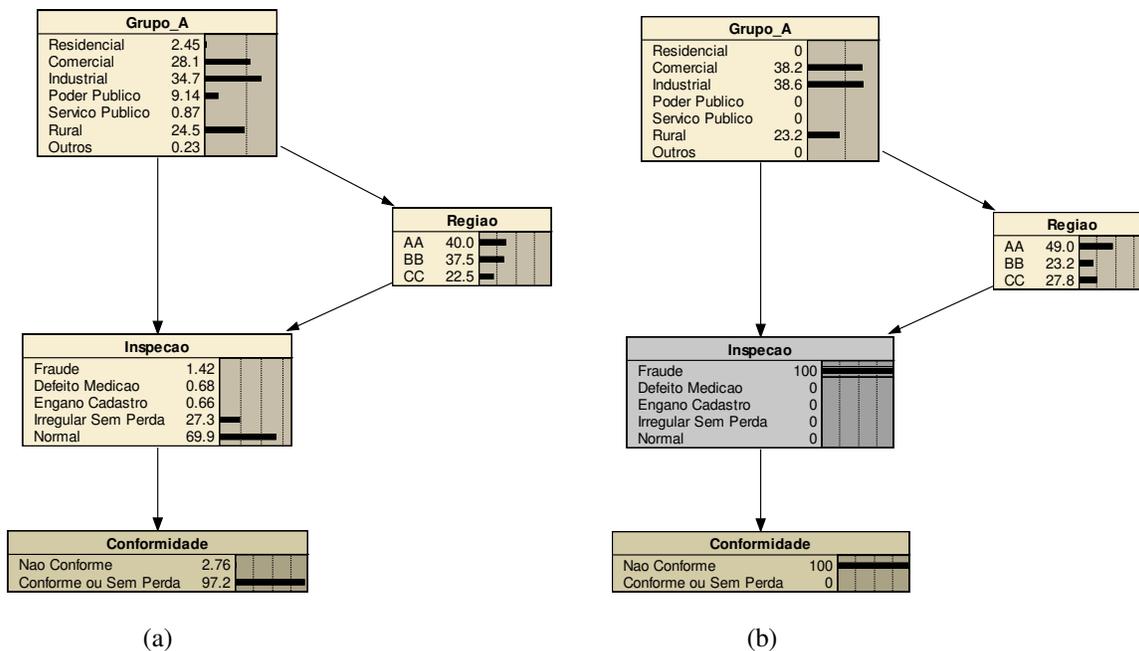


Figura 5.3 – RB do grupo A, (a) não conformidade 2,76%, e (b) instanciação do estado *Fraude*.

Assim, dentre os 1.718 clientes do grupo A estima-se que existam 47 unidades não conformes, predominando a irregularidade fraude.

São também determinadas as crenças quanto a PC anual de energia por classe e região sendo, por exemplo, a perda esperada no grupo A, classe comercial de

$$PC_{\text{comercial}} = (5,9 + 5,9 + 5,5 + 3,4) \cdot 168,0 \sim 3.471 \text{MWh/ano}$$

Tabela 5.10– Estimativa (já arredondada) de unidades por classe e média anual de consumo.

Grupo A	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total	Consumo (MWh/ano)
Residencial	0	2	0	15	25	42	10,8
Comercial	9	0	11	140	323	483	168,0
Industrial	9	6	0	182	399	596	669,6
Poder Público	0	0	0	24	133	157	21,6
Serviço Público	0	1	0	3	11	15	216,0
Rural	6	3	0	105	307	421	14,8
Outros	0	0	0	1	3	4	10,8
Total	24	12	11	470	1.201	1.718	

O total de perda estimada para o grupo A é de 13.828 MWh/ano, dos quais 6.011 MWh na região AA e 4.705 na região BB. O motivo fraude seria responsável por 57,4% consumo irregular.

5.2.2.2 - Clientes do Grupo B

Considerando que dentre os 583.760 consumidores do grupo B 91,3% estão na classe residencial e outros 6,2% na comercial se preferiu subdividir estas duas classes.

Assim, a residencial foi subdividida em quatro subclasses conforme a faixa de média de consumo mensal: aquelas unidades com consumo inferior a 50 kWh/mês, consumo desde 51 a 200 kWh/mês, desde 201 a 500, e com consumo acima de 500 kWh/mês. A classe comercial foi trabalhada separando-se os consumidores com média mensal até 1 mil kWh daqueles com média acima deste valor. A Tabela 5.11 mostra os resultados das inspeções de todas as unidades da amostra.

Tabela 5.11 - Resultados da amostra inspecionada no grupo B.

	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total
Resid. Cons.< 50 kWh/mês –Reg.AA	3	0	0	18	39	60
Resid. Cons.< 50 kWh/mês –Reg.BB	3	0	1	21	37	62
Resid. Cons.< 50 kWh/mês –Reg.CC	3	1	0	23	37	64
Resid. Cons. 51 a 200 –Reg.AA	3	2	1	27	38	71
Resid. Cons. 51 a 200 –Reg.BB	3	1	0	18	47	69
Resid. Cons. 51 a 200 –Reg.CC	4	1	1	21	39	66
Resid. Cons. 201 a 500 –Reg.AA	2	1	1	14	23	41
Resid. Cons. 201 a 500 –Reg.BB	1	1	1	11	25	39
Resid. Cons. 201 a 500 –Reg.CC	2	0	1	12	22	37
Resid. Cons. > 501 kWh –Reg.AA	2	0	1	9	11	23
Resid. Cons. > 501 kWh –Reg.BB	1	1	0	6	13	21
Resid. Cons. > 501 kWh –Reg.CC	1	1	0	3	14	19
Comer. Cons. ≤ 1 mil kWh –Reg.AA	1	1	0	14	19	35
Comer. Cons. ≤ 1 mil kWh –Reg.BB	2	1	0	12	14	29
Comer. Cons. ≤ 1 mil kWh –Reg.CC	0	1	0	10	16	27
Comer. Cons.> 1 mil kWh –Reg.AA	1	1	1	15	15	33
Comer. Cons.> 1 mil kWh –Reg.BB	1	1	0	9	16	27

	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total
Comer. Cons.> 1 mil kWh –Reg.CC	1	0	0	7	14	22
Industrial – Região AA	1	1	0	4	10	16
Industrial – Região BB	0	0	1	4	9	14
Industrial – Região CC	0	0	1	6	3	10
Poder público – Região AA	1	1	0	6	12	20
Poder público – Região BB	0	1	0	9	10	20
Poder público – Região CC	0	0	1	5	14	20
Serviço público – Região AA	0	0	0	2	7	9
Serviço público – Região BB	0	0	0	1	7	8
Serviço público – Região CC	0	0	0	2	6	8
Rural – Região AA	1	1	1	9	10	22
Rural – Região BB	1	0	1	6	13	21
Rural – Região CC	0	1	0	6	15	22
Outros – Região AA	0	0	0	1	3	4
Outros – Região BB	0	0	1	0	3	4
Outros – Região CC	0	0	0	1	2	3
Total	38	19	14	312	563	946

Após a conclusão das inspeções, e feitas as consistências, são calculadas as estimativas de não conformidade e das PC de modo semelhante ao que feito para o Grupo A. Como visto na Figura 5.4 (a) a não conformidade no grupo B é 7,17%, estimando-se que existam 41.852 unidades irregulares (27.219 com fraude, 9.012 com defeito na medição e 5.621 com engano no banco de dados). Cabe destacar que os percentuais dos estados das variáveis *Grupo_B* e *Região* dizem respeito a como se distribuem as frequências das UC, por exemplo, 3,48% delas são da classe comercial com consumo médio mensal até 1.000 kWh.

Pelos números da Figura 5.4 (a) a fraude representa 65,04% da não conformidade. Isto é uma inferência probabilística, ou seja, busca-se saber a participação da fraude dentro do total de não conformidade. Poder-se-ia instanciar o estado *Não conforme* do nó *Conformidade* na RB, ou calcular usando a expressão (3.22) relativa a esta inferência, desta forma

$$P(\text{fraude}|\text{não conforme}) = \frac{4,66}{4,66 + 1,54 + 0,96} = \frac{4,66}{7,17} = 0,6504$$

A crença é que a maior não conformidade esteja na região AA (7,6%), e a menor em BB (6,4%). Na Figura 5.4 (b) há a instância simultânea do estado *não conforme* do nó *Conformidade* e do estado *Região AA* da variável *Região*, e observa-se a participação das não conformidades por classe e tipo nesta região, por exemplo, a participação de fraude nas irregularidades é 59,6%, portanto menor que a média da empresa (65,0%).

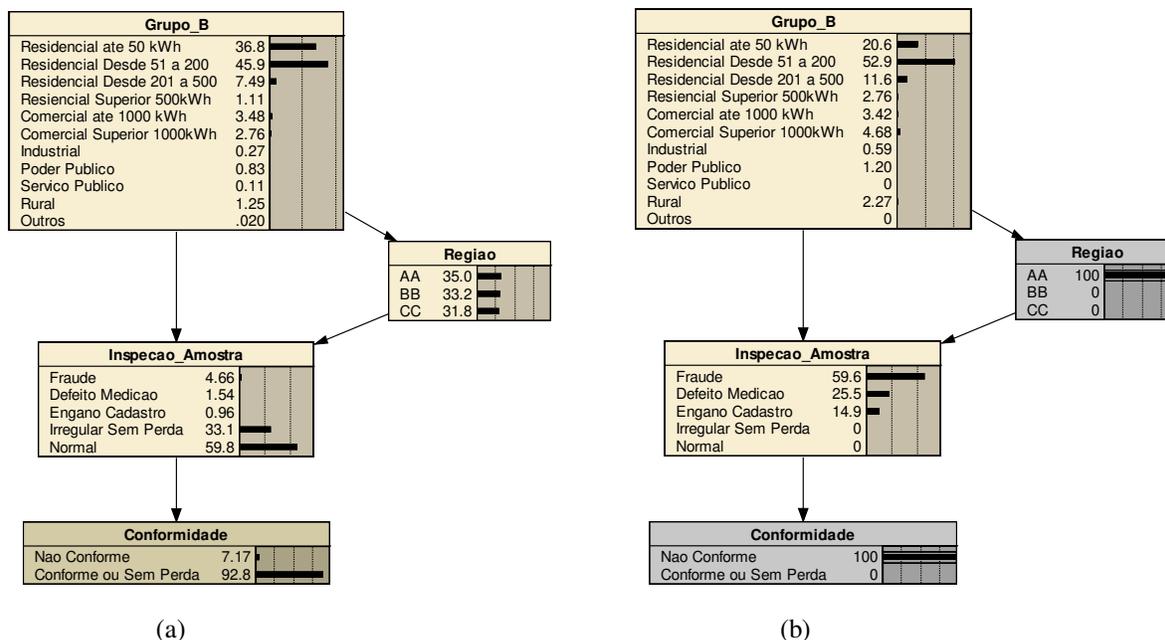


Figura 5.4 – RB relativa ao grupo B, (a) não conformidade 7,17%, e (b) instância simultânea.

No Grupo B, admitindo para as unidades irregulares o mesmo consumo médio anual das unidades conformes de acordo com as classes e faixa de consumo do grupo, a PC é de 130.778 MWh/ano, dos quais 64.026 MWh (ou 49,0%), na região AA. Conforme as causas, 51,1% da perda comercial de energia deve-se a fraude e 30,0% a defeito nos equipamentos da medição.

5.2.2.3 - Clientes do grupo de Iluminação Pública (IP)

O consumo anual da iluminação pública na área de concessão corresponde a 85.426 MWh. O nó *Inspeção* tem três estados, um relativo à situação de conformidade (*Normal*) e dois que caracterizam não conformidade e são as principais causas das perdas comerciais:

- *Lâmpadas com operação irregular* – é o balanço da energia cujo faturamento difere da situação real em virtude de existirem lâmpadas acesas durante o dia (perda positiva) e lâmpadas apagadas durante a noite (perda negativa); para este estado a investigação de

campo compila os dados de forma a permitir fazer o balanço avaliando como está tal saldo de perdas, se positivo ou negativo.

- *Erro no cadastro* - a desatualização cadastral existe devido a extensões de rede com novos pontos instalados ou em função da troca da potência da lâmpada, ambos sem que a Prefeitura informe à empresa.

Para iluminação pública a amostra envolve um número mínimo de pontos de IP dentro das quadriculas sorteadas aleatoriamente, sendo o número efetivo de pontos de IP da amostra um pouco maior, valores anteriormente indicados na Tabela 5.6. Feitas as investigações em campo a incidência dos três estados da *Inspeção*, aconteceu conforme mostra a Tabela 5.12, quadro “Pontos da amostra”. A partir dos pontos são calculados os respectivos dados de consumo inclusive os percentuais de não conformidade por região e total, também apresentados na Tabela 5.12, quadro “Energia anual da amostra”.

Tabela 5.12 – Pontos e energia anual da amostra de IP, além da não conformidade por região.

		AA	BB	CC	Total
Pontos da amostra	Quantidade	292	214	174	680
	Lâmpadas com operação irregular	2	4	2	8
	Erro no cadastro	5	2	2	9
Energia anual da amostra (MWh/ano)	Encontrada normal	167,82	119,34	95,81	385,2
	Lâmpadas com operação irregular	1,18	2,30	1,13	4,6
	Erro no cadastro	2,94	1,15	1,13	5,2
Não conformidade (%)		2,40	2,80	2,30	2,50

Com base na proporcionalidade entre consumo da área amostrada e consumo total da Região, bem como nos índices de não conformidade por região e total da empresa, são estimadas as perdas comerciais de energia. Acredita-se que a PC anual atinja 2.136 MWh, dos quais 1.015 devido ao balanço de lâmpadas com operação irregular e 1.121 MWh devido a erro ou não atualização de cadastro. Em valores absolutos a perda de energia anual é maior na região AA (840 MWh), e menor na CC (532 MWh).

Inserindo no programa que trabalha com RB os dados das TPC originárias das inspeções pode-se inferir que a crença da não conformidade total em IP é de 2,50%, dos quais 1,31% devido ao estado *erro no cadastro*, conforme apresenta a Figura 5.5 (a).

Na Figura 5.5 (b) é instanciado o estado *Pontos da região BB*. Na referida região a não conformidade é de 2,80% e as PC somam 764 MWh/ano.

