



MODELO PARA REPRESENTAÇÃO DE SÉRIES TEMPORAIS
CORRELACIONADAS COM APLICAÇÃO NA AVALIAÇÃO DA
CONFIABILIDADE POR SIMULAÇÃO MONTE CARLO NÃO-SEQUENCIAL

Julio Alberto Silva Dias

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientadora: Carmen Lucia Tancredo Borges

Rio de Janeiro
Setembro de 2014

MODELO PARA REPRESENTAÇÃO DE SÉRIES TEMPORAIS
CORRELACIONADAS COM APLICAÇÃO NA AVALIAÇÃO DA
CONFIABILIDADE POR SIMULAÇÃO MONTE CARLO NÃO-SEQUENCIAL

Julio Alberto Silva Dias

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM
CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof^a. Carmen Lucia Tancredo Borges, D.Sc.

Prof. Djalma Mosqueira Falcão, Ph.D.

Prof. Glauco Nery Taranto, Ph.D.

Dr. Mario Veiga Ferraz Pereira, Ph.D.

Prof. Mauro Augusto da Rosa, Ph.D.

RIO DE JANEIRO, RJ - BRASIL
SETEMBRO DE 2014

Dias, Julio Alberto Silva

Modelo para Representação de Séries Temporais Correlacionadas com Aplicação na Avaliação da Confiabilidade por Simulação Monte Carlo Não-Sequencial/ Julio Alberto Silva Dias. – Rio de Janeiro: UFRJ/COPPE, 2014.

XIV, 174 p.: il.; 29,7 cm.

Orientador: Carmen Lucia Tancredo Borges

Tese (doutorado) – UFRJ/ COPPE/ Programa de Engenharia Elétrica, 2014.

Referências Bibliográficas: p. 168-174.

1. Confiabilidade de Sistemas de Potência. 2. Simulação Monte Carlo Não-Sequencial. 3. Geração Eólica. I. Borges, Carmen Lucia Tancredo. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MODELO PARA REPRESENTAÇÃO DE SÉRIES TEMPORAIS
CORRELACIONADAS COM APLICAÇÃO NA AVALIAÇÃO DA
CONFIABILIDADE POR SIMULAÇÃO MONTE CARLO NÃO-SEQUENCIAL

Julio Alberto Silva Dias

Setembro/2014

Orientador: Carmen Lucia Tancredo Borges

Programa: Engenharia Elétrica

Neste trabalho é proposto um novo modelo para a avaliação da confiabilidade composta por Simulação Monte Carlo Não Sequencial, que possibilita a representação de elementos variantes no tempo (carga, geração eólica, vazões afluentes, etc.) estatisticamente dependentes, preservando a correlação entre estas variáveis. A proposta é baseada na combinação de modelos não paramétricos, que dispensam a caracterização a priori das funções de densidade de probabilidade das variáveis aleatórias envolvidas, com a aplicação do coeficiente de informação máxima, que permite mapear relacionamentos não lineares entre as variáveis, e estruturas de rede Bayesianas, que permitem tratar a questão da alta dimensionalidade quando são representadas múltiplas séries temporais. O modelo proposto permite a obtenção de índices de confiabilidade com a mesma precisão da Simulação Monte Carlo Sequencial, porém com custo computacional da ordem do da Simulação Monte Carlo Não Sequencial, sendo ainda flexível para a construção e simulação de processos estocásticos que são difíceis de serem representados a partir das abordagens tradicionais de séries temporais, como a introdução de variáveis exógenas no processo, ou ainda, para representar o relacionamento entre variáveis com diferentes níveis de discretização, como é o caso da geração eólica e das vazões afluentes de um sistema hidrelétrico.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

A MODEL TO REPRESENT CORRELATED TEMPORAL SERIES APPLIED
IN RELIABILITY EVALUATION BY NON-SEQUENTIAL
MONE CARLO SIMULATION

Julio Alberto Silva Dias

September/2014

Advisor: Carmen Lucia Tancredo Borges

Department: Electrical Engineering

In this thesis we propose a new method to reliability evaluation that applies non Sequential Monte Carlo Simulation with statistic dependent time varying elements (load, wind generation, river inflows, etc.) preserving the correlation among these variables. The proposed model combines non-parametrical models that dismiss the prior probability density function characterization of the random variables involved, with the maximum information coefficient that endue the representation of nonlinear relationships among these variables, as well as the use of Bayesian Network framework to deal with the high dimensionality of multivariate time series. The proposed model allows the obtainment of reliability indices that have the same precision as the ones obtained from Sequential Monte Carlo simulation in addition to reducing considerable the process computational time. Furthermore, the model is flexible to stochastic processes construction and simulation which representation would be impossible in the conventional approach, as the introduction of exogenous variables in the process or the representation of variables with different levels of discretization that should be used, for example, in order to maintain the correlation between wind generation and inflows in a hydroelectric system.

SUMÁRIO

ÍNDICE DE FIGURAS.....	ix
ÍNDICE DE TABELAS	xiii
1 Introdução.....	1
1.1 Modelos Estocásticos para Representação de Séries Temporais na Confiabilidade	3
1.2 Motivação e Objetivos do Trabalho.....	7
1.3 Revisão Bibliográfica.....	8
1.4 Trabalhos Publicados	11
1.5 Organização do Trabalho	12
2 Avaliação da Confiabilidade por Simulação Monte Carlo.....	13
2.1 Introdução	13
2.1.1 Níveis Hierárquicos.....	13
2.2 Representação de Componentes por Modelos de Markov.....	16
2.3 Simulação Monte Carlo Sequencial.....	20
2.4 Simulação Monte Carlo Não Sequencial	21
2.4.1 Índices de Confiabilidade Composta	22
2.4.2 Convergência dos Índices de Confiabilidade	23
2.5 Representação de Elementos Variantes no Tempo	24
2.5.1 Simulação Monte Carlo Pseudo-Sequencial	25
2.5.2 Simulação Monte Carlo Pseudo-Cronológica	26
2.6 Considerações Finais	29
3 Estimação da Densidade de Probabilidade, Medidas de Dependência e Redes Bayesianas	30
3.1 Introdução	30
3.2 Função de Densidade de Probabilidade	31

3.2.1	Estimação Não Paramétrica da Função de Densidade de Probabilidade – Uso do Histograma.....	32
3.2.2	Estimação de Densidade por Funções Kernel - <i>Kernel Density Estimation</i>	35
3.3	Densidades de Probabilidade Conjunta e Dependência Estatística entre Variáveis Aleatórias	40
3.3.1	Correlação de Pearson.....	41
3.3.2	Informação Mútua (<i>Mutual Information</i>) e Coeficiente de Informação Máxima (<i>Maximum Coefficient of Information</i>)	43
3.4	Rede Bayesiana.....	45
4	Análise Estatística do Comportamento do Vento para Geração Eólica	50
4.1	Introdução	50
4.2	Características da Potência Elétrica Gerada	52
4.3	Modelo Estocástico da Turbina	53
4.4	Efeito Esteira.....	55
4.5	Análise Estatística de Séries de Vento.....	59
4.5.1	Séries Horárias de Vento para um Horizonte de um Ano	60
4.5.2	Estimação da Densidade de Probabilidade	62
4.5.3	Estimação da Auto-correlação	66
4.6	Análise Estatística de Séries de Geração Eólica	68
5	Proposta de Modelo para Representação de Elementos Variantes no Tempo Estatisticamente Dependentes	73
5.1	Introdução	73
5.2	Modelo para Representação de Elementos Variantes no Tempo.....	79
5.3	Amostragem dos Estados do Modelo na Simulação Estocástica.....	85
5.4	Representação da Incerteza na Direção do Vento.....	89
5.5	Representação de Variáveis Exógenas.....	89
5.6	Estimação de Índices de Frequência e Duração.....	96
5.6.1	Modelo de Transição de Estados.....	96

5.6.2 Aplicação do Modelo para o Cálculo dos Índices de Frequência e Duração 99

6	Resultados	102
6.1	Introdução	102
6.2	Dados de Vento.....	103
6.3	Sistema IEEE-RTS	106
6.3.1	Caso 1 - RTS + 27 Usinas Eólicas – Velocidade do Vento	110
6.3.2	Caso 2 - RTS + 27 Usinas Eólicas – Velocidade e Direção do Vento 122	
6.3.3	Caso 3 - RTS + 27 Usinas Eólicas – Variável Climática Exógena .	129
6.3.4	Caso 4 – Sistema Brasileiro	133
7	Conclusões e Continuação do Trabalho	150
7.1	Considerações Gerais.....	150
7.2	Desenvolvimentos Futuros.....	153
I	Kernel Density Estimation	155
I.1	Medidas de Desempenho.....	155
I.2	Estimação do Parâmetro de Alisamento.....	156
I.3	Estimação de Densidade com Suporte para Variáveis com Limite.....	157
I.4	Proposta de Transformação para Estimação de Densidade com Suporte para Variáveis com Limite	159
II	Rede Bayesiana.....	162
II.1	Estimação de Estrutura de Redes Bayesianas pelo algoritmo Hill Climbing	162
II.2	Estimação dos Parâmetros de Redes Bayesianas pelo Método MLE (Maximum Likelihood Estimation – Estimação da Máxima Verossimilhança)	165
	Referências Bibliográficas	168

ÍNDICE DE FIGURAS

Figura 1 – Níveis Hierárquicos para a Confiabilidade.....	14
Figura 2 – Representação de um Sistema no NH1.....	14
Figura 3 – Representação de um Sistema Hipotético no NH2.....	15
Figura 4 – Representação de um Sistema Hipotético no NH3.....	16
Figura 5 – Exemplos de Modelos a Múltiplos Estados.....	18
Figura 6 – Modelo Markoviano não agregado de Múltiplos Níveis	27
Figura 7 – Variação do histograma com o comprimento	34
Figura 8 –Funções de distribuição de probabilidade acumulada para variável	35
Figura 9 –Estimação de Função de Densidade por Histograma e Kernel Density Estimation[43]	36
Figura 10 – Exemplo de Estimação com Kernel Gaussiano[44]	38
Figura 11 – Estimações subsuavizada e supersuavizada	39
Figura 12 –Exemplo de variáveis aleatórias linearmente dependentes.....	42
Figura 13 – Exemplo de rede Bayesiana.....	47
Figura 14 –Evolução de Número de Usinas Eólicas[51]	51
Figura 15 –Geração no Período de Fev/2013 até Fev2014[51]	51
Figura 16 – Curvas de Potência de Turbinas com Controles Distintos[52].....	53
Figura 17 – Elementos que Compõem a Turbina Eólica[52].....	53
Figura 18 – Diagrama de Espaço de Estados de uma Usina com 1 ou 2 Turbinas[52]	54
Figura 19 – Efeito Esteira em um Parque Eólico.....	55
Figura 20 – Histograma do Ângulo de Incidência do Vento (Direção do Vento)	57
Figura 21 – Densidade de Probabilidade 2D para a Série de Vento s-210 (Módulo e Direção)	57
Figura 22 – Séries Horárias de Vento para um Horizonte de um Ano	61
Figura 23 – Séries para as Três Primeiras Semanas do Horizonte.....	62

Figura 24 – Variação da Distribuição de Weibull de Acordo com os Parâmetros[52]	63
Figura 25 – Funções de Densidade de Probabilidade Estimadas para as Séries...	64
Figura 26 – Comparações Entre as Densidades não Paramétricas e as Paramétricas	65
Figura 27 – Auto-Correlações Parciais Estimadas para as Séries.....	67
Figura 28 – Densidade de Probabilidade de uma Série de Geração Eólica do Nordeste.....	69
Figura 29 – Autocorrelação Parcial para a Série de Geração Eólica do Nordeste	69
Figura 30 – Densidade de Probabilidade de uma Série de Geração Eólica do Sul	70
Figura 31 – Autocorrelação Parcial para a Série de Geração Eólica do Nordeste	71
Figura 32 – Densidade de Probabilidade Conjunta das Séries de Geração do Sul: Litoral x Interior	71
Figura 33 – Transformação de Variáveis Aleatórias em Variáveis Normalmente Distribuídas.....	75
Figura 34 – Geração de Amostras a Partir das Variáveis Aleatórias Transformadas	76
Figura 35 – Visão Geral do Modelo.....	80
Figura 36 – Função de Densidade de Probabilidade Estimada com Suporte em $[0, \infty)$	81
Figura 37 – Função de Probabilidade Acumulada Estimada	82
Figura 38 – Função Inversa da Probabilidade Acumulada Estimada	82
Figura 39 – Função de densidade de probabilidade da variável transformada uniforme	83
Figura 40 – Processo de Estimação e Geração de Amostras a partir do Modelo Proposto	88
Figura 41 – Incorporação da Representação de Variáveis Exógenas a Visão Geral do Processo.....	91
Figura 42 – Incorporação da Representação de Dependência Estatística com Vazões Afluentes.....	93

Figura 43 – Formulação Alternativa – Representação da Dependência Estatística com as Energias Afluentes	95
Figura 44 – Rede Bayesiana Conceitual para a Transição de Estados.....	97
Figura 45 – Incorporação do Modelo de Transição de Estados a Visão Geral do Processo.....	101
Figura 46 - Mapa Contendo a Distribuição Geográfica dos Postos de Medição[57]	105
Figura 47 - Representação Unifilar do Sistema IEEE-RTS[62]	108
Figura 48 - Geração da Turbina em Função da Velocidade do Vento.....	109
Figura 49 – Comparação das Convergências da LOLF para o Caso 1	113
Figura 50 - Densidade de Probabilidade do Corte de Carga para Estados com Déficit – Sem Geração.....	115
Figura 51 - Densidade de Probabilidade do Corte de Carga para Estados com Déficit – Caso 1	115
Figura 52 - Corte de Carga Médio por Barra – Caso 1	117
Figura 53 - Redução Percentual de Corte de Carga Médio por Barra – Caso 1 .	118
Figura 54 – Densidade de Probabilidade para as Variáveis Aleatórias de Vento	119
Figura 55 – Densidades de Probabilidade Conjunta (Densidades 2D) para Algumas da Séries	121
Figura 56 – Histograma da Diferença Absoluta dos MICs (Histórico e Sintética)	122
Figura 57 – Função de Eficiência das Usinas pela Direção do Vento	124
Figura 58 - Densidade de Probabilidade do Corte de Carga para Estados com Corte de Carga – Casos 1 e 2.....	126
Figura 59 – Densidades de Probabilidade para as Variáveis Aleatórias de Direção do Vento	127
Figura 60 – Densidade de Probabilidade Conjunta para a Série de Vento s-225 (Módulo e Direção)	128
Figura 61 – Densidade de Probabilidade Conjunta para Séries de Direção.....	129
Figura 62 – Densidade de Probabilidade para as Variáveis Aleatórias de Vento	132
Figura 63 – Visão Geral do Sistema de Transmissão do Sistema Brasileiro	133

Figura 64 – Demanda do Sistema Brasileiro para o Ano de 2015.....	136
Figura 65 – Barras do Sistema Brasileiro com Maiores Fatores de Participação	136
Figura 66 – Séries Normalizadas da Eólica do Nordeste e das ENAs do Sudeste e do Sul.....	137
Figura 67 – Barras com Maiores Probabilidades de Perda de Carga.....	139
Figura 68 – Barras com Maiores EPNS.....	140
Figura 69 – Densidade de Probabilidade do Corte de Carga para o Sistema Brasileiro	140
Figura 70 – Probabilidade Acumulada do Corte de Carga para Estados com Déficit.....	141
Figura 71 – Probabilidade Acumulada do Corte de Carga para Estados com Déficit – Escala até 500 MW.....	141
Figura 72 – Distribuição de Corte de Carga Médio por Sistema – Sem Eólicas	143
Figura 73 – Densidade de Probabilidade do Corte de Carga para o Sistema Brasileiro – Com Eólica x Sem Eólica	144
Figura 74 – Densidade de Probabilidade do Corte de Carga para o Sistema Brasileiro – Com Eólica x Sem Eólica – Escala de Corte até 40 MW.....	145
Figura 75 – Probabilidade Acumulada do Corte de Carga para Estados com Déficit – Escala até 500 MW – Com Eólica x Sem Eólica	146
Figura 76 – Probabilidade Acumulada do Corte de Carga para Estados com Déficit – Escala até 100 MW – Com Eólica x Sem Eólica	146
Figura 77 – Distribuição de Corte de Carga Médio por Sistema – Com Eólicas	148
Figura 78 – Comparação de Valores Médios de Corte de Carga.....	148
Figura 79 – Redução Percentual do Corte de Carga Médio por Sistema.....	149
Figura 80 – Comparação de Estimativa sem Definição de Suporte e Com Definição de Suporte	158
Figura 81 – Função sigmoide para diferentes constantes.....	160
Figura 82 – Função de densidade de probabilidade estimada com suporte em $[0, \infty)$	161
Figura 83 – Estimação de rede Bayesiana por <i>hill climbing</i>	165

ÍNDICE DE TABELAS

Tabela 1 – Comparações <i>MIC x MIC-R2</i> para Séries de Módulo e Direção do Vento	58
Tabela 2 – Comparações <i>MIC x MIC-R2</i> para Correlação Espacial entre Séries de Direção	58
Tabela 3 – Comparações <i>MIC x MIC-R2</i> para Auto-Correlação de <i>lag-1</i> de Séries de Direção.....	59
Tabela 4 – Estações de Medição de Velocidade de Vento.....	104
Tabela 5 – Características Básicas do Sistema IEEE-RTS	106
Tabela 6 – Dados das Usinas do Sistema IEEE-RTS	106
Tabela 7 – Dados das Linhas de Transmissão e Transformadores do Sistema IEEE-RTS.....	107
Tabela 8 – Distribuição das Usinas Eólicas ao Longo do RTS.	109
Tabela 9 – Índices de Confiabilidade para o Sistema – Sem Eólicas	110
Tabela 10 – Índices de Confiabilidade para o Sistema – Caso 1	111
Tabela 11 – Intervalos de Confiança para os Índices da SMC Sequencial – Caso 1	111
Tabela 12 – Erros Percentuais para os Índices de Confiabilidade do Sistema – Caso 1	111
Tabela 13 – Tempo Desprendido para as Atividades de Estimação do Modelo Proposto.....	114
Tabela 14 – Tempo Despendido para as Etapas da Simulação do Modelo Proposto	114
Tabela 15 – Referência para os Valores de <i>p-value</i>	119
Tabela 16 – Valores Adotados de Eficiência por Direção	123
Tabela 17 – Índices de Confiabilidade para o Sistema – Caso 2	124
Tabela 18 – Intervalos de Confiança para os Índices da SMC Sequencial – Caso 2	124

Tabela 19 – Erros Percentuais para os Índices de Confiabilidade do Sistema – Caso 2	125
Tabela 20 – Comparação dos Índices de Confiabilidade entre os Casos 1 e 2 ...	125
Tabela 21 – Medições para o <i>Oceanic Niño Index (ONI)</i>	130
Tabela 22 – Índices Considerados para o <i>Oceanic Niño Index (ONI)</i>	131
Tabela 23 – Índices de Confiabilidade para o Sistema	131
Tabela 24 – Características Básicas da Representação do Sistema Brasileiro....	134
Tabela 25 – Usinas Eólicas do Sistema Brasileiro.....	135
Tabela 26 – Correlações entre a Geração Eólica e a Energia Afluente por Sistema	138
Tabela 27 – Riscos de Déficit para o Sistema Brasileiro	138
Tabela 28 – Índices de Confiabilidade para o Sistema Brasileiro, sem Eólicas .	139
Tabela 29 – Probabilidades dos Estados Mais Severos de Corte de Carga do Sistema Brasileiro.....	142
Tabela 30 – Índices de Confiabilidade para o Sistema Brasileiro, com Eólicas .	143
Tabela 31 – Probabilidades dos Estados Mais Severos de Corte de Carga do Sistema Brasileiro – Com Eólica x Sem Eólica	147

Capítulo 1

Introdução

A avaliação da confiabilidade de sistemas de potência é um estudo que mensura a capacidade do sistema, ou elemento do sistema, de desempenhar a sua função de atendimento ao consumidor de forma adequada, isto é, dentro das condições operativas padronizadas, de forma segura, econômica e confiável, dadas algumas condições de incerteza. Idealmente o fornecimento de energia seria contínuo, porém os elementos que compõe o sistema estão sujeitos a falhas que desencadeiam perturbações que podem afetar os consumidores, ocasionando a interrupção do fornecimento de energia elétrica.

De forma geral, a confiabilidade de sistemas de potência pode ser avaliada sob dois aspectos distintos: a adequação e a segurança do sistema. O aspecto da segurança é mais complexo por envolver as características dinâmicas do sistema, isto é, mensura a capacidade do sistema em responder a distúrbios dinâmicos ou transitórios e manter margens aceitáveis de segurança. A capacidade para avaliar a confiabilidade no aspecto da segurança é ainda muito limitada atualmente, devido principalmente à complexidade associada com a modelagem do comportamento dinâmico do sistema[1]. Já na avaliação da adequação, a resposta a uma falha do sistema é considerada do ponto de vista estático, que não incluem a dinâmica e a resposta aos distúrbios transitórios. Em termos gerais, a avaliação no aspecto da adequação mensura a capacidade do sistema em suportar a ocorrência de contingência sem precisar cortar carga, sem levar em conta se alguma instabilidade transitória tornaria a transição do estado normal para o estado em contingência infactível[2].

A confiabilidade de sistemas de potência é medida a partir de um conjunto de indicadores estabelecidos que basicamente se dividem em medidas de probabilidade, energia não suprida, frequência e duração de ocorrências de falhas. A maioria das técnicas aplicadas para a avaliação da confiabilidade de sistemas de potência enfocam apenas a adequação, e, desta maneira, estes índices de confiabilidade são, em um sentido mais estrito, índices de adequação do sistema, e não índices de confiabilidade de uma forma mais geral.

Durante as fases de planejamento e operação dos sistemas elétricos, grandes investimentos devem ser feitos visando a melhoria destes indicadores, aumentando-se a confiabilidade dos sistemas. Por outro lado, existem restrições econômicas que devem ser respeitadas para que o sistema não venha a se tornar demasiadamente caro, ou até inviável, uma vez que estes investimentos são refletidos nas tarifas do consumidor final. Estas restrições nos investimentos fazem com que os sistemas operem ainda mais perto dos seus limites, demandando análises mais precisas dos riscos de interrupção ou degradação da qualidade do suprimento de energia elétrica[2].

Os métodos de avaliação da confiabilidade de sistemas de potência podem ser enquadrados em dois grupos principais, o grupo dos métodos determinísticos e o dos métodos probabilísticos.

Na análise determinística, a partir de um determinado critério estabelecido, são eleitas a priori algumas configurações e contingências, e verificado se o sistema é capaz de atender a estas situações.

Os critérios mais usuais são:

1. **N-1:** analisa-se se o sistema é capaz de suportar todas as contingências simples em regime permanente, sem ocorrência de perda de carga ou violações operativas;
2. **N-2:** analisa-se se o sistema é capaz de suportar todas as combinações de contingências duplas em regime permanente, sem ocorrência de perda de carga ou violações operativas;

Os critérios determinísticos apresentam uma maior simplicidade de uso e facilidade de interpretação dos resultados, uma vez que existe uma associação direta da simulação implementada com a prática operativa do sistema.

Em contrapartida, os métodos determinísticos não consideram as probabilidades de ocorrência das contingências, ignorando a natureza aleatória de operação do sistema, das disponibilidades energéticas e das falhas dos equipamentos, resultando em soluções que podem ser excessivamente conservativas e caras.

Na análise probabilística, a aleatoriedade dos eventos do sistema é considerada, permitindo que se quantifiquem os riscos associados às alternativas operacionais.

As falhas nos sistemas de potência podem ocorrer ao nível da geração, transmissão e distribuição, podendo ser ainda falhas individuais ou múltiplas. Adiciona-se a este cenário de incertezas o fato de que os sistemas de potência são constituídos por inúmeros elementos cuja operação normal também agrega alguma incerteza intrínseca. Entre estes se destacam a demanda elétrica do sistema, que apresenta diferentes graus de

incerteza para os diferentes horizontes de análise, e a geração hidrelétrica, que depende das vazões afluentes dos rios que alimentam os reservatórios.

Atualmente, com a grande penetração da geração eólica na matriz elétrica dos sistemas de potência por todo o mundo, passou-se a ter uma nova fonte de incerteza neste conjunto, devido a esta forma de geração ser modulada pela dinâmica dos ventos, cujo comportamento é extremamente variável e de difícil previsão.

Devido a esta combinação de incertezas, a avaliação da confiabilidade pela abordagem probabilística é desejável em detrimento da abordagem determinística.

Na abordagem probabilística, a avaliação da confiabilidade pode ser realizada por meio de duas representações distintas: espaço de estados e representação cronológica. Na representação por espaço de estados, os estados do sistema podem ser selecionados através de duas técnicas: enumeração de estados e Simulação Monte Carlo Não-Sequencial, utilizando a amostragem de estados. Já na representação cronológica, por meio da Simulação Monte Carlo Sequencial, os estados operativos dos componentes são sequencialmente amostrados por vários períodos de tempo, simulando a cronologia do processo estocástico de operação do sistema. A abordagem sequencial tende a produzir resultados mais precisos que a não sequencial quando existe a representação de elementos variantes no tempo, tais como as curvas de demanda e a geração eólica, especialmente se esses elementos forem estatisticamente dependentes. Por outro lado, a abordagem sequencial demanda um alto custo computacional, podendo se tornar proibitiva para sistemas de grande porte. Já a abordagem não sequencial possui um custo computacional muito inferior à sequencial, porém apresenta uma grande limitação para representar elementos variantes no tempo, principalmente porque o método de Simulação Monte Carlo Não Sequencial convencional não considera a dependência estatística existente entre as variáveis aleatórias, podendo sua aplicação indevida levar a resultados bastante imprecisos.

1.1 Modelos Estocásticos para Representação de Séries Temporais na Confiabilidade

Os métodos de avaliação de confiabilidade baseados em representação cronológica têm em comum a necessidade de se conhecer a priori as séries temporais que descrevem os cenários dos elementos variantes no tempo. Isto é, os cenários de demanda, vazões afluentes, geração eólica, ou quaisquer outros elementos variantes no tempo, precisam ser previamente estabelecidos para que possam ser aplicados à simulação estocástica do sistema. Devido a isto, os modelos empregados para construir os cenários destas séries temporais são tão críticos quanto o próprio método de

simulação, pois, independentemente se o método de simulação é cronológico, não cronológico ou híbrido, a responsabilidade de obtenção de cenários mais adequados, isto é, que proporcionem análises preditivas mais coerentes com a realidade, é destes modelos.

Uma abordagem usual na avaliação da confiabilidade de sistemas de potência é a representação destas séries por meio do emprego de um modelo de simulação dos dados históricos, no qual, pressupondo-se que as séries representam processos estocásticos estacionários, os cenários das séries temporais são definidos a partir da repetição dos dados observados no histórico. Estas séries de observações são então combinadas com os modelos estocásticos dos elementos convencionais do sistema, tais como o estado operativo das unidades geradoras e das linhas de transmissão, produzindo os estados do sistema que serão avaliados na análise da adequação.

O emprego de cenários construídos a partir da repetição dos dados históricos dos elementos variantes no tempo é uma alternativa de fácil compreensão e implementação na simulação estocástica. Uma de suas principais vantagens é que, na presença de muitos elementos variantes no tempo, garante que os efeitos da correlação entre as variáveis serão capturados pela simulação, assim como este tipo de representação garante também a preservação de qualquer outra medida estatística verificada nas observações históricas.