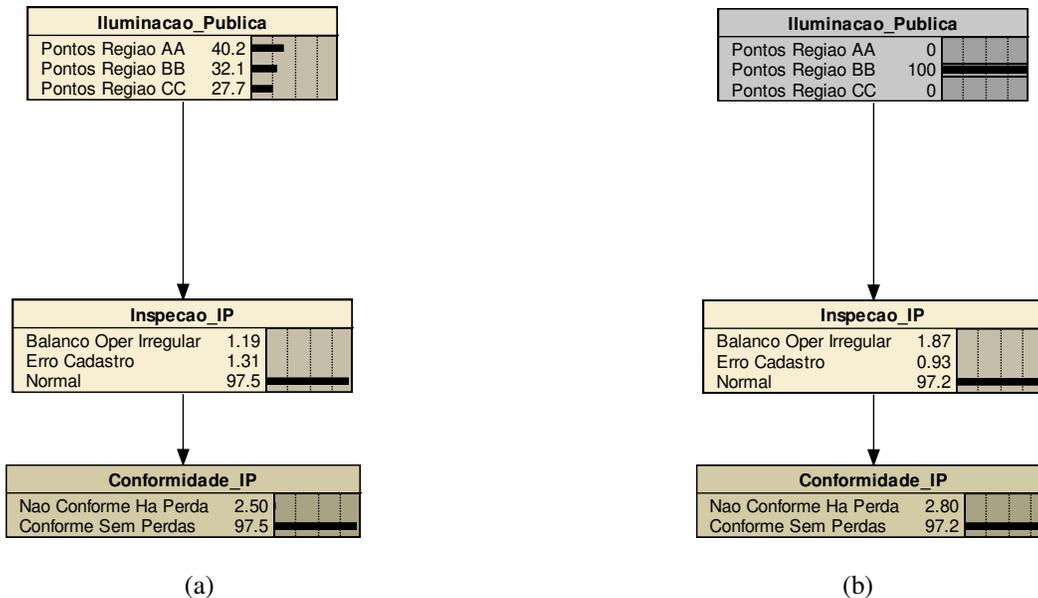


Figura 5.5 – RB do grupo IP, (a) não conformidade 2,50%, (b) instanciado o estado *Pontos da Região BB*.

5.2.3 – Conformidade e estimativa das PC para os grupos de não clientes

5.2.3.1 – Grupo dos Desligados (ex-clientes)

São tratados apenas os ex-clientes supridos em baixa tensão porque se admite que os existentes em alta sejam muito poucos, e que pela importância a empresa sempre realize inspeções anuais de todos de forma que não exista perda dentre ex-clientes em alta tensão. Na Figura 4.5 (c) é mostrada a RB para este grupo e com as informações disponíveis no banco de dados da distribuidora (Tabela 5.2) são preenchidas as tabelas de probabilidade dos nós *Desligados_Classe* e *Região*. Por exemplo, a probabilidade da unidade ser comercial tal que seja do grupo Desligados é

$$P(\text{comercial}|\text{Desligados}) = 9097/116767 = 0,077907$$

Como o grupo estudado é Desligados esta probabilidade condicional é indicada apenas como $P(\text{comercial})$, e corresponde ao elemento da segunda linha da tabela das probabilidades do nó de entrada *Desligados_classe*, que é

$$P(\text{Desligados_classe}) = \begin{bmatrix} 0,889652 \\ 0,077907 \\ 0,005695 \\ 0,009695 \\ 0,001876 \\ 0,014456 \\ 0,000719 \end{bmatrix}$$

A probabilidade de uma unidade da classe comercial deste grupo pertencer à região BB é

$$P(\text{região BB}|\text{comercial, Desligados}) = 3059/9097 = 0,336265$$

Usando a nomenclatura simplificada a probabilidade condicional antes calculada é

$$P(\text{região BB}|\text{comercial, Desligados}) = P(\text{região BB}|\text{comercial}) = 0,336265$$

A partir dos dados da Tabela 5.2 são calculadas todas as probabilidades condicionais para suprir o nó Região, uma matriz 7x3 visto que são 7 classes e 3 regiões. Por exemplo,

$$P(\text{região}|\text{Classe}) = \begin{bmatrix} 0,330885 & 0,341734 & 0,327381 \\ 0,370892 & 0,336265 & 0,292844 \\ 0,369925 & 0,341353 & 0,288722 \\ 0,349823 & 0,321555 & 0,328622 \\ 0,301370 & 0,374429 & 0,324201 \\ 0,348934 & 0,367891 & 0,283175 \\ 0,464286 & 0,321429 & 0,214286 \end{bmatrix}$$

Para o nó *Inspeção*, que é filho de *Desligados_Classe* e *Região*, a TPC é uma matriz 21x5 pois cada linha combina os estados dos nós pais e cada coluna é um estado da variável *Inspeção*. A TPC é determinada com as probabilidades *a posteriori* em razão das inspeções realizadas nas unidades da amostra. A Tabela 5.13 mostra os resultados das inspeções; assim a probabilidade da unidade se encontrar auto-religada tal que seja da região BB e da classe comercial é

$$P(\text{autoreligada}|\text{regiãoBB, comercial}) = 3/23 = 0,130435$$

Não explicitando toda a matriz de probabilidade do nó *Inspeção*, mas apenas a quinta linha

$$\begin{aligned} P(\text{Inspeção}|\text{Classe} = \text{comercial, região} = \text{BB}) &= p(\text{Inspeção}|\text{comercial, regiãoBB}) = \\ &= (0,130435 \quad 0,043478 \quad 0,173913 \quad 0,260870 \quad 0,391304) \end{aligned}$$

A variável *Conformidade* é filha de *Inspeção*, e seus dois estados se relacionam assim: para a unidade auto-religada há certeza de perda ou não conformidade, e para todos os demais estados (demolida, desabitada, regular e desligada) há a certeza da unidade conforme.

Tabela 5.13– Resultados das inspeções do grupo Desligados.

Estados dos pais	Auto-religada	Demolida	Desabitada	Regular	Desligada	Total
Residencial AA	3	4	15	6	21	49
Residencial BB	5	2	10	8	26	51
Residencial CC	5	1	18	5	28	57
Comercial AA	2	1	5	9	7	24
Comercial BB	3	1	4	6	9	23
Comercial CC	1	3	0	6	12	22
Industrial AA	0	0	4	1	4	9
Industrial BB	0	1	2	1	5	9
Industrial CC	1	1	1	0	5	8
Poder Público AA	0	0	10	1	1	12
Poder Público BB	0	0	3	2	6	11
Poder Público CC	0	0	4	2	5	11
Serviço Púb. AA	0	0	0	0	5	5
Serviço Púb. BB	1	0	1	3	1	6
Serviço Púb. CC	0	0	0	1	4	5
Rural AA	1	2	2	4	6	15
Rural BB	1	5	4	1	4	15
Rural CC	1	2	1	2	7	13
Outros AA	0	0	1	1	2	4
Outros BB	0	0	0	2	1	3
Outros CC	0	0	1	0	1	2
Total	24	23	86	61	160	354

Calcular a probabilidade de unidades com perdas desde que estejam na classe comercial, região BB significa aplicar a regra de cadeia para o estado auto-religada, isto é

$$\begin{aligned}
 P(\text{perda} | \text{Inspeção} = \text{autoreligada}, \text{região} = \text{BB}, \text{Classe} = \text{comercial}) &= \\
 &= p(\text{perda} | \text{autoreligada}, \text{regiãoBB}, \text{comercial}) =
 \end{aligned}$$

$$= 1 \times 0,10465 \times 0,336265 \times 0,077907 = 0,003418$$

A quantidade esperada de unidades com perda na referida classe e região é o produto desta probabilidade pelo universo (116.767), ou 399 unidades. Assim são feitas todas as estimativas, chegando-se à crença de que existam 9.524 unidades auto-religadas, das quais 2.425 na região AA, 3.934 em BB. A não conformidade (p) no grupo Desligados é

$$p_{\text{Desligados}} = 100 \frac{9524}{116767} = 8,16\%$$

A Figura 5.6 (a) apresenta a RB do grupo já com todas as TPC dos nós: a não conformidade tem o mesmo valor da probabilidade do estado auto-religado do nó *Inspecção*, pois este é o único que caracteriza perda comercial.

Instancia-se o estado *Região BB* na Figura 5.6 (b), e vê-se que o índice de não conformidade (9,87%) é maior que a média da empresa.

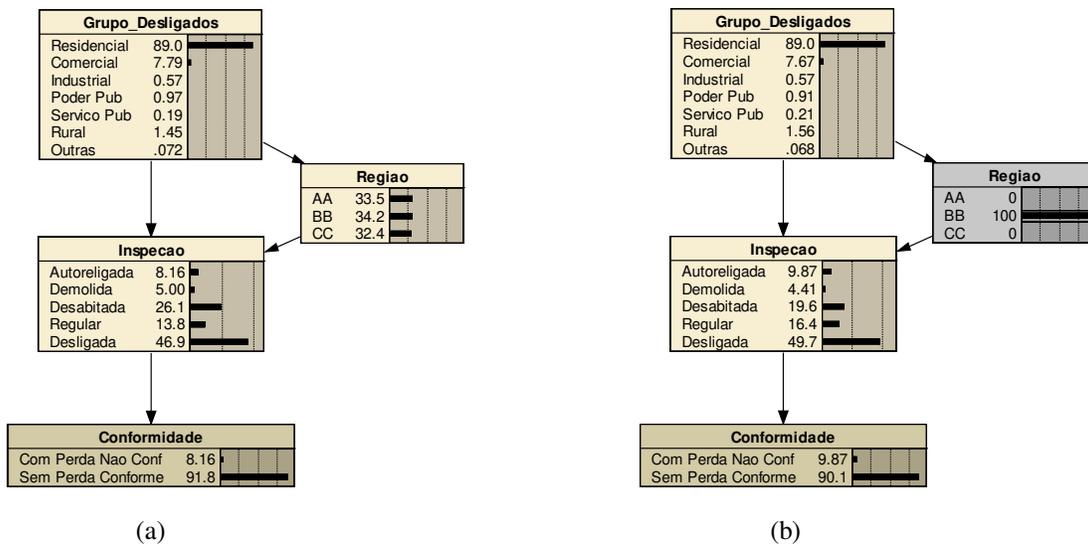


Figura 5.6 – RB relativa aos Desligados, (a) não conformidade 8,16%, e (b) estado *Região BB* instanciado.

A partir da estimativa da quantidade de unidades irregulares é determinada a PC. Foram usadas as mesmas médias de consumo por classe dos consumidores regulares do grupo B, que são

$$\text{Consumos(Desligados_classe)} = \begin{bmatrix} \text{residencial} \\ \text{comercial} \\ \text{industrial} \\ \text{poder público} \\ \text{serviço público} \\ \text{rural} \\ \text{outros} \end{bmatrix} = \begin{bmatrix} 1,425 \\ 21,254 \\ 48,192 \\ 3,780 \\ 19,339 \\ 3,312 \\ 5,796 \end{bmatrix} \quad [\text{MWh/ano}]$$

Portanto a PC dentre as unidades desligadas da classe comercial na região BB é

$$PC(\text{regiãoBB, comercial}) = 399 \times 21,254 = 8.480 \text{ MWh/ano}$$

Estima-se a PC do grupo em 31.046 MWh/ano (13.840 na região BB e 8.102 MWh/ano em CC).

5.2.3.2 – Grupo das Ligações Clandestinas (LC)

Neste grupo foram feitas correlações com as quantidades de unidades residenciais em baixa tensão por região ao dimensionar a amostra, e para que se pudesse falar em não conformidade já que a rigor nas inspeções por quadrículas cada LC encontrada é uma não conformidade. Os dados das inspeções da amostra são mostrados nas duas primeiras linhas da Tabela 5.14, e na terceira linha é calculada a não conformidade percentual. As duas linhas do quadro “estimativas por região” na tabela são relativas às quantidades de LC e às perdas anuais por região.

Tabela 5.14 – Dados das inspeções das amostras e estimativa de LC por região.

		Região AA	Região BB	Região CC	Total
Inspeção da amostra	UC residenciais regulares na área inspecionada	739	727	712	2178
	LC encontradas	50	45	34	129
	Não conformidade (%)	6,77	6,19	4,78	5,92
Estimativas por região	Quantidade de LC	12.289	11.020	8.270	31.579
	PC (MWh/ano)	17.507	15.699	11.782	44.988

A não conformidade (p) para a região AA é

$$P_{\text{perdaAA}} = 100 \frac{50}{739} = 6,77\%$$

A crença com respeito às quantidades de LC por região é uma extrapolação sendo mantidos os índices de não conformidade encontrados, agora referidos ao número de unidades residenciais do grupo B. Já as perdas de energia foram calculadas considerando que uma unidade com LC esteja consumindo a mesma energia média anual que uma unidade residencial regular.

5.2.4 – Validação da hipótese inicial

Após todos os índices de não conformidade terem sido determinados faz-se uma comparação com os valores da hipótese inicial (H_0) para se realizar a validação. As não conformidades e as estimativas de PC por grupo estão na Tabela 5.15. Fez-se o teste de validação qui-quadrado, com

grau de liberdade quatro. O valor do qui-quadrado amostral calculado é de 11,63 que comparado ao limite crítico de 9,49 conduz à rejeição de H_0 . Vê-se que a estimativa total deste diagnóstico é 222.777 MWh/ano, distante das PC totais, logo 40.991 MWh são atribuídos a “outras”.

Tabela 5.15 – Não conformidade e perda por grupo.

	Grupo A	Grupo B	IP	LC	Desligados	Outras
Hipótese inicial (%)	8,0	10,8	3,0	3,8	3,8	---
Não conformidade encontrada	2,76	7,17	2,50	5,92	8,16	---
PC (MWh/ano)	13.828	130.778	2.136	44.988	31.046	40.991

O valor das perdas “outras” pode ser positivo ou negativo correspondendo a estas duas situações:

- Outras a identificar – quando o somatório das PC nos grupos é menor que a PC total, daí “outras” ser positivo. Isto significa que pode haver imprecisão nos cálculos das PT, que há outras perdas a identificar, ou ainda ser devido à margem de erro na metodologia do diagnóstico com a definição das amostras;
- Outras devido a perdas superestimadas – situação na qual o somatório das PC nos grupos é maior que a PC, daí que as PT ou as PC podem ter sido superestimadas, ou mesmo devido à margem de erro do diagnóstico.