Em contraste à utilização da simulação dos dados históricos, está a abordagem de simulação de cenários sintéticos, obtidos a partir da estimação de modelos estocásticos para os elementos variantes no tempo. Os modelos mais comuns para este tipo de abordagem são os modelos auto-regressivos (AR) e suas derivações (ARMA, ARp, ARMAp, ARIMA, SARIMA, etc.), bastante difundidos e aplicados na literatura para a representação de séries temporais de vazões afluentes, [3] e [4], demanda elétrica, por exemplo, [5] e [6], e geração eólica, por exemplo, [7] e [8].

Entre os motivos mais comuns para se empregar modelos estocásticos na representação de séries temporais em detrimento de se aplicar diretamente os cenários históricos, estão:

1. Necessidade de filtragem das séries temporais

Os principais motivos para se filtrar séries temporais são a remoção de tendências e não estacionariedades da série, o tratamento de erros e ausência de medidas, e a suavização das séries[9].

Nestes casos, as séries históricas podem não ser diretamente aplicáveis, mas é possível a estimação de modelos estocásticos a partir dos dados das séries históricas que

irão preservar as principais características do processo estocástico original, mas expurgando os problemas observados nos dados.

2. Representação de maior variabilidade nos cenários simulados

Esta representação pode ser necessária, por exemplo, quando o histórico disponível é reduzido, e a geração de cenários sintéticos a partir da aplicação dos modelos estocásticos é uma alternativa para complementar a análise.

3. Representação de variáveis exógenas às séries temporais

Um exemplo bastante comum deste tipo de representação é a geração de cenários de demanda elétrica considerando como entrada um conjunto de variáveis exógenas macroeconômicas, uma vez que correlação entre o crescimento da demanda e o crescimento econômico de um país ou de uma região[10].

Para as variáveis aleatórias que possuem forte dependência com o clima, como é o caso do vento, das vazões afluentes e da própria demanda elétrica, o emprego de modelos estocásticos apresenta outra vantagem, que é a flexibilidade para a incorporação de informação adicional fornecida por variáveis exógenas ao processo. Esta representação de informações exógenas em simulações estocásticas também tem sido uma área de estudo que tem recebido bastante atenção recentemente, com o objetivo de tornar as análises mais aderentes aos cenários efetivamente realizados. Um exemplo bem atual deste tipo de abordagem é a consideração de índices ENSO (El Nino Southern Oscillation indices) [11] na construção de modelos estocásticos para representação das incertezas nas vazões afluentes e da dinâmica dos ventos. A partir de uma base histórica de medições para estes índices, e dada uma previsão conhecida para estes índices para um horizonte conhecido, é possível estimar modelos estocásticos que irão gerar cenários muito mais plausíveis para o horizonte de estudo do que aqueles que seriam gerados por um modelo que fosse indiferente a esta informação.

Para ilustrar a relevância atual do tema, alguns exemplos de estudos recentes nesta linha de pesquisa, aplicado na análise de potencial eólico de regiões brasileiras, são as referências [12] e [13].

Embora a utilização de modelos estocásticos proporcione maior flexibilidade de representação, algumas dificuldades adicionais são introduzidas.

Primeiramente, a abordagem usualmente utilizada para este tipo de representação é a abordagem paramétrica, que requer que as variáveis aleatórias sejam enquadradas previamente em alguma família de distribuição de probabilidade, sendo todo o desenvolvimento e ajuste dos modelos realizados sobre esta premissa[14]. Para as variáveis aleatórias representadas em estudos de sistemas de potência, tem-se, por

exemplo, que as famílias de distribuição mais adotadas são a distribuição log-normal, para representar vazões afluentes dos rios, [3] e [4], e a distribuição de Weibull, para representar a velocidade do vento [15]. Uma desvantagem deste tipo de abordagem é que muitas vezes uma mesma distribuição não é a melhor alternativa para caracterizar o conjunto de todas as variáveis daquele tipo representadas no sistema.

Em [16] é apresentado uma análise neste sentido para abordagem paramétrica dos modelos de vazões afluentes. Ao longo deste trabalho serão apresentadas análises mais detalhadas da desvantagem da abordagem paramétrica nos modelos estocásticos de vento.

Um segundo problema é que os modelos estocásticos de séries temporais, em geral, representam a dependência estatística entre os elementos variantes no tempo por meio da correlação linear [17] e [18]. Conforme será mais bem discutido ao longo deste trabalho, a correlação linear não é uma medida suficientemente precisa para explicar o relacionamento entre as variáveis aleatórias de interesse deste trabalho, devido à presença de relações não lineares, principalmente quando além da representação do módulo da velocidade do vento, também se faz necessária a representação da incerteza na direção do vento.

Ainda associado ao problema da representação da dependência estatística entre os elementos variantes no tempo, o aumento da dimensionalidade das variáveis representadas tende a fazer com que o número de parâmetros do modelo cresça exponencialmente. A rigor, um modelo auto-regressivo multivariado precisa considerar as correlações cruzadas entre as variáveis [18], ou seja, precisa representar a dependência de cada variável em um determinado instante em relação não apenas a ela mesma, mas também em relação às outras variáveis, nos instantes anteriores, o que torna a representação ainda mais complexa.

Outro ponto de desvantagem que pode ser citado é a dificuldade para se representar a dependência estatística entre elementos variantes no tempo que possuem níveis diferentes de discretização. Um exemplo disto é a representação da correlação entre a geração eólica e as vazões afluentes dos rios. A incerteza intrínseca das vazões afluentes possui uma dinâmica muito mais lenta que a incerteza da geração eólica, sendo usualmente as vazões medidas e modeladas semanalmente ou mensalmente, enquanto que a geração eólica requer representação horária, ou até medidas em intervalos menores de tempo.

Por fim, no que tange aos métodos de simulação estocástica utilizados na avaliação da confiabilidade, tem-se a desvantagem adicional de ser necessária a geração a priori de todas as amostras de séries sintéticas utilizadas durante a Simulação Monte

Carlo. Embora não seja uma desvantagem tão significativa quanto os outros pontos enumerados, o conjunto de dados sintéticos gerados pode ser bem volumoso, dependendo do número de variáveis representadas, o que afeta diretamente alguns aspectos computacionais práticos da implementação do modelo, como limitações de memória RAM, uso extensivo de acesso I/O a disco, e outras questões relacionadas à eficiência da implementação computacional do modelo.

1.2 Motivação e Objetivos do Trabalho

Dentro do contexto apresentado, este trabalho propõe um novo modelo estocástico para representação da incerteza dos elementos variantes no tempo, preservando a dependência estatística entre eles, na avaliação da confiabilidade pela Simulação Monte Carlo Não Sequencial, com foco principal na modelagem da geração eólica, embora não limitado a este tipo de variável.

As principais características do modelo proposto, que juntas constituem a motivação para o desenvolvimento deste modelo, são:

1. Representação não paramétrica dos elementos variantes no tempo;

A abordagem convencional paramétrica exige que se assumam a priori famílias de distribuição de probabilidade para os elementos variantes no tempo, o que pode não ser trivial ou até mesmo possível à medida que se representam variáveis aleatórias de diferentes naturezas. Neste trabalho, é proposta uma representação mais geral, baseada na abordagem não paramétrica de representação de funções de distribuição de probabilidade, de modo a contornar esta limitação.

2. Representação de dependência estatística não linear entre as variáveis;

As abordagens multivariadas para os modelos estocásticos restringem-se, em geral, à representação apenas da correlação linear entre as variáveis, não havendo representação dos relacionamentos estatísticos não lineares. Desta forma, neste trabalho é proposta também a representação deste tipo de relacionamento, de modo que qualquer dependência estatística entre os elementos variantes no tempo seja reproduzida na simulação estocástica dos sistemas.

3. Representação de sistemas de alta dimensionalidade, sem que o modelo esteja sujeito à maldição da dimensionalidade;

A maldição da dimensionalidade faz referência ao fato de que há um aumento exponencial da complexidade de um problema devido ao aumento de sua dimensionalidade. Este é o principal problema de se representar a dependência

estatística entre variáveis aleatórias em um sistema de grande porte. Desta maneira, neste trabalho este ponto é tratado, sendo proposta a aplicação de uma abordagem que permite evitar este aumento descontrolado da complexidade do modelo na representação de sistemas de alta dimensionalidade.

4. Flexibilidade para representação de variáveis exógenas nos processos estocásticos dos elementos variantes no tempo;

Motivado principalmente pela forte dependência que os elementos variantes no tempo em sistemas de potência têm com o clima, neste trabalho é proposto um modo de introduzir informações de variáveis exógenas ao processo no modelo estocástico.

5. Flexibilidade para representação de dependência estatística entre variáveis com níveis de discretização diferentes.

Motivado principalmente pela necessidade de se capturar a correlação entre as séries de vazões afluentes e as séries de geração eólica, neste trabalho também é proposta uma maneira de se incorporar esta representação no modelo estocástico.

Destaca-se, ainda, que o algoritmo proposto para Simulação Monte Carlo Não Sequencial dispensa a necessidade de se gerar previamente séries sintéticas para os elementos variantes no tempo representados, assim como também não é necessária a prévia geração de séries de estados operativos para os componentes convencionais, como é exigido nos métodos tradicionais.

1.3 Revisão Bibliográfica

Dentre os primeiros trabalhos que propuseram a aplicação da Simulação Monte Carlo para avaliação da confiabilidade de sistemas de potências, destacam-se [19], [20], [21], [22], [23], nos quais se começou a explorar a confiabilidade no nível da geração, transmissão e sub-transmissão, bem como a representação de eventos múltiplos de falhas.

Em [24] é apresentada uma proposta de modelo para representação de uma usina eólica em um estudo de confiabilidade, sendo construída uma função de distribuição de probabilidade para a representação da incerteza da geração total da usina. A carga do sistema também é representada a partir de uma função de distribuição de probabilidade, sem representação de correlação entre as duas variáveis.

Em [25] é apresentada uma proposta de modelo para representação de usinas eólicas, considerando a possibilidade de dependência estatística entre elas, sendo esta modelada a partir da correlação linear. A partir da definição desta correlação, o trabalho

apresenta o desenvolvimento analítico das distribuições de probabilidade para se amostrar a geração das usinas.

Em [26] é apresentada uma proposta para avaliação da confiabilidade da geração em múltiplas áreas, sendo a dependência estatística das cargas destas áreas modeladas a partir da correlação linear.

Em [27] é apresentada uma proposta para representação de cargas variantes no tempo na avaliação da confiabilidade, a partir da utilização da Simulação Monte Carlo Sequencial. É aplicada uma técnica de redução da variância com o objetivo de reduzir o tempo necessário para a convergência da simulação.

Em [28] é apresentado o método da Simulação Monte Carlo Pseudo-Sequencial, que permite a obtenção dos índices de confiabilidade com a mesma precisão da Simulação Sequencial, mas com um esforço computacional bastante reduzido. O método consiste em amostrar os estados de adequação de forma não sequencial, mas avaliando de forma sequencial os estados subsequentes aos estados de falha do sistema. A carga variante no tempo é representada a partir de uma série cronológica que é combinada com as séries sintéticas dos demais elementos do sistema.

Em [29] é apresentado o método da Simulação Monte Carlo Pseudo-Cronológica, que difere da representação da Simulação Monte Carlo Pseudo-Sequencial principalmente no modo como representa o comportamento das cargas do sistema e de suas áreas, uma vez que propõe para este fim a utilização de um modelo de Markov não-agregado a múltiplos níveis. O principal ganho de desempenho da simulação pseudo-cronológica está associada a utilização do método de estimação dos índices de frequência e duração pelo processo de transição um passo a frente, que requer apenas a avaliação da adequação de um estado adicional para cada estado de falha.

Em [30] é apresentado um método para se obter a representação da geração eólica através da representação equivalente por meio de um modelo de geração convencional. O método no entanto é aplicado apenas a usinas eólicas individualmente, sem representar a correlação entre diferentes usinas eólicas.

Em [31] é apresentada uma proposta para a representação da geração eólica na Simulação Monte Carlo Sequencial, aplicando-se o modelo *ARMA* para representar as séries de vento, mas sendo assumido que estas possuem distribuição de probabilidade normal, de modo que a dependência estatística entre as usinas seja representada a partir da correlação linear.

Em [32] e em [33] são apresentadas propostas de representação de usinas eólicas na avaliação da confiabilidade baseadas em cadeias de Markov. Em [32] é proposta a

representação da correlação entre diferentes usinas a partir da representação dos estados conjuntos das usinas nesta cadeia de Markov, o que limita a aplicação do método a sistemas com número reduzido de usinas.

Em [34] é proposto o método de Simulação Monte Carlo Quase-Sequencial para a avaliação da confiabilidade de sistemas com representação de elementos variantes no tempo, incluindo as cargas variantes no tempo e a geração eólica, não sendo, no entanto, apresentado nenhum tratamento para a representação da dependência estatística entre estes elementos. O algoritmo é baseado na Simulação Monte Carlo Pseudo-Cronológica, mas com uma proposta mais simples de implementação.

Em [35] é proposta a aplicação do método de Simulação Monte Carlo com Entropia Cruzada na avaliação da confiabilidade da geração. O método consiste em utilizar uma função de densidade de probabilidade de *importance sampling* auxiliar para otimizar as amostras dos eventos raros, de modo a se reduzir a variância das estimativas dos índices de confiabilidade, acelerando o processo de convergência. Neste trabalho apenas as cargas variantes no tempo e as unidades convencionais são modeladas, sem a representação da geração eólica.

Em [36] é proposta a aplicação do método de Simulação Monte Carlo com Entropia Cruzada na avaliação da confiabilidade composta, sendo observado um ganho considerável de custo computacional em relação às demais abordagens na literatura, devido a diminuição da variância da estimação na convergência do processo. A representação da geração eólica consiste na utilização de três séries temporais pré-definidas, caracterizadas em cenário médio, favorável e desfavorável, sendo também pré-definidas três diferentes probabilidades para a ocorrência destes cenários.

Em relação aos objetivos deste trabalho, destacam-se na bibliografia revisada duas categorias de publicação. A primeira categoria corresponde aos trabalhos que propõem métodos e algoritmos que têm como objetivo possibilitar a representação dos aspectos cronológicos, mas com esforço computacional inferior a Simulação Monte Carlo Sequencial. Estes trabalhos não enfocam o desenvolvimento dos modelos estocásticos para a representação dos elementos variantes no tempo, sendo os cenários estabelecidos pelas séries sintéticas previamente gerados para o algoritmo de simulação. Desta maneira, os problemas elencados para a representação da dependência estatística entre os elementos variantes no tempo, bem como as demais características estatísticas das séries temporais, não são efetivamente tratados.

A segunda categoria corresponde aos trabalhos que propõem efetivamente o desenvolvimento de modelos estocásticos para a representação dos elementos variantes no tempo de interesse deste trabalho. No entanto, estas propostas apresentam limitações

devido à necessidade de se arbitrar as funções de densidade de probabilidade para caracterizar os elementos variantes no tempo, a impossibilidade de se representar relacionamentos não lineares entre as variáveis envolvidas, e ainda, não tratam o problema da dimensionalidade dos sistemas, que pode se tornar crítico na representação de muitos elementos variantes no tempo estatisticamente dependentes.

Desta maneira, as principais contribuições do modelo proposto neste trabalho são no sentido de contornar estas limitações: utilizar uma abordagem aderente à Simulação Monte Carlo Não Sequencial de modo a se manter o reduzido custo computacional, mas provendo uma solução mais flexível e geral para a representação dos elementos variantes do tempo.

1.4 Trabalhos Publicados

Os seguintes trabalhos foram publicados durante a realização desta pesquisa:

Borges, C. L. T., Dias, J. A. S., “Impact of Large Penetration of Correlated Wind Generation on Power System Reliability”, In: Hossain, J., Mahmud, A., Large Scale Renewable Power Generation, 1 ed, Springer Singapore, pp. 273-287. 2014.

Borges, C. L. T. ; Dias, J. A. S. “Representation of Wind and Load Correlation in Non-Sequential Monte Carlo Reliability Evaluation.” In: Billinton, Roy; Karki, Rajesh; Verma, A. K.. (Org.). Reliability and Risk Evaluation of Wind Integrated Power Systems. 1ed. New Delhi: Springer, pp. 91-106, 2013.

Borges, C. L. T., Dias, J. A. S., “A Non Parametric Stochastic Model for River Inflows Based on Kernel Density Estimation”, In: 13th International Conference on Probabilistic Methods Applied to Power Systems, Durham, 2014.

Dias, J. A. S., Borges, C. L. T., “Representation of Wind and Load Correlation in Non-Sequential Monte Carlo Reliability Evaluation.” In: 12th International Conference on Probabilistic Methods Applied to Power Systems, Istanbul, 2012.

Dias, J. A. S., Borges, C. L. T. “Object Oriented Model for Composite Reliability Evaluation Including Time Varying Load and Wind Generation.” In: 11th International Conference on Probabilistic Methods Applied to Power Systems, Cingapura, 2010.

Dias, J. A. S., Borges, C. L. T., “Modelo Orientado a Objetos para Avaliação da Confiabilidade Composta por Simulação Monte Carlo com Representação da Geração Eólica”, Controle & Automação, v. 20, n. 3, p. 359-372, 2009.

1.5 Organização do Trabalho

Este trabalho foi organizado com a seguinte estrutura:

No Capítulo 2 é apresentada a formulação básica da Simulação Monte Carlo, tanto Sequencial quanto Não Sequencial, e introduzidas as metodologias da Simulação Monte Carlo Pseudo-Sequencial e Simulação Monte Carlo Pseudo-Cronológica, principais metodologias de “benchmarking” para o modelo proposto.

No Capítulo 3 são apresentados os conceitos básicos, ferramentas estatísticas e formulações fundamentais necessários para a construção do modelo proposto, destacando-se a apresentação das formulações do método de estimação de função de densidade de probabilidade não paramétrica por *kernel density estimation* e das redes Bayesianas.

No Capítulo 4 é apresentada uma análise estatística sobre os elementos variantes no tempo de maior interesse para este trabalho, a geração eólica, destacando-se as propriedades que tornam interessante a aplicação do modelo proposto neste trabalho.

No Capítulo 5 é apresentada toda formulação e desenvolvimento do modelo proposto.

No Capítulo 6 são apresentadas as simulações e resultados do modelo proposto aplicado a um sistema teste da literatura e ao sistema real brasileiro.

No Capítulo 7 são apresentadas as conclusões e sugestões de continuidade da pesquisa desenvolvida.

Capítulo 2

Avaliação da Confiabilidade por Simulação Monte Carlo

2.1 Introdução

A avaliação da confiabilidade de sistemas de potência é um estudo que mensura a capacidade do sistema, ou componente do sistema, de desempenhar a sua função de atendimento ao consumidor de forma adequada, isto é, dentro das condições operativas padronizadas. A fim de atingir este objetivo, planos de investimento e estratégias de operação precisam prever a possibilidade de ocorrência de falhas dos componentes e outros distúrbios no sistema, para estimar se a qualidade e continuidade do suprimento, medidas por vários índices de confiabilidade, permanecerão dentro de níveis satisfatórios[1].

Neste capítulo são apresentados os conceitos básicos da avaliação de confiabilidade de sistemas de potência e os métodos clássicos da literatura, baseados na Simulação Monte Carlo, utilizados nesta análise.

2.1.1 Níveis Hierárquicos

Na análise de confiabilidade, o sistema elétrico pode ser representado em três níveis hierárquicos, conforme apresentado na Figura 1, dependendo da abrangência do estudo:

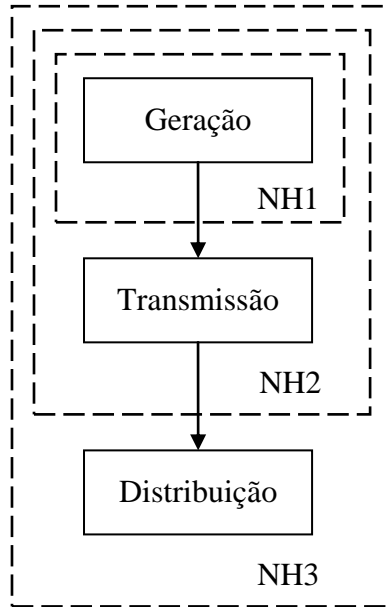


Figura 1– Níveis Hierárquicos para a Confiabilidade

1. **Nível Hierárquico 1 (NH1):** abrange o estudo da confiabilidade a nível de geração, onde o principal interesse é a determinação da capacidade estática da geração. Nesta consideração, toda a carga e toda a geração são concentradas em uma única barra e as limitações da transmissão são ignoradas, conforme apresentado na Figura 2.

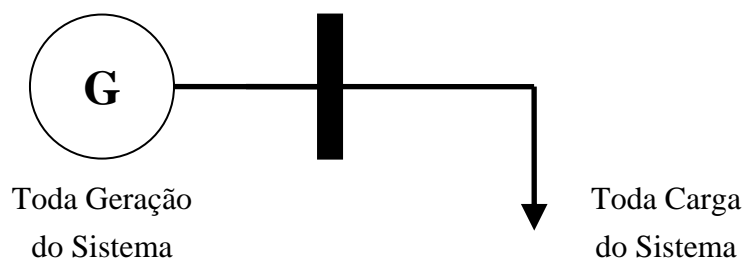


Figura 2– Representação de um Sistema no NH1

2. **Nível Hierárquico 2 (NH2):** abrange o estudo da confiabilidade a nível de geração e transmissão, sendo a distribuição representada como pontos de cargas equivalentes, conforme o sistema hipotético apresentado na Figura 3. Nesta representação são analisados os efeitos das falhas de geração, transmissão e a

composição de ambos. Este estudo é também chamado de confiabilidade composta e será abordado neste trabalho.

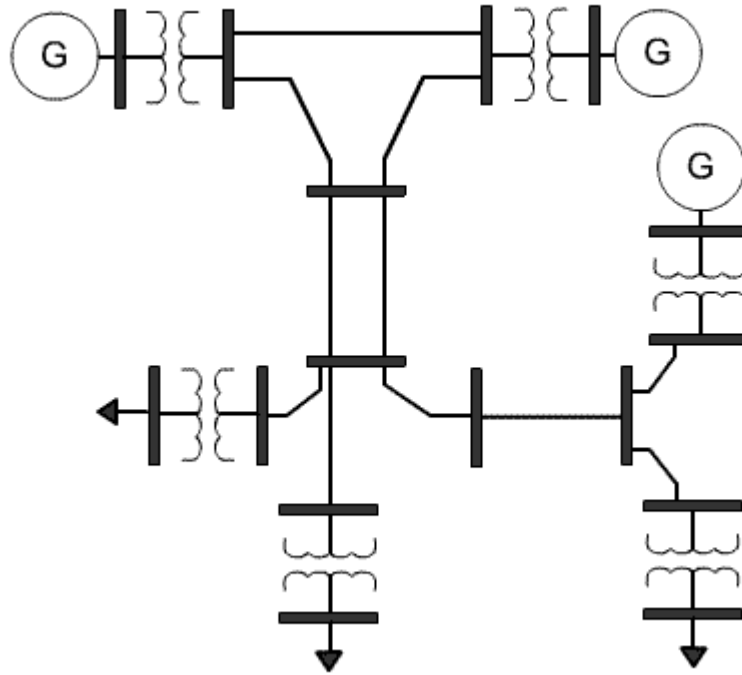


Figura 3– Representação de um Sistema Hipotético no NH2

3. **Nível Hierárquico 3 (NH3):** abrange o estudo da confiabilidade a nível de geração, transmissão e distribuição, conforme apresentado na Figura 4, onde o elemento *GD* representa a possibilidade da modelagem de geração distribuída no estudo. Em geral esta representação mais completa é evitada em sistemas reais, devido à elevada dimensão do problema, sendo a confiabilidade da distribuição tratada separadamente dos sistemas de geração e transmissão.

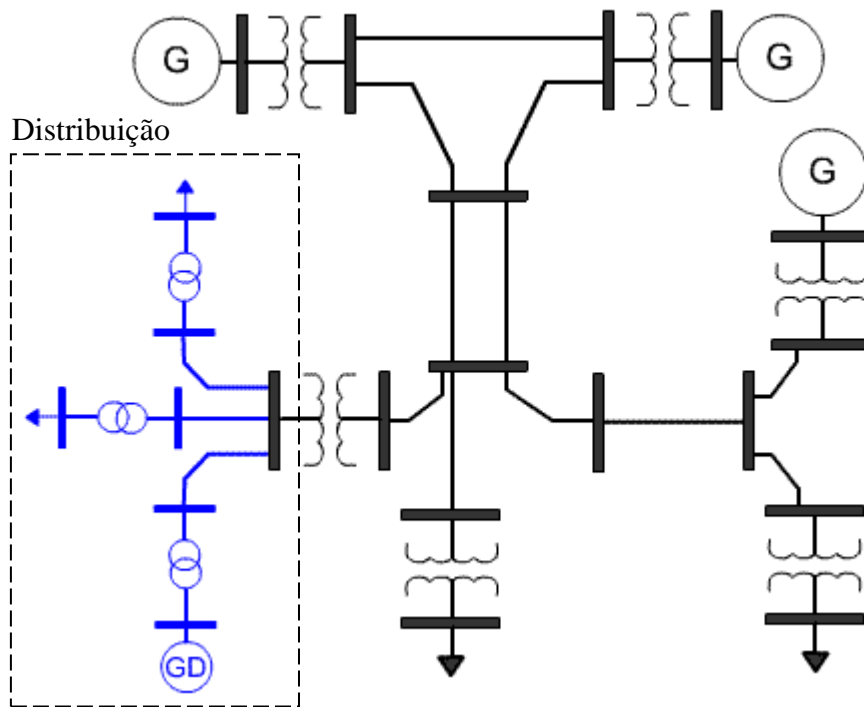


Figura 4– Representação de um Sistema Hipotético no NH3

2.2 Representação de Componentes por Modelos de Markov

O sistema elétrico de potência é composto por um conjunto dos mais variados componentes, tais como geradores, cargas, transformadores, linhas de transmissão, entre outros. Um estado do sistema consiste na realidade da combinação dos estados de operação de cada um destes componentes [37]. Conforme será apresentado a seguir, processos de Markov têm sido utilizados na avaliação de confiabilidade para modelar o comportamento de equipamentos e sistema.

Os processos de Markov são processos estocásticos que apresentam a particularidade, chamada de memória markoviana, em que os estados passados são irrelevantes para a predição do estado futuro, desde que o estado atual seja conhecido. Considerando-se os instantes ordenados de forma crescente: $t_1 < t_2 < \dots < t_n$, associados a um conjunto de variáveis aleatórias: X_1, X_2, \dots, X_n , tem-se:

$$F_{X_n|X_1X_2\dots X_{n-1}} = F_{X_n|X_{n-1}} \quad (2.1)$$

Se a probabilidade de transição entre estados independe do instante t inicial, mas apenas do intervalo de tempo Δt considerado, o processo de Markov é caracterizado como um processo homogêneo.

Os processos de Markov homogêneos também possuem a característica de serem processos estacionários, isto é, a probabilidade de transição de um estado para o outro é constante durante todo o processo estocástico.

Problemas de confiabilidade são normalmente modelados como discretos no espaço e contínuos no tempo. Os componentes permanecem em um dos estados, até que uma transição para outro estado ocorra, onde eles residirão até a próxima transição[37].

A modelagem a múltiplos estados é o caso mais geral para a representação de modelos estocásticos de componentes, sendo utilizado para a representação de usinas com vários patamares de geração, curvas de carga ou ainda componentes que possuam modos de saídas dependentes, como pode ser visto nos exemplos apresentados na Figura 5.

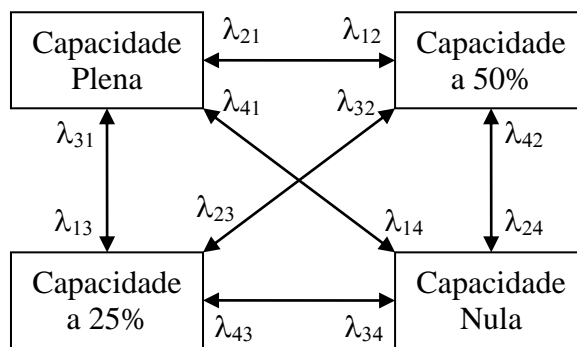
A modelagem de componentes pelo método de Markov considera que as probabilidades de transição entre os estados são constantes durante todo o processo. Considerando-se i e j como sendo estados quaisquer do sistema, estas probabilidades podem ser definidas como:

$$P[X(t + \Delta t) = j | X(t) = i] = p_{ij}(t + \Delta t) \quad (2.2)$$

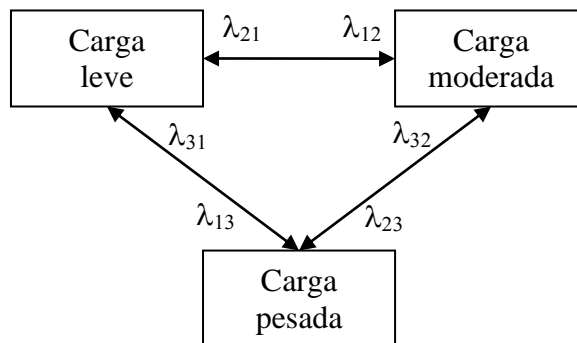
A soma das probabilidades de transição de um estado para outro qualquer, incluindo ele mesmo, deve ser igual a 1, logo:

$$p_{ii}(\Delta t) + \sum_{j \neq i} p_{ij}(\Delta t) = 1 \quad (2.3)$$

a) Representação de uma usina com 4 estados de geração:



b) Representação de carga em 3 patamares:



c) Representação de saídas dependentes para dois componentes:

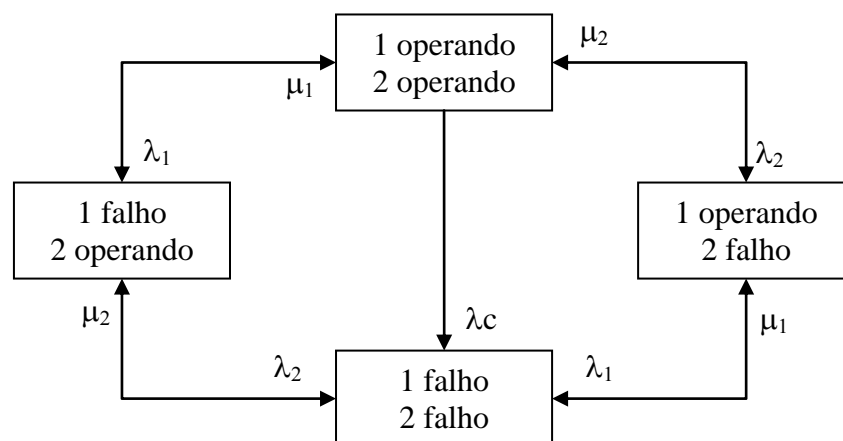


Figura 5– Exemplos de Modelos a Múltiplos Estados

Definindo-se q_{ij} e q_{ii} como sendo as intensidades de transição, constantes durante todo o processo:

$$q_{ij} = \lim_{\Delta t \rightarrow 0} \frac{p_{ij}(\Delta t)}{\Delta t} \quad (2.4)$$

$$q_{ii} = \lim_{\Delta t \rightarrow 0} \frac{1 - p_{ii}(\Delta t)}{\Delta t} \quad (2.5)$$

As probabilidades de transição podem ser definidas utilizando-se as intensidades de transição:

$$p_{ij}(t + \Delta t) = q_{ij}\Delta t \quad (2.6)$$

$$p_{ii}(t + \Delta t) = 1 - q_{ii}\Delta t \quad (2.7)$$

A partir de (2.3), (2.4) e (2.5), tem-se:

$$q_{ii} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \sum_{i \neq j} p_{ij}(\Delta t) = \sum_{i \neq j} q_{ij} \quad (2.8)$$

As probabilidades não condicionadas de estados de $X(t)$ podem ser calculadas através de:

$$p_i(t) = P[X(t) = i] \quad (2.9)$$

Se as probabilidades de estados são conhecidas no instante t , as probabilidades em $t + \Delta t$ podem ser calculadas através de:

$$p_i(t + \Delta t) = p_i(t)p_{ii}(\Delta t) + \sum_{j \neq i} p_j(t)p_{ji}(\Delta t) \quad (2.10)$$

Definindo em termos das intensidades de transição:

$$p_i(t + \Delta t) = p_i(t)(1 - q_{ii}\Delta t) + \sum_{j \neq i} p_j(t)q_{ji}\Delta t \quad (2.11)$$

Reorganizando os termos:

$$\frac{p_i(t + \Delta t) - p_i(t)}{\Delta t} = -p_i(t)q_{ii} + \sum_{j \neq i} p_j(t)q_{ji} \quad (2.12)$$

Tomando-se $\Delta t \rightarrow 0$:

$$\dot{p}_i = \frac{dp_i(t)}{dt} = -p_i(t)q_{ii} + \sum_{j \neq i} p_j(t)q_{ji} \quad (2.13)$$

A equação (2.13) pode ser escrita de forma matricial, agrupando todos os estados em vetores:

$$\underline{\dot{p}}(t) = \underline{p}(t)A \quad (2.14)$$

Onde A é denominada Matriz de Intensidade de Transição, independente do tempo, sendo definida como:

$$A = \begin{bmatrix} -q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & -q_{22} & \cdots & q_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ q_{n1} & q_{n2} & \cdots & -q_{nn} \end{bmatrix} \quad (2.15)$$

Para as aplicações de interesse deste trabalho, consideram-se importantes apenas as probabilidades de longo prazo ou probabilidades de regime permanente. O vetor de probabilidades de longo prazo \underline{p} é calculado resolvendo-se:

$$\underline{p}A = 0 \quad (2.16)$$

A resolução direta de (2.16) não é possível, pois as n equações que constituem o sistema não são linearmente independentes. Uma nova equação, linearmente independente das demais, é dada por:

$$\sum_{i=1}^n p_i = 1 \quad (2.17)$$

Desta forma o vetor \underline{p} pode ser calculado resolvendo-se:

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}^T \begin{bmatrix} -q_{11} & q_{12} & \cdots & q_{1n} & 1 \\ q_{21} & -q_{22} & \cdots & q_{2n} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ q_{n1} & q_{n2} & \cdots & -q_{nn} & 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ 1 \end{bmatrix}^T \quad (2.18)$$

2.3 Simulação Monte Carlo Sequencial

Na Simulação Monte Carlo Sequencial os estados amostrados preservam as características cronológicas do sistema, existindo a preocupação de se representar corretamente as transições entre os estados consecutivos do sistema, bem como de se reproduzir as curvas de carga e geração, quando existentes.

O cálculo dos índices de confiabilidade pode ser expresso pela avaliação da seguinte expressão:

$$\bar{E}(G) = \frac{1}{N} \sum_{k=1}^N G(y_k) \quad (2.19)$$

onde N é o número de séries sintéticas simuladas, y_k é a série sintética composta pelos estados amostrados sequencialmente dentro do período k e G é uma função específica para o cálculo dos índices de confiabilidade para a série y_k .

O termo $\bar{E}(G)$ fornece uma estimativa dos índices de confiabilidade calculados sobre todas as séries sintéticas simuladas.

Assumindo-se que y_k é anual, um algoritmo conceitual para avaliação da confiabilidade composta a partir da Simulação Monte Carlo Sequencial:

1. Gere uma série sintética anual de estados do sistema y_k , amostrando-se sequencialmente no tempo os estados e a duração dos estados do sistema;
2. Analise a adequação do estado x da série y_k a partir de uma função de avaliação $F(x)$, acumulando-se os resultados relativos ao ano k ;
3. Calcule os índices de confiabilidade anuais relativos ao ano pela avaliação da função $G(y_k)$ sobre os valores acumulados no passo 2;
4. Atualize a estimativa de $\bar{E}(G)$, o valor esperado dos índices de confiabilidade do processo, baseado no resultado obtido no passo 3;
5. Se a precisão das estimativas é aceitável, o processo é dito convergido, caso contrário retorne ao passo 1.

A série sintética anual do passo 1 do algoritmo é gerada pela combinação dos processos de transição de estados dos componentes do sistema e da variação cronológica dos elementos variantes no tempo. O processo de transição de estados dos componentes é obtido pela amostragem sequencial da distribuição de probabilidade da duração dos estados.

2.4 Simulação Monte Carlo Não Sequencial

Na Simulação Monte Carlo Não-Sequencial os estados do sistema são selecionados por amostragem do espaço de estados, não se preocupando com a cronologia dos eventos. Cada componente do sistema elétrico possui um modelo estocástico que descreve os seus possíveis estados e probabilidades de operação.

Em [38] é sugerido um algoritmo conceitual para avaliação da confiabilidade composta a partir da SMC Não Sequencial:

1. Selecione um estado do sistema $x \in X$ correspondente a um ponto de operação, isto é, disponibilidade de componentes, níveis de carga e geração, etc;
2. Calcule o valor de uma função de avaliação $F(x)$;

3. Atualize a estimativa $\bar{E}(F)$, o valor esperado dos índices de confiabilidade do processo, baseado no resultado obtido no passo 2;
4. Se a precisão das estimativas é aceitável, o processo é dito convergido, caso contrário retorne ao passo 1.

Na abordagem não-sequencial, os estados do sistema selecionados no passo 1 do algoritmo, dependem da combinação dos estados de todos os seus componentes, sendo estes determinados pelo sorteio da probabilidade de cada componente aparecer em cada um de seus respectivos estados. Estas probabilidades são descritas pelo vetor \underline{p} , obtido a partir da equação (2.18).

O estado do componente pode ser amostrado pelo método de transformação inversa, através da geração de um número aleatório U_j distribuído uniformemente entre $[0,1]$, de modo que o estado do componente será dado por:

$$x_j = \begin{cases} 1 & \text{se } 0 \leq U_j < P_1 \\ 2 & \text{se } P_1 \leq U_j < P_2 \\ \vdots & \quad \quad \quad \vdots \\ n & \text{se } P_{n-1} \leq U_j \leq 1 \end{cases} \quad (2.20)$$

2.4.1 Índices de Confiabilidade Composta

Os índices básicos na avaliação da confiabilidade composta de sistemas são:

1. LOLP (Loss of Load Probability) – Probabilidade de perda de carga:

$$LOLP = \sum_{i \in S} p_i \quad (2.21)$$

p_i -> probabilidade do estado i do sistema;

S -> conjunto de todos os estados do sistema associados com corte de carga.

2. LOLE (Loss of Load Expectation) – Número esperado de horas de déficit de potência (horas/ano):

$$LOLE = \sum_{i \in S} p_i T \quad (2.22)$$

T -> período de tempo de análise (em geral, 1 ano, $T = 8760$).

3. EPNS (Expected Power Not Supplied) – Duração média de perda de carga (MW):

$$EPNS = \sum_{i \in S} C_i p_i \quad (2.23)$$

C_i -> corte de carga do sistema no estado i .

4. EENS (Expected Energy Not Supplied) – Valor esperado de energia não suprida (MWh):

$$EENS = T \sum_{i \in S} C_i p_i \quad (2.24)$$

5. LOLF (Loss of Load Frequency) – Frequência de perda de carga (ocorrências/ano):

$$LOLF = \sum_{i \in S} (F_i - f_i) \quad (2.25)$$

F_i -> frequência de saída do estado i do sistema;

f_i -> porção de f_i que corresponde a não ultrapassar a barreira entre o conjunto de estados com perda de carga e o conjunto de estados sem perda de carga.

6. LOLD (Loss of Load Duration) – Duração média de perda de carga (horas):

$$LOLD = \frac{LOLE}{LOLF} \quad (2.26)$$

Estes índices podem ser calculados a nível de sistema ou individualmente para cada barra.

Os índices LOLP, LOLE, EPNS e EENS constituem o subconjunto dos índices de probabilidade e energia enquanto que os índices LOLF e LOLD constituem o subconjunto dos índices de frequência e duração.

2.4.2 Convergência dos Índices de Confiabilidade

A maneira padrão de se testar a convergência dos índices de confiabilidade é calcular a incerteza em torno da estimativa do índice:

$$V(F) = \frac{1}{N} \sum_{i=1}^N (F(x_i) - \bar{E}(F))^2 \quad (2.27)$$

Sendo a variância da estimativa da expectância dada por:

$$V(\bar{Q}) = \frac{1}{N} V(F) \quad (2.28)$$

O coeficiente de variação da estimativa é então uma medida da sua incerteza relativa, valendo:

$$\alpha = \frac{\sqrt{V(\bar{Q})}}{\bar{Q}} \quad (2.29)$$

Conforme o desenvolvimento apresentado em [39], a partir da expressão (2.28) é possível estabelecer os limites superior e inferior da estimativa do valor esperado da função-teste. Pelo Teorema Central do Limite sabe-se que, para uma amostra suficientemente grande, $E(F)$ tende a possuir uma distribuição de probabilidades Normal, independente de qual for a verdadeira distribuição $P(x)$ dos estados. Logo, é possível calcular os limites superior e inferior para o valor verdadeiro, com um determinado grau de confiança. Por exemplo, pode-se afirmar, com 95% de confiança, que o valor verdadeiro de $E(F)$ está no intervalo da seguinte expressão:

$$\left[E(F) - 1.96\sqrt{V(E(F))}, E(F) + 1.96\sqrt{V(E(F))} \right] \quad (2.30)$$

2.5 Representação de Elementos Variantes no Tempo

A introdução de elementos variantes no tempo dificulta consideravelmente a modelagem dos problemas de confiabilidade devido ao menos a dois fatores principais: a correlação que pode existir entre estas variáveis e a grande incerteza acerca destes elementos. A correlação entre as variáveis pode ser tanto espacial quanto temporal, e em geral estes dois efeitos são mutuamente observados na representação dos elementos variantes no tempo de um sistema de potência: demandas, vazões, geração eólica, etc. As incertezas a respeito dos elementos variantes no tempo exige que diferentes cenários sejam avaliados. A explícita representação de diferentes cenários para os elementos variantes no tempo advém do fato de que provavelmente o real perfil destas variáveis seria desconhecido para o horizonte de estudo, e, portanto, uma análise mais rigorosa precisaria ponderar o impacto dos diferentes cenários de evolução destas variáveis nas estatísticas finais do estudo.

As duas dificuldades adicionais citadas acima não são um real problema para a abordagem sequencial da Simulação Monte Carlo, mas para a abordagem não sequencial tradicional podem se tornar, e em geral se tornam, um limitante para a obtenção de resultados com um nível aceitável de precisão.

Com o objetivo de obter a precisão da Simulação Monte Carlo Sequencial na representação de elementos variantes no tempo com o custo computacional da

Simulação Monte Carlo Não Sequencial, alguns métodos são propostos na literatura, destacando-se a Simulação Monte Carlo Pseudo-Sequencial[28] e a Simulação Monte Carlo Pseudo-Cronológica[29], especificamente pela implementação do método de estimação dos índices de frequência e duração pelo processo de transição um passo a frente.

2.5.1 Simulação Monte Carlo Pseudo-Sequencial

A simulação pseudo-sequencial é um método híbrido em que a simulação não-sequencial é utilizada para selecionar os estados de falha do sistema e a simulação sequencial é aplicada às subsequências de falha, ou interrupções contínuas de energia, formadas pelos estados de falha vizinhos ao estado selecionado [40].

Este método tem como objetivo conservar a flexibilidade e a precisão da simulação sequencial e associá-la ao reduzido esforço computacional requerido pela simulação não sequencial no cálculo dos índices de confiabilidade. Para tal, foram consideradas duas características da simulação sequencial:

- O esforço computacional requerido para a construção das séries sintéticas de operação do sistema que é desprezível comparada com a análise de cada estado da série;
- A maioria dos estados não contribui para o cálculo dos índices de confiabilidade.

O procedimento da avaliação da confiabilidade através da simulação pseudo-sequencial é realizado nos seguintes passos [40]:

a) gere diversas séries sintéticas de operação do sistema a partir do mesmo processo utilizado pela simulação sequencial com amostragem da duração dos estados dos componentes do sistema;

b) sorteie uma série sintética de operação dentre todas as séries geradas no passo (a) que estão igualmente distribuídas;

c) sorteie um tempo t na série selecionada no passo (b);

d) avalie o estado x associado ao tempo t selecionado no passo (c). Se este estado é de falha, prossiga para o passo (e); caso contrário, retorne ao passo (b);

e) obtenha uma subsequência de falha (interrupção de energia) através das simulações *forward* e *backward*:

- execute uma simulação cronológica no sentido direto do tempo começando do estado selecionado até que seja encontrado um estado de sucesso, ou seja, avalie os próximos estados da série enquanto um estado de sucesso não é encontrado;

- execute uma simulação cronológica no sentido inverso do tempo começando do estado selecionado até que seja encontrado um estado de sucesso, ou seja, avalie os estados da série anteriores ao selecionado enquanto um estado de sucesso não é encontrado;

f) calcule as estimativas dos índices de confiabilidade, como na Simulação Monte Carlo Não Sequencial. Se a precisão destas estimativas for aceitável, pare; caso contrário, retorne ao passo (b).

A simulação pseudo-sequencial utiliza, para o cálculo da maioria dos índices de confiabilidade, as mesmas funções-teste usadas pela simulação não-sequencial. No entanto, a função-teste do índice de frequência de perda de carga (*LOLF*) deve ser definida em função da duração da interrupção determinada no passo (e) do algoritmo acima, como mostra a expressão (2.31).

$$F_{LOLF}(x) = \begin{cases} 0 & \text{se } x \text{ é um estado de sucesso} \\ \frac{1}{D_I} & \text{se } x \text{ é um estado de falha} \end{cases} \quad (2.31)$$

onde, D_I representa a duração da interrupção.

A eficiência computacional deste método está no fato de que as simulações *forward* e *backward* são realizadas apenas quando um estado de falha é sorteado. Contudo, existe, em relação a simulação não-sequencial, um esforço computacional extra que pode ser representado pelo número de simulações adicionais realizadas pelo algoritmo pseudo-sequencial. Este número é proporcional ao produto $LOLP \times (LOLD + 1)$, onde $LOLD + 1$ é o número médio de simulações em torno de cada estado de falha se cada estado do sistema tiver uma duração média de uma hora.

2.5.2 Simulação Monte Carlo Pseudo-Cronológica

Este método é uma variação da Simulação Monte Carlo Pseudo-Sequencial. A diferença entre estas metodologias está na representação dos comportamentos das cargas do sistema e de suas áreas (subsistemas). A simulação pseudo-cronológica utiliza o modelo de Markov não-agregado a múltiplos níveis, apresentado a seguir.

O principal ganho de desempenho da simulação pseudo-cronológica está associada a utilização do método de estimação dos índices de frequência e duração pelo processo de transição um passo a frente, que será também abordado.

Assim como a simulação pseudo-sequencial, a simulação pseudo-cronológica também foi originalmente formulada para a representação da carga. No entanto, a extensão do método para considerar os demais elementos variantes no tempo é trivial.

2.5.2.1 Modelo de Markov Não-Agregado a Múltiplos Níveis

Para manter informações cronológicas a respeito da carga, o modelo, ilustrado na Figura 6, é composto por um conjunto de T níveis múltiplos, conectados na mesma ordem em que aparecem no histórico da carga.

O modelo utiliza uma taxa de transição constante $\lambda L = 1/\Delta T$, onde ΔT representa a unidade de tempo usada para discretizar o período T . Para cada uma das m áreas consideradas, é fornecido o nível de carga por intervalo de tempo. Como exemplo, $L_h(A_m)$ corresponde ao nível de carga da hora h apresentado pela área m .

Como é ilustrado na Figura 6, quando a carga da área 1 transita do nível 1 para o nível 2, isto é, de $L_1(A_1)$ para $L_2(A_1)$, o mesmo ocorre para as demais áreas, ou seja, $L_1(A_2)$ transita para $L_2(A_2)$, ... e $L_1(A_m)$ transita para $L_2(A_m)$. Visto que todas as taxas de transição são iguais, as cargas permanecerão, em média, ΔT horas (por exemplo, 1 hora) em cada nível, assim como o período de análise terá, em média, T horas (por exemplo, 8760 horas).

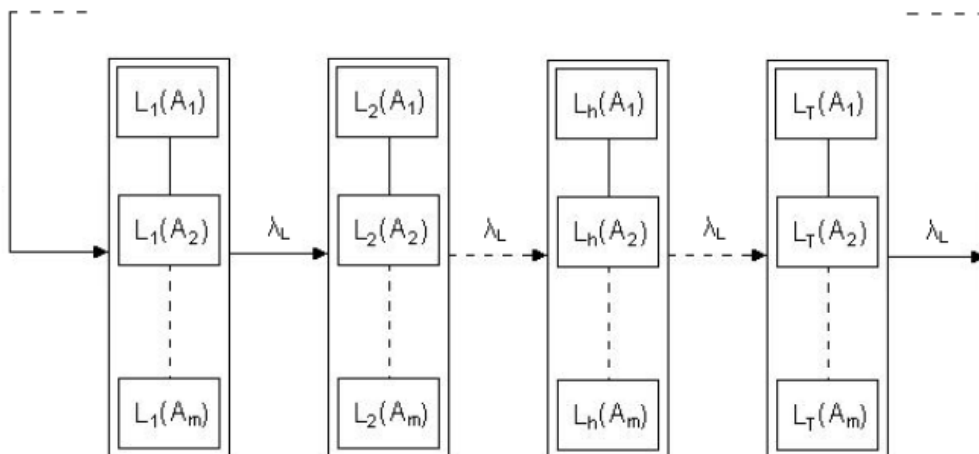


Figura 6– Modelo Markoviano não agregado de Múltiplos Níveis

2.5.2.2 Método de Estimação dos Índices de Frequência e Duração pelo Processo de Transição um Passo a Frente

A principal vantagem deste método é que este prescinde a hipótese de coerência do sistema.

A função teste da LOLF adotada neste método é definida como:

$$F_{LOLF}(x^k) = \begin{cases} 0 & \text{se } x^k \text{ é estado de sucesso} \\ \lambda_k^{out} & \text{se } x^k \text{ é estado de falha e } x^m \text{ é estado de sucesso} \\ 0 & \text{se } x^k \text{ é estado de falha e } x^m \text{ é estado de falha} \end{cases} \quad (2.32)$$

Onde,

λ_k^{out} é o somatório das taxas de transição do estado de falha x^k para todos os estados diretamente ligados a ele;

x^m é um estado qualquer, obtido a partir da transição de um único componente em x^k .

Considerando que cada transição de estado do sistema corresponde a transição individual de um único componente, cuja taxa de transição entre os estados inicial e final, a e b , vale λ_{ab} , pode-se definir a frequência com que o sistema transita de x^k para x^m como:

$$f_{km} = P(x^k) \times \lambda_{ab} = P(x^k) \times \lambda_{km} \quad (2.33)$$

Dado que a frequência de saída do estado x^k é dada por:

$$f_k^{out} = P(x^k) \times \lambda_k^{out} \quad (2.34)$$

A probabilidade de ocorrência da transição km , pode ser obtida a partir da razão entre as frequências f_{km} e f_k^{out} :

$$P_{km} = \frac{f_{km}}{f_k^{out}} = \frac{\lambda_{km}}{\lambda_k^{out}} \quad (2.35)$$

Sendo λ_k^{out} obtido pelo somatório de todas as possíveis transições individuais de cada componente do sistema:

$$\lambda_k^{out} = \sum_{i=1}^{MT} \lambda_{ki} \quad (2.36)$$

Onde,

MT é o conjunto de estados para os quais o sistema pode transitar a partir de x^k .

Desta maneira, para cada estado de falha do sistema x^k , é construída uma função de distribuição de probabilidade acumulada, utilizada para amostrar o estado, correspondente a um passo a frente de x^k , que deve ser avaliado, x^m .

2.6 Considerações Finais

Os métodos híbridos, Simulação Monte Carlo Pseudo-Sequencial e Simulação Monte Carlo Pseudo-Cronológico, diferem da Simulação Monte Carlo Não Sequencial tradicional, principalmente, por não representarem os elementos variantes no tempo por meio de suas distribuições de probabilidade, e sim pelo mesmo conjunto de séries cronológicas aplicadas à Simulação Sequencial. Além disso, os algoritmos conceituais destes dois métodos não abrangem os procedimentos para a geração destas séries cronológicas, sendo, em geral, ou adotadas as próprias séries históricas dos elementos variantes no tempo, ou então utilizadas séries sintéticas geradas a partir de modelos estocásticos externos a simulação. Desta maneira, embora estes métodos de simulação garantam a preservação da cronologia das séries temporais utilizadas, a qualidade dos cenários gerados, que compõe os cenários de adequação avaliados durante a simulação, não depende do método de simulação, e sim destes procedimentos externos utilizados para fazer a estimação e geração das séries sintéticas.

Em relação a estes métodos, a principal diferença do modelo proposto é que ele define de fato um modelo estocástico para os elementos variantes no tempo, que propõe alternativas para se contornar algumas limitações dos modelos estocásticos tradicionais, aplicados na geração das séries sintéticas que alimentam tanto a Simulação Sequencial quanto os demais métodos híbridos.

Adicionalmente, por corresponder a uma representação das distribuições de probabilidade dos elementos variantes no tempo, como será apresentado posteriormente, o modelo pode ser incorporado diretamente no algoritmo da Simulação Monte Carlo Não Sequencial, sem a necessidade de se gerar previamente cenários de séries temporais.

Capítulo 3

Estimação da Densidade de Probabilidade, Medidas de Dependência e Redes Bayesianas

3.1 Introdução

O objetivo deste capítulo é introduzir conceitos e apresentar formulações referentes aos principais conceitos estatísticos que são diretamente aplicadas no modelo proposto neste trabalho, ou ainda que sirvam de base para a análise posterior do desempenho do modelo. São apresentados três tópicos distintos, pouco inter-relacionados, mas de igual importância para a construção do modelo proposto.

O primeiro tópico refere-se à caracterização de variáveis aleatórias por meio de funções de densidade de probabilidade estimadas de forma não paramétrica. A relevância deste tópico se dá pelo fato de que a representação do processo estocástico do vento implica no surgimento de variáveis cujas densidades de probabilidade são de difícil parametrização, isto é, pode ser impossível a eleição de uma família de função de densidade de probabilidade para prover um ajuste satisfatório das variáveis aleatórias, principalmente quando se considera a representação não apenas da incerteza do módulo da velocidade do vento, mas também da direção. Desta maneira, a abordagem não paramétrica se torna desejável, sendo apresentado para tal fim o método de estimação de função de densidade de probabilidade por *kernel density estimation*[41].

O segundo tópico refere-se às medidas de dependência estatística entre variáveis aleatórias. A importância deste tópico advém do fato de que os elementos variantes no tempo que compõem um sistema de potência apresentam diferentes graus de dependência estatística entre si e a construção de um modelo efetivo para a representação destas variáveis em uma simulação estocástica precisa levar em conta este relacionamento. Neste contexto, são apresentados a medida de dependência estatística mais comumente abordada na literatura, a correlação de *Pearson*, os motivos que fazem com que esta correlação possa não ser uma boa medida de dependência na

representação das incertezas modeladas neste trabalho, e, principalmente, as medidas menos usuais, porém fundamentais para o modelo proposto, a informação mútua e o coeficiente de informação máxima.