A rejeição da hipótese inicial significa que as estimativas iniciais de não conformidade estão distantes das médias das amostras. Estão acima para os grupos A e B enquanto para os grupos Desligados e LC são inferiores. Logo, seguindo a metodologia define-se uma nova hipótese.

5.3 – Diagnóstico sob nova hipótese

5.3.1 Nova hipótese, amostras e validação

Faz-se uma nova hipótese contendo as estimativas de não conformidade dos grupos em valores mais próximos daqueles determinados pós-pesquisa, e objetivando reduzir as perdas outras se adotou uma menor margem de erro. Para a nova hipótese (H'_0) os índices de não conformidade estimados são valores mais próximos daqueles encontrados na hipótese rejeitada, entretanto deve valer a condição de contorno que o somatório dos cinco grupos corresponda ao total de PC.

Assim, os índices de não conformidade esperados nesta nova hipótese são 4,0% para o grupo A, 3,0% para IP, 9,9% no grupo B, 8,0% para Desligados e 6,6% para LC, podendo-se escrever

$$H'_0 = (p_A = 4,0\%, p_B = 9,9\%, p_{IP} = 3,0\%, p_D = 8,0\%, p_{LC} = 6,6\%)$$

Com intervalo de confiança de 95% e erro de 3,0% resultam para as novas amostras aleatórias: 516 inspeções no grupo A, 1.504 para o grupo B e 1.047 para os desligados (349/por região). Face ao menor erro, as quadrículas a vistoriar em IP devem representar um mínimo de 1,1% do número de pontos de iluminação/região, e no grupo de LC as zonas a inspecionar devem conter ao menos 1,1% dos consumidores residenciais em baixa tensão existentes nas regiões.

Após estas novas inspeções são refeitas as TPC da RB e determinados os índices de não conformidade. Os resultados significam uma não conformidade encontrada (NC_E) de

$$NC_E = (p_A = 3,84\%, p_B = 8,14\%, p_{IP} = 2,87\%, p_D = 8,98\%, p_{LC} = 6,39\%)$$

A nova hipótese é aceita, pois na validação o valor do qui-quadrado amostral é de 0,50, portanto inferior ao limite. Logo, o novo diagnóstico é estatisticamente válido. O mesmo está descrito nos itens seguintes e os dados das inspeções das amostras pesquisadas compõem o Anexo B.

5.3.2 Verificação da conformidade e estimativa das PC para os grupos de clientes

A crença é que a não conformidade no grupo A é de 3,84% correspondendo à existência de 66 unidades irregulares, Figura 5.7 (a).

As principais causas da não conformidade são fraude (47,0%), defeito na medição (34,8%) e erro no banco de dados (18,2%). Para a fraude a expectativa é que existam 12 destas irregularidades na região AA, 12 na BB e 7 na CC. A região BB apresentou menor índice de não conformidade (3,57%), e a AA a maior, como mostra a Figura 5.7 (b).

Admitindo para as unidades irregulares os mesmos consumos médios anuais dos consumidores das respectivas classes estimam-se as PC em 21.305 MWh/ano, dos quais 8.285 MWh/ano são devido a defeito na medição e 11.592 MWh/ano em consequência de fraude. A maior concentração das perdas está na região AA (10.392 MWh/ano).

No grupo B as probabilidades do nó de entrada e *Região* são as mesmas que na RB referente à hipótese inicial e agora se altera a TPC do nó *Inspeção*.

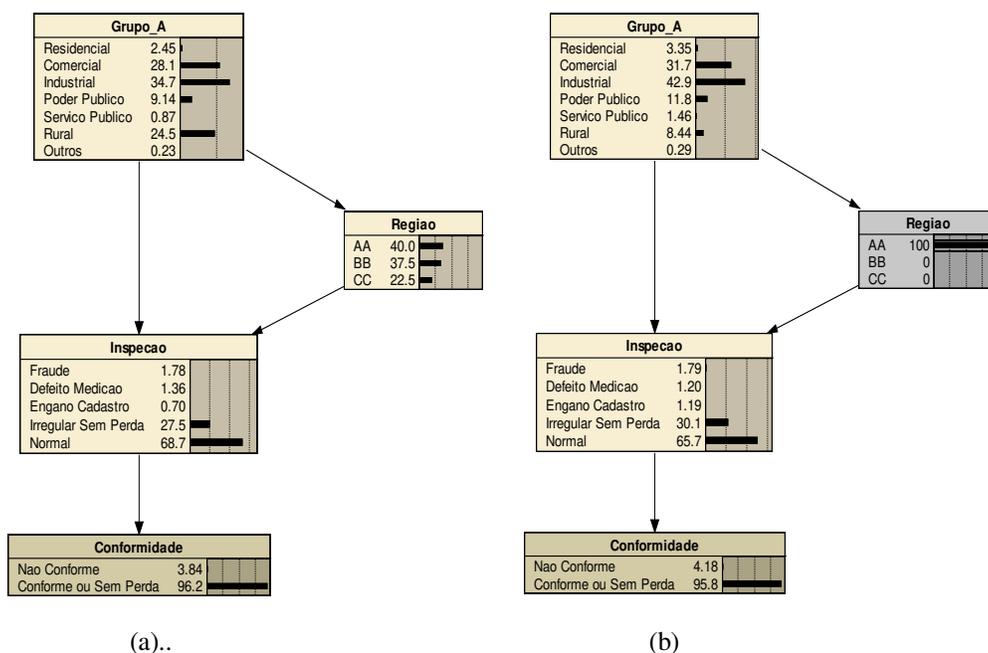


Figura 5.7 – RB do grupo A, (a) não conformidade 3,84%, e (b) estado *Região AA* instanciado.

De acordo com as inspeções calculam-se as probabilidades, sendo esperada não conformidade de 8,14%, ver Figura 5.8 (a). As principais causas são fraude e defeito na medição. Instancia-se o estado *Não conforme* na Figura 5.8 (b) e vê-se como as irregularidades se distribuem por região e classe de consumo. Podem ser feitas outras instanciações de acordo com o que se deseje observar.

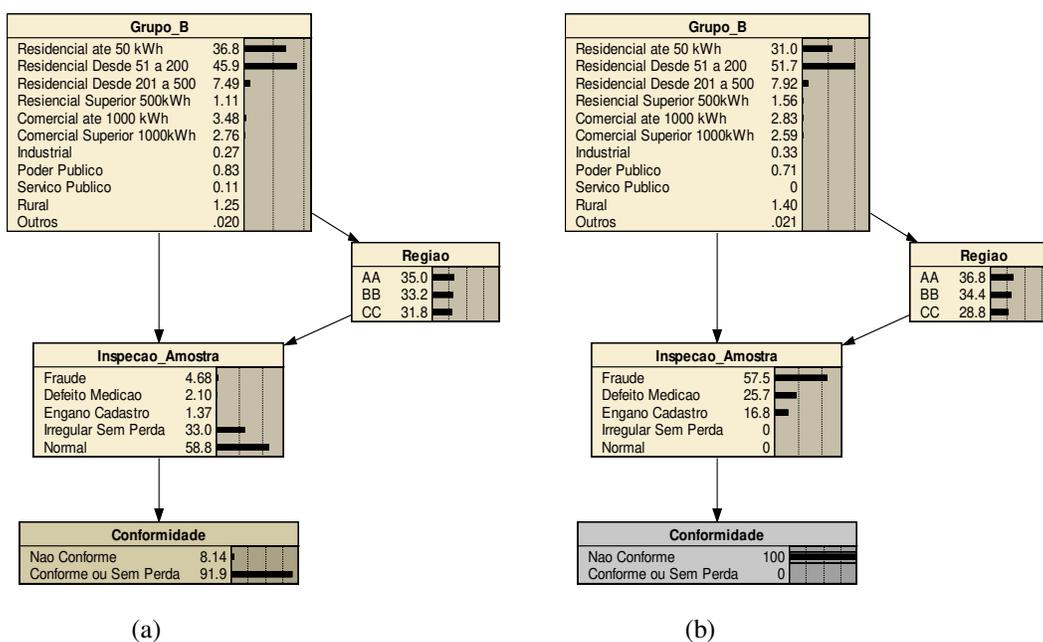


Figura 5.8 – RB do grupo B, (a) não conformidade 8,14%, e (b) instanciação do estado *Não Conforme*.

Crê-se que existam 27.324 unidades do grupo B com fraude (4,68%), sendo 9.428 na região AA.

Estima-se maior número de UC com fraude (13.244) na classe residencial com média de consumo mensal desde 51 a 200 kWh.

A PC calculada para o grupo B é 136.577 MWh/ano, dos quais 65.908 MWh na região AA e 43.994 em BB. Estima-se em 66.945 MWh/ano as perdas devido a fraude, 42.426 por defeito na medição, e 27.207 MWh em função de erros no banco de dados.

Com respeito à IP, feitas as vistorias nas quadrículas da amostra a não conformidade esperada é 2,87%, e a estimativa de PC é de 2.454MWh/ano. Os resultados destas inspeções estão no Anexo B e o procedimento para determinação da não conformidade e estimativas de perda é análogo ao que foi feito na primeira hipótese, descrito em 5.2.2.3. O motivo principal são os pontos de iluminação com engano no banco de dados que causam PC anual de 1.390 MWh.

Por exemplo, na região AA há 59.486 pontos e na área da amostra foram inspecionados 748 pontos que correspondem ao consumo anual de 440,4 MWh. Foram encontrados 4,12 MWh/ano de perda devido a *Balanço de funcionamento irregular* e outros 8,24 MWh/ano causados por *Engano no banco de dados*, isto é devido a 14 pontos não devidamente cadastrados. Logo a expectativa de não conformidade em AA é 2,81%.

Determinou-se a maior não conformidade na região BB (3,11%), entretanto o maior valor de perdas comerciais está em AA (983 MWh/ano).

5.3.3 Verificação da conformidade e estimativa das PC para os grupos de não clientes

Para estes dois grupos, Desligados e LC, o diagnóstico é similar ao descrito nos itens 5.2.3.1 e 5.2.3.2 aplicando-se os dados das novas inspeções que constam do Anexo B.

Dentre as 116.767 unidades desligadas foram inspecionadas 1.047 das quais 89 estavam auto-religadas. Tais resultados são usados na nova TPC do nó *Inspeção*, pois para as demais variáveis da RB as probabilidades são as mesmas. Após a simulação, a não conformidade no grupo Desligados é 8,98%, ver Figura 5.9 (a). Isto corresponde a 10.489 unidades auto-religadas, com maior incidência nas classes residencial (88,2%) e comercial (10,3%), e na região BB (34,2%).

Instanciações podem ser facilmente feitas segundo aquilo que se queira analisar. Na Figura 5.9 (b) instancia-se a Região BB: a não conformidade nesta área é de 9,01%, e há um menor percentual de unidades abandonadas (18,9% contra 22,7% da empresa, Figura 5.9 (a)).

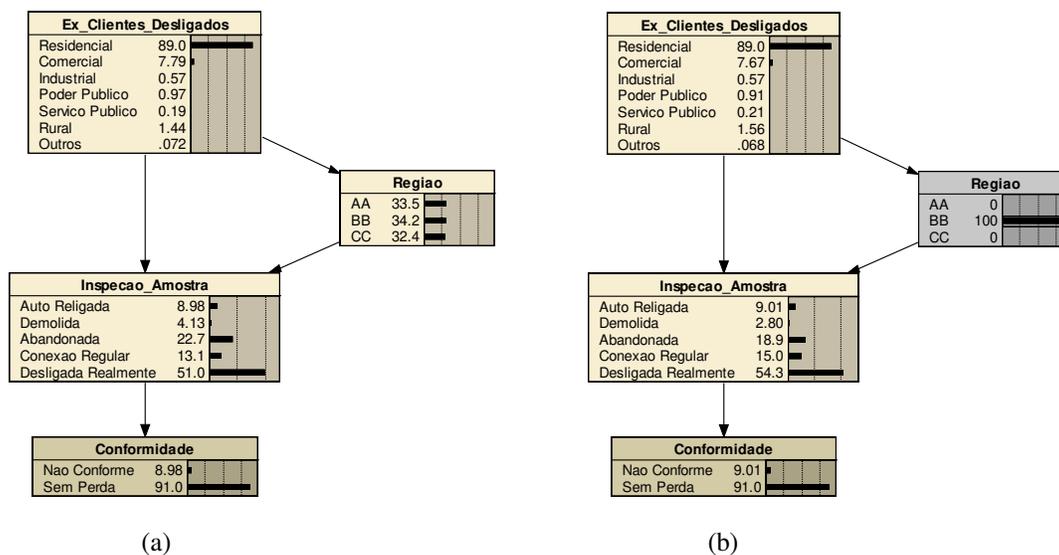


Figura 5.9 – RB do grupo Desligados, (a) não conformidade 8,98%, e (b) instanciação do estado Região BB.

Considerado o consumo médio anual da unidade irregular igual ao das respectivas classes dos clientes do grupo B, estimam-se as PC no grupo em 37.812 MWh/ano (14.822 na região AA).

Para as LC, após as vistorias das quadrículas da amostra e feitas as correlações anteriormente citadas verifica-se uma não conformidade de 6,39%, isto é, o total de LC deve atingir tal percentual em relação aos consumidores residenciais em baixa tensão.

Para a estimativa das perdas de energia considerou-se que o consumo de cada LC é o mesmo que a média do consumidor residencial. Embora se saiba que em geral aqueles conectados irregularmente têm menor renda e consumo, há o consumo excessivo e o aumento das PT devido ao uso de condutores inadequados e emendas da pior qualidade, daí ter sido considerada a mesma média. A PC estimada para as LC é 48.414 MWh/ano (16.918 MWh/ano na região BB).

5.3.4 Estimativa das perdas e inferências

Face aos índices de não conformidade encontrados de 3,84%, 8,14%, 2,87%, 8,98%, e 6,39%, respectivamente para os grupos A, B, IP, Desligados e LC, a nova hipótese é validada, o que foi descrito antecipadamente no item 5.3.1. As estimativas feitas para as PC por região e grupo são

mostradas na Tabela 5.16, na qual a última coluna são valores em percentual referidos ao total das perdas do grupo em relação à energia comprada. Como o total das PC da empresa é 263.768 MWh/ano resulta em 17.205 MWh/ano de perdas “*outras*”, positivas, a identificar.

Tabela 5.16 - Perdas Comerciais (MWh/ano): avaliação por região aplicado o diagnóstico.