Por fim, no último tópico são apresentados os conceitos e as formulações referentes às redes Bayesianas, cuja aplicação no modelo proposto tem como principal objetivo garantir um tratamento à dimensionalidade do modelo estocástico, evitando o crescimento exponencial do número de parâmetros do modelo com o aumento do número de elementos variantes no tempo representados.

Posteriormente, no Capítulo 5, será apresentado como estes tópicos se relacionam para compor o modelo proposto para representação de um processo estocástico dos elementos variantes no tempo na avaliação de confiabilidade por simulação não sequencial.

3.2 Função de Densidade de Probabilidade

O comportamento de uma variável aleatória contínua X pode ser caracterizado através da sua função de densidade de probabilidade, $f_X(x)$.

A partir do conhecimento de $f_X(x)$, a probabilidade de uma variável aleatória contínua pertencer a um dado intervalo pode ser definido pela integral da sua função de densidade de probabilidade sobre este intervalo.

Formalmente, define-se[42]:

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (3.1)$$

Para qualquer intervalo $[a, b] \subseteq \mathfrak{R}$.

Uma definição importante na caracterização de uma variável aleatória, derivada da função de densidade de probabilidade, e que desempenha um papel fundamental na análise da confiabilidade sob o enfoque probabilístico, conforme apresentado mais a frente, é a função de probabilidade acumulada, $F_X(x)$, dada por[42]:

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (3.2)$$

A função $F_X(x)$ descreve a probabilidade de uma variável aleatória contínua X assumir um valor menor ou igual a x .

Se $F_x(x)$ é estritamente crescente e contínua, então existe uma função inversa, $F_x^{-1}(y)$, para $y \in [0,1]$, denominada função de distribuição inversa, ou função quantil, para a qual:

$$F_x(x) = y \Leftrightarrow F_x^{-1}(y) = x \quad (3.3)$$

Adicionalmente, se uma variável aleatória contínua Y segue uma distribuição uniforme $U[0,1]$, então a transformação $Z = F_x^{-1}(Y)$, resultará em uma variável aleatória contínua Z , cuja densidade de probabilidade é caracterizada por $f_x(x)$.

Prova [42]:

Se $F_x^{-1}(Y)$ segue uma distribuição $f_x(x)$, então:

$$P(F_x^{-1}(Y)) = P(F_x^{-1}(Y) \leq x) \quad (3.4)E$$

Como $F_x(x)$ é monotônica, pode-se aplicá-la aos dois lados da expressão, resultando em:

$$P(F_x^{-1}(Y) \leq x) = P(Y \leq F_x(x)) \quad (3.5)$$

Como Y segue uma distribuição uniforme $U[0,1]$, tem-se que:

$$P(Y \leq F_x(x)) = F_x(x) \quad (3.6)$$

Uma vez que a distribuição uniforme se caracteriza por:

$$P(U \leq u) = U, \forall U \in [0,1] \quad (3.7)$$

Analogamente, no sentido inverso, se a transformação $Y = F_x(X)$ for aplicada, resultará em uma variável aleatória contínua Y , cuja densidade de probabilidade é caracterizada pela distribuição uniforme $U[0,1]$, sendo esta transformação conhecida como transformação da integral de probabilidade.

As duas transformações apresentadas são fundamentais no modelo proposto neste trabalho, conforme será posteriormente discutido. Devido a isso, a correta estimação da função de densidade de probabilidade é crítica no desenvolvimento do modelo.

3.2.1 Estimação Não Paramétrica da Função de Densidade de Probabilidade – Uso do Histograma

No que tange à estimação de funções de densidade de probabilidades de variáveis aleatórias contínuas, a abordagem mais comum é a paramétrica, na qual se supõe que as variáveis aleatórias são bem caracterizadas por alguma família de densidade de probabilidade conhecida, tais como a distribuição normal, log-normal, Weibull, etc., e

se aplica alguma técnica específica para obtenção dos parâmetros que melhor caracterizam os dados a partir da distribuição paramétrica.

No lado oposto está a abordagem não paramétrica, que dispensa esta necessidade de rotulação inicial da função de densidade, permitindo, teoricamente, maior flexibilidade para lidar com um maior número de situações.

O histograma é tido como o tipo de estimador não paramétrico mais simples para uma função de densidade de probabilidade de uma variável aleatória, sendo construído a partir da pré-definição do comprimento de um intervalo (h), de uma posição inicial (a_0) e distribuição das observações ao longo dos intervalos subsequentes formados por estes elementos, de maneira que, definindo-se H_k como sendo o número total de observações no k -ésimo intervalo, tal como:

$$H_k = \# \{X_i | X_i \in (a_{k-1}, a_k)\} \quad (3.8)$$

Onde,

$\# \{X_i | X_i \in (a_{k-1}, a_k)\}$ é a quantidade de observações de X_i , dado que $X_i \in (a_{k-1}, a_k)$

$$a_k = a_0 + kh$$

A função de densidade de probabilidade estimada será dada por:

$$\hat{p}(x) = \frac{1}{kh} \sum_{k=1}^n H_k 1_{(a_{k-1}, a_k)}(x) \quad (3.9)$$

Onde,

$1_{(a_{k-1}, a_k)}(x)$ é uma função que vale 1 para $x \in (a_{k-1}, a_k)$ e 0, caso contrário.

Embora o histograma seja uma abordagem que permite facilmente a construção de funções de distribuição de probabilidade diretamente a partir das observações, ele apresenta algumas importantes limitações, dentre elas a dependência da estimativa com o comprimento do intervalo e posição inicial definida, e o fato destes não constituírem uma função contínua.

Para uma melhor ilustração das limitações do histograma, na Figura 7 são apresentados diferentes histogramas para um mesmo conjunto de observações de uma variável aleatória. Este conjunto foi gerado sintetizando-se dois terços das amostras a partir de uma distribuição normal, de média $\mu = 1$ e desvio padrão $\sigma = 0.5$, e um terço das amostras a partir da distribuição normal, de média $\mu = 3$ e desvio padrão $\sigma = 0.5$, resultando desta forma em uma distribuição bimodal.

Os histogramas foram estimados variando-se o comprimento, h , e considerando-se como posição inicial, a_0 , o valor mínimo observado.

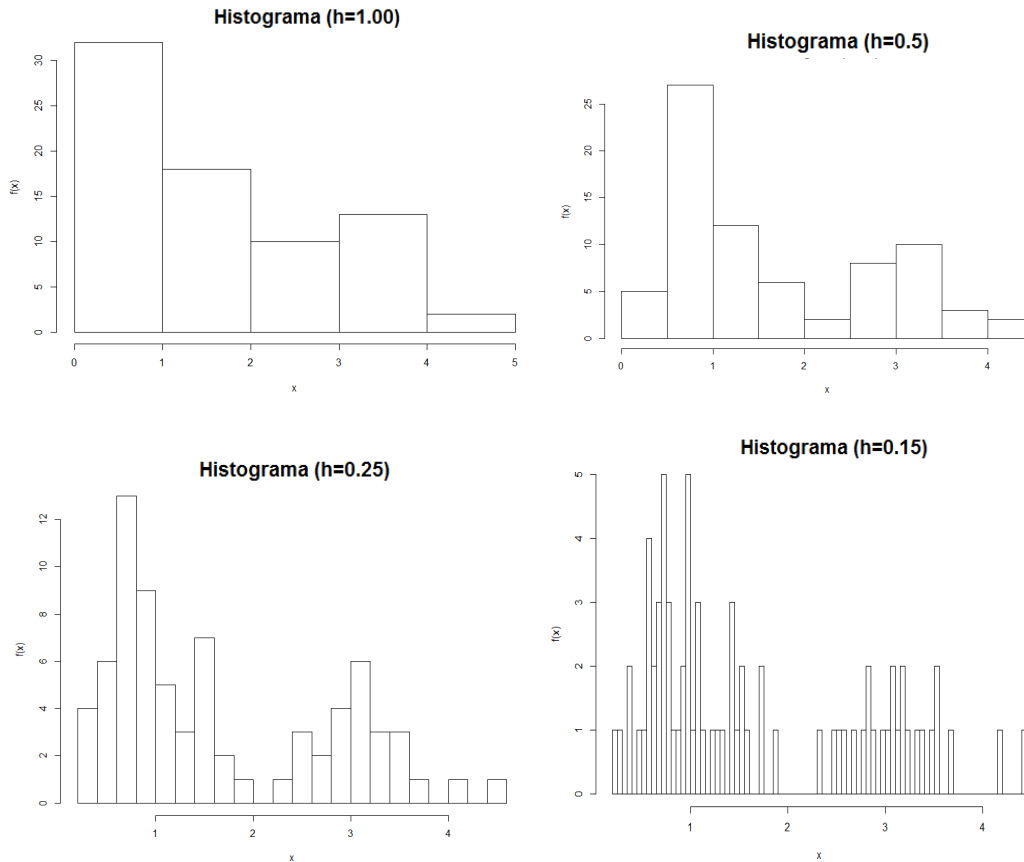


Figura 7 – Variação do histograma com o comprimento

O primeiro histograma, estimado com $h=1$, não consegue caracterizar a característica bimodal da distribuição, reproduzindo de maneira pobre a distribuição arbitrada. No segundo histograma, estimado com $h=0.5$, já se observa a característica bimodal da distribuição, no entanto este histograma não traduz corretamente a simetria no em torno de cada uma das modas. Simetria esta que deveria existir, uma vez que as amostras são provenientes da distribuição normal, que é simétrica. No terceiro, e, principalmente, no quarto histograma, já é possível se observar valores nulos para alguns dos intervalos do histograma. No entanto, estes valores nulos, que indicam que a probabilidade de se obter uma amostra para esta distribuição nestes intervalos é nula, estão em desacordo com a distribuição real utilizada para gerar os histogramas.

Um problema adicional é o efeito alternado observado de aumento e diminuição da distribuição de probabilidade à medida que se avança nos intervalos, o que também é incoerente com a distribuição arbitrada.

Na Figura 8 são apresentadas as funções de distribuição de probabilidade acumulada obtidas a partir dos histogramas estimados, para $h = 0.5$ e $h = 0.25$. Observa-se que as funções não são contínuas e nem estritamente crescentes, condições necessárias para a validade da relação apresentada na equação (3.3). Como consequência, a transformação da integral de probabilidade não é aplicável.

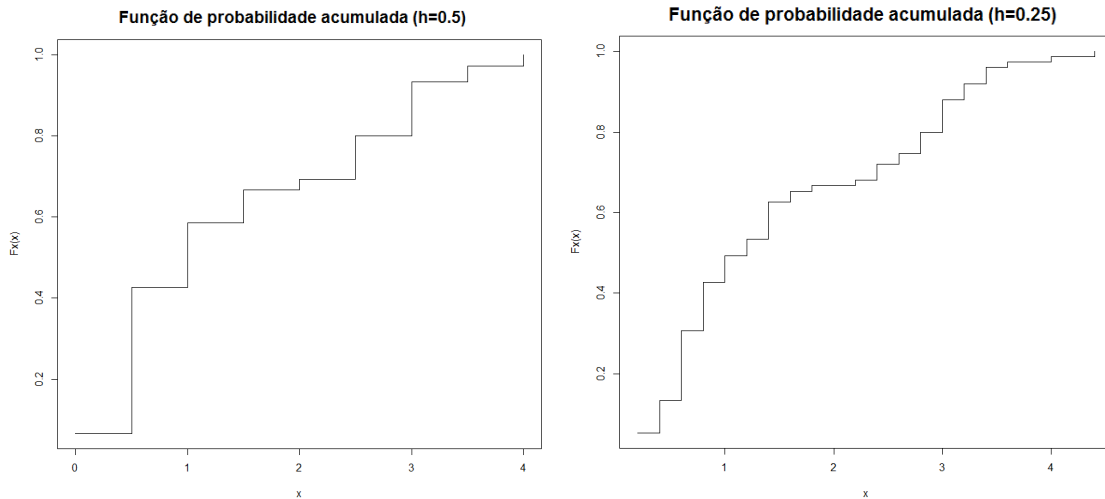


Figura 8—Funções de distribuição de probabilidade acumulada para variável

3.2.2 Estimação de Densidade por Funções Kernel - *Kernel Density Estimation*

O método de estimação de função de densidade por *kernel density estimation*[41], assim como o histograma, é uma abordagem não paramétrica que permite construir uma estimativa para a função de densidade de probabilidade de uma determinada variável aleatória diretamente a partir das suas observações.

O processo de estimação de função de densidade por *kernel density estimation* pertence a classe de problemas de suavização (ou alisamento) de dados, quando é necessário inferir uma população a partir de um número finito de dados[43].

Na Figura 9, extraída de [43], é apresentado um exemplo de comparação entre a estimação da função de densidade de uma amostra real por histograma e por *kernel density estimation*. Para a estimação do histograma foram utilizados três diferentes parâmetros de intervalo, resultando em três histogramas bastante diferentes, pois, conforme dito anteriormente, um dos problemas do histograma é esta grande dependência com o parâmetro de intervalo. O último gráfico corresponde a estimativa por *kernel density estimation*, que tem como objetivo produzir a função contínua que melhor aproximaria a real densidade de probabilidade da amostra.

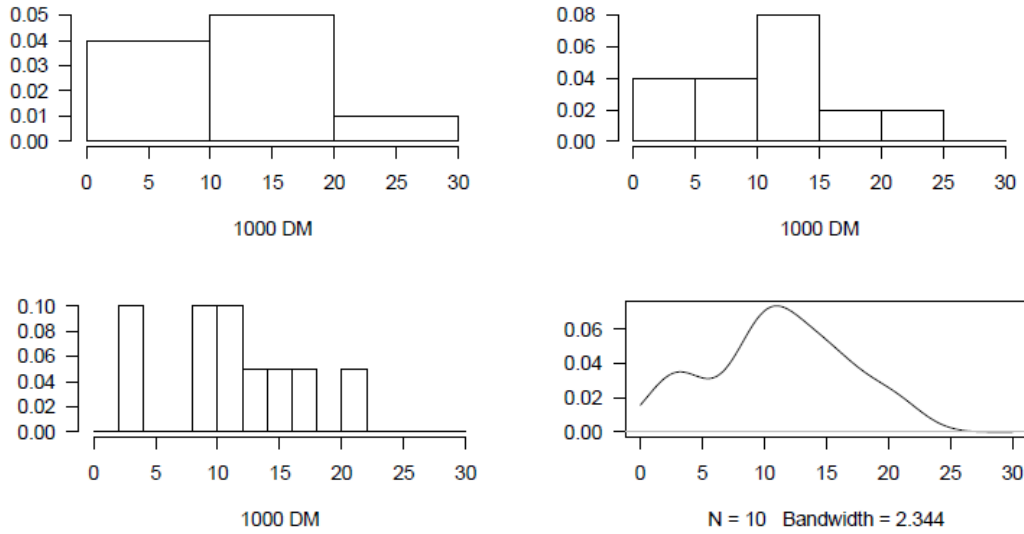


Figura 9–Estimação de Função de Densidade por Histograma e Kernel Density Estimation[43]

Dada uma variável aleatória X , cuja função de densidade de probabilidade é dada por $f(x)$, tem-se que a seguinte aproximação é válida:

$$P(x-h < X < x+h) = \int_{x-h}^{x+h} f(t)dt \approx 2hf(x) \quad (3.10)$$

Desta forma:

$$f(x) \approx \frac{1}{2h} P(x-h < X < x+h) \quad (3.11)$$

Uma estimativa para esta função pode ser dada por:

$$\hat{f}(x) = \frac{1}{2h} \frac{\sum_i 1_{[x-h, x+h]}(x_i)}{n} \quad (3.12)$$

Onde x_i corresponde a i -ésima observação de X , n é o número de observações e $1_A(x)$ é a função indicadora, definida por:

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \quad (3.13)$$

Alternativamente, esta equação pode ser representada como:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n w(x-x_i, h) \quad (3.14)$$

Onde:

$$w(t, h) = \begin{cases} \frac{1}{2h} & |t| < h \\ 0 & |t| \geq h \end{cases} \quad (3.15)$$

De maneira mais geral, esta função pode ser escrita da seguinte forma:

$$w(t, h) = \frac{1}{h} K\left(\frac{t}{h}\right) \quad (3.16)$$

Onde K é a função chamada *kernel*, e sendo h o parâmetro de alisamento. A função *kernel* define a forma da função de densidade de probabilidade enquanto que o parâmetro de alisamento define o quão suave é esta forma.

A equação (3.15) corresponde a uma das funções *kernel* mais tradicionais, a retangular, dada pela equação:

$$K(t) = \begin{cases} \frac{1}{2} & |t| < 1 \\ 0 & |t| \geq 1 \end{cases} \quad (3.17)$$

Outra função *kernel* bastante utilizada, e mais importante para este trabalho, é a *kernel* gaussiana, dada por:

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)t^2} \quad (3.18)$$

Na Figura 10, extraída de [44], é apresentado um exemplo de estimação de densidade de probabilidade utilizando funções *kernel* gaussianas. Pode-se visualizar a função de densidade estimada como uma combinação de Gaussianas de mesmo valor de desvio padrão, mas com médias diferentes, centradas nos valores das observações.

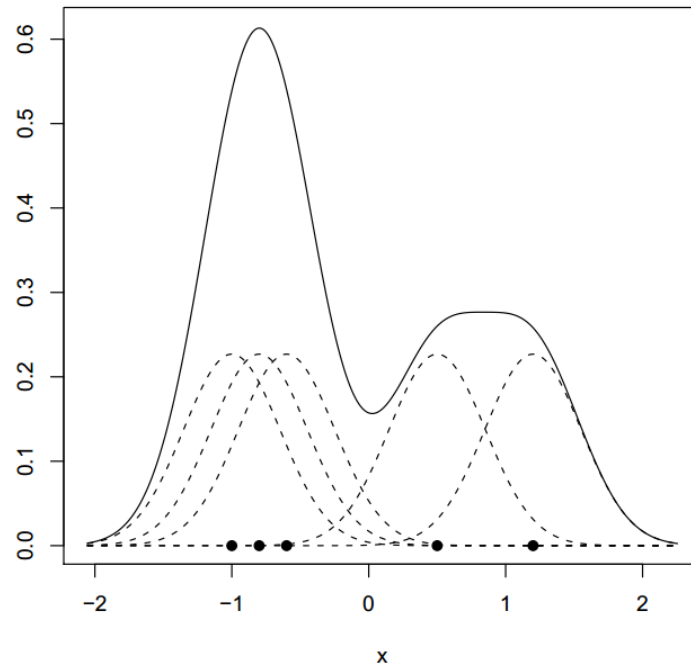


Figura 10– Exemplo de Estimação com Kernel Gaussiano[44]

Para um conjunto de observações para o qual se deseja obter uma estimativa da função densidade de probabilidade, dada a escolha de um tipo de função *kernel*, o problema passa a ser determinar o parâmetro de alisamento. Esta determinação é crítica, uma vez que se for feita incorretamente, poderá tornar a estimativa subsuavizada ou supersuavizada. A Figura 11 mostra duas funções de distribuição de probabilidade obtidas para o mesmo conjunto de dados utilizados para a obtenção da distribuição apresentada na Figura 10. O gráfico superior corresponde à estimativa utilizando um parâmetro de alisamento muito menor do que o parâmetro ideal, enquanto o inferior corresponde à estimativa utilizando um parâmetro de alisamento muito maior do que o ideal. Observa-se que o parâmetro menor levou a uma distribuição ruidosa, pouco suave, enquanto que o parâmetro maior levou a uma distribuição suave demais, na qual nem se identifica a característica bimodal da distribuição.

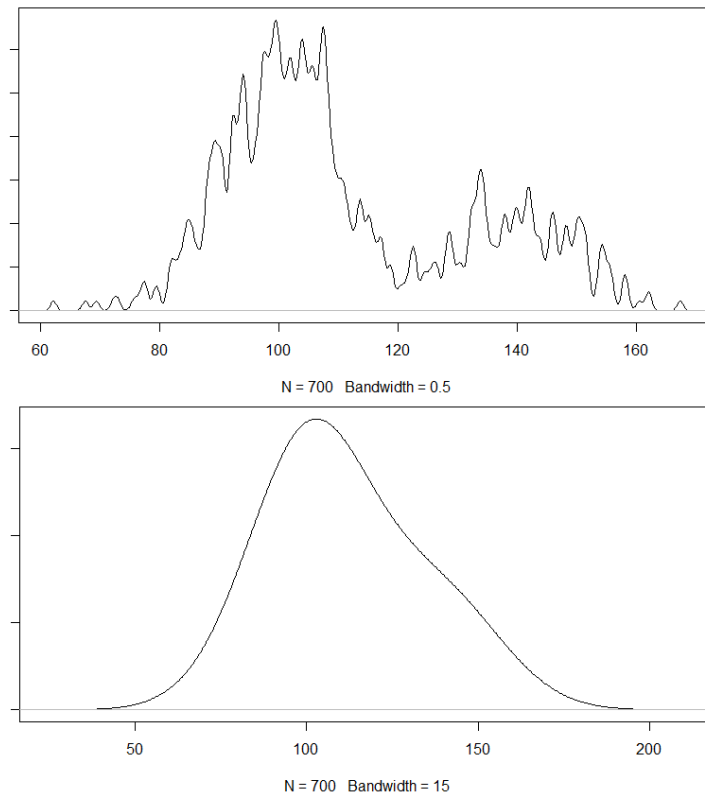


Figura 11– Estimações subsuavizada e supersuavizada

A escolha do parâmetro de alisamento é essencialmente um problema de otimização no qual se deseja minimizar alguma, ou algumas, medidas de erros da estimativa, conforme apresentado em maiores detalhes no Apêndice I.

A natureza de algumas variáveis aleatórias pode ser tal que a densidade de probabilidade estimada exista apenas em um domínio limitado, $X \in [L, U]$, onde L e U são respectivamente os limites inferior e superior da variável X . Isto é verdade, por exemplo, na representação da incerteza da velocidade do vento, uma vez que o módulo da velocidade nunca pode ser negativo. Para lidar com esta condição, o método pode ser modificado, incorporando-se o conceito de estimação com suporte, que nada mais é do que a inclusão de uma restrição explícita de domínio no processo de estimação da função de densidade de probabilidade.

A importância desta consideração se deve ao fato das funções de densidades de probabilidade serem diretamente utilizadas na geração de cenários, pelo modelo proposto neste trabalho, e desta maneira se garante que não serão amostradas velocidades negativas para as séries de ventos representadas.

3.3 Densidades de Probabilidade Conjunta e Dependência Estatística entre Variáveis Aleatórias

Estendendo o conceito de densidade de probabilidade para o caso bidimensional, define-se como função de densidade de probabilidade conjunta de duas variáveis, X_1 e X_2 , a função $f_{X_1, X_2}(x_1, x_2)$, tal que [42]:

$$P(X_1 \leq s, X_2 \leq t) = \iint_{x_1 \leq s, x_2 \leq t} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \quad (3.19)$$

E que satisfaça as seguintes condições:

$$f_{X_1, X_2}(x_1, x_2) \geq 0 \quad (3.20)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = 1 \quad (3.21)$$

Assim como para o caso unidimensional, a função de densidade de probabilidade conjunta $f_{X_1, X_2}(x_1, x_2)$ é informação suficiente para caracterizar o comportamento conjunto das variáveis aleatórias X_1 e X_2 , independentemente de como estas variáveis estão relacionadas.

De forma mais geral, o conceito de densidade de probabilidade conjunta pode ser aplicada para n variáveis aleatórias. Destacando-se, no entanto, que esta função torna-se mais complexa à medida que se aumenta a dimensão das variáveis, e a menos de casos particulares triviais, como a função de densidade gaussiana multivariável, se torna impossível a definição de uma expressão racional para a caracterização da função de densidade de probabilidade conjunta de n variáveis.

A necessidade da caracterização do comportamento de um conjunto de variáveis aleatórias por uma função de densidade de probabilidade conjunta está diretamente associada ao conceito de dependência estatística entre as variáveis. Se as variáveis são estatisticamente independentes, o comportamento conjunto das variáveis pode ser descrito simplesmente pelas densidades de probabilidades marginais, uma vez que o processo estocástico conjunto é dado pela combinação de cada processo estocástico de forma individual e independente. Quando existe dependência estatística entre as variáveis, este relacionamento precisa fazer parte da informação que descreve as variáveis, através do conhecimento da probabilidade conjunta.

O grau de relacionamento entre variáveis aleatórias pode ser aferido através de uma medida de dependência estatística, isto é, um parâmetro associado a um par de variáveis aleatórias que codifica em seu valor a intensidade da dependência estatística entre estas variáveis. Dentre as medidas existentes, a mais comum é a correlação de *Pearson*, também conhecida como correlação linear.

No que se refere à estimação de modelos e amostragem de séries sintéticas em simulações estocásticas, o sucesso do emprego da correlação de *Pearson* é justificado pela facilidade de se estimar as correlações de um conjunto de observações e de posteriormente se produzir amostras multivariadas que preservem esta correlação. No entanto, esta medida apresenta algumas limitações e contra indicações com respeito à natureza das variáveis aleatórias abordadas neste trabalho, conforme será posteriormente discutido. Desta maneira, outras medidas mais efetivas, precisam ser introduzidas.

Nesta seção serão apresentadas, além da própria correlação de *Pearson*, as duas medidas fundamentais para o modelo proposto: a informação mútua e o coeficiente de informação máxima.

3.3.1 Correlação de Pearson

A correlação de Pearson, também conhecida como correlação linear, é a medida mais comum de dependência estatística entre duas variáveis aleatórias, sendo definida como a razão entre a covariância bivariada e os desvios padrões das variáveis aleatórias:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (3.22)$$

Para uma amostra finita de n observações conjuntas de X e Y , pode-se definir a correlação da amostra como:

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.23)$$

O coeficiente de correlação de Pearson varia entre -1 e 1. Um valor de coeficiente igual a 1 implica que uma equação linear pode descrever perfeitamente o relacionamento entre X e Y , sendo que todos os pares das possíveis observações conjuntas de (X, Y) estariam sobre a reta descrita por esta equação linear, para a qual Y cresce quando X cresce. Um valor de coeficiente igual a -1 implica que a relação entre as duas variáveis é inversa, quando X cresce, Y decresce, sendo que esta relação também pode ser perfeitamente descrita por uma equação linear. Um valor de coeficiente igual a zero implica que não existe relacionamento linear entre as variáveis, o que, porém, não implica necessariamente que as variáveis sejam estatisticamente independentes.

Na Figura 12 são apresentados quatro gráficos X-Y de duas variáveis aleatórias, sendo os valores das correlações de Pearson, respectivamente, 0,85, 0,65, 0,15 e 0,05.

Pode-se observar a forte relação linear entre as variáveis nos dois primeiros gráficos pela presença de uma tendência de crescimento conjunto entre as variáveis. Nos demais gráficos, para os quais o coeficiente de correlação é baixo, não se observa este comportamento.

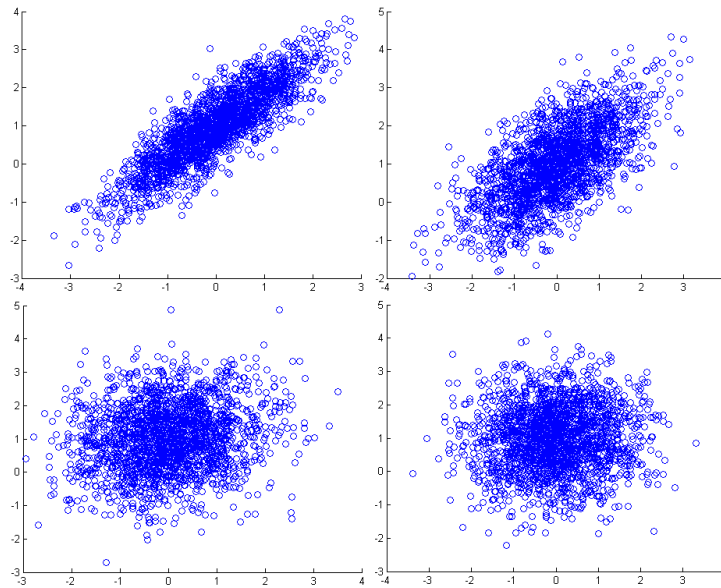


Figura 12–Exemplo de variáveis aleatórias linearmente dependentes

Embora a correlação de *Pearson* seja umas das medidas de dependência estatística mais usuais para as mais diversas aplicações estatísticas, e esteja muito presente na análise de séries temporais, a rigor, o cálculo do coeficiente de correlação de *Pearson* parte da suposição que as seguintes condições são verdadeiras:

- Relacionamento entre os dados é linear;
- Distribuição normal bivariada;

Desta forma, a medida da intensidade da dependência estatística entre duas variáveis aleatórias pode ser comprometida, caso estas condições não sejam satisfeitas. Adicionalmente, outras desvantagens da correlação de *Pearson* é que esta é uma medida muito sensível à assimetria da distribuição dos dados e a presença de *outliers*, e não é aplicável para se medir a dependência estatística entre variáveis aleatórias discretas, ou ainda, a dependência entre variáveis contínuas e discretas.

Conforme será posteriormente discutido, as variáveis aleatórias de interesse para este trabalho, principalmente aquelas associadas aos processos estocásticos do vento, apresentam, entre outros, as seguintes características:

- Distribuição bivariada não normal:

A densidade de probabilidade marginal de uma série de vento não é bem caracterizada por uma distribuição normal. Quando se observa a densidade de probabilidade conjunta entre duas séries de vento, este tipo de caracterização é ainda menos aderente aos dados.

- Forte presença de assimetria nas distribuições das variáveis aleatórias:

Sendo uma das causas para a não normalidade das séries de vento, se observa uma característica assimétrica nas densidades de probabilidade que pode ser menor ou maior dependendo da série.

- Combinação de variáveis contínuas e discretas:

Embora este aspecto não seja explorado neste trabalho, existem abordagens para a representação de séries de direção do vento que consideram que esta é uma variável discreta, cujas medidas podem assumir os valores descritos pelos pontos cardeais (norte, sul, leste e oeste) e colaterais (nordeste, noroeste, sudeste e sudoeste). A representação da incerteza da direção do vento neste caso tem como consequência a necessidade de definição de processos estocásticos que combinam variáveis contínuas e discretas.

3.3.2 Informação Mútua (*Mutual Information*) e Coeficiente de Informação Máxima (*Maximum Coefficient of Information*)

A informação mútua é um conceito que vem da teoria da informação e, em contraste com a correlação de *Pearson*, é sensível a qualquer tipo de dependência entre as variáveis, de maneira que só é nula para duas variáveis aleatórias estritamente independentes [45].

A informação mútua entre duas variáveis aleatórias contínuas, x e y é definida pela seguinte expressão [45]:

$$MI(x, y) = \iint f(x, y) \log_2 \frac{f(x, y)}{f(x)f(y)} dx dy \quad (3.24)$$

Onde, $f(x)$ e $f(y)$ são as densidades marginais da densidade de probabilidade conjunta $f(x, y)$.

A obtenção da informação mútua para variáveis aleatórias contínuas não é trivial, principalmente pela dificuldade de se estimar o termo $f(x, y)$ a partir de um conjunto finito de observações destas variáveis. Um método simples empregado na prática consiste na discretização dos dados, superpondo a um *grid* retangular o gráfico $x - y$ das amostras, associando então os valores dos pares (x, y) às suas respectivas células no *grid*. A informação mútua pode então ser estimada a partir da distribuição de probabilidade discretizada $p(X, Y)$ como:

$$MI(x, y) \approx MI(X, Y) = \sum_{x, y} p(X, Y) \log_2 \frac{p(X, Y)}{p(X)p(Y)} \quad (3.25)$$

Onde, $p(x)$ e $p(y)$ são as distribuições marginais das variáveis discretizadas, associadas a distribuição de probabilidade conjunta $p(X, Y)$.

Uma propriedade importante da informação mutua é que ela é preservada caso seja aplicada qualquer função de transformação inversível as variáveis aleatórias [45].

Um problema da informação mútua é que esta é uma medida que não satisfaz as condições de equitabilidade[46]. Embora a equitabilidade seja um conceito que não possua uma formalização, sendo por vezes tratada como um conceito heurístico[47], define-se a equitabilidade de uma medida estatística como a capacidade dela fornecer pontuações similares para relacionamentos de diferentes tipos, mas com iguais níveis de ruídos[46]. Por exemplo, sejam três variáveis aleatórias, X , Y e Z , tais que Y corresponde a uma função linear de X mais um certo nível de ruído e Z corresponde a uma função senoidal de X , mais o mesmo nível de ruído, uma medida de dependência estatística entre estas variáveis será dita satisfazer as condições de equitabilidade se a pontuação medida entre X e Y , e X e Z , for similar.

A consequência prática da ausência de equitabilidade na informação mútua é que a comparação entre duas medidas distintas de dependência estatística pode ser tendenciosa[46]. Isto é, se por um lado é possível se mensurar o quanto dois pares de variáveis aleatórias são mais ou menos linearmente relacionados pela comparação de suas correlações de *Pearson*, a mesma comparação pode não ser válida para a informação mútua, uma vez que esta medida varia com a natureza das incertezas, clusters de discretização, etc.

Uma medida de dependência derivada da informação mútua que satisfaz as condições de equitabilidade é o coeficiente de informação máxima (*maximun coefficient of information - MIC*).

O MIC é definido como[46]:

$$MIC(x, y) = \max_{xy < B(|D|)} \frac{MI^*(D, x, y)}{\log_2 \min \{x, y\}} \quad (3.26)$$

Onde D é a configuração de discretização (*bins* escolhidos) e $B(|D|)$ uma função que estabelece um limite para o número de diferentes configurações de *bins* avaliadas. Usualmente $B = N^{0.6}$ [46], onde N é o número total de amostras de (x, y) .

A obtenção do MIC pode ser interpretada como um problema de otimização no qual se determina a discretização ótima de $x \rightarrow X$ e $y \rightarrow Y$, isto é, quais os *bins* que devem ser utilizados, que maximiza a relação entre a informação mútua de X e Y , e uma medida da complexidade destes *bins*, dada pelo logaritmo do menor dos *bins*, X ou Y .

Graficamente, pode se entender o cálculo do MIC como a sobreposição de vários *grids* bidimensionais ao gráfico da dispersão de x e y , subdividindo cada um destes *grids* em partições que encapsulam o relacionamento entre as variáveis. A partir da computação da informação mútua proveniente de cada uma destas diferentes configurações de *grid*, e após a normalização desta informação mútua, para garantir uma comparação justa para diferentes dimensões de X e Y , o MIC corresponderá a maior informação mútua normalizada.

O fato de $MI(X, Y)$ ser um valor que necessariamente está no intervalo: $0 \leq MI(X, Y) \leq \log_2(\min(|X|, |Y|))$, implica que $MIC(x, y)$ é sempre um valor compreendido entre 0 e 1.

Em [46] é definido ainda a medida $MIC-R2$. Como o MIC é em geral aproximadamente igual ao quadrado da correlação para relacionamentos lineares, pode-se definir uma medida de não linearidade do relacionamento entre duas variáveis, indicada como $MIC-R2$, como sendo a diferença entre o MIC e a correlação de *Pearson* [46]. A medida $MIC-R2$ será próxima de zero para relacionamentos lineares e será grande para relações não lineares que apresentam valores altos de MIC .

3.4 Rede Bayesiana

As redes Bayesianas são modelos concebidos a partir da combinação de teoria de probabilidade com teoria dos grafos, podendo eficientemente, e de forma compacta, representar a distribuição de probabilidade conjunta de variáveis n -dimensionais, constituindo uma alternativa atrativa para se representar a distribuição de probabilidade conjunta de um número elevado de variáveis aleatórias[48].

As redes Bayesianas são caracterizadas por dois componentes que a descrevem qualitativa e quantitativamente:

a) Uma estrutura de grafo, direcionado e acíclico (DAG), $G = \langle N, E \rangle$, onde os nós, $N = \{X_1, X_2, \dots, X_n\}$, deste grafo representam as variáveis aleatórias do problema e cujas arestas, $E \subseteq N \times N$, descrevem as dependências e independências condicionais entre estas variáveis;

b) Um conjunto de parâmetros, Θ , associados às arestas do grafo, usualmente probabilidades condicionais, que descrevem as distribuições de probabilidades.

Uma propriedade fundamental das redes Bayesianas é que um nó X_i é condicionalmente independente de qualquer outro nó $X_j \notin pa(X_i)$, dado $pa(X_i)$, onde $pa(X_i)$ é o conjunto de nós pais do nó associado à variável aleatória X_i em G .

Mais formalmente, tem-se que [48]:

$$X_i \perp X_j | pa(X_i) \forall X_j \notin pa(X_i) \quad (3.27)$$

E

A partir destas distribuições de probabilidade condicionais é possível reconstruir a distribuição de probabilidade conjunta das variáveis aleatórias, de modo que:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i)) \quad (3.28)$$

E

A equação (3.28) ilustra a modularidade das redes Bayesianas, que é a característica que as tornam atrativas para a representação da dependência estatística de um conjunto grande de variáveis, cuja função tem sua complexidade aumentada exponencialmente com o aumento do número de variáveis, em várias funções de menor complexidade, dado a quebra da distribuição da probabilidade conjunta.

O principal objetivo da estrutura das redes Bayesianas é representar de maneira aproximada, mas suficientemente precisa, as relações de independência condicional entre as variáveis aleatórias.

Para se definir conceitualmente as rede Bayesianas é necessário a introdução do conceito de *d-separation*, apresentado sucintamente a seguir.

Sejam S , T e V três subconjuntos de nós no grafo direcionado e acíclico, G , e seja p um caminho qualquer entre os nós em S e os nós em T , que representa uma sucessão de arcos, independente das suas direções. Então, diz-se que V bloqueia p , se existe algum nó Z em p que satisfaz uma das seguintes condições[49]:

1. Existem arcos que convergem para Z , ao longo de p , e nem Z e nenhum de seus descendentes estão em V .

2. Não existem arcos que convergem para Z , ao longo de p , e Z está em V .

Além disso, diz-se que V constitui uma d -separation entre S e T , que é escrito como $(S \perp_G T|V)$, se V bloqueia todos os caminhos de qualquer nó em S para um nó em T .

Um grafo direcionado e acíclico G é dito ser uma representação I -map de uma distribuição de probabilidade P , se todas as independências condicionais apresentadas em G , através das regras da d -separation, são válidas em P [49].

Isto é, para todo X, Y e Z , tem-se que:

$$(X \perp_G Y|Z) \Rightarrow (X \perp_p Y|Z) \quad (3.29)$$

Finalmente, o grafo G é dito ser uma rede Bayesiana se este corresponde ao mínimo I -map de P . Isto é, nenhum arco pode ser removido de G sem que este perca sua propriedade de ser uma representação I -map da distribuição de probabilidade P .

Na Figura 13, é apresentado um exemplo no qual a variável aleatória F é função das variáveis estatisticamente dependentes: A, B, C, D e E .

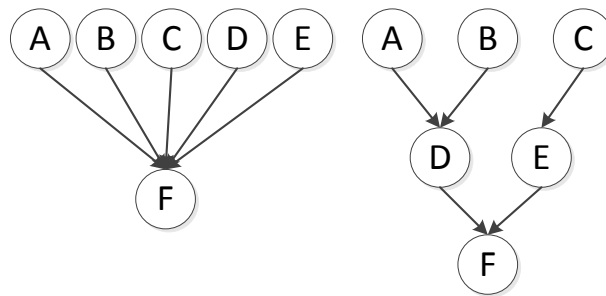


Figura 13– Exemplo de rede Bayesiana

A rigor, a distribuição de probabilidade de F , deveria levar em consideração a distribuição de probabilidade conjunta de A, B, C, D e E , sendo dada por:

$$P(F) = f(P(A, B, C, D, E)) \quad (3.30)$$

Mas supondo-se que a estrutura de rede apresentada do lado direito da figura seja adequada para representar as dependências estatísticas das variáveis, esta função poderia ser escrita, a partir da equação (3.28), como:

$$P(F) = f(P(D|A, B)P(A)P(B)P(E|C)P(C)) \quad (3.31)$$

Enquanto na representação exata havia uma única função de distribuição de probabilidade, cuja complexidade dependia de cinco variáveis aleatórias, na

representação através da rede Bayesiana a função mais complexa, $P(D|A,B)$, é condicionada a apenas duas variáveis aleatórias.

Estimar a estrutura da rede Bayesiana que melhor representa a distribuição de probabilidade conjunta de um conjunto de variáveis aleatórias é um processo de complexidade NP-Completo, por se tratar de um problema combinatório no qual, a priori, todas as possíveis combinações de conexões entre os nós N deveriam ser avaliadas, de modo a se obter a estrutura de grafo ótima. Na prática, para um conjunto grande de variáveis, esta avaliação completa é computacionalmente inviável, sendo usual o emprego de métodos que aplicam alguma heurística para reduzir o espaço de busca e encontrar uma estrutura de grafo, que, se teoricamente não é a estrutura ótima, é ao menos suficientemente adequada para a representação desejada.

Um método de larga aplicação na estimação das redes Bayesianas é baseado no algoritmo de *hill climbing*, que consiste em explorar recursivamente os pontos vizinhos a uma determinada configuração, adotando no passo seguinte aquela configuração para a qual existe um maior ganho na métrica adotada de avaliação da adequação da rede.

Este método pertence a classe de métodos de estimação de estrutura de redes Bayesianas baseados em pontuação (métodos *score-based*), sendo apresentado com maiores detalhes no Apêndice II.

O conhecimento a priori a respeito da natureza dos dados, como, por exemplo, as relações de causalidade entre as variáveis, pode ser incorporado ao algoritmo de estimação de estrutura de redes Bayesianas através de restrições conhecidas como restrições *whitelisting* e restrições *blacklisting*. Estas restrições correspondem a conjuntos de arcos pré-definidos que devem de forma garantida estar presentes ou estar ausentes na estrutura estimada.

Esta abordagem é importante na elaboração do modelo de transição de estados, conforme será visto mais a frente, na qual algumas restrições relacionadas a cronologia de eventos precisam ser representadas via restrições *blacklisting*, e também na representação de variáveis exógenas ao processo estocástico do vento, na qual os estados destas variáveis não são determinadas pelo modelo estocástico.

No algoritmo de estimação abordado neste trabalho, *hill climbing*, a modelagem das restrições *blacklisting* é obtida pela simples eliminação no espaço de busca dos candidatos vizinhos a configuração atual que introduzam algum relacionamento proibido entre as variáveis.

Estimar os parâmetros de redes Bayesianas a partir de uma topologia conhecida é um problema bem mais simples do que o problema da estimação da topologia. Um dos

métodos mais empregados para este fim é o método da estimação da máxima verossimilhança (MLE – Maximum Likelihood Estimation), utilizado no modelo desenvolvido neste trabalho.

No Apêndice II é apresentada com maiores detalhes a formulação deste método.

Capítulo 4

Análise Estatística do Comportamento do Vento para Geração Eólica

4.1 Introdução

Embora o modelo proposto seja flexível para a representação de qualquer elemento variante no tempo, o principal foco deste trabalho é a representação da geração eólica. Desta forma, neste capítulo serão apresentadas as principais características da geração eólica, os principais componentes que integram e impactam na geração das usinas eólicas, e por fim, algumas análises estatísticas de diferentes séries de vento, de modo a prover um melhor entendimento empírico do comportamento do vento que justificam as premissas adotadas no modelo desenvolvido neste trabalho.

A geração eólica é um tipo de energia obtido pela conversão da energia cinética contida no vento em energia mecânica, através das turbinas eólicas. Esta energia mecânica é então convertida em eletricidade a partir dos geradores acoplados a estas turbinas eólicas. A geração eólica é considerada uma das formas de energia solar, uma vez que o sol é a fonte primária que causa influência direta nos fenômenos climáticos do planeta. O vento se forma a partir da não uniformidade da distribuição do calor pela superfície da Terra.

Dentre as opções de fontes alternativas de energia disponíveis atualmente, a energia eólica é considerada uma das mais populares, pelo fato da sua produção ser limpa, não envolver custo de operação, no que tange a fonte de energia, e os sistemas serem economicamente comparáveis com fontes convencionais de combustível fóssil[50].

Se inicialmente a geração eólica era limitada a pequenas unidades para atendimento da demanda em pontos isolados da rede, atendimento residencial e a comunidades, ou ainda para injeção de uma pequena parcela de potência diretamente na rede de distribuição (geração distribuída), atualmente o cenário já é bastante diferente, tendo crescido a instalação em larga escala de grandes fazendas eólicas, fornecedoras de percentuais consideráveis de energia a rede básica.

Na Figura 14 é apresentada a evolução recente do número de usinas eólicas operando no país, que somavam 91 no total em fevereiro de 2014, sobrepondo-se a evolução da capacidade de operação, que alcançou um total de 2250 MW, enquanto que na Figura 15 é apresentada a participação de cada região na geração eólica total do sistema brasileiro para o período compreendido entre fevereiro de 2013 e fevereiro de 2014.

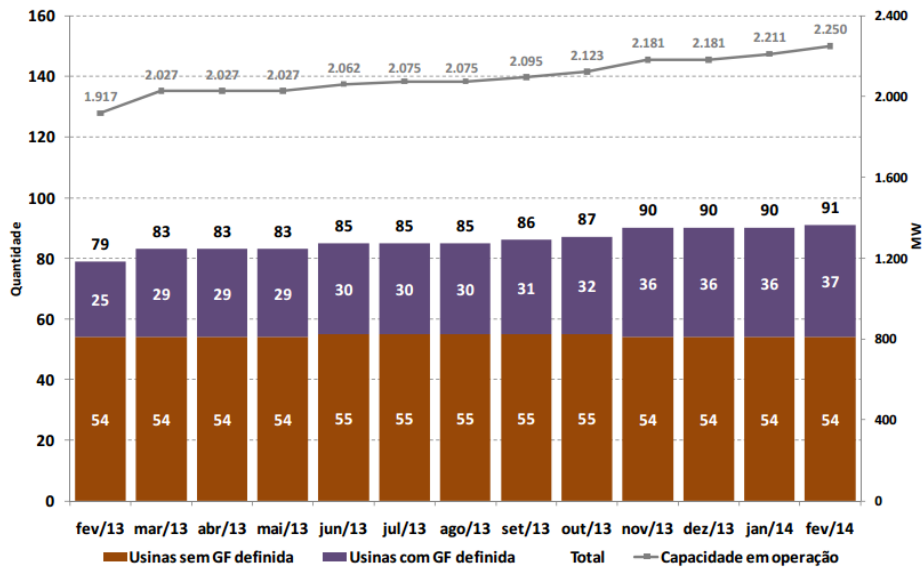


Figura 14–Evolução de Número de Usinas Eólicas[51]

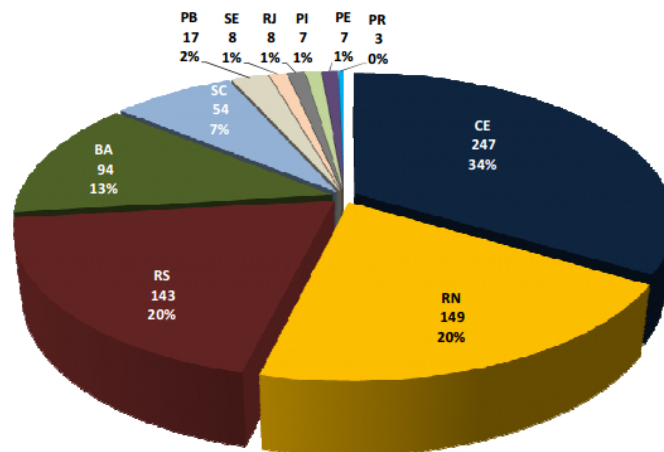


Figura 15–Geração no Período de Fev/2013 até Fev2014[51]

4.2 Características da Potência Elétrica Gerada

A potência extraída pela turbina eólica de um fluxo de ar de densidade ρ , movendo-se a uma velocidade v , perpendicular a uma seção transversal de área A de um cilindro varrido pelas hélices da turbina, pode ser expressa como [52]:

$$P = \frac{1}{2} C_p \rho A v^3 \quad (4.1)$$

onde C_p é o coeficiente de potência da turbina. Este coeficiente representa a eficiência aerodinâmica da turbina e mostra quanto de energia do vento a turbina é capaz de converter em potência mecânica, para cada velocidade de vento. Como a geração é proporcional ao cubo da velocidade do vento, as variações de geração de potência podem ser muito grandes.

Os controles de estol e de passo propiciam um certo nível de controle de geração das turbinas. Velocidades muito altas podem provocar danos nas pás e nos outros componentes da turbina. Por essa razão, o dispositivo de controle limita o funcionamento da turbina até uma velocidade de corte, mais comumente conhecida como *cut out*. Esta velocidade está em torno de 20 a 25 m/s. Por outro lado, velocidades baixas fazem a turbina gerar potências muito baixas e são insuficientes para a partida do gerador. A velocidade inicial de funcionamento da turbina, conhecida como velocidade de acionamento ou *cut in*, está em torno de 3 a 5 m/s. Dessa forma, a operação da turbina fica limitada ao intervalo entre a velocidade de acionamento e a velocidade de corte. Para valores fora desse intervalo, a potência gerada é nula.

A curva de potência de uma turbina eólica representa a sua característica de operação $P(v)$. A Figura 16, extraída de [52], mostra a curva de potência de duas turbinas comerciais, com controles tipo passo e estol. Observa-se que para controle do tipo passo, a geração permanece constante e com o valor igual à potência nominal entre v_{nom} e $v_{cut - out}$. Já para controle do tipo estol, há uma variação da geração em torno do valor nominal.

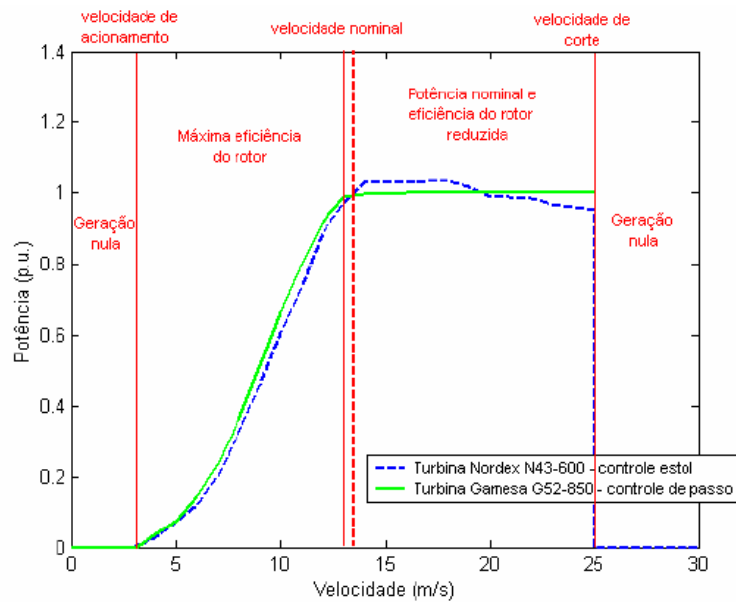


Figura 16– Curvas de Potência de Turbinas com Controles Distintos[52]

4.3 Modelo Estocástico da Turbina

Para um modelo mais preciso da geração eólica para estudos de confiabilidade, é necessário também considerar a representação do comportamento da turbina, que também está sujeita a falhas, agregando mais alguns graus de incerteza ao problema de confiabilidade.

O modelo da turbina eólica deve considerar a possibilidade de falha em um dos quatro elementos que compõem a turbina, que são apresentados em um diagrama de blocos simplificado na Figura 17:

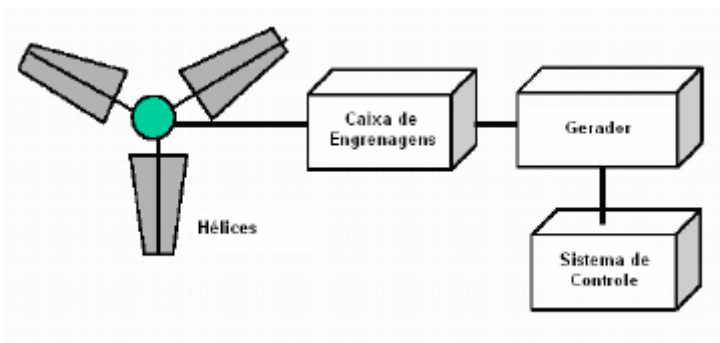


Figura 17– Elementos que Compõem a Turbina Eólica[52]

1. **Hélices:** responsável pela captação da energia do vento, é o elemento de maior contribuição para a falha da turbina, dada a exposição desta às condições do

vento, sendo a taxa de falha típica deste componente dada por 21,5 falhas por cada 20 anos[53];

2. **Caixa de Engrenagens:** responsável pela transmissão da energia mecânica produzida no rotor para o gerador. Como o rotor possui uma baixa velocidade de rotação em relação à velocidade dos geradores convencionais, elas são necessárias para adaptação dos dois sistemas. A taxa de falha típica deste componente é de 4,45 falhas por cada 20 anos[53];
3. **Gerador:** responsável pela transformação da energia mecânica do rotor em energia elétrica, é o elemento de menor contribuição para a falha da turbina, sendo a taxa de falha típica deste componente dada por 3,2 falhas por cada 20 anos[53];
4. **Sistema de Controle:** entre os diversos sistemas de controle de uma turbina, destaca-se a monitoração da velocidade e a orientação do rotor, sendo a taxa de falha típica deste componente como um todo dada por 7,65 falhas por cada 20 anos[53];

A turbina eólica pode ser modelada como um componente a dois estados: em operação e em falha, sendo que o valor total da taxa de falha é dado pela soma dos valores das taxas de cada componente [33]. No caso dos valores típicos apresentados, a taxa de falha da turbina estaria em torno de 2 falhas/ano. Para a modelagem de uma usina que possua N turbinas eólicas, é necessária a consideração da possibilidade de transição entre cada um dos estados operativos dentre todas as turbinas, o que resulta em um total de 2^N estados operativos. Para um melhor entendimento, a Figura 18 apresenta o diagrama de estados para uma usina composta de uma e duas turbinas, com taxas de falha e reparo distintas.

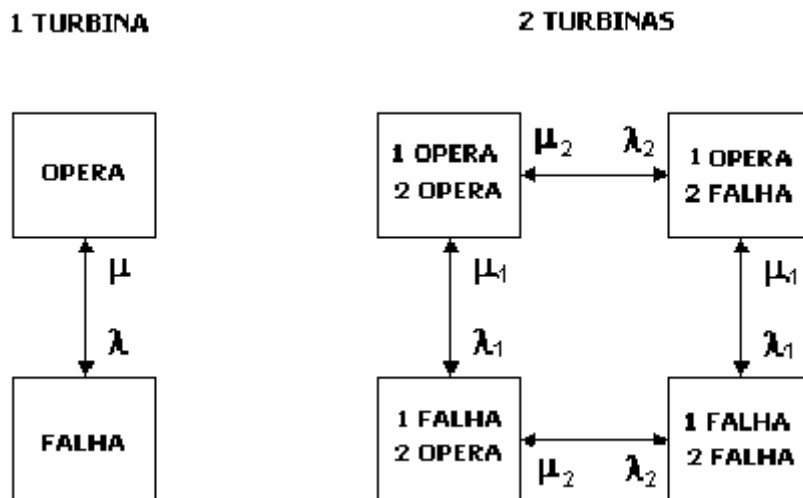


Figura 18– Diagrama de Espaço de Estados de uma Usina com 1 ou 2 Turbinas[52]

4.4 Efeito Esteira

A instalação de um elevado número de turbinas em uma fazenda eólica favorece a geração de perdas aerodinâmicas nas turbinas, resultando em uma geração de energia menor do que aquela nominalmente prevista para a unidade geradora, devido ao sombreamento provocado pelas outras turbinas, como pode ser visto na Figura 19. Este efeito é conhecido como efeito esteira e embora historicamente ele venha sendo desconsiderado na avaliação da confiabilidade de sistemas, o crescente aumento da penetração deste tipo de energia na matriz energética dos sistemas faz com que estas perdas comecem a se tornar significativas, requerendo cada vez mais que os modelos representem com um maior detalhamento este comportamento nas análises e simulações.