Grupos	Região AA	Região BB	Região CC	Total	Total (%)
A	10.392	7.027	3.886	21.305	0,79
B	65.908	43.994	26.675	136.577	5,04
Iluminação Pública	983	848	622	2.454	0,09
Desligados	14.822	12.723	10.267	37.812	1,39
Ligação Clandestina	17.920	16.918	13.576	48.414	1,78
Total	110.025	81.510	55.027	246.562	9,09

A região AA concentra 41,7% da PC e a região BB 30,9%. As maiores perdas comerciais ocorrem nos grupos B e de Ligação Clandestina. A PC em IP é até inferior a “*outras*”.

Tomando-se as perdas globais da empresa (17,3%), inclusive as PT, a repartição se dá conforme ilustrado na Figura 5.10. Vêm-se quão significativas são as PC nos grupos Desligados e LC.

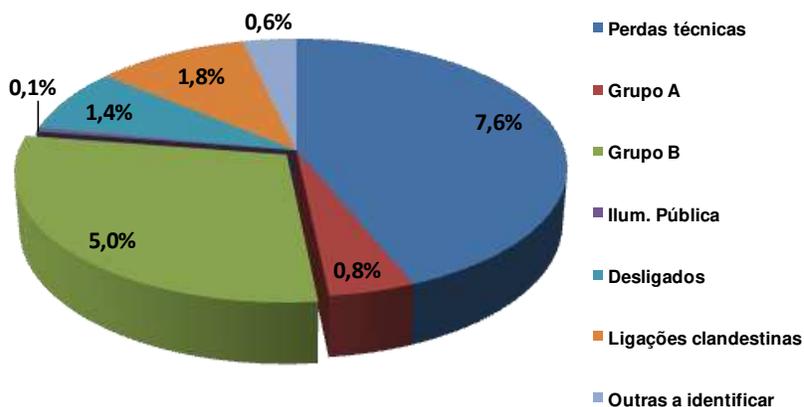


Figura 5.10 – Distribuição da perda global da empresa que é de 17,3%.

Na Tabela 5.16 são mostrados valores médios das estimativas de perdas, mas a margem de erro da pesquisa foi 3%. Esta margem foi relativa a não conformidade, entretanto pode-se colocar, de modo aproximado, que a mesma se mantém no consumo. Assim, por exemplo, no grupo A, a PC

estaria entre 20.666 MWh/ano e 21.944 MWh/ano, com valor médio de 21.305 MWh/ano. Como houve incremento no tamanho da amostra o erro se reduziu e neste caso foi de 2,85%.

As perdas técnicas e comerciais por região estão mostradas na Figuras 5.11.

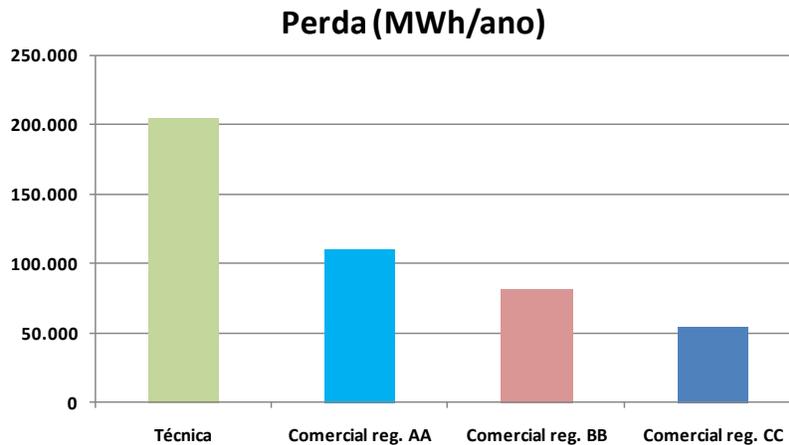


Figura 5.11 – Perdas técnica e comercial por região.

Na Figura 5.12 há a repartição das PC conforme as causas: 25,4% delas são devido a fraude em baixa tensão, as ligações clandestinas representam 18,4%, os defeitos da medição do grupo B 16,1%, e as PC dentre as unidades desligadas 14,3% do total.

Destaca-se que esta metodologia de diagnóstico usando redes Bayesianas possibilita obter muitas informações, e as instanciações são fáceis e rápidas.

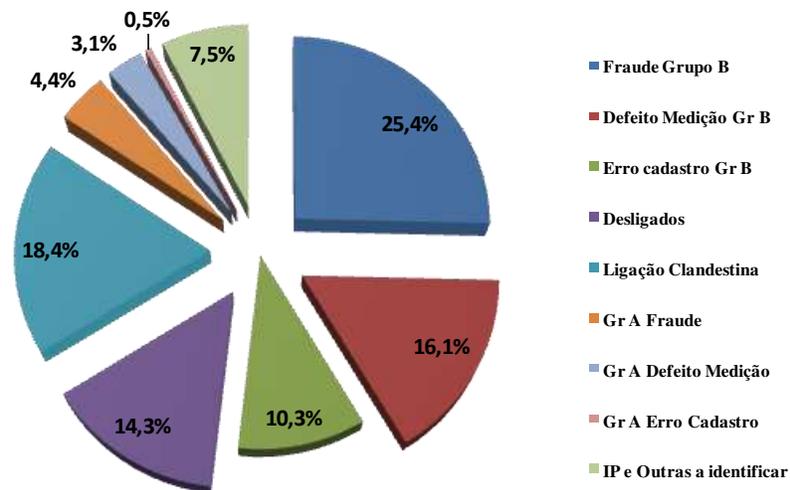


Figura 5.12 – Perda comercial conforme as causas.

Desejando-se inferir outras avaliações oportunas para a elaboração de estratégias de combate às PC, como por exemplo, se a perda no Grupo B em determinada região se deve mais a fraude ou a enganos no cadastro e erros da medição, devem ser realizadas instâncias.

A seguir alguns exemplos de conclusões do diagnóstico a partir de instâncias:

a) No grupo B

- as classes industrial e comercial respondem por 6,5% das UC mas são responsáveis por 47,8% da energia perdida no grupo ou 24,8% da perda comercial;
- as UC residenciais com média de consumo acima de 200 kWh/mês correspondem a 8,6% das unidades do grupo, mas detêm 18,2% das perdas do grupo;
- dos 66.945 MWh perdidos por fraude no grupo B a região AA tem uma participação de 42,9% e a região BB de 33,0%;
- a região AA concentra 50,5% das perdas devido a defeito na medição, e 57,8% devido a erro em cadastro.

b) No grupo Desligados

- do total de perda comercial de energia da empresa as unidades auto-religadas representam 14,3%;
- dentre os ex-clientes irregulares 39,2% devem estar na região AA;

c) No grupo A

- Este grupo é responsável por 8,1% da perda comercial da distribuidora.

d) Nos grupos de IP e ligação clandestina

- A perda em IP é pequena, mas as ligações clandestinas significam 18,4% da perda comercial de energia, estando mais concentradas nas regiões AA e BB.

5.4 – Análise econômica das ações e definição de um programa eficiente para redução das perdas comerciais

Finalizadas as etapas anteriores devem ser definidas ações que irão compor o programa de redução das PC. As ações podem ser diferenciadas regionalmente mas é essencial que sejam calculadas suas relações custo/benefício de modo a definir aquelas mais eficientes, otimizando recursos financeiros e humanos.

No item anterior foram resumidas inferências entretanto outras ainda foram feitas obtendo-se estimativas por classe de consumo e região. Por exemplo, estima-se que dentre as 16.133 UC comerciais do grupo B com média mensal acima de 1.000 kWh existam 1.233 com perdas totalizando 47.690 MWh/ano, e dentre as 267.745 UC residenciais com média mensal entre 51 e 200 kWh crê-se que há 24.549 irregulares correspondendo a uma perda de 37.077 MWh/ano.

Dentre as possíveis ações para um biênio destacam-se:

- Inspeccionar todas as UC industriais do grupo B (1.562). Devem existir 159 unidades não conformes nas quais a PC é 7.656 MWh/ano;
- Inspeccionar todas as UC comerciais do grupo B (36.445); o retorno é 57.667 MWh/ano;
- Ainda no grupo B, inspeccionar todas as UC residenciais com consumo médio mensal acima de 500 kWh (6.482), devendo-se encontrar 742 irregulares e recuperar 6.472 MWh/ano;
- Vistoriar todas as 116 mil unidades desligadas, pois a crença é que 8,98% estejam auto-religadas, o que significa recuperar 37.812 MWh/ano;
- Inspeccionar todos os 1.718 clientes supridos em alta tensão, sendo esperado encontrar 66 unidades irregulares e um retorno de 21.305 MWh/ano;
- Efetuar campanhas de retirada das LC buscando apoio da polícia e órgãos judiciais para que haja penalidade a este tipo de roubo;
- Empenhar-se junto aos órgãos de governo para prover a legalização ou retirada em definitivo das habitações irregulares para que se normalizem as ligações de energia;
- Examinar as áreas com LC e analisar a viabilidade de colocação de rede anti-furto nas mesmas;
- Reavaliar as inspeções que resultaram em defeito na medição especialmente nas unidades do grupo B e na região AA. Confirmado o diagnóstico identificar as causas efetivas;
- Efetuar capacitação dos empregados envolvidos com o cadastro das unidades salientando a importância da atualização e de que não haja engano quando da implantação de novos clientes no sistema. Isto poderia se iniciar com a região AA para depois ser estendido às demais regiões;
- Acompanhar mais de perto as não conformidades dos grupos A e B por motivos internos, desenvolvendo uma auditoria no processo de cadastramento das unidades no banco de

dados de faturamento e no dimensionamento dos equipamentos de medição. Isto auxiliaria na decisão por ações mais apropriadas.

Como forma de priorizar algumas destas ações são calculadas as suas relações custo/benefício, para o que se considerou: custo médio de venda R\$254,71/MWh em janeiro de 2009 (ANEEL,2010), juros de 12% ao ano, composição do custo da equipe de inspeção do grupo B que é de R\$154,13 por inspeção. Para se chegar a este custo admitiu-se a turma composta por um eletricista e um técnico com salários de R\$1.500,00 e R\$3.000,00 respectivamente, aluguel de veículo a R\$1.800,00/mês, que são percorridos 3 mil km/mês, que o investimento nos equipamentos da turma totaliza R\$20mil com vida útil de três anos, e como média a execução de 4 inspeções/dia, o que resulta em 968 inspeções anuais por equipe (22 dias úteis mensais em 11 meses). Deste modo os custos anuais da turma são:

▪ Pessoal (encargos de 100%)	R\$108.000,00
▪ Veículo, aluguel	R\$ 21.600,00
▪ Equipamentos, custo anual equivalente	R\$ 8.326,98
▪ Combustível (R\$2,66 o litro e média 8,5km/l)	R\$ 11.265,88
▪ Total anual, valor presente	R\$149.192,86

Considerou-se ainda que a inspeção de uma unidade consumidora rural custa 30% mais que as demais do grupo B, e a inspeção da unidade do grupo A custa 50% mais. A vistoria em uma unidade desligada custa R\$51,33.

Na apuração da relação custo/benefício, o custo é o valor presente total da ação. Por exemplo, para a ação “inspecionar todas as unidades do grupo A” é o produto do número de unidades a vistoriar pelo custo médio unitário da inspeção. O benefício é o valor presente da energia recuperada nos três anos seguintes à implementação da ação, sem considerar qualquer cobrança retroativa (cabível nos casos de fraude). Não é recomendável um horizonte superior a cinco anos na avaliação dos benefícios face à dinâmica que envolve a distribuição de energia elétrica.

Deste modo foram calculadas relações entre custo e benefício para algumas ações, como mostrado na Tabela 5.17, que apresenta ainda o total de unidades envolvidas, o percentual de UC com perdas e energia anual recuperada.

A distribuidora pode definir um plano de redução das PC priorizando as ações conforme a relação custo/benefício. As ações de treinamento de profissionais da ligação e a colocação de rede anti-

furto em zonas específicas não foram analisadas economicamente visto que para tais ações a empresa pode melhor avaliar visto que conhece melhor seus custos de treinamento e de rede.

Tabela 5.17 – Relações custo/benefício e outras estimativas para algumas ações de inspeção.

Ações de inspeção	Total de UC	UC com perdas (%)	Energia recuperada (MWh/ano)	Relação custo/benefício
Todas as UC do Grupo A	1.718	3,8	21.305	0,022
Desligados não residenciais	12.885	9,6	24.624	0,039
Industriais do Grupo B	1.562	10,2	7.656	0,046
Comerciais > 1 mil kWh/mês Grupo B	16.133	7,6	47.691	0,076
Residenciais > 500 kWh/mês Grupo B	6.482	11,4	6.472	0,225
Comerciais < 1 mil kWh/mês Grupo B	20.312	6,6	9.976	0,458
Residenciais Gr. B consumo de 201 a 500 kWh/mês	43.722	8,6	18.345	0,536
Desligados residenciais	103.882	8,9	13.188	0,590
Poder público do Grupo B	4.873	6,9	1.270	0,863
Rurais do Grupo B	7.282	9,1	2.200	0,968
Residenciados Gr. B consumo de 51 a 200 kWh/mês	267.745	9,2	37.077	1,624

No caso específico deste estudo todas as ações viáveis contidas na Tabela 5.17 poderiam compor um plano de redução das perdas comerciais para dois anos. Isto significaria a inspeção de 1.718 UC do grupo A, 100.366 UC do grupo B e a vistoria de 116.767 ex-clientes desligados, e com tal plano a perda comercial seria reduzida em 131.423 MWh/ano ou 57,9% da PC atual.

Deste modo, mantidas as condições de mercado e a perda técnica, após o terceiro ano (as ações se refletem por completo um ano depois) esta distribuidora não mais apresentaria 17,3% de perda global e sim 12,3%, e nesta nova situação as perdas técnicas representariam 8,0% e as não técnicas 4,3%. O índice de situação de perdas da empresa cairia de 128% para 54,1%. Este índice

poderia ser até menor pois recalculando as PT estas devem crescer já que a energia vendida aumentaria.

A ação não viável da Tabela 5.17 diz respeito às unidades residenciais do grupo B com média de consumo mensal entre 51 e 200 kWh. Tanto para estes como para os residenciais da faixa de consumo médio inferior a 50 kWh/mês, poderiam ser feitas novas pesquisas por amostragem, especialmente nas regiões AA e BB seguindo o exemplo sugerido na Figura 5.13 e com isto posteriormente poderiam ser tratados diferentemente municípios, tipos de ligação e se a mesma atende casa ou apartamento.