Em grandes fazendas eólicas offshore a perda média observada na potência gerada é da ordem de 10% a 20% da produção [54].



Figura 19– Efeito Esteira em um Parque Eólico

Segundo [55], o efeito esteira é um problema a ser considerado em todo o ciclo de vida das fazendas eólicas:

Durante a fase de planejamento, é preciso se obter uma boa estimativa para a localização de cada unidade aerogeradora na fazenda eólica, de modo a maximizar a potência gerada e minimizar o desgaste das unidades. Neste processo, o efeito esteira é

representado, por meio de modelos computacionais, em problemas de otimização de layout, de modo a melhorar ainda mais o custo benefício do empreendimento.

Durante a fase de financiamento, os agentes financeiros requerem estimativas precisas tanto da produção anual de energia da fazenda eólica como do custo de manutenção. Estes dois números estão significativamente relacionados com a qualidade da análise do efeito esteira. Uma redução na incerteza dos modelos computacionais do efeito esteira tem como consequência uma redução no risco dos investimentos, tornando os empreendimentos de geração eólica mais competitivos.

Finalmente, durante a fase de operação, quanto mais precisa a previsão da produção das fazendas eólica, mais precisas serão as análises energéticas e elétricas da operação do sistema, permitindo uma operação mais segura e econômica.

As características elencadas são um problema adicional para a representação das dinâmicas dos ventos por meio de um modelo estocástico. Primeiramente porque o modelo precisa representar conjuntamente as incertezas do módulo e da direção do vento, e este relacionamento não é tão facilmente caracterizado. Adicionalmente, a própria função de densidade de probabilidade marginal da direção do vento não é uma função de fácil caracterização, como pode ser observado na Figura 20, onde são apresentados histogramas para os ângulos de incidência do vento, que definem as direções, para quatro séries de medições (s-210, s-225, s-240 e s-248).

Em relação às séries históricas correspondentes ao posto de medição s210, na Figura 21 é apresentada uma estimativa para a função de densidade de probabilidade conjunta estimada para as variáveis correspondentes, função de densidade 2D, obtidas a partir de um método baseado em estimação *kernel density*, para duas variáveis [56].

As áreas mais escuras indicam as áreas de maiores densidades da função de densidade de probabilidade conjunta estimada, isto é, as coordenadas (velocidade, direção) para as quais se tem a maior concentração de observações.

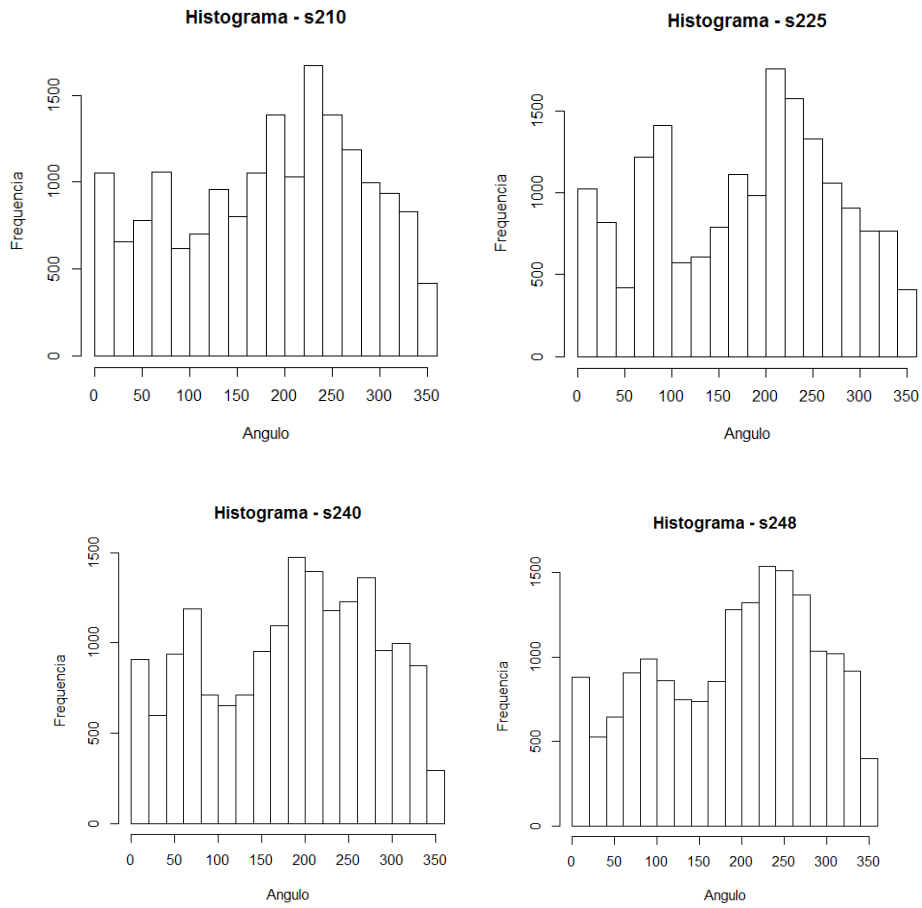


Figura 20– Histograma do Ângulo de Incidência do Vento (Direção do Vento)

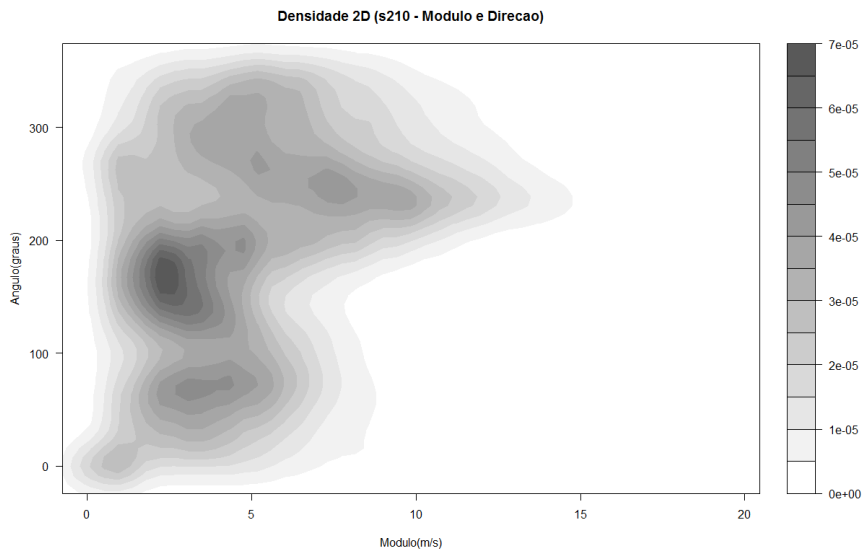


Figura 21– Densidade de Probabilidade 2D para a Série de Vento s-210 (Módulo e Direção)

Visualmente é possível perceber que o relacionamento entre as duas dimensões é descrito por uma função bidimensional complexa, de difícil caracterização paramétrica, uma vez que os contornos da função de densidade são bastante irregulares. Nota-se ainda que não é adequado medir este relacionamento por meio de uma medida linear. Esta afirmação fica mais evidente quando se compara a medida de correlação de Pearson, $r = 0,209$, com o coeficiente de informação máxima das duas variáveis, $MIC = 0,601$, resultando no índice $MIC-R2 = 0,557$.

Estendendo a análise, na Tabela 1 são apresentadas comparações para outras séries históricas, a partir dos índices MIC e $MIC-R2$, onde o último quantifica a imprecisão da caracterização da dependência estatística através da correlação linear, conforme discutido na seção 3.3.2.

A proximidade dos índices MIC e $MIC-R2$, indica que a correlação linear não é uma boa medida de dependência para as variáveis abordadas.

<i>Série</i>	<i>MIC</i>	<i>MIC-R2</i>
s-210	0,601	0,557
s-225	0,494	0,452
s-229	0,602	0,587
s-235	0,620	0,609
s-240	0,543	0,513

Tabela 1– Comparações MIC x $MIC-R2$ para Séries de Módulo e Direção do Vento

A mesma análise pode ser aplicada para comparar tanto a correlação espacial entre as diferentes séries históricas de direções do vento, como a auto-correlação de $lag-1$ de cada série.

Em ambos os casos também se verifica que a correlação linear é inadequada para representar a dependência estatística das variáveis, como pode ser observado na Tabela 2 e na Tabela 3.

<i>Série Direção - 1</i>	<i>Série Direção - 2</i>	<i>MIC</i>	<i>MIC-R2</i>
s-210	s-225	0,605	0,266
s-210	s-229	0,592	0,322
s-225	s-229	0,755	0,327
s-225	s-235	0,533	0,259
s-235	s-240	0,674	0,272

Tabela 2– Comparações MIC x $MIC-R2$ para Correlação Espacial entre Séries de Direção

<i>Série Direção - 1</i>	<i>MIC</i>	<i>MIC-R2</i>
s-210	0,730	0,261

s-225	0,749	0,322
s-229	0,716	0,320
s-235	0,670	0,355
s-240	0,768	0,217

Tabela 3– Comparações *MIC* x *MIC-R2* para Auto-Correlação de *lag-1* de Séries de Direção

A partir da análise empírica apresentada, pode-se esperar que para a obtenção de uma representação mais precisa da incerteza das variáveis elencadas, módulo e direção do vento, o modelo estocástico precisa considerar informações adicionais àquelas codificadas na correlação linear. No limite, o modelo precisa ser capaz de reproduzir os mesmos relacionamentos medidos pelo coeficiente de informação máxima.

Conforme será apresentado no Capítulo 5, esta representação é um dos principais objetivos do modelo proposto neste trabalho.

4.5 Análise Estatística de Séries de Vento

Nesta seção algumas estatísticas importantes associadas a séries reais de vento são avaliadas, com o objetivo de caracterizar melhor o comportamento estocástico do vento e justificar tanto a motivação como a aplicabilidade do modelo desenvolvido neste trabalho.

Foi utilizada a base de dados do Projeto KNMI HYDRA, provido pelo *Royal Netherlands Meteorological Institute (KNMI)* [57], por ser uma base aberta, com um histórico consolidado de medidas horárias de velocidade e direção do vento para dezenas de estações na Holanda, tendo como principal atrativo o fato de ser bastante completa.

Em comparação com outras bases de dados de séries de vento, principalmente em comparação com muitas das séries medidas no sistema brasileiro, esta é uma base que apresenta um horizonte de medições bastante extenso, havendo muitas séries com mais de 20 anos de medições reais. Reúne mais de 30 estações de medições, o que é uma excelente característica para estudar a representação de dependência estatística de alta dimensionalidade, principalmente pelo fato do relacionamento entre as diferentes estações ser bastante heterogêneo, sendo que existem medições tanto para postos offshore, quanto para postos em terra.

4.5.1 Séries Horárias de Vento para um Horizonte de um Ano

Para possibilitar uma primeira visão geral do comportamento horário do vento, na Figura 22 são apresentadas seis séries horárias de vento para o horizonte de um ano, correspondentes aos postos de medição: s-210, s-225, s-242, s-249, s-269 e s-380.

Observa-se nestas séries o comportamento bastante variável, típico das séries de vento, com frequentes transições dos estados próximos dos valores mínimos para os estados próximos dos valores máximos em um curto espaço de tempo.

Visualmente também é possível perceber que existe uma forte correlação entre as séries, principalmente quando se compara as séries s-210 e s-225, que visualmente são bastante similares.

Na Figura 23 são apresentadas as mesmas séries, enquadrando apenas as primeiras três semanas do horizonte. Para cada série, as quatro semanas são sobrepostas em uma escala de 168 horas, sendo cada semana representada por um tipo diferente de linha.

Comparar as séries para os mesmos tipos de linha implica na comparação das séries para os mesmos instantes de tempo, isto é, as séries de linha cheia correspondem a mesma semana para todas as séries, assim como as séries de linha tracejada e as séries de linha pontuada. A partir desta comparação, é possível se observar melhor a alta correlação entre as séries, dada a similaridade do perfil das séries, em especial, como já anteriormente observado, para as séries s-210 e s-225 e também para as séries s-269 e s-380.

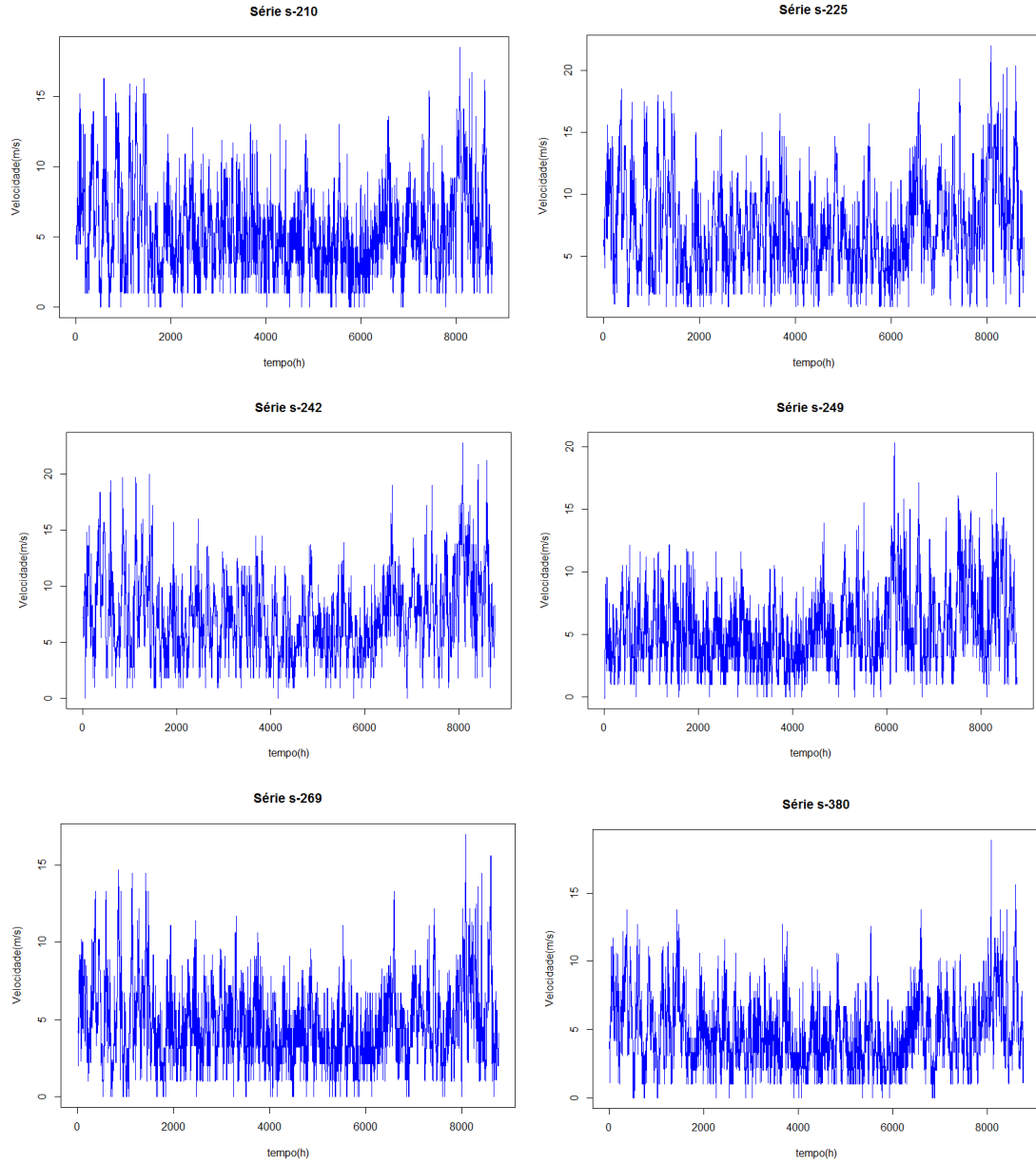


Figura 22– Séries Horárias de Vento para um Horizonte de um Ano

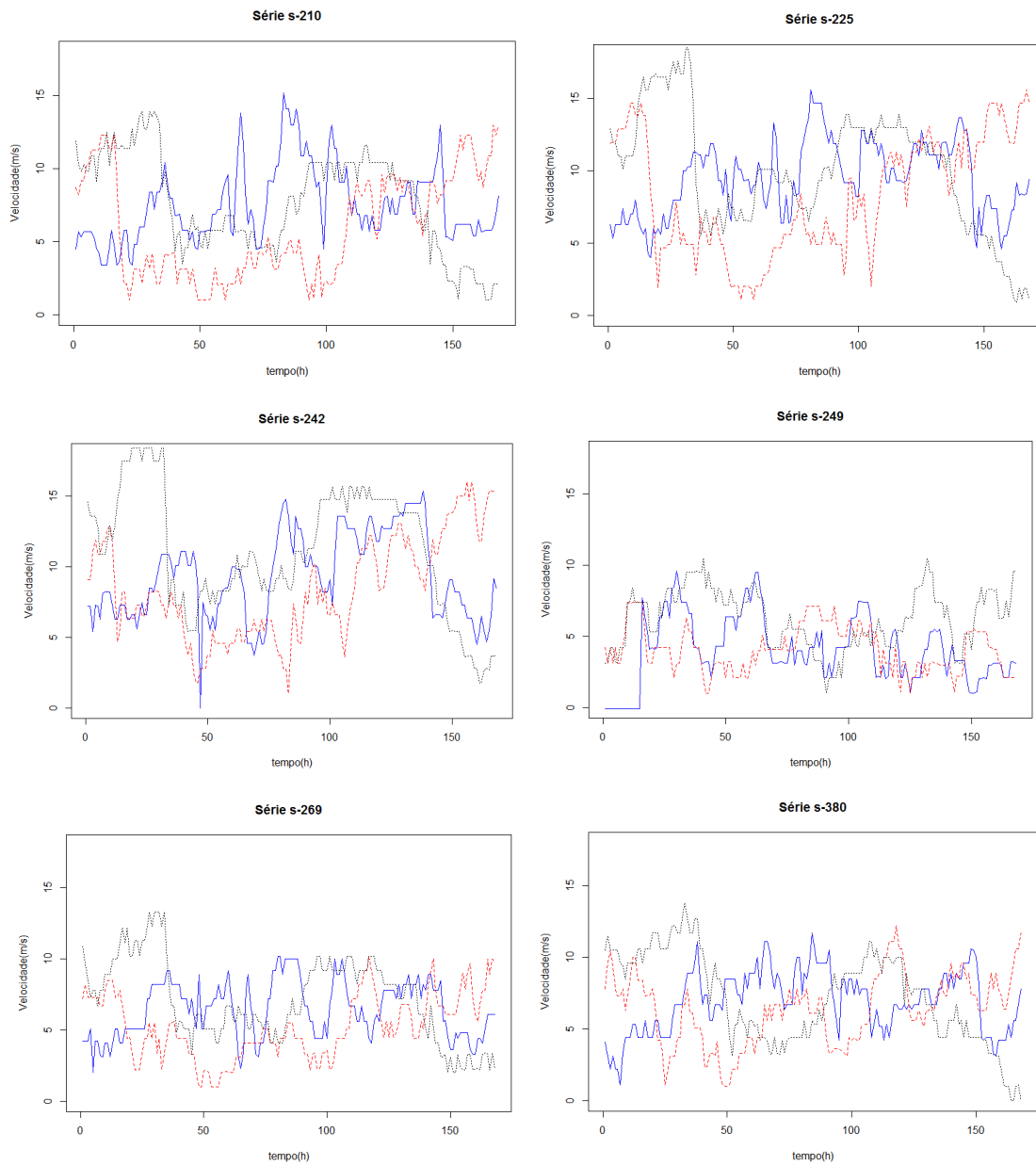


Figura 23– Séries para as Três Primeiras Semanas do Horizonte

4.5.2 Estimação da Densidade de Probabilidade

Na literatura sobre o assunto, é comum a representação das distribuições de probabilidade das séries de vento a partir de um ajuste em cima da distribuição de Weibull [15], tanto para a representação da distribuição consolidada do ano, como para a representação da distribuição segregada por mês.

Esta distribuição é dada por:

$$f(x) = \frac{k}{c} \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right] \quad (4.2)$$

Onde c e k são respectivamente os parâmetros de escala e de forma.

Quanto maior o valor do parâmetro k , mais estreita e mais pontuda é a curva de weibull, conforme está representado na Figura 24, extraída de [52]. Isso significa que existe uma menor variação da velocidade de vento. O fator de escala está relacionado ao valor médio da velocidade, quanto maior o valor de c , maior será o valor de velocidade média.

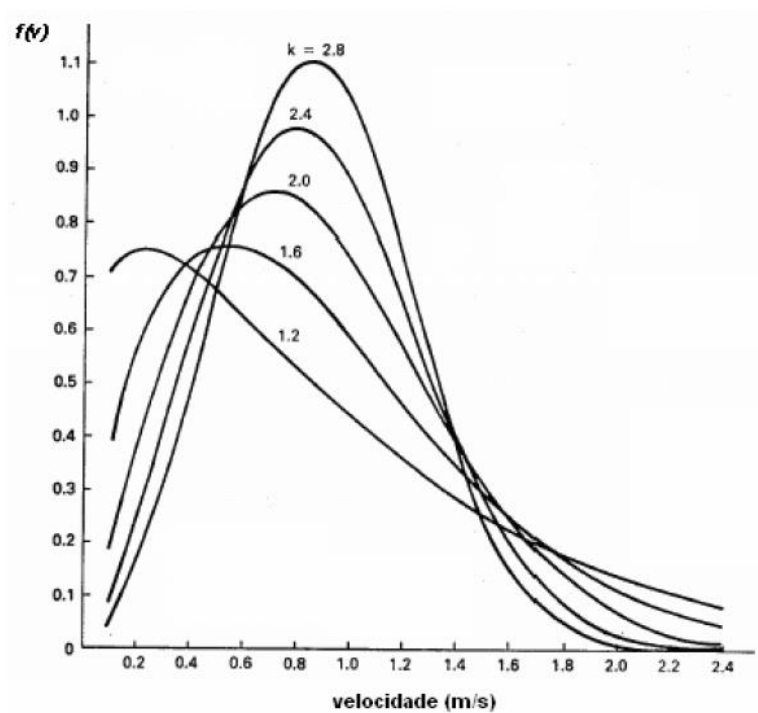


Figura 24– Variação da Distribuição de Weibull de Acordo com os Parâmetros[52]

Para os modelos auto-regressivos lineares de séries temporais, adotar a abordagem paramétrica e assumir a distribuição de weibull, apresenta a vantagem de possibilitar a derivação de uma expressão matemática para a transformação da série original na sua equivalente, com distribuição normal, conforme discutido em [58].

Conforme já dito anteriormente, os modelos propostos neste trabalho utilizam a abordagem não paramétrica na estimação das funções de distribuição de probabilidade, baseando-se no método de estimação por *kernel density estimation*.

Na Figura 25 são apresentadas as funções de densidade de probabilidade para as séries de vento, estimadas de maneira independente para cinco anos consecutivos e para o conjunto total de observações.

Observa-se que as funções de densidade de probabilidade não variam muito ano a ano, e por consequência a função de densidade de probabilidade de todo o conjunto de observações das séries é também bastante similar a cada uma das densidades individualizadas por ano. Isto é uma característica importante na definição de modelos para a simulação estocástica de séries de vento, uma vez que sugere uma boa plausibilidade para os cenários sintéticos gerados por um modelo estimado a partir dos dados históricos.

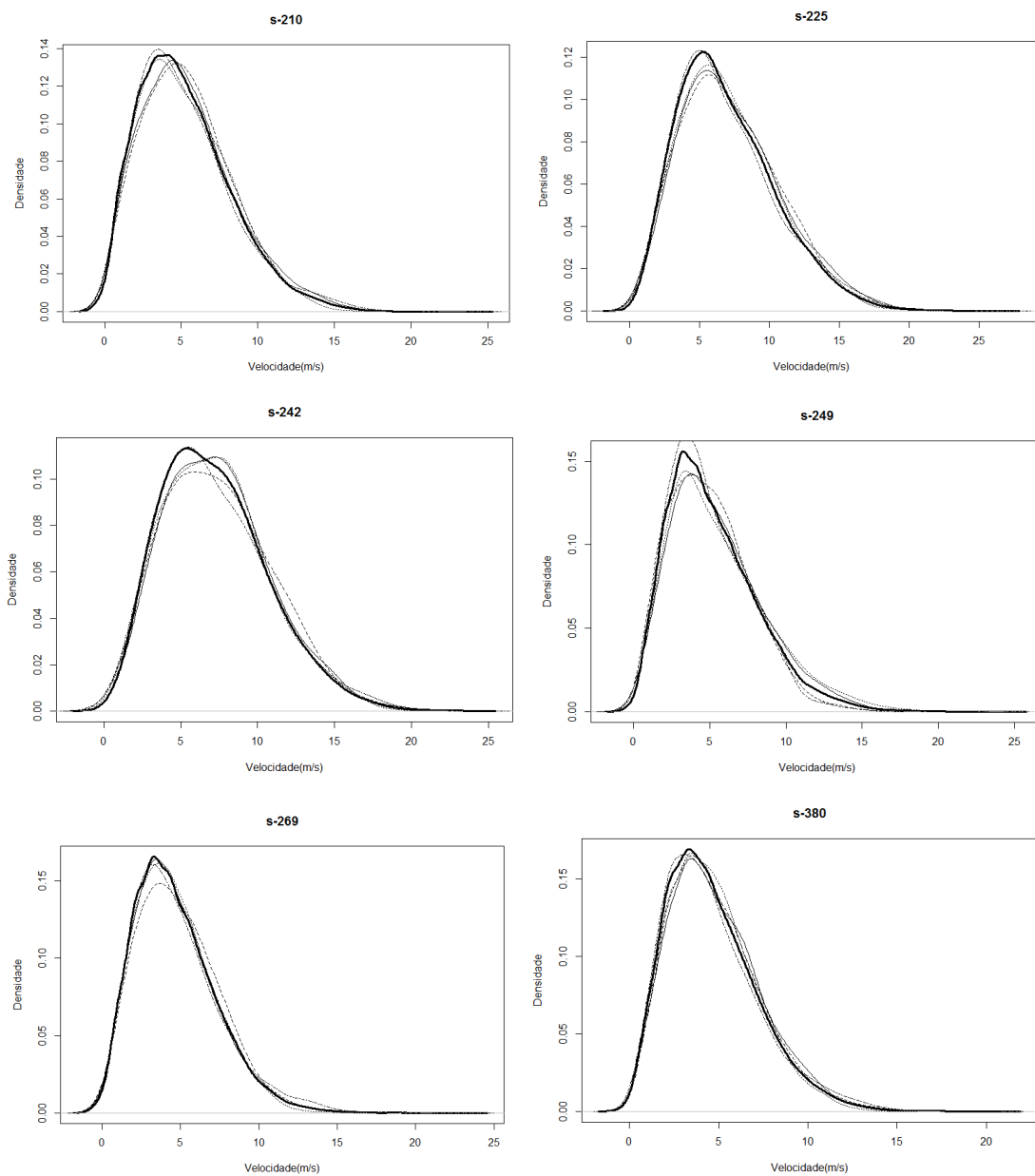


Figura 25– Funções de Densidade de Probabilidade Estimadas para as Séries

Um ponto interessante a ser avaliado neste tópico é a comparação entre as estimativas paramétricas e não paramétricas da função de densidade de probabilidade da velocidade do vento. Na Figura 26 esta comparação é apresentada para as séries analisadas, tendo sido utilizada família de distribuição Weibull para o ajuste das funções de densidade de probabilidade paramétricas.

A linha cheia representa a densidade estimada de forma não paramétrica e a linha tracejada representa a densidade obtida a partir do ajuste dos parâmetros da distribuição Weibull.

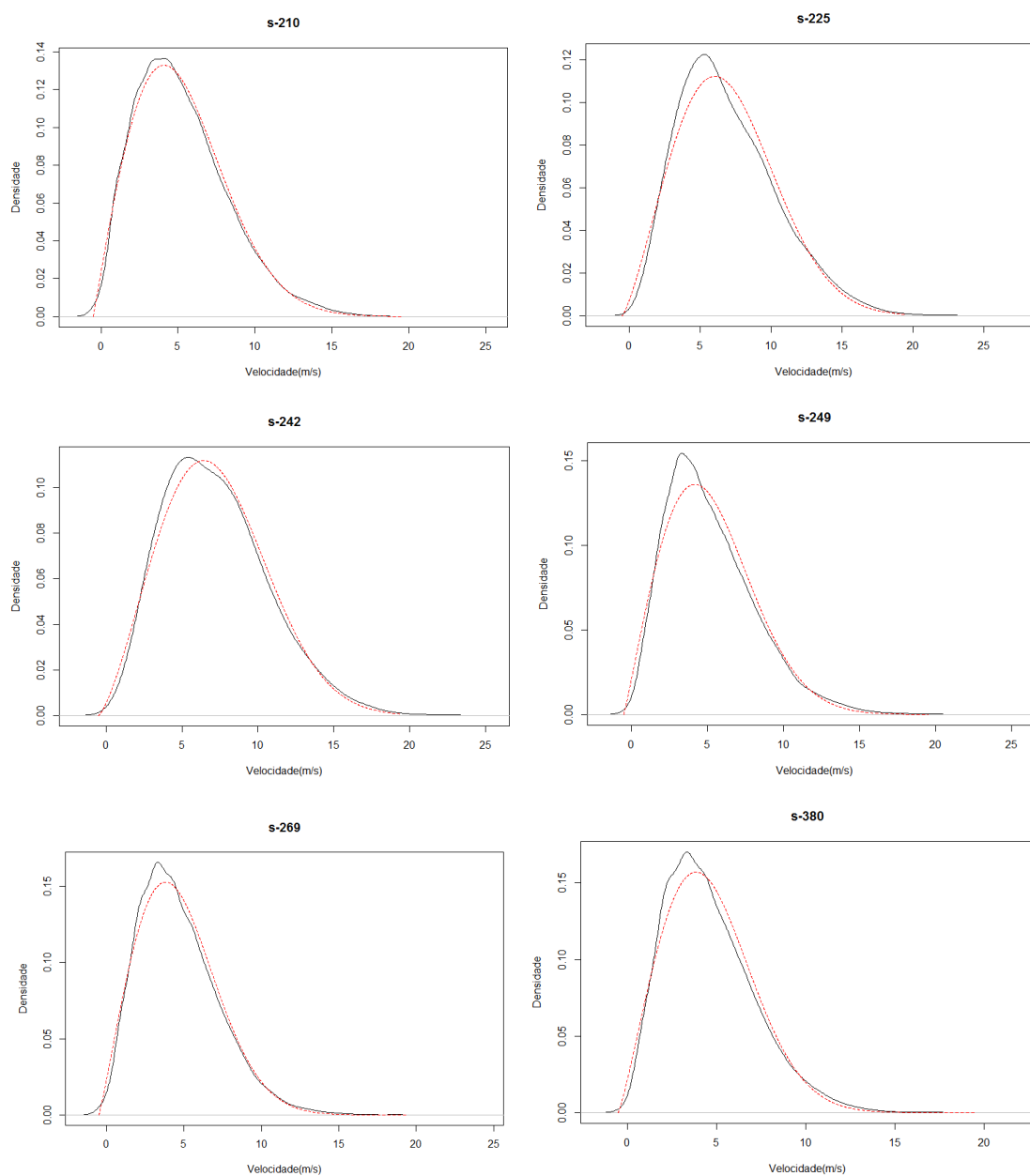


Figura 26– Comparações Entre as Densidades não Paramétricas e as Paramétricas

Observa-se que para as séries analisadas existe uma boa similaridade entre os dois tipos de densidade de probabilidade, havendo apenas um leve deslocamento da moda de uma distribuição em comparação à outra. Isto sugere que, se por um lado a distribuição Weibull parece ser uma família adequada para a representação das séries apresentadas, por outro, mesmo sem nenhuma informação a respeito desta ou qualquer outra família de distribuição de probabilidade, o estimador não paramétrico é capaz de identificar a função de densidade de probabilidade de forma bastante precisa.

4.5.3 Estimação da Auto-correlação

A auto-correlação, também conhecida na análise de séries temporais como correlação temporal, mensura o grau de dependência estatística da variável aleatória em um instante qualquer em relação a esta mesma variável nos instantes anteriores. Isto é, mensura o acoplamento temporal de uma série.

Na avaliação da confiabilidade, a representação da auto-correlação é importante para obtenção dos índices baseados em frequência e duração das falhas, uma vez que os índices baseados em probabilidade e energia dependem apenas das probabilidades estacionárias dos estados dos componentes. Para a obtenção de uma estimativa precisa dos índices baseados em frequência e duração, no entanto, é necessário que os modelos estocásticos representem corretamente o acoplamento temporal dos componentes.

Na análise de séries temporais, é usual o emprego de uma medida derivada da auto-correlação, denominada auto-correlação parcial, que avalia a auto-correlação de uma série temporal para um determinado atraso, *lag*, eliminando da medida a dependência linear do atraso *lag* com o atraso *lag+1*. Esta abordagem permite uma visão mais correta a respeito de qual a ordem, isto é, qual a extensão do passado que o modelo estocástico precisa conhecer, a partir do instante atual, para que ele seja capaz de representar corretamente a dinâmica da variação da série temporal[59].

Na Figura 27 são apresentadas as auto-correlações parciais para as séries de vento, para um atraso máximo de trinta horas, *lag-30*.

Observa-se que invariavelmente as séries apresentam uma auto-correlação bastante significativa para *lag-1*, mas extremamente baixa para os demais atrasos. Esta característica, que é muito frequentemente observada em séries temporais de vento, permite a abordagem de representação da dinâmica do vento por meio de uma cadeia de markov, uma vez que esta característica se coaduna com a propriedade de memória markoviana, premissa básica para que um processo estocástico possa ser representado

por meio de cadeias de markov. Esta propriedade será explorada na construção do modelo estocástico para representar as transições dos estados dos elementos variantes no tempo, apresentado no Capítulo 5, permitindo a obtenção de índices de frequência e duração.

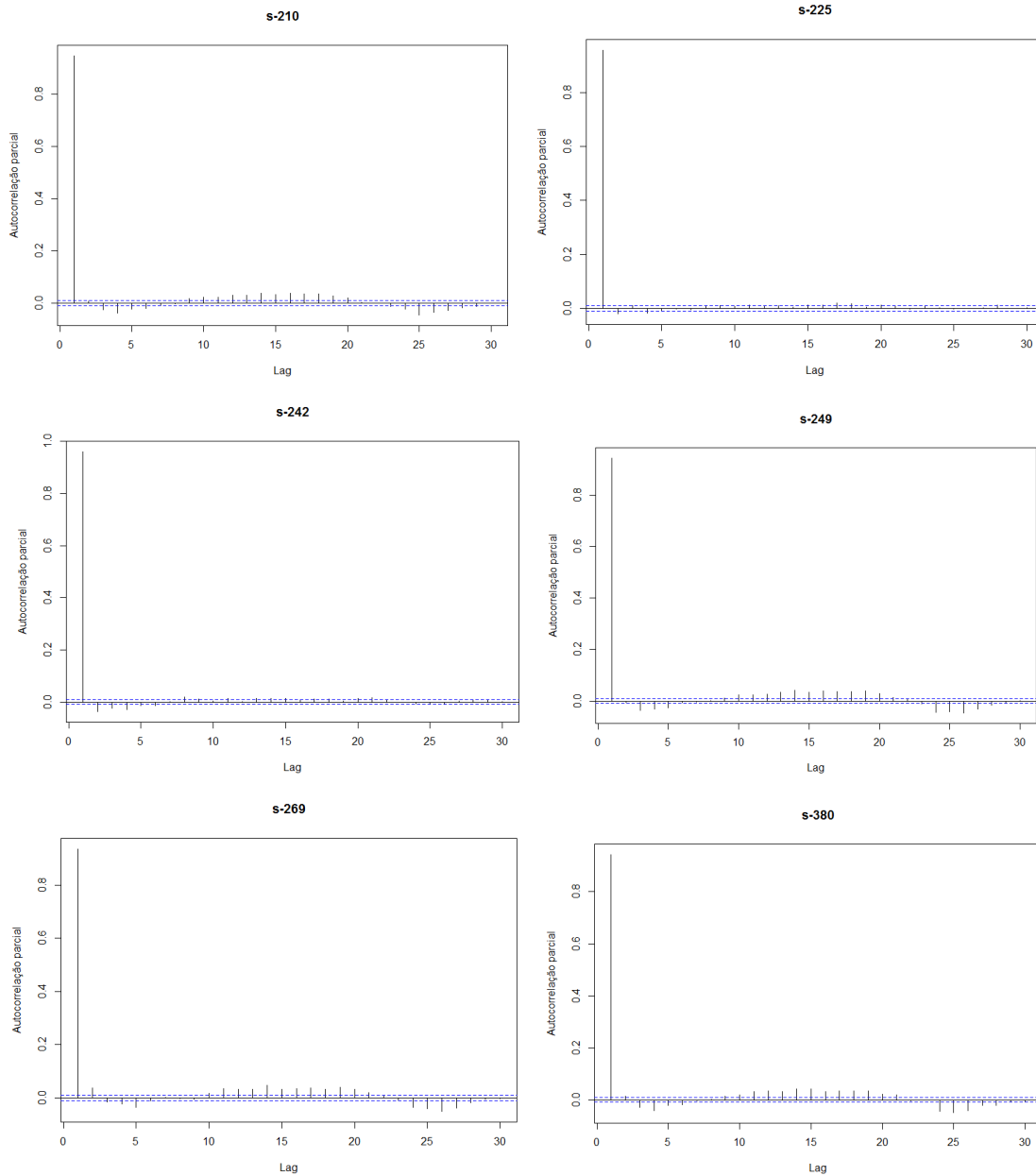


Figura 27– Auto-Correlações Parciais Estimadas para as Séries

4.6 Análise Estatística de Séries de Geração Eólica

Por conveniência do estudo ou limitação de dados disponíveis em algumas análises, os modelos que são empregados para a geração de cenários de geração eólica na simulação estocástica não são construídos a partir das séries históricas de velocidade e direção do vento, mas sim diretamente a partir das séries históricas de geração das turbinas, ou ainda, das gerações das usinas, já agregando na potência final das usinas a geração de cada turbina.

As variáveis aleatórias que representam a incerteza destas séries de geração tendem a ter características ainda mais difíceis de representar através de um modelo estocástico, quando comparadas às variáveis aleatórias das séries de vento, uma vez que existem não linearidades adicionais inseridas no processo, resultando em uma maior distorção das densidades de probabilidade e no aumento da complexidade do relacionamento entre as variáveis. Estas não linearidades se originam, principalmente, devido as curvas de potência das turbinas, que, conforme anteriormente apresentado, possuem uma característica não linear.

Ao contrário das séries temporais de vento, para as quais é discutido na literatura o emprego de famílias de distribuição para caracterizar o comportamento destas variáveis aleatórias, no caso das séries de geração eólica, não é possível, em geral, adotar nenhum tipo conhecido de família de distribuição que seja adequada para esta caracterização.

Na Figura 28 é apresentada a densidade de probabilidade estimada para uma série de geração eólica do Nordeste. A série corresponde a uma variação do fator de capacidade que é aplicado a um conjunto de usinas do Nordeste. Desta maneira, a geração final de cada uma destas usinas corresponde à aplicação deste percentual a capacidade nominal de geração de cada usina.

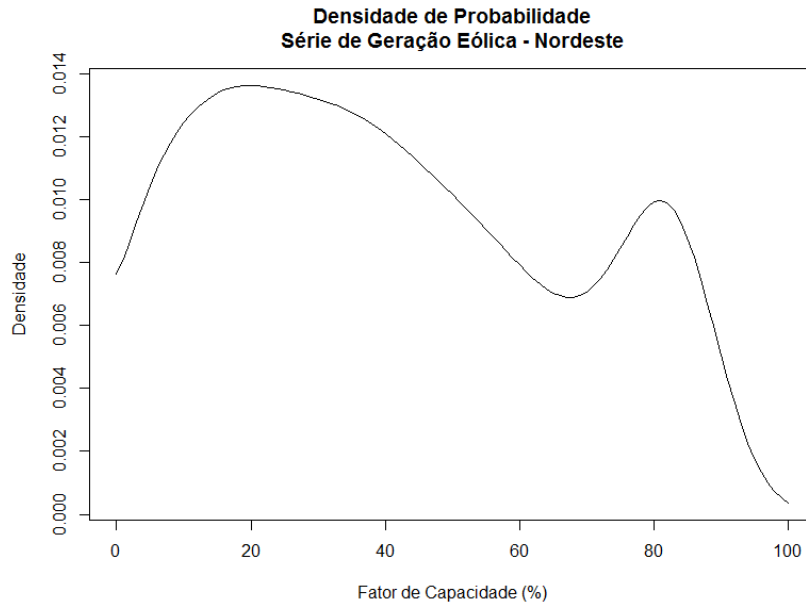


Figura 28– Densidade de Probabilidade de uma Série de Geração Eólica do Nordeste

Observa-se que não é viável a caracterização desta variável aleatória por meio de uma família conhecida de distribuição, principalmente devido à característica bimodal da distribuição.

Na Figura 29 é apresentada a autocorrelação parcial para esta série de geração. Este gráfico revela que, de forma similar às séries de vento, a predominância do atraso de primeira ordem sobre os demais permite que o processo estocástico da geração eólica também seja representado por meio de um modelo de cadeias de Markov.

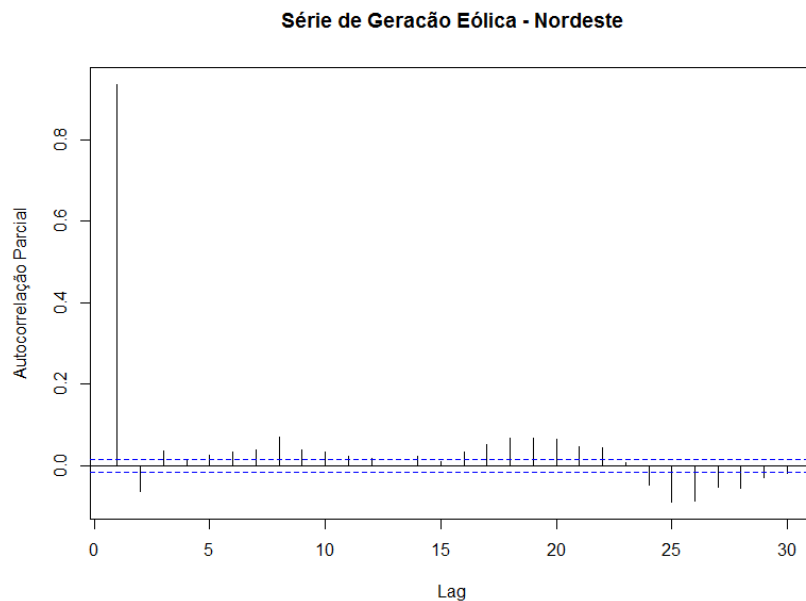


Figura 29– Autocorrelação Parcial para a Série de Geração Eólica do Nordeste

Adicionalmente, na Figura 30 é apresentada a densidade de probabilidade estimada para uma série de geração eólica do Sul.

Para esta série, também se verifica a característica bimodal da densidade de probabilidade, embora o formato seja diferente da série do Nordeste. Conforme já observado para a série anterior, não é correta a aplicação da abordagem paramétrica neste caso, com o arbítrio de uma família conhecida de distribuição.

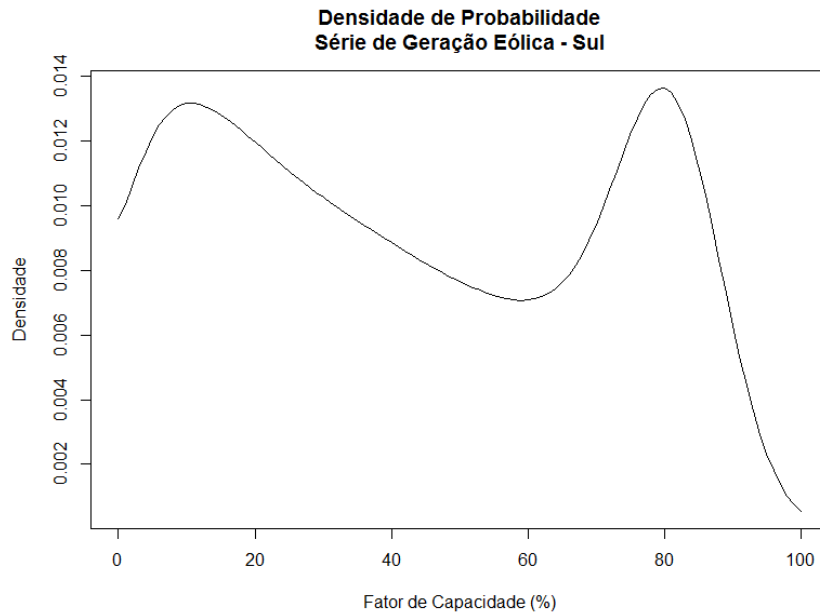


Figura 30– Densidade de Probabilidade de uma Série de Geração Eólica do Sul

O gráfico de autocorrelação parcial para esta série é apresentado na Figura 32, o qual reitera a possibilidade de representação do processo estocástico da geração eólica por meio de um modelo de cadeias de Markov.

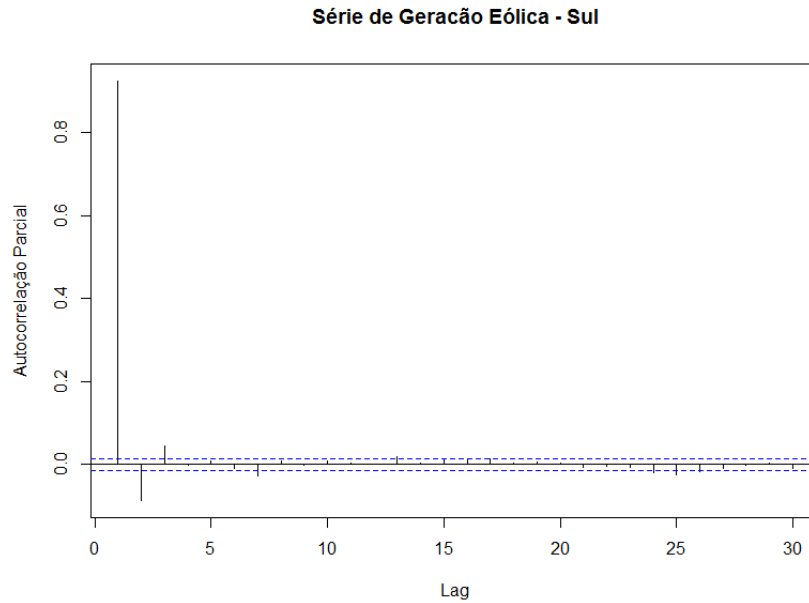


Figura 31– Autocorrelação Parcial para a Série de Geração Eólica do Nordeste

Para ilustrar o quanto o relacionamento entre as variáveis aleatórias tende a ser mais complexo na representação direta das séries de geração, na Figura 32 é apresentada a função de densidade de probabilidade conjunta estimada para as séries de geração eólica das usinas do Sul localizadas no litoral e no interior da região.

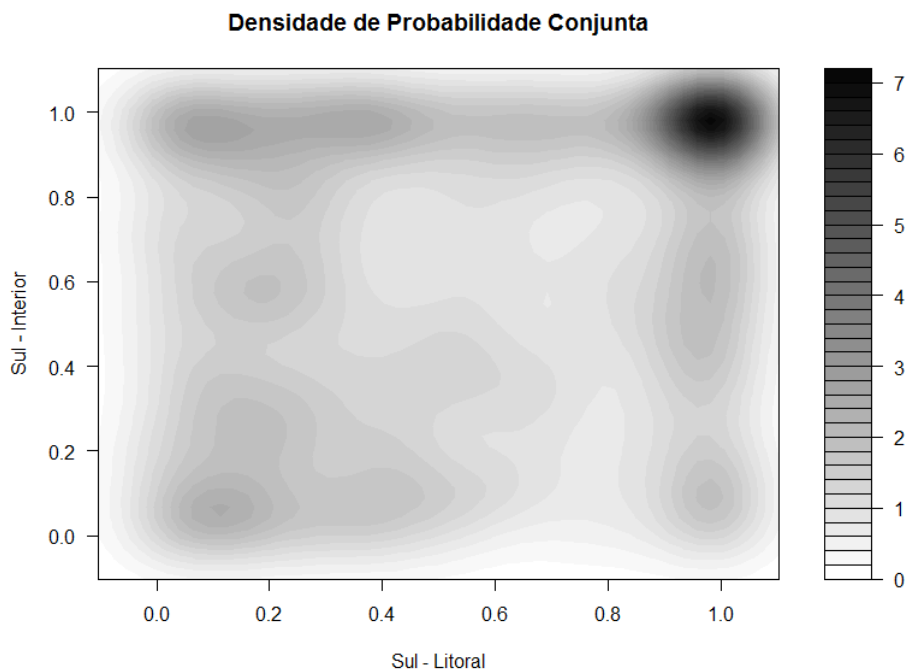


Figura 32– Densidade de Probabilidade Conjunta das Séries de Geração do Sul: Litoral x Interior

O principal ponto a se destacar, é que o gráfico indica uma alta probabilidade de coincidência das duas gerações nos seus níveis mais altos. No entanto, essa coincidência não é mantida para os níveis mais baixos de geração, uma vez que a densidade indica que há uma maior dispersão da densidade para os valores inferiores de geração. Este comportamento fortemente não linear é impossível de ser caracterizado pela correlação de *Pearson*.

Esta observação pode ser quantificada a partir do medidor *MIC-R2*, que vale 0,10 para este par, valor que está próximo do *MIC* medido, que vale 0,13.

Como consequência deste fato, a adoção de um modelo de representação da geração eólica que represente a dependência estatística por meio da correlação de *Pearson* tenderia a desprezar este forte acoplamento entre as variáveis para os valores mais altos de geração, uma vez que esta informação não é capturada pela correlação linear.

Capítulo 5

Proposta de Modelo para Representação de Elementos Variantes no Tempo Estatisticamente Dependentes

5.1 Introdução

No processo de amostragem da Simulação Monte Carlo, os estados do sistema correspondem a combinação dos estados de todos os seus componentes, sendo que na Simulação Monte Carlo Não Sequencial estes são determinados a partir da probabilidade de cada componente estar em cada um de seus respectivos estados.

Considerando-se um sistema com m componentes, pode-se representar um estado deste sistema pelo vetor $\underline{x} = [x_1 \ x_2 \ \dots \ x_k \ \dots \ x_m]$, onde x_k representa o estado do k -ésimo componente.

Os modelos utilizados na Simulação Monte Carlo Não-Sequencial partem do princípio de que os estados dos elementos são estatisticamente independentes. Assim, definindo-se $P(x_k)$ como a probabilidade de ocorrência associada ao k -ésimo componente, a probabilidade de ocorrência de i -ésimo estado do sistema $P(\underline{x}^i) = p_i$, é dada por:

$$P(\underline{x}^i) = p_i = \prod_{k=1}^m P(x_k) \quad (5.1)$$

Esta independência estatística entre os elementos do sistema permite que, mesmo para sistemas de grande porte, o processo de amostragem seja de baixo grau de complexidade, uma vez que a amostra do estado de cada elemento depende apenas da função de distribuição de probabilidade do próprio elemento, conforme apresentado no Capítulo 2.

A introdução de variáveis estatisticamente dependentes no sistema, implica no surgimento de probabilidades conjuntas na equação, tal como:

$$P(\underline{x}^i) = p_i = P(v_1, v_2, \dots, v_w) \cdot \prod_{k=1}^m P(x_k) \quad (5.2)$$

Onde v_1, v_2, \dots, v_w representam as W variáveis estatisticamente dependentes do sistema e $P(v_1, v_2, \dots, v_w)$ representa a função de distribuição de probabilidade conjunta que caracteriza estas variáveis.

Estas probabilidades conjuntas impedem a determinação precisa dos índices de confiabilidade utilizando a abordagem convencional da Simulação Monte Carlo Não Sequencial, uma vez que esta assume a hipótese de eventos independentes para os estados dos elementos.

Desta maneira, para que seja possível aplicar a Simulação Monte Carlo Não Sequencial na avaliação da confiabilidade de um sistema com elementos estatisticamente dependentes, é necessário que o algoritmo conceitual, apresentado na seção 2.3, seja modificado, de modo a incorporar a informação contida na probabilidade conjunta destes elementos.

Uma abordagem possível é a representação explícita da probabilidade conjunta das variáveis estatisticamente dependentes, isto é, a representação explícita da função $P(v_1, v_2, \dots, v_w)$ no modelo de Simulação Monte Carlo Não Sequencial[60].

Um caso trivial para a aplicação deste tipo de abordagem é quando as variáveis são contínuas e atendem as condições de normalidade, podendo ser caracterizadas conjuntamente por uma única função de densidade de probabilidade normal multivariada, tal que:

$$(v_1, v_2, \dots, v_w) \sim N(\mu, \Sigma) \quad (5.3)$$

Onde, $\mu = [E[v_1], E[v_2], \dots, E[v_w]]$ é o vetor de médias das variáveis representadas e $\Sigma = [Cov[v_i, v_j]]$, $i = 1, 2, \dots, w$, $j = 1, 2, \dots, w$ é a matriz de covariância.

A obtenção de amostras de variáveis aleatórias estatisticamente dependentes, caracterizadas a partir de uma função de distribuição normal multivariada, é um assunto já bastante discutido na literatura, por exemplo, [25], existindo alternativas bem eficientes de algoritmos para esta aplicação, por exemplo, [61], não existindo desta forma maiores dificuldades para se aplicar a Simulação Monte Carlo Não Sequencial neste caso.