Não só para esta faixa como especialmente para os consumidores comerciais, industriais e rurais é recomendável também o acompanhamento do histórico de consumo, do faturamento mínimo, dentre outros, de modo a melhor direcionar futuras inspeções. É recomendável avaliar o uso de ferramentas da inteligência artificial na seleção de unidades a inspecionar com base no acompanhamento dos dados de leitura e faturamento, visando aumentar o índice de acerto na identificação de irregularidades. DURÃES et al (2006) descrevem algumas destas ferramentas.

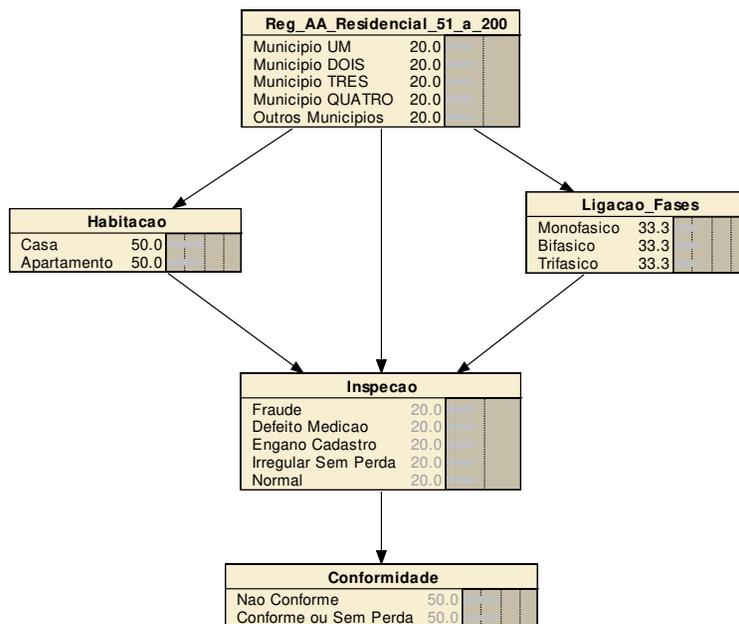


Figura 5.13 – Modelo de RB para investigação em uma faixa de consumo específica das UC residenciais.

Destaca-se que como o diagnóstico e o plano de ação para redução são dinâmicos, outras rotinas seriam implementadas, podendo-se alterar os rumos em função de fatos novos. O universo é dinâmico, há novas ligações, novos clientes desligados, novos tipos de fraude podem surgir,

novos empregados na distribuidora, enfim estas e outras ações, caracterizadas por uma presença constante da empresa, contribuirão para redução do índice das perdas comerciais.

Concluindo-se deve também ser estimado o custo de um diagnóstico como este proposto. As quantidades de inspeções a realizar nos grupos A, B e Desligados foram respectivamente 516, 1.504 e 1.047 unidades.

Os custos estimados por tipo de inspeção estão colocados anteriormente neste item, inclusive o acréscimo para unidades rurais. Considerando os custos das investigações nos grupos de Ligações Clandestinas e IP juntos o mesmo que o das inspeções dos desligados significa que o dispêndio com as inspeções totalizaria cerca de R\$ 461 mil. Assumindo os custos de engenharia do projeto da ordem de R\$308 mil, o valor total da aplicação do diagnóstico em uma empresa deste tipo, cerca de 580 mil consumidores, atinge R\$769.000,00 (setecentos e sessenta e nove mil reais). Ressalte-se que este valor não deve se alterar muito para empresas de maior porte já que as amostras a averiguar não crescem nas mesmas proporções dos clientes. O prazo para execução de todo o diagnóstico é aproximadamente cinco meses, desde que a distribuidora realmente se empenhe na realização das inspeções em campo.

A empresa distribuidora pode realizar este diagnóstico de perdas comerciais apenas em uma das suas regionais desde que conheça o total de PC da referida área. Como a metodologia é flexível, para esta empresa usada no estudo de caso, na hipótese de se conhecer *a priori* como se reparte o total das perdas comerciais entre as suas três regiões (o que corresponderia a haver medições totalizadoras nas regionais), o diagnóstico poderia ser feito inicialmente apenas em uma das regiões, a AA por exemplo. Os modelos de RB a usar seriam semelhantes aos anteriormente empregados no diagnóstico da empresa, e a Figura 5.14 apresenta como seria a rede para o grupo Desligados (nó do topo é Desligados_Reg_AA). Neste caso deveria ser adequada a variável regional; na mencionada figura tal variável é denominada *Zonas_Região_AA*, cujos estados seriam cinco: *Reg. Metropolitana_KK*, *Município_XX*, *Município_YY*, *Município_ZZ* e *Demais_Localidades*. Altera-se obviamente a dimensão da matriz de probabilidades da variável *Inspeção_Amostra* que agora teria dimensão 35x5.

Deste modo as quantidades de inspeções a realizar seriam menores (504 no grupo B, 172 no grupo A e 349 nos Desligados, além de menor área a pesquisar para IP e LC), daí que os custos

seriam da ordem de R\$390.000,00 (trezentos e noventa mil reais), podendo o prazo de execução se reduzir a quatro meses.

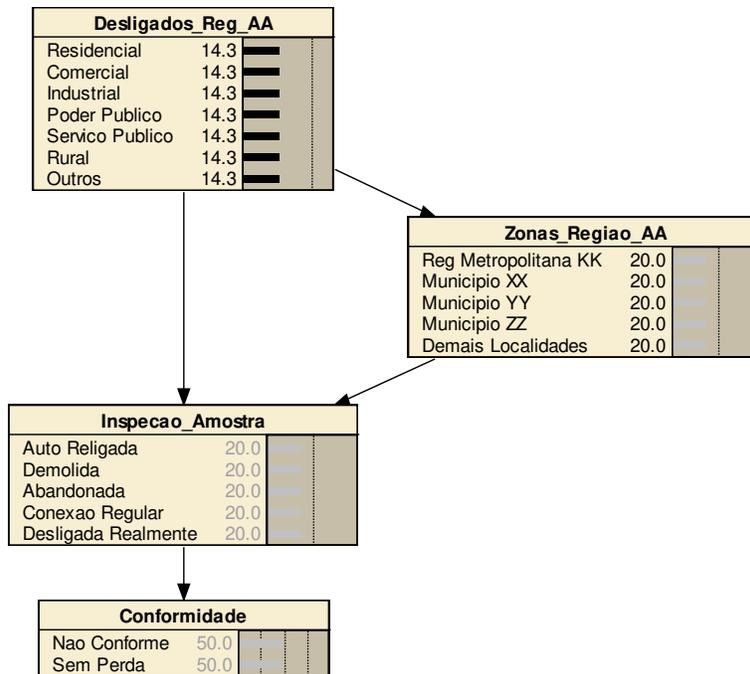


Figura 5.14 – Modelo de RB para diagnóstico das PC na Região AA, grupo *Desligados*.

6. CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresenta uma pesquisa das perdas comerciais em distribuidoras de energia elétrica objetivando elaborar uma metodologia para diagnóstico das mesmas usando redes Bayesianas e investigações em campo referentes a conformidade.

Foram realizados estudos das perdas técnicas e comerciais, revisão bibliográfica das PC, e nos Capítulos Quatro e Cinco foram apresentadas a metodologia proposta para o diagnóstico e o estudo de caso, respectivamente.

Este capítulo traz no primeiro item as principais conclusões, no segundo os trabalhos que podem ser continuados e ao final as publicações feitas durante a pesquisa.

6.1 – Conclusões

A finalidade do diagnóstico é obter uma prospecção global das perdas comerciais, possibilitando que ações para redução das PC possam ser avaliadas economicamente. Com os resultados apresentados no Capítulo Cinco para o estudo de caso completo, puderam-se verificar as relações benefício/custo para as várias ações propostas. Isto permite a elaboração de um plano de redução das perdas comerciais consistente e eficiente.

Estas são as principais conclusões:

- A metodologia proposta para o diagnóstico das perdas comerciais utiliza as redes Bayesianas e análise de conformidade. Esta metodologia permite estimar os quantitativos referentes às principais causas e segundo a distribuição regional dentro as áreas da empresa distribuidora;
- O modelo de rede Bayesiana por Grupos usado na investigação das perdas comerciais entre os clientes e entre os não clientes, bem como as sub-redes empregadas em cada um dos Grupos, mostraram-se adequados face aos resultados obtidos;

- As probabilidades condicionais *a posteriori* necessárias para suprir a RB podem ser obtidas de inspeções em campo realizadas em amostras aleatórias, calcadas na investigação da conformidade;
- Um modelo similar ao proposto no capítulo quatro foi adaptado e testado satisfatoriamente com dados reais oriundos de inspeções realizadas pela COELBA no Grupo B (baixa tensão), na área da Região Metropolitana de Salvador, servindo para demonstrar também a sua flexibilidade;
- A aplicação completa de um diagnóstico empregando dados contidos no Relatório do CODI e alguns outros estimados, mostrou ser possível identificar problemas específicos e quantificar as perdas, permitindo a avaliação econômica de diversas das ações para combate às perdas comerciais. Conseqüentemente pode-se elaborar um eficiente programa de redução destas perdas.
- O indicador proposto relacionando perdas técnicas e comerciais permite comparar as distribuidoras. Sua utilização possibilitaria ao órgão regulador avaliar as empresas quanto ao controle do processo de perdas e quanto às exigências de diagnósticos e programas de ação de redução distintos conforme a situação das mesmas.

Face aos resultados obtidos conseguiu-se demonstrar ter alcançado o objetivo maior qual seja o estabelecimento de uma metodologia para realização do diagnóstico das perdas comerciais.

A apresentação de dados mundiais referentes ao aumento das perdas no setor elétrico, a quantificação do problema no Brasil cujo valor relativo a tais perdas pode alcançar 7,9 bilhões de reais ao ano (ANEEL, Tarifas Médias, 2010), além da exemplificação de situações similares experimentadas por outros países em desenvolvimento como Malásia, Chile, Colômbia, México, Turquia, Índia, Paquistão, Tailândia, Filipinas, Argentina, Peru, Venezuela e África do Sul, são sempre relevantes. A colocação deste tema em congressos e seminários certamente contribui para motivar a área acadêmica, especialmente do Norte e Nordeste do país, para o estudo das perdas comerciais. Isto pode ser considerado um objetivo secundário alcançado.

Embora as perdas técnicas não tenham sido o foco principal do trabalho, em função das pesquisas realizadas concluiu-se que os métodos probabilísticos que as quantifiquem por intervalo ou pela média e desvio padrão são mais adequados aos cálculos por segmento. Isto se justifica porque há muita incerteza nos parâmetros e a diversidade é maior quanto mais se está próximo do

consumidor final. Verificou-se também que há aspectos em geral não abordados nas metodologias usadas, os quais podem contribuir para o aumento das perdas técnicas como o uso do intervalo de integração da demanda de 15 minutos, a não consideração do envelhecimento dos medidores eletromecânicos, as elevadas correntes de partida destes medidores face às pequenas cargas hoje utilizadas, além das maiores perdas nos transformadores de distribuição recuperados em oficinas de reparo.

6.2 – Trabalhos futuros

Em função dos limites da pesquisa, principalmente no que diz respeito à completa realização do diagnóstico em uma distribuidora com dados mais atuais que aqueles do estudo de caso usado no Capítulo Cinco e devido ao foco nas perdas comerciais algumas investigações devem prosseguir. Vislumbram-se ainda outras questões que merecem aprofundamento, podendo conduzir a estes trabalhos futuros:

- A aplicação completa da metodologia em uma distribuidora deve ser concluída bem como outras devem ser buscadas. Até mesmo o custo da metodologia por enquanto é estimado, podendo-se melhor avaliar com a efetiva aplicação.
- Não houve tempo suficiente para que fosse bem caracterizado o consumo de energia elétrica em unidades irregulares nem em consumidores carentes. Isto pode ser objeto de pesquisa específica buscando-se determinar médias de consumo anual das unidades com perdas tanto devido à origem interna quanto externa e por classe de consumidor, inclusive estendendo-se a estudos tarifários e relativos a subsídios atualmente existentes para camadas mais carentes da população.
- Aqui não foram explorados quais fatores conduzem ao furto e à fraude de energia, examinando-se causas como renda, tarifa, inadimplência, políticas habitacionais, condições habitacionais, presença do governo, além das medidas punitivas e ações da empresa e da polícia no combate às contravenções. É o tipo de estudo que requererá a incorporação de profissionais de outras áreas como as ciências sociais, face à dimensão do problema. Entretanto crê-se que seria possível desenvolver modelos de rede Bayesiana para análise de risco quanto ao roubo e ao furto de energia.

- Procurar avaliar economicamente benefícios que a princípio parecem intangíveis a exemplo daqueles decorrentes das retiradas de ligações clandestinas, de ações moralizadoras como incentivo à denúncia, ou mesmo as inserções na mídia divulgando o combate das irregularidades ou o programa de redução de perdas da empresa.

6.3 – Publicações

Durante o período desta pesquisa foram estudadas as perdas técnicas, comerciais e as redes Bayesianas. A Tabela 6.1 apresenta as publicações decorrentes. Foi feita a submissão de artigo para divulgação em revista, estando em análise.

Tabela 6.1 – Artigos publicados em seminários e congressos.