Com o intuito de prover um melhor entendimento desta afirmação, é apresentada a seguir a modificação do algoritmo conceitual da Simulação Monte Carlo Não Sequencial para incorporar a representação explícita desta função de densidade de probabilidade conjunta normal multivariada:

1. Selecione um estado do sistema $\underline{x} \in X$ correspondente a um ponto de operação, isto é, disponibilidade de componentes, níveis de carga e geração, etc;
 - a. *Amostre os estados dos elementos convencionais, representados por variáveis aleatórias estatisticamente independentes, a partir das suas próprias funções de distribuições de probabilidade;*
 - b. *Amostre os estados dos elementos representados por variáveis aleatórias estatisticamente dependentes, a partir da função de densidade de probabilidade normal multivariada que as caracteriza, $N(\mu, \Sigma)$.*
2. Calcule o valor de uma função de avaliação $F(\underline{x})$;
3. Atualize a estimativa $\bar{E}(F)$, o valor esperado dos índices de confiabilidade do processo, baseado no resultado obtido no passo 2;
4. Se a precisão das estimativas é aceitável, o processo é dito convergido, caso contrário retorne ao passo 1.

Embora este procedimento seja simples, dificilmente as variáveis aleatórias que representam os elementos dos sistemas de potência atendem às condições de normalidade para que possam ser caracterizadas a partir da densidade de probabilidade normal multivariada. E ainda, conforme anteriormente discutido, os principais elementos variantes no tempo representados nos sistemas de potência apresentam grandes assimetrias nas suas distribuições, sendo imprecisas as suas representações pela distribuição normal.

Para se contornar esta limitação, um procedimento adotado na prática é a aplicação de funções de transformação às variáveis aleatórias, que permitem que sejam obtidas novas variáveis que podem ser individualmente caracterizadas como variáveis aleatórias normalmente distribuídas, conforme apresentado na Figura 33.

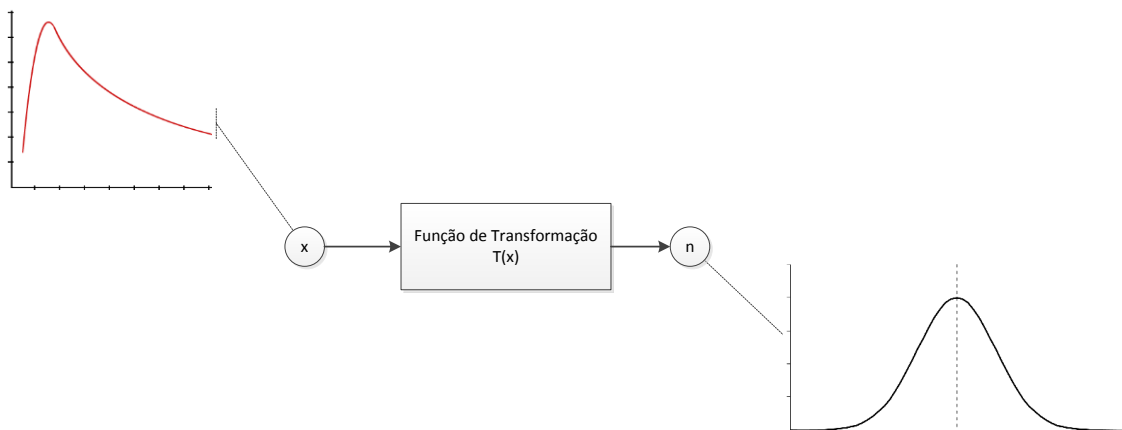


Figura 33– Transformação de Variáveis Aleatórias em Variáveis Normalmente Distribuídas

Após essa transformação, a amostragem dos estados é realizada a partir da densidade de probabilidade destas variáveis transformadas normais, conforme descrito no algoritmo anterior. Ao final do processo, são aplicadas as funções inversas da transformação, que produzem, enfim, as amostras com as mesmas distribuições das variáveis originais, conforme apresentado na Figura 34.

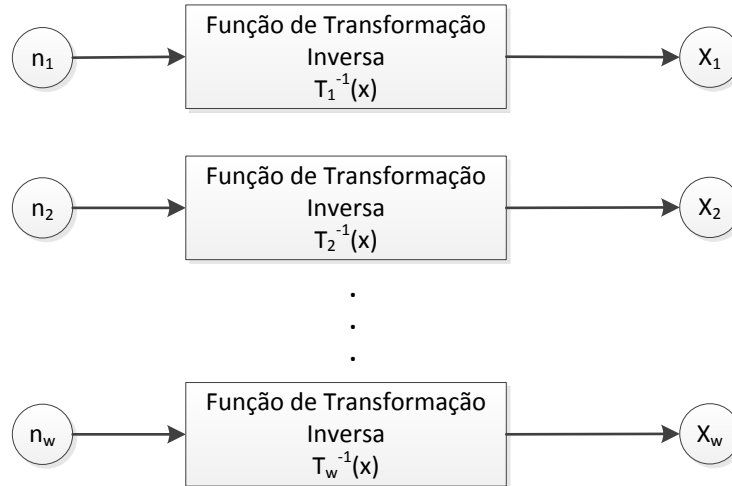


Figura 34– Geração de Amostras a Partir das Variáveis Aleatórias Transformadas

Este procedimento é bastante comum na estimação de modelos estocásticos auto-regressivos para representação de vazões afluentes[3], para o qual se arbitra, em geral, que as variáveis aleatórias que descrevem as vazões afluentes seguem uma distribuição de probabilidade log-normal, o que implica na adoção da função logarítmica como função de transformação e, conseqüentemente, na adoção da função exponencial como função inversa da transformação.

Uma primeira dificuldade que se tem para se aplicar este método é que é necessário se determinar as funções de transformações, e estas variam de acordo com a família de distribuição de probabilidade arbitrada e com os parâmetros de ajuste das variáveis aleatórias a estas famílias de distribuição. No entanto, os principais pontos negativos desta abordagem estão relacionados às limitações da correlação linear, discutidas na seção 3.3.1. Primeiramente, ao se aplicar uma transformação a qualquer conjunto de variáveis aleatórias, só se garante a preservação da correlação linear das variáveis originais para o caso em que a função de transformação é linear. Esta condição nunca é verdadeira, uma vez que a transformação linear não produz variáveis aleatórias normais a menos que as variáveis originais já sejam normalmente distribuídas [42].

Em segundo lugar, o fato de cada variável aleatória transformada atender individualmente às condições de normalidade não implica necessariamente que

conjuntamente elas possam ser caracterizadas a partir de uma distribuição de probabilidade normal multivariada [42].

Desta maneira, embora esta abordagem seja usual, é importante frisar que, a rigor, existem alguns erros conceituais que são desprezados quando se adota este tipo de processo. As principais consequências destes erros conceituais para o modelo estocástico são a geração de amostras multivariadas com distorção da correlação linear estimada a partir dos dados históricos, e impossibilidade da representação dos relacionamentos estatísticos não lineares.

Uma abordagem possível para este problema, apresentada em [60], é a decomposição da função de distribuição de probabilidade conjunta destas variáveis em funções de distribuição de probabilidades condicionais, através da aplicação da regra da cadeia, de modo que:

$$P(v_1, v_2, \dots, v_w) = P(v_1) \cdot P(v_2|v_1)P(v_3|v_2, v_1) \dots P(v_w|v_{w-1}, \dots, v_2, v_1) \quad (5.4)$$

Desta forma, constrói-se funções de distribuição de probabilidade condicionais correspondentes aos termos: $P(v_2|v_1)$, $P(v_3|v_2, v_1)$, etc., que podem ser utilizadas no algoritmo da Simulação Monte Carlo Não Sequencial para se gerar amostras para as variáveis estatisticamente dependentes.

Embora esta abordagem possa diminuir a complexidade das funções de distribuição de probabilidade representadas, em relação à alternativa de se representar explicitamente uma única função de distribuição de probabilidade, ainda assim a aplicação desta abordagem está limitada à representação de poucas variáveis, pois o número de parâmetros das funções de distribuição de probabilidade cresce exponencialmente com o aumento do número de variáveis das quais ela depende.

Face às limitações discutidas acima, o objetivo principal do modelo proposto neste trabalho é prover uma nova alternativa para a definição de uma função de densidade de probabilidade conjunta para a caracterização das variáveis aleatórias que representam os elementos variantes no tempo na Simulação Monte Carlo Não Sequencial.

Dada a dificuldade de aplicação da abordagem paramétrica na estimação desta função de densidade de probabilidade conjunta para representação dos elementos variantes no tempo, conforme já discutido, a premissa inicial do modelo proposto é a adoção da abordagem não paramétrica. Uma característica adicional é que as variáveis aleatórias associadas aos elementos variantes no tempo abordados neste trabalho são contínuas. Isso dificulta a aplicação do método não paramétrico mais simples, que consiste na caracterização da função de distribuição de probabilidade por meio de um

histograma das séries históricas, conforme discutido na seção 3.2.1. Em contraste, o método *kernel density estimation*, descrito na seção 3.2.2, é uma alternativa flexível que permite a estimação não paramétrica de funções de densidade de probabilidade de variáveis contínuas, sendo, por este motivo, adotado no modelo proposto para a estimação da função de densidade de probabilidade conjunta para a representação dos elementos variantes no tempo.

Uma limitação inicial da aplicação do método *kernel density estimation* neste contexto é que a sua formulação é válida apenas para se representar funções de densidade de probabilidade de uma única variável aleatória, o caso unidimensional, e a função que precisa ser estimada para que seja possível a aplicação da Simulação Monte Carlo Não Sequencial é a função de densidade de probabilidade conjunta das variáveis aleatórias estatisticamente dependentes, correspondente ao caso multidimensional. Conforme discutido no Capítulo 3, as densidades de probabilidade conjuntas representam de forma mais correta as variáveis aleatórias porque, além de caracterizar as variáveis aleatórias individualmente, também caracterizam os relacionamentos estatísticos não lineares entre estas variáveis.

Conforme também discutido no Capítulo 3, a preservação dos relacionamentos estatísticos não lineares entre as variáveis aleatórias está diretamente relacionada à preservação da informação mútua entre estas variáveis. No caso da representação de elementos variantes no tempo, este aspecto não é trivial, devido as variáveis aleatórias serem contínuas e a informação mútua ser uma medida que varia com a discretização arbitrada para os dados, por esta ser uma medida que não atende às condições de equitabilidade, como discutido na seção 3.3.2. O coeficiente de informação máxima, também apresentado na seção 3.3.2, introduz uma abordagem que permite tratar a informação mútua de maneira que ela atenda às condições de equitabilidade, tendo como subproduto a determinação da forma como os dados devem ser discretizados, de modo a se maximizar a representação do relacionamento estatístico entre as variáveis aleatórias. Esta maximização é obtida devido à estimação do coeficiente de informação máxima corresponder à formulação de um problema de otimização para determinação da máxima informação mútua que pode ser medida entre as variáveis aleatórias.

Face ao exposto, a hipótese adotada para a construção do modelo proposto é que a função de densidade de probabilidade conjunta que se deseja estimar pode ser decomposta em duas componentes. A primeira componente descreve as variáveis contínuas de maneira independente, por meio de estimativas não paramétricas das funções de densidade de probabilidade marginais das variáveis, e a segunda componente codifica a dependência estatística entre estas variáveis, por meio da preservação da informação mútua entre as variáveis discretizadas, obtidas a partir da determinação da

discretização que maximiza a representação dos relacionamentos não lineares entre as variáveis aleatórias contínuas.

Para sistemas de grande porte, a representação das informações mútuas entre todas as variáveis ainda é um problema que pode estar sujeito à maldição da dimensionalidade. Para lidar com este problema, é proposta a aplicação de uma representação em rede Bayesiana da dependência estatística entre as variáveis discretizadas, de modo a evitar que a complexidade do modelo cresça exponencialmente com o aumento do número de variáveis, conforme discutido na seção 3.4.

5.2 Modelo para Representação de Elementos Variantes no Tempo

Conforme estabelecido na seção anterior, a premissa do modelo proposto neste trabalho é que a função de densidade de probabilidade conjunta que caracteriza os elementos variantes no tempo pode ser decomposta em duas componentes: a representação em um domínio contínuo das variáveis aleatórias, por meio das suas densidades de probabilidade marginais, e a representação em um domínio discreto do relacionamento não linear entre estas variáveis. Embora, a princípio, fosse possível a proposição de um algoritmo para se estimar estas duas componentes diretamente das variáveis aleatórias, é necessário introduzir um processo de transformação de variável neste procedimento. A necessidade desta transformação está relacionada ao objetivo principal da definição desta função de densidade de probabilidade conjunta, que é possibilitar a posterior geração de amostras sintéticas para os elementos variantes no tempo, na fase de amostragem de estados da Simulação Monte Carlo Não Sequencial.

Na seção 3.2 foi apresentada a transformação integral de probabilidade, que possibilita a obtenção de variáveis aleatórias caracterizadas por qualquer função de densidade de probabilidade, a partir de variáveis aleatórias uniformemente distribuídas. Com a aplicação desta transformação, é possível a geração de amostras de qualquer variável aleatória a partir de um algoritmo de geração de amostras para variáveis uniformemente distribuídas. Por este motivo, a representação proposta é construída sobre um conjunto de variáveis uniformemente distribuídas, obtidas a partir da transformação das variáveis aleatórias originais.

A visão geral das etapas para estimação e simulação do modelo proposto está apresentada no diagrama da Figura 35.

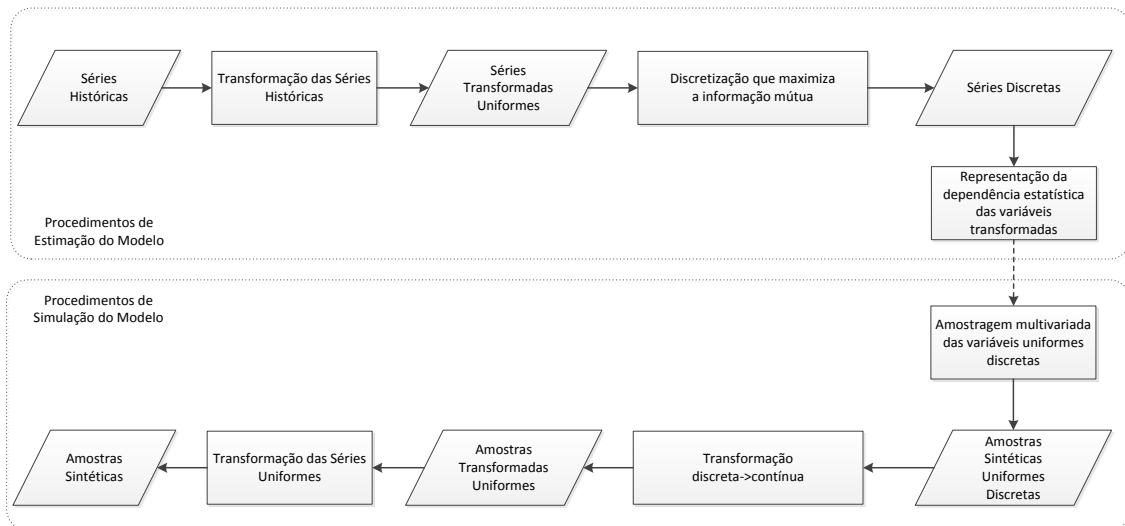


Figura 35– Visão Geral do Modelo

Os passos necessários para a execução destas etapas são apresentados de forma mais bem detalhada a seguir:

1. Transformação das Séries Históricas

A necessidade de se trabalhar com séries transformadas uniformes, visando a posterior aplicação da inversa da transformação integral de probabilidade para a obtenção de amostras sintéticas dos elementos temporais, requer que inicialmente as funções de densidade de probabilidade das variáveis aleatórias sejam estimadas, pois é a partir destas funções que se pode obter a função de probabilidade acumulada e a sua inversa.

Uma vez obtidas as funções de probabilidade acumulada, estas podem ser aplicadas as séries históricas, de modo a produzir as séries transformadas uniformes.

Desta forma, os passos para consolidação desta etapa são:

- a. Estimação não paramétrica das densidades de probabilidade marginais dos elementos variantes no tempo:

O modelo proposto adota uma abordagem não paramétrica, na qual se emprega o método *kernel density estimation* para obtenção de estimativas para estas densidades de probabilidade marginais, conforme apresentado na seção 3.2.2. Aplicando-se este método de estimação às séries históricas de vento, individualmente, obtém-se para cada variável aleatória v_j uma estimativa da densidade de probabilidade $\hat{f}_x(v_j)$.

Por definição, a variável aleatória que descreve o processo estocástico da velocidade do vento não pode ser negativa e por este motivo, de acordo com o discutido

na seção 3.2.2, a sua densidade de probabilidade deve possuir um suporte definido pelo intervalo $[0, \infty)$.

Na literatura, para se definir um suporte neste intervalo é comum a adoção da função de transformação logarítmica[43]. Neste trabalho é proposta a utilização de uma função de transformação que garante o atendimento a esta restrição de domínio, mas sem distorcer tanto a densidade de probabilidade original. Esta proposta é apresentada de forma detalhada no Apêndice I .

Na Figura 36 são apresentadas as funções de densidade de probabilidade estimadas para uma série de vento, considerando a representação de suporte, representada pela linha cheia, e sem considerar a representação de suporte, representada pela linha pontilhada.

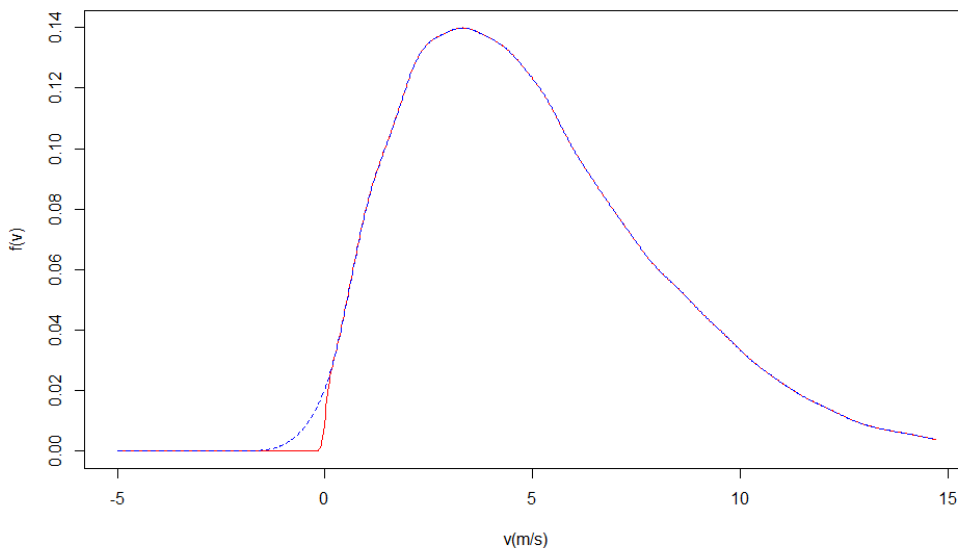


Figura 36– Função de Densidade de Probabilidade Estimada com Suporte em $[0, \infty)$

- b. Obtenção das funções de probabilidade acumulada das séries e as suas respectivas inversas:

A função de probabilidade acumulada de uma variável aleatória contínua correspondente a uma série temporal de vento, $F_x(v_j)$, é definida pela integral da função de densidade de probabilidade, conforme apresentado no Capítulo 3.

Como a função $F_x(v_j)$ é monotonicamente crescente e, por consequência, biunívoca, pode-se afirmar ainda que para cada variável aleatória existe uma função inversa da função de probabilidade acumulada, $F_x^{-1}(v_j)$.

Para efeito de ilustração, para uma mesma série de vento, são apresentadas as funções de probabilidade acumulada, na Figura 37, e a respectiva inversa, na Figura 38.

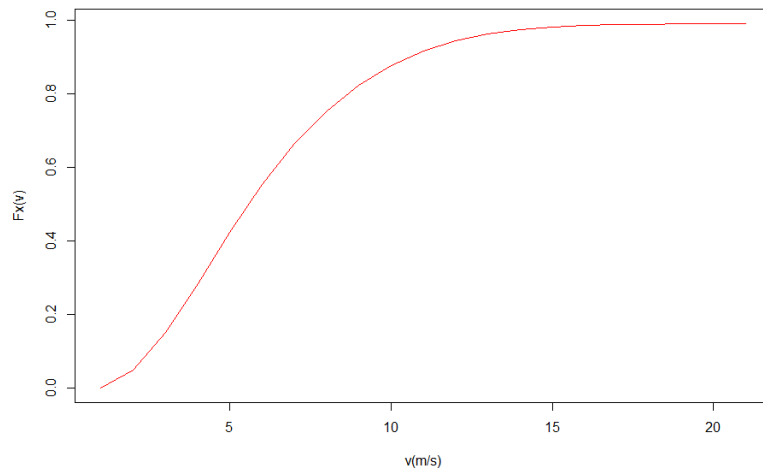


Figura 37– Função de Probabilidade Acumulada Estimada

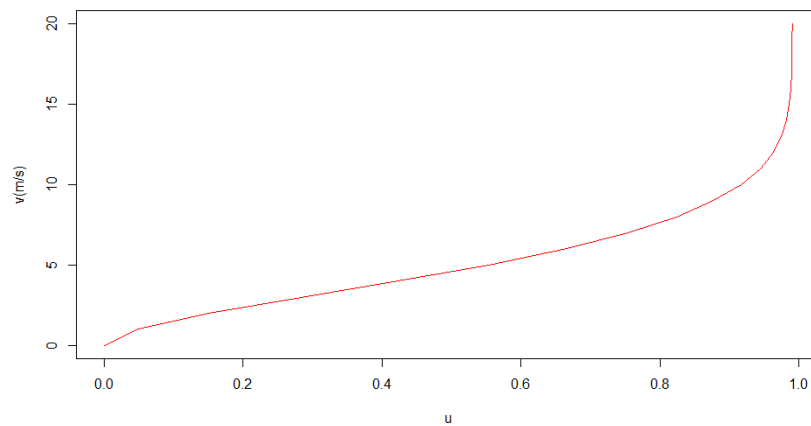


Figura 38– Função Inversa da Probabilidade Acumulada Estimada

- c. Transformação das séries de vento em novas séries temporais uniformemente distribuídas:

Dada a determinação da função de probabilidade acumulada, pode se obter um novo conjunto de séries temporais, u_1, u_2, \dots, u_w , cujas densidades de probabilidade são uniformes pela aplicação da transformação integral de probabilidade às séries históricas.

Esta transformação é dada pela equação (5.5).

$$\begin{cases} u_1 = \hat{F}_{x_1}(v_1) \\ u_2 = \hat{F}_{x_2}(v_2) \\ \vdots \\ u_w = \hat{F}_{x_w}(v_w) \end{cases} \quad (5.5)$$

Se as funções $\hat{F}_{x_j}(v_j)$ são boas estimativas para as distribuições de probabilidade acumulada das variáveis aleatórias associadas às séries históricas, então as séries transformadas u_j terão distribuição de probabilidade aproximadamente uniforme, com intervalos definidos em $[0,1]$, devido a esta ser a principal propriedade da transformação integral de probabilidade, conforme discutido na seção 3.2.

Para uma melhor ilustração desta propriedade, na Figura 39 é apresentada uma estimativa de função de densidade de probabilidade, obtida pela aplicação do próprio método *kernel density estimation* à variável aleatória transformada. Observa-se que mesmo que o *kernel* gaussiano não seja o mais adequado para a estimativa deste tipo de função, ainda assim é possível visualizar o comportamento uniforme da função na maior parte do domínio.

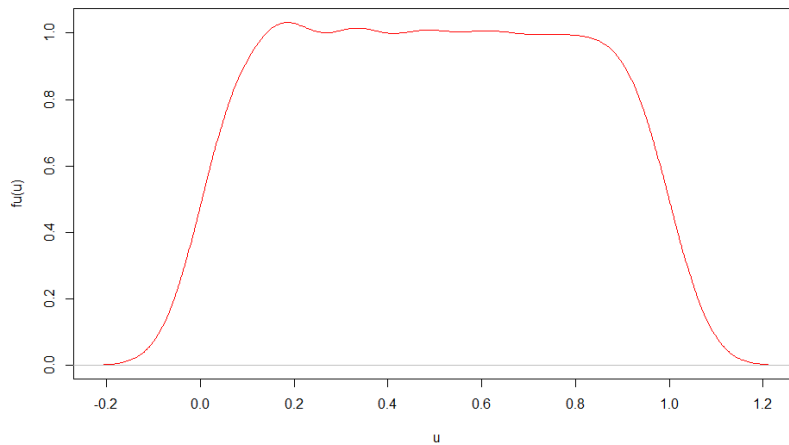


Figura 39– Função de densidade de probabilidade da variável transformada uniforme

2. Discretização que maximiza a informação mútua.

A representação da dependência estatística entre os elementos temporais requer que os relacionamentos não lineares entre estas variáveis aleatórias sejam capturados pelo modelo. A proposta é que a representação desta dependência estatística seja realizada a partir da preservação da informação mútua entre as variáveis. No entanto, para que esta abordagem seja possível, é necessário que as variáveis sejam representadas de forma discreta.

Para a obtenção desta discretização, é proposta a utilização da medida de dependência estatística dada pelo coeficiente de informação máxima das variáveis aleatórias, uma vez que o problema da determinação do coeficiente de informação máxima produz como resultado, entre outras indicações, o modo como as variáveis devem ser discretizadas para se maximizar a informação mútua entre elas.

Desta maneira, os passos para obtenção desta discretização são:

- a. Determinação dos *bins* a serem utilizados para a discretização das séries:

Determina-se o coeficiente de informação máxima, apresentado na seção 3.3.2, de modo a se determinar, como um subproduto do processo de obtenção do *MIC*, os clusters C , que devem ser utilizados.

- b. Obtenção das séries discretas:

Definido o número de clusters, C , as séries discretas, L_j , são construídas de modo que, para todo $j \in W$, para um instante qualquer, t :

$$\begin{cases} L_j = l_1, \text{ se } 0 \leq u_j < \frac{1}{C} \\ L_j = l_2, \text{ se } \frac{1}{C} \leq u_j < \frac{2}{C} \\ \vdots \\ L_j = l_C, \text{ se } \frac{C-1}{C} \leq u_j \leq 1 \end{cases} \quad (5.6)$$

Conforme discutido na seção 3.3.2, se a discretização de u_j é adequada, tem-se que:

$$I(L_i, L_j) \approx I(u_i, u_j) \forall i, j \in [1, W] \quad (5.7)$$

3. Representação da dependência estatística das variáveis transformadas

Por fim, o mapeamento da dependência estatística das variáveis é realizado a partir da estimação da rede Bayesiana, que codifica as probabilidades condicionais, a partir de uma topologia de grafo, impedindo que a complexidade do modelo cresça exponencialmente, a medida que se aumenta o número de variáveis representadas.

Os passos para a estimação da rede Bayesiana são:

a. Estimaco da estrutura da rede Bayesiana;

A partir do *dataset* de observaces simultneas $[\bar{L}_1 \ \bar{L}_2 \ \dots \ \bar{L}_w]$, obtm-se uma estimativa para a estrutura do grafo da rede Bayesiana, $G = \langle N, E \rangle$, constituda por um conjunto de ns (N) e arestas (E), aplicando-se, por exemplo, um mtodo de estimaco baseado em pontuaco, *score-based*, associado a um algoritmo de busca local, *hill climbing*, com a pontuaco definida pelo critrio *BIC* (*Bayesian Information Criterion*), tendo sido estes tpicos desenvolvidos na seo 3.4.

b. Estimaco dos parmetros da rede Bayesiana;

Finalmente, a partir da estrutura de grafo, $G = \langle N, E \rangle$, obtida no passo anterior, e do *dataset*, $[\bar{L}_1 \ \bar{L}_2 \ \dots \ \bar{L}_w]$, os parmetros da rede Bayesiana, Θ , podem ser estimados a partir da maximizaco da verossimilhana dos parmetros da rede em relao aos dados, mtodo *MLE*.

Ao final destes passos, a funo de densidade de probabilidade conjunta que representa os elementos variantes no tempo  caracterizada pelo conjunto de parmetros $T = \langle \hat{F}_x^{-1} \ G \ \