Título	Evento	Local e data
Perdas técnicas adicionais nos segmentos rede secundária, ramais e medidores	SBSE – Seminário Brasileiro de sistemas elétricos	Belo Horizonte, abril de 2008
Perdas Técnicas: aspectos geralmente não abordados e sugestões à Nota Técnica 035 da ANEEL	XVIII SENDI - Seminário Nacional de Distribuição de Energia Elétrica	Recife, outubro de 2008
Proposta de uma metodologia simplificada para cálculo das perdas técnicas	XVIII SENDI - Seminário Nacional de Distribuição de Energia Elétrica	Recife, outubro de 2008
Uso de rede Bayesiana na identificação das perdas não técnicas	XIII ERIAC - Encuentro Regional Iberoamericano de CIGRÉ	Puerto Iguazu, Argentina, maio de 2009
Energy losses: aspects not considered and the use of probabilistic methods	20 th CIRED - International Conference on Electricity Distribution	Praga, Rep. Checa, junho de 2009
Diagnosis of nontechnical energy losses using Bayesian networks	20 th CIRED - International Conference on Electricity Distribution	Praga, Rep. Checa, junho de 2009
Proposição de um modelo para quantificar as perdas comerciais	XX SNPTEE – Seminário Nacional de Produção e Transmissão de Energia Elétrica	Recife, novembro de 2009

Título	Evento	Local e data
Identificação das perdas comerciais através de rede Bayesiana	XX SNPTEE – Seminário Nacional de Produção e Transmissão de Energia Elétrica	Recife, novembro de 2009
Prospecção das perdas não técnicas em distribuição: estimativa por região e origem	XVIII CBA – Congresso Brasileiro de Automática	Bonito, setembro de 2010
Diagnóstico das perdas não técnicas empregando redes Bayesianas	CIDEL - Congreso Internacional de Distribución Eléctrica	Buenos Aires, Argentina, setembro de 2010

ANEXO A

APLICAÇÃO DA METODOLOGIA NA ESTIMATIVA DAS PERDAS COMERCIAIS DOS CLIENTES DO GRUPO B DE UMA CONCESSIONÁRIA

No item 4.2.3 colocou-se de modo resumido como foram usados os dados de uma concessionária, no caso a COELBA, para aplicação da metodologia como estudo de caso em empresa.

Como dito tais dados são públicos, referem-se ao grupo B da Região Metropolitana de Salvador, e fazem parte da dissertação de mestrado de DANTAS (2006). Alguns dados estão no item 4.2.3 e a rede Bayesiana foi adaptada, por não se dispor de todos os dados como desejado.

Como o trabalho original usou um número maior de estados na variável inspeção, a correspondência necessária está mostrada na Tabela 4.5. Também as figuras do modelo de RB usado, e da rede após a simulação estão apresentados no item 4.2.3.

Aqui se faz uma complementação com a apresentação de algumas tabelas de probabilidade e figuras da RB, para elucidar a aplicação, mostrando inclusive outra forma de obter as TPC, agora a partir das tabelas de probabilidades conjuntas e usando o conceito de marginalização.

A Tabela A.1, apresenta o número de unidades por classe, a probabilidade da UC pertencer a uma classe, e da amostra inspecionada em campo traz o total e o resumo das ocorrências.

Tabela A.1 - Consumidores por classe, probabilidade e dados da amostra inspecionada (DANTAS, 2006).

Estados Da “Classe”	Quantidade de UC	Probabilidade P(classe)	Total da amostra	Ocorrências nas inspeções (diagnóstico)			
				Fraude	Defeito	Irregular sem Perda	Normal
Residencial	749.957	0,908270	201	18	12	169	2
Comercial	70.567	0,085463	86	5	1	73	7
Industrial	3.190	0,003863	33	1	0	28	4
Poder Público	1.819	0,002203	30	0	1	26	3
Serv.Público	113	0,000137	12	0	0	8	4
Rural	22	0,000027	7	0	0	7	0
Consumo Próprio e Outros	30	0,000036	5	0	0	3	2
Total	825.698	1,0000000	374	24	14	314	22

Com os dados da Tabela A.1 calcula-se a probabilidade da UC pertencer às várias classes. Por exemplo, a probabilidade para a unidade comercial é

$$P(\text{comercial}) = \frac{70567}{825698} = 0,085463$$

Caso seja usado um dos softwares comerciais os dados da variável *Classe_Grupo_B* são aquelas probabilidades indicadas na Tabela A.1.

Para a variável *Inspeção* deve ser buscada a TPC, aqui determinada usando a tabela de probabilidade conjunta.

A Tabela A.2 mostra os dados das probabilidades conjuntas dos estados das variáveis *Inspeções* e *Classe Grupo B*, que são calculados a partir de uma tabela semelhante referente aos resultados das inspeções, cujo original está em DANTAS (2006), agrupados conforme Tabela A.1.

Para facilitar, a última linha da Tabela A.2 corresponde ao somatório da coluna que, conforme expressão (3.11) é o fator de marginalização significando a probabilidade de pertencer a um determinado estado dentro da *Classe Grupo B*, mas para as inspeções realizadas.

Assim, por exemplo, o fator de marginalização (FM) para a classe comercial é

$$FM(\text{Comercial}) = \sum_{\text{Inspeções } i}^{10} P(\text{Comercial}, \text{Inspeções}_i) = 0,22995$$

Com os dados da tabela de probabilidade conjunta da inspeção (Tabela A.2) e aplicando-se o fator de marginalização podem ser realizados em uma planilha os cálculos das probabilidades condicionais e com isto monta-se toda a TPC necessária para o nó *Inspeção*. Por exemplo, a probabilidade de uma UC ter ligação direta tal que pertença à classe comercial é

$$P(\text{LigaçãoDireta}|\text{Comercial}) = \frac{P(\text{ligaçãoDireta}, \text{Comercial})}{FM(\text{Comercial})} = \frac{0,00267}{0,22995} = 0,011628$$

Ou percentualmente 1,16%.

A Tabela A.3 apresenta os dados (TPC) na forma requerida pelo software Netica® (NORSYS, 2008) relativa ao nó *Inspeções*, cujo pai é *Classes Grupo B*. Em relação ao modo de entrada dos dados realmente no software, foram trocadas linhas com colunas por uma questão de facilitar a apresentação da tabela.

Tabela A.2 – Probabilidades conjuntas entre estados das variáveis *Inspeções* e *Classe_Grupo_B*.

	Estado da <i>Inspeção</i>	Classes_Grupo_B						
		Resi- dencial	Comer- cial	Indus- trial	Poder Público	Serv. Públ.	Rural	Outros
Exter- no Fraude	Ponte no Bloco	0,01070	0,00267	0,00267	0,00000	0,00000	0,00000	0,00000
	Ligação Direta	0,00000	0,00267	0,00000	0,00000	0,00000	0,00000	0,00000
	Interrup. Bob. Pot.	0,00267	0,00267	0,00000	0,00000	0,00000	0,00000	0,00000
	Desvio Antes medidor	0,01337	0,00267	0,00000	0,00000	0,00000	0,00000	0,00000
	Desvio Embutido	0,01070	0,00267	0,00000	0,00000	0,00000	0,00000	0,00000
	Medidor Avariado Sem selo	0,01070	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Interno Defei- to empresa	Medidor Defeituoso	0,02941	0,00267	0,00000	0,00267	0,00000	0,00000	0,00000
	Medidor Danificado	0,00267	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Sem Perda	Irregular Sem perda	0,45187	0,19519	0,07487	0,06952	0,02139	0,01872	0,00802
	Normal	0,00535	0,01872	0,01070	0,00802	0,01070	0,00000	0,00535
Fator marginalização		0,53743	0,22995	0,08824	0,08021	0,03209	0,01872	0,01337

Do mesmo modo, a probabilidade de uma unidade residencial ter sido encontrada com irregularidade, mas sem perda é determinada como

$$P(\text{IrregularSemPerda}|\text{Residencial}) = \frac{0,45187}{0,53743} = 0,840798$$

Ou percentualmente 84,08%.

Tabela A.3 – Probabilidades condicionais da variável *Inspecções*, conforme fornecido ao software Netica®.

	Estado da <i>Inspecção</i>	Classes_Grupo_B						
		Resi- dencial	Comer- Cial	Indus- trial	Poder Público	Serv. Públ.	Rural	Outros
Exter- no Fraude	Ponte no Bloco	0,01990	0,01163	0,03030	0,00000	0,00000	0,00000	0,00000
	Ligação Direta	0,00000	0,01163	0,00000	0,00000	0,00000	0,00000	0,00000
	Interrup. Bob. Pot.	0,00498	0,01163	0,00000	0,00000	0,00000	0,00000	0,00000
	Desvio Antes medidor	0,02488	0,01163	0,00000	0,00000	0,00000	0,00000	0,00000
	Desvio Embutido	0,01990	0,01163	0,00000	0,00000	0,00000	0,00000	0,00000
	Medidor Avariado Sem selo	0,01990	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Interno Defei- to empresa	Medidor Defeituoso	0,05473	0,01163	0,00000	0,03333	0,00000	0,00000	0,00000
	Medidor Danificado	0,00498	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Sem Perda	Irregular Sem perda	0,84080	0,84884	0,84848	0,86667	0,66667	1,00000	0,60000
	Normal	0,00995	0,08140	0,12121	0,10000	0,33333	0,00000	0,40000

Depois de inserir todos os dados na RB, feita a simulação o resultado é mostrado na Figura 4.7. Acredita-se que a não conformidade é de 14,2% (percentual de UC com perda).

Usando o mesmo programa instanciou-se o estado *Fraude* do nó *Inspecção Agrupada*, ver Figura A.1. A crença é que 94,1% das fraudes está presente em UC residenciais, ocorrendo mais desvio antes da medição (27,3%), ponte no bloco de terminais (22,2%) e desvio embutido (22,1%).

Na Figura A.2 instanciam-se simultaneamente os estados *Residencial* e *Fraude*: o resultado não é muito diferente do anterior já que *Residencial* é a grande maioria dentro do Grupo B. Pode-se prosseguir fazendo outras averiguações.

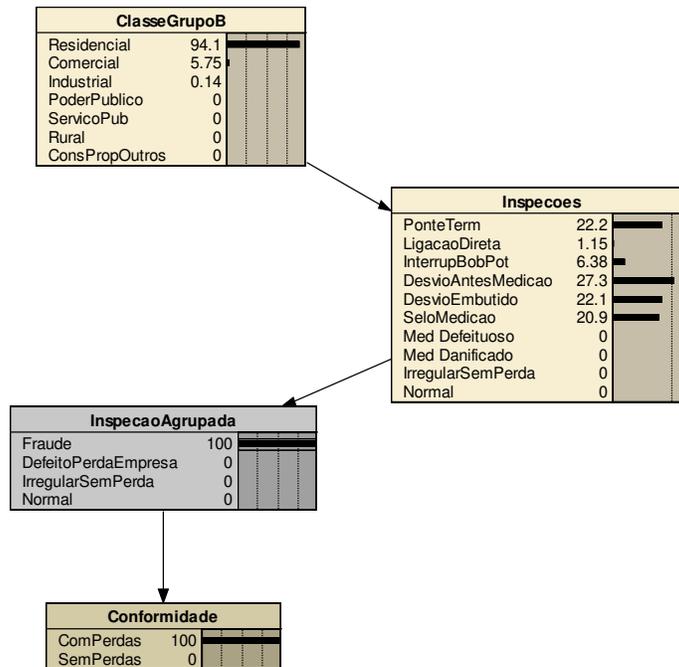


Figura A.1 – RB utilizada, instanciado o estado *Fraude* do nó *Classe_Grupo B*.

Para estimar as quantidades de unidades com perda conforme os motivos, os cálculos são feitos à parte dos softwares comerciais, pois os resultados destes são apenas probabilidades.

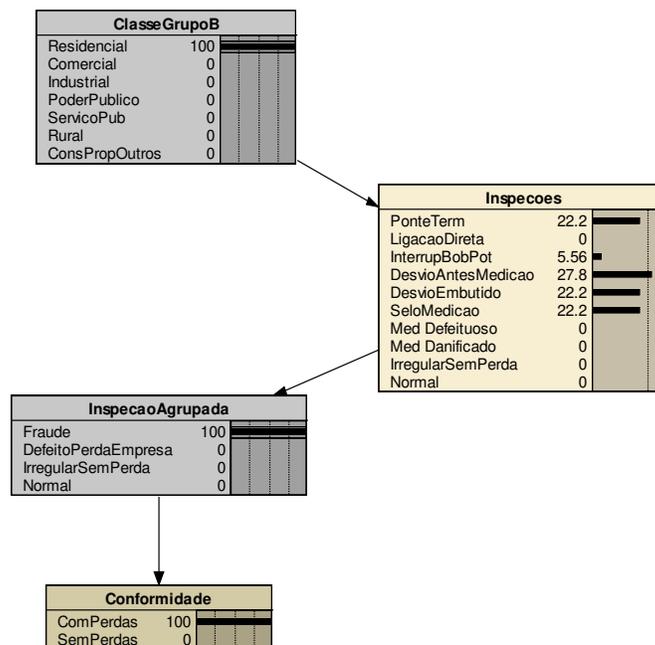


Figura A.2 – RB apresentando a instanciação simultânea dos estados *Fraude* (do nó *Inspeção_Agrupada*) e *Residencial* (do nó *Classe_Grupo B*).

Assim, nas estimativas destes quantitativos usa-se a regra de cadeia (3.30). São usados os valores das probabilidades condicionais mostrados na Tabela A.3 e as probabilidades indicadas na Tabela A.1; este valor vezes o universo (total de unidades, também mostrado na Tabela A.1) conduz à estimativa desejada. Exemplificando, a estimativa de unidades comerciais com ligação direta é

$$UC_{\text{Ligação Direta, Comerciais}} = P(\text{Ligação Direta} | \text{Comercial}) \cdot P(\text{comercial}) \cdot 825698 = \\ = 0,01163 \cdot 0,085463 \cdot 825698 = 820,5 \sim 821$$

A aplicação destes cálculos por estado da inspeção e classe resulta nas estimativas da Tabela A.4.

Tabela A.4 – Estimativa de unidades por estado da inspeção e classes, para o Grupo B, Salvador – (dados de DANTAS, 2006).

	Estado da Inspeção	Classes_Grupo_B						
		Residencial	Comercial	Industrial	Poder Público	Serv. Públ.	Rural	Outros
Externo Fraude	Ponte no Bloco	14.925	821	97	0	0	0	0
	Ligação Direta	0	821	0	0	0	0	0
	Interrup. Bob. Pot.	3.731	820	0	0	0	0	0
	Desvio Antes medidor	18.656	820	0	0	0	0	0
	Desvio Embutido	14.924	821	0	0	0	0	0
	Medidor Avariado Sem selo	14.925	0	0	0	0	0	0
Interno Defeito empresa	Medidor Defeituoso	41.042	820	0	61	0	0	0
	Medidor Danificado	3.731	0	0	0	0	0	0
Sem Perda	Irregular Sem perda	630.561	59.900	2.706	1.576	75	22	18
	Normal	7.462	5.744	387	182	38	0	12
Total de unidades		749.957	70.567	3.190	1.819	113	22	30

Deste modo as estimativas de UC conformes e não conformes estão na Tabela A.4. A crença é que há 117.015 UC com perdas, das quais 111.934 na classe residencial, 4.923 na comercial, 97 na industrial e 61 em poder público. Desejando-se conhecer o percentual de unidades não conformes, somam-se as oito primeiras linhas da Tabela A.4 e se calcula o percentual deste valor em relação ao universo. Logo, a não conformidade (p) do grupo B é

$$P_{\text{GrupoB}} = \frac{117015}{825698} = 0,1417$$

Em valor percentual é de 14,17%.

O agrupamento que foi realizado na rede diz respeito a separar causas externas (os vários motivos de fraude) das internas. Portanto, do total de 14,17% de não conformidade, a responsabilidade da causa fraude é a concentração dos estados correspondentes às seis primeiras linhas da Tabela A.4, logo a probabilidade devido à fraude é

$$P_{\text{Fraude}} = \frac{71360}{825698} = 0,0864$$

Ou percentualmente 8,64%.

Daí, a participação das causas de responsabilidade interna da companhia corresponde ao complemento, ou seja, a probabilidade percentual das linhas sete e oito, neste caso 5,6%. Isto é o que se vê na RB simulada no software e mostrada na Figura 4.7.

Com base no consumo médio anual de cada uma das classes fez-se a estimativa da perda anual de energia que é de 185.393 MWh (ou 11,14% da energia vendida no Grupo B). Tal estimativa supera aquela indicada no original (DANTAS, 2006), que é um total de 152.068 MWh, tendo as justificativas sido apresentadas no item 4.2.3.

ANEXO B

TABELAS REFERENTES ÀS INSPEÇÕES NOS DIVERSOS GRUPOS USADOS NO DIAGNÓSTICO SOB A NOVA HIPÓTESE

Aqui estão apresentadas as tabelas com os resultados das inspeções em todos os cinco grupos, clientes e não clientes, relativos à condição da nova hipótese. Estes dados serviram de base para a preparação das tabelas de probabilidade condicional usadas no item 5.3.

A Tabela B.1 traz os dados das inspeções da amostra do grupo A, e a Tabela B.2 mostra os dados referentes às inspeções do grupo B.

Tabela B.1 - Resultados da nova amostra inspecionada no grupo A.

Estados dos pais	Fraude	Defeito medição	Engano cadastro	Irregular s/ perda	Normal	Total
Residencial Reg.AA	0	0	0	2	4	6
Residencial Reg.BB	0	0	0	1	2	3
Residencial Reg.CC	0	1	0	0	3	4
Comercial Reg.AA	1	1	1	18	32	53
Comercial Reg.BB	1	1	1	10	31	44
Comercial Reg.CC	2	0	0	15	24	41
Industrial Reg.AA	2	1	0	21	48	72
Industrial Reg.BB	1	1	0	21	33	56
Industrial Reg.CC	1	1	0	9	32	43
Poder Público Reg.AA	0	0	1	4	15	20
Poder Público Reg.BB	0	0	0	1	11	12
Poder Público Reg.CC	0	0	0	3	11	14
Serviço Púb. Reg.AA	0	0	0	1	4	5
Serviço Púb. Reg.BB	0	0	0	1	2	3
Serviço Púb. Reg.CC	0	1	0	0	1	2
Rural Reg.AA	0	0	0	5	9	14
Rural Reg.BB	1	0	0	14	37	52
Rural Reg.CC	0	1	0	10	57	68
Outros Reg.AA	0	0	0	0	2	1
Outros Reg.BB	0	0	0	1	1	2
Outros Reg.CC	0	0	0	0	0	0
Totais	9	8	3	137	359	516

Tabela B.2 – Resultados da nova amostra inspecionada no grupo B.

Estados dos pais	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total
Resid. Cons.< 50 kWh/mês –Reg.AA	6	1	1	38	73	119
Resid. Cons.< 50 kWh/mês –Reg.BB	8	1	2	46	85	142
Resid. Cons.< 50 kWh/mês –Reg.CC	6	3	1	60	93	163
Resid. Cons. 51 a 200 –Reg.AA	8	6	3	66	97	180
Resid. Cons. 51 a 200 –Reg.BB	9	5	2	48	112	176
Resid. Cons. 51 a 200 –Reg.CC	9	3	3	52	101	168
Resid. Cons. 201 a 500 –Reg.AA	2	1	1	14	23	41
Resid. Cons. 201 a 500 –Reg.BB	1	1	1	11	25	39
Resid. Cons. 201 a 500 –Reg.CC	2	0	1	12	23	38
Resid. Cons. ≥ 501 kWh –Reg.AA	2	0	1	9	11	23
Resid. Cons. ≥ 501 kWh –Reg.BB	1	1	0	6	13	21
Resid. Cons. ≥ 501 kWh –Reg.CC	1	1	0	3	14	19
Comer. Cons. ≤ 1 mil kWh –Reg.AA	1	1	0	14	19	35
Comer. Cons. ≤ 1 mil kWh –Reg.BB	2	1	0	12	14	29
Comer. Cons. ≤ 1 mil kWh –Reg.CC	0	1	0	10	16	27
Comer. Cons.> 1 mil kWh –Reg.AA	1	1	1	15	15	33
Comer. Cons.> 1 mil kWh –Reg.BB	1	1	0	9	17	28
Comer. Cons.> 1 mil kWh –Reg.CC	1	0	0	7	14	22
Industrial – Região AA	1	1	0	4	10	16
Industrial – Região BB	0	0	1	4	9	14
Industrial – Região CC	0	0	1	6	3	10
Poder público – Região AA	1	1	0	6	12	20
Poder público – Região BB	0	1	0	9	10	20
Poder público – Região CC	0	0	1	5	14	20
Serviço público – Região AA	0	0	0	2	7	9
Serviço público – Região BB	0	0	0	1	7	8
Serviço público – Região CC	0	0	0	2	6	8
Rural – Região AA	1	1	1	9	10	22
Rural – Região BB	1	0	1	6	13	21
Rural – Região CC	0	1	0	6	15	22
Outros – Região AA	0	0	0	1	3	4
Outros – Região BB	0	0	1	0	3	4

Estados dos pais	Fraude	Defeito medição	Engano cadastro	Irregular s/perda	Normal	Total
Outros – Região CC	0	0	0	1	2	3
Total	65	33	23	494	889	1.504

A Tabela B.3 apresenta os dados das inspeções de quadrículas de iluminação pública.

Tabela B.3 - Resultados das novas quadrículas inspecionadas no grupo IP (pontos e energia).

		AA	BB	CC	Total
Pontos da amostra	Quantidade	748	546	484	1.778
	Lâmpadas com operação irregular	7	9	6	22
	Erro no cadastro	14	8	7	29
Energia anual da amostra (MWh/ano)	Encontrada normal	428,1	303,5	265,5	997,1
	Lâmpadas com operação irregular	4,12	5,16	3,38	12,7
	Erro no cadastro	8,24	4,59	3,95	16,8
Não conformidade (%)		2,81	3,11	2,69	2,87

As Tabelas B.4 e B.5 mostram os dados relativos às pesquisas de não clientes, grupos Desligados e Ligações Clandestinas (LC).

Tabela B.4 – Resultados da nova amostra inspecionada no grupo Desligados.

Estados dos pais	Auto-religada	Demolida	Desabitada	Regular	Desligada	Total
Residencial AA	23	21	69	31	133	277
Residencial BB	25	6	53	40	156	280
Residencial CC	27	4	76	27	150	284
Comercial AA	4	1	5	10	7	27
Comercial BB	3	1	4	6	11	25
Comercial CC	2	3	0	6	14	25
Industrial AA	0	0	4	1	4	9

Estados dos pais	Auto-religada	Demolida	Desabitada	Regular	Desligada	Total
Industrial BB	0	1	2	1	5	9
Industrial CC	1	1	1	0	6	9
Poder Público AA	0	0	10	1	1	12
Poder Público BB	0	0	3	2	6	11
Poder Público CC	0	0	4	2	5	11
Serviço Púb. AA	0	0	0	0	5	5
Serviço Púb. BB	1	0	1	3	1	6
Serviço Púb. CC	0	0	0	1	4	5
Rural AA	1	2	2	4	6	15
Rural BB	1	5	4	1	4	15
Rural CC	1	2	1	2	7	13
Outros AA	0	0	1	1	2	4
Outros BB	0	0	0	2	1	3
Outros CC	0	0	1	0	1	2
Total	89	47	241	141	529	1.047

Tabela B.5 - Resultados das novas quadrículas inspecionadas no grupo LC e estimativas por região.

		Região AA	Região BB	Região CC	Total
Inspeção da amostra	UC residenciais regulares na área inspecionada	2.166	2.084	1.999	6.249
	LC encontradas	150	139	110	399
	Não conformidade (%)	6,93	6,67	5,50	6,39
Estimativas por região	Quantidade de LC	12.578	11.875	9.530	33.983
	PC (MWh/ano)	17.920	16.918	13.576	48.414

REFERÊNCIAS BIBLIOGRÁFICAS

ABRADEE – Associação Brasileira de Distribuidores de Energia Elétrica, disponível em www.abradee.org.br; Perdas comerciais, relatório 08-05. Acesso em 07.05.07.

AGÊNCIABRASIL, disponível em <<http://www.agenciabrasil.gov.br>>, notícias de 27.02.07. Acesso em 19.10.09.

AGROSOFT, disponível em <<http://www.agrosoft.org.br/agropag/211699.htm>>. Acesso em 30.9.09.

ALENCAR, M. S. Probabilidade e processos estocásticos, 1ª. ed. Érica, São Paulo, Brasil, 2008.

ALMEIDA, M. A. S. Metodologia de identificação de perdas não-técnicas em unidades consumidoras por atividade de consumo. [dissertação de mestrado], UNIFACS, Universidade Salvador, 2006, Salvador, BA.

ALVES, T. M. M. Combate às perdas comerciais, experiência CEMIG, 2003, disponível em <http://www.abradee.org.br/downloads/1_Worshop_Furtos/CEMIG.pdf>. Acesso em 07.05.07.

ANEEL – Agência Nacional de Energia Elétrica, página <www.aneel.gov.br>:

- Nota Técnica 035/SRD/ANEEL, 2007, “Metodologia e procedimentos para o estabelecimento de regulamentação para apuração de perdas técnicas no segmento distribuição de energia elétrica”, de 22.07.2007. Acesso em 02.09.08;
- Nota Técnica 348/SRE/ANEEL, 2007, “Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica”, de 12.12.2007. Acesso em 28.08.09.
- PRODIST – Procedimentos de distribuição de energia elétrica no sistema elétrico nacional, Módulo 1 – Introdução, p.47, de 27.08.2007. Acesso 15.04.08.
- PRODIST – Procedimentos de distribuição de energia elétrica no sistema elétrico nacional, Módulo 7 – Perdas técnicas regulatórias. Acesso 15.04.08.
- Tarifas médias de fornecimento no Brasil e por região. Acesso em 04.12.2010.

ARANHA NETO, E.A.C., COELHO, J., BETTIOL, A.L., BARCELOS, S.M. Combate às perdas não técnicas no Brasil. In: CLADE – Congreso Latinoamericano de distribución eléctrica Mar Del Plata, Argentina, septiembre, 2008.

ARANHA NETO, E.A.C., ISSICABA, D., COELHO J., BETTIOL A. L., CARNIATO, A., MARCELINO, R., BARCELOS, S. M. Cálculo de perdas técnicas e comerciais em sistemas de

distribuição: comparação entre metodologia e conjunto de medições em um projeto piloto”. In: XIII ERIAC – Encontro Regional Iberoamericano do CIGRÉ, Puerto Iguazu, maio de 2009.

ARAÚJO, A.C.M. Perdas e inadimplência na atividade de distribuição de energia elétrica no Brasil [tese], abril de 2007. UFRJ – Universidade Federal do Rio de Janeiro, Rio de Janeiro.

BACELAR, J. Equações de fatores de perdas. Anais do IV Seminário de Engenharia de Potencia, IEEE, Chile, Aug.1994, pp.173-176;

- Perdas no sistema de distribuição da COELBA, Relatório 214R, COELBA, 09/1994.

BASTOS, P. R. F. M., FERREIRA, N., SOUZA, B. A. Proposta de uma metodologia simplificada para cálculo das perdas técnicas. Anais do XVIII SENDI, Seminário Nacional de Distribuição de Energia Elétrica, Recife-PE, outubro de 2008.

BASTOS, P. R. F. M., FERREIRA, N., SOUZA, B. A. “Energy losses: aspects not considered and the use of probabilistic methods”. In: 20th International Conference on Electricity Distribution, CIRED, paper 0567, session 4. Praha, June, 2009

BERASATEQUI, J. L., BILBAO, J., TORRES, E., EGUÍA P., SAENZ J. R., Determination of energy losses. In: Conference Publication, CIRED, 18-21, June, 2001.

BORDALO, U.A., RODRIGUES, A.B., SILVA, M.G., A new methodology for probabilistic short-circuit evaluation with application in power quality analysis. IEEE Transactions on Power Systems, V.21, n.2, p.474-479, May, 2006.

BRAZ, HELON D. de M., Algoritmos genéticos para configuração ótima de redes de distribuição de energia elétrica [dissertação], Campina Grande, 2003. Universidade Federal de Campina Grande, UFCG, 105 pp.

CALMEYER J.E., FOURIE, J.W. A statistical method to minimize electrical energy losses in a local electricity distribution network. In: IEEE Africon, 2004, pp.667-673.

CANALRURAL, disponível em <<http://www.canalrural.com.br>>, acesso em 30.09.09.

CCON- Comitê Coordenador de Operações Norte e Nordeste. Determinação de perdas em sistemas de distribuição. Anais do XI SENDI, Blumenau, setembro de 1992.

CHIEN, CHEN-FU, CHEN, S.L., LIN, Y.S. Using Bayesian network for fault location on distribution feeder. IEEE Transactions on Power Delivery, V.17, n.3, July 2002.

CIPOLI, et al. Metodologia para avaliação e medição das perdas técnicas, comerciais e totais da distribuição. Anais do I Congresso de Inovação Tecnológica de Energia Elétrica, Brasília, 2001.

CODI – Comitê de Distribuição. Método para determinação, análise e otimização das perdas técnicas em sistemas de distribuição. Documento técnico CODI-3.2.19.34.0, agosto de 1996.

COMETTI, E.S., PERIM, G.T., DIAS, H.B.P., VAREJÃO, F.M. Uma abordagem baseada no conhecimento para identificação de perdas elétricas. In: CBEE – Congresso Brasileiro de Eficiência Energética, Vitória, E.S., 2007.

DANTAS, P. R. P. Avaliação de perdas de energia elétrica não técnicas, metodologia aplicada ao município de Salvador-Ba, [dissertação de mestrado]. Universidade Salvador, UNIFACS. Salvador, BA, 2006.

DAS, B. Consideration of input parameter uncertainties in load flow solution of three-phase unbalanced radial distribution system. IEEE Transactions on Power Systems, V.21, n.3, 2006.

DURÃES, R. L., ALMEIDA, P. E. M., ALBERTO, B. L. A. Inteligência Computacional nas Distribuidoras de Energia Elétrica: evolução tecnológica, aplicações e impactos na redução das perdas não-técnicas. In: Anais do XVII SENDI, Seminário Nacional de Distribuição de Energia Elétrica. Belo Horizonte, MG, agosto de 2006.

EDF – Electricité de France. International Symposium Management Stories. Collection published by International Distribution EDF, Paris, France, 1998 (exemplares avulsos):

No.1 – Improving corporate performance in the energy sector the different types of operation;

No.14 – Power distribution system losses;

No.15 – Reduction of non-technical losses in power distribution systems.

ELLER, N.A. Arquitetura de informação para o gerenciamento de perdas comerciais de energia elétrica. 115 pp., [tese de doutorado em engenharia da produção]. Universidade Federal de Santa Catarina, Florianópolis, SC, 2003.

FERNANDES, C.M.A.P. Redução de perdas técnicas em redes primárias de distribuição de energia elétrica, [tese de doutorado], junho de 2003. UNICAMP, Campinas, SP, 2003.

FERREIRA, H. M. Uso de ferramentas de aprendizado de máquina para prospecção de perdas comerciais em distribuição de energia elétrica, [dissertação de mestrado], UNICAMP, Campinas, SP, janeiro de 2008.

FLATEN, D. L. Distribution system losses calculated by percent loading. IEEE Transactions on Power Systems, V. 3, n. 3, August, 1988.

FRANCISCO, E de R., FAGUNDES E. B. Estudo geoestatístico entre variáveis sócio-econômicas e perda de energia. In: I Encontro Nacional sobre Geoprocessamento no Setor Elétrico, Brasília, dezembro de 2006.

FRÓES DA SILVA, L. L., Iluminação pública no Brasil: aspectos energéticos e institucionais, [dissertação de mestrado]. COPPE/UFRJ, Rio de Janeiro, 2006.

GALLUS, Jr. W.A., ZHANG, J., PU, J., MCCALLEY, J. D., STER, H. A Bayesian approach for short-term transmission line thermal overload risk assessment. IEEE Transactions on Power Delivery, V 17, n.3, July, 2002.

GELMAN, A., CARLIN, J. B., STERN, H. S., RUBIN, D. B., Bayesian data analysis, 2nd edition, Chapman & Hall/CRC, USA, 2004.

GIMENEZ, F. E. S. Estrategia para la identificación y control de pérdidas através de medidores totalizadores em La agencia regional Alto Paraná. In: XIII ERIAC, Encuentro Regional Iberoamericano de CIGRÉ, Puerto Iguazu, Argentina, maio de 2009.

GONEN, T. Electric power distribution system engineering. McGraw Hill, Singapore, 1987, pp.52-59.

GRAINGER, J.J. KENDREW, T. J. Evaluation of technical losses on electric distribution systems. In: Electricity Distribution, CIRED, may 1989.

GUSTAFSON, M.W. Demand, energy and marginal electric systems losses. IEEE Transactions on Power Apparatus and Systems, V. PAS-102, n. 9, September, 1983.

GUSTAFSON, M.W., BAYLOR, J. S.. Approximating the system losses equation, IEEE Transactions on Power Systems, V. 4, n. 3, pp.850-855, August, 1989.

HADI, A.S., GUTIERREZ, J. M., CASTILLO, E. Sistemas Expertos y Modelos de Redes Probabilísticas. Academia Española de Ingenieria, 1998, disponível na página <<http://personales.unican.es/gutierjm/BookCGH.html>>. Acesso em 09.09.2008.

HUGIN®, <<http://www.hugin.com>>. Comercializa o software HUGIN que simula redes Bayesianas. Acesso em 05.06.08 e em 06.12.08.

IBGE –Instituto Brasileiro de Geografia e Estatística, <<http://www.ibge.gov.br>>. Acesso em 30.9.09.

INMETRO, Instituto Nacional de Metrologia, Normatização e Qualidade Industrial, Portaria 088, de abril de 2006.

JARDINI, J.A. et al. Curvas diárias de carga – base de dados estabelecida com medições de campo. In: CIRED, Argentina, 1996.

JENSEN, F. V., NIELSEN T. D. Bayesian networks and decision graphs. 2nd edition, Springer Science, New York, 2007.

JIANG, R., TAGARIS, H., LACHSZ, A., JEFFREY, M. Wavelet based feature extraction and multiple classifiers for electricity fraud detection. In: IEEE Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. In Conference Proceedings, pp.2251-2256, 2002.

KANG, M.S., CHEN, C.S., LIN, C.H., HUANG, C.W., KAO, M.F.. A Systematic loss analysis of Taipower distribution systems. IEEE Transactions on Power Systems, V.21, n.3, August 2006, pp.1062-1068.

LASSO, H., ASCANIO, C., GUGLIA, M. A model for calculating technical losses in the secondary energy distribution network. In: Transmission&Distribution conference and exposition: Latin America, Caracas, Venezuela, August 2006.

MADSEN A.L., KJAERULFF U.B. Bayesian Networks and Influence Diagrams: a guide to construction and analysis. Springer Science+Business Media, New York, 2008.

MATOS, L. N. Utilização de redes bayesianas como agrupador de classificadores locais e global. [tese de doutorado], outubro de 2004, UFCG. Campina Grande, PB, 2004.

MARQUES, M., MATOS, M., FIDALGO, J. N. Estimation and allocation of losses in distribution networks. In: 9 CHLIE, Congreso Hispano-Luso de Ingeniería Eléctrica, Marbella, Espanha, junho de 2005.

MÉFFE, A. Metodologia para cálculo de perdas técnicas por segmento do sistema de distribuição, [dissertação de mestrado], 2001. Universidade de São Paulo – USP. São Paulo.

MÉFFE, A., OLIVEIRA C. C. B. Technical loss calculation by distribution system segment with corrections from measurements. In: 20th International Conference on Electricity Distribution. CIRED, Prague, June 2009, paper 752.

MIKIÉ, O. M. Variance-based energy loss computation in low voltage distribution networks, IEEE Transactions on Power Systems, V. 22, n. 1, February 2007, pp.179-185.

MME, Ministério de Minas e Energia, Brasil. Balanço Energético Nacional de 2008, ano-base 2007, disponível em <<http://www.mme.gov.br>>. Acesso em 20.01.09.

NADIRA, R., BENCHLUCH, S., DORTOLINA, C. A. A novel approach to computing distribution losses. In: IEEE, PES Transmission and Distribution Conference, V. 2, pp.659-663, 2003.

NADIRA, R., DORTOLINA, C.A., The loss that is unknown is no loss at all: a topdown/bottom-up approach for estimating distribution losses. IEEE Transactions on Power Systems, Vol.20, No.2, pp.1119-1125. May, 2005.

NEAPOLITAN, R. E., Learning Bayesian Networks. Pearson Education, Inc., New Jersey, USA, 2004.

NEAPOLITAN, R. E. Probabilistic Methods for Bioinformatics. Elsevier Inc., San Francisco, USA, 2009. Disponível em <www.neiu.edu/~reneapol/informaticsbook.pdf>. Acesso em 24.09.08.

NORSYS®, <<http://www.norsys.com>>. Comercializa o software NETICA. Acesso em 05.06.08.

ORILLAZA, J. R. C., DEL MUNDO, R., MIRAS, J. A. C. Development of models and methodology for the segregation of distribution system losses for regulation. IEEE Region 10, Conference, 1-4, TENCON, 2006.

PEARL, J., Probabilistic Reasoning in intelligent systems: networks of plausible inference. First published 1988. Revised 2nd printing. Morgan Kaufmann Publishers, Inc., San Francisco, 1991.

PEARL, J. Causality: models, reasoning, and inference. Cambridge University Press. First published 2000, 9th printing, New York, 2009.

PENIN, Carlos A. de S. Combate, prevenção e otimização das perdas comerciais de energia elétrica, [tese], 2008, Universidade de São Paulo – USP, São Paulo, 2008.

POVEDA, M. A new method to calculate power distribution losses in environment of high unregistered loads. Transmission and Distribution Conference, 1999 IEEE, New Orleans, USA, April 1999, pp.609-614.

POURRET O., NAIN P., MARCOT B. Bayesian networks: a practical guide to applications. John Wiley & Sons Ltd. England, 2008.

QUEIROZ, L. M. O. Algoritmos genéticos híbridos para redução de perdas técnicas em redes primárias de distribuição considerando variações de demanda. [dissertação de mestrado], junho de 2005. UNICAMP, Campinas, SP, 2005.

QUEIROZ, L., LYRA C., CAVELLUCCI, C. Evaluation of technical losses estimation in LV power distribution systems. In: 20th International Conference on Electricity Distribution. CIRED, paper 0552, Prague, June 2009.

RAJIÈÉ D., TALESKI, R. Energy summation method for energy loss computation in radial distribution network. IEEE Transactions on Power Systems, V. 11, n. 2, May 1996.

RAMOS, G., TORRES, ROGNON, J-P. Análisis de confiabilidad de sistemas industriales aplicando redes Bayesianas, considerando aspectos de PQ e seguridad – Caso de Estudio Sistema IEEE493. IEEE Latin America Transactions, V. 5, n. 8, December 2007.

RAO, P. S. N., DEEKSHIT, R. Energy loss estimation in distribution feeders. IEEE Transactions on Power Delivery, V.21, n.3, July 2006.

RIOS, R. L. F., HENRIQUES, A. L. Controle Metrológico de Medidores com mais de 20 Anos em Serviço. Anais do SENDI – Seminário Nacional de Distribuição de Energia Elétrica, Belo Horizonte, agosto de 2006.

ROCHA, J. E. N., FIGUEIREDO, K., LAZO, J. G. L. et al, Metodologia para detecção de irregularidades e prevenção de perdas comerciais Anais do SENDI – Seminário Nacional de Distribuição de Energia Elétrica, Belo Horizonte, agosto de 2006.

RUSSELL, S., NORVIG, P. Inteligência Artificial. Quinta tiragem da tradução da segunda edição. Elsevier, Rio de Janeiro, 2004.

SILVEIRA, M., DANTAS, P. R. P., VALENTE, A., BASTOS, P. R. F. M., MICHELI, I. Metodologia para identificação das perdas não técnicas - Matriz de perdas. In: Congreso Internacional de Distribución Electrica – CIDEL, Buenos Aires, Argentina, novembro de 2006.

SILVEIRA, A., MORAES, R. T. B. Otimizando Resultados com a Gestão das Perdas Não-Técnicas de Energia Elétrica. Anais do XVII SENDI Seminário Nacional de Distribuição de Energia Elétrica, Belo Horizonte, agosto de 2006.

SIQUEIRA C.A., ARAUJO, A.C.M., Considerações sobre as perdas na distribuição de energia elétrica no Brasil. Anais do XVII SENDI Seminário Nacional de Distribuição de Energia Elétrica, Belo Horizonte, agosto de 2006.

SMITH, T. B. Electricity Theft: a comparative analysis. Energy Policy, Elsevier, V. 32, Issue 18, December 2004, pp.2067-2076.

STRAUCH, M. Metodologia probabilística para cálculo de perdas técnicas em redes secundárias de distribuição [dissertação]. Universidade Salvador - UNIFACS, Salvador, BA, 2002.

SUN, D. I. H., ABE, S., SHOULTS, R. R., CHEN, M. S., EICHENBERGER, P., FARRIS, D. Calculation of energy losses in a distribution system. IEEE Transactions on Power Apparatus and Systems, V. PAS-99, n. 4, July/August, 1980, pp.1347-1356.

URANI A., FONTES, A., FRANCO S. Efeito do ambiente sócio-econômico sobre as perdas não técnicas na distribuição de energia elétrica. Relatório do IETS – Instituto de Estudos do Trabalho e Sociedade para a LIGHT-Rio (disponível na página www.aneel.gov.br, dentre as contribuições quando da Audiência Pública 052/ANEEL/2007), abril de 2008. Acesso em 17.08.09.

VALENTE, A.L. C., ALMEIDA, M. A. S., DANTAS, P. R. P., SILVEIRA, M., BASTOS, P. R. F. M. Sistema de auxílio à identificação de perdas comerciais. In: Anais do III Congresso de Inovação Tecnológica de Energia Elétrica (CITENEL), Florianópolis, 2005.

VALENZUELA, L., MONTAÑA, C. Una experiencia exitosa en el control de pérdidas No-Técnicas. In: Anais do V CIERTEC – Seminário Internacional sobre Gestión de Pérdidas, Maceió, agosto de 2005.

VARDEMAN, S. B. Statistics for engineering problem solving. IEEE Press. PWS Publishing Company. Boston, 1994.

VIEIRALVES, E. X. Proposta de uma Metodologia para Avaliação das Perdas Comerciais dos Sistemas Elétricos: O Caso Manaus. [Dissertação de mestrado], agosto de 2005. UNICAMP, Campinas, SP, 2005.

YAN, W., LIGUO, Z., LIMIN, H., YONGLI, Z. Bayesian network based time-sequence simulation for power system reliability assessment. In: 7th Mexican International Conference on Artificial Intelligence, Atizapan de Zaragoza, México, October 2008.

YIN, X., HUANG, D., XIE, L., QIAN, W. Systems reliability analysis and fault diagnosis based on Bayesian networks. In ISA 2009 – International Workshop on Intelligent System and Application, Wuhan, China, May 2009.

YONGLI, Z., LIMIN, H., JINLING L. Bayesian networks – Based approach for power systems fault diagnosis. IEEE Transactions on Power Delivery, V. 21, n. 2, April, 2006, pp.634-639.

ZAR, J. H. Biostatistical Analysis. Prentice-Hall International, Inc. London, 1974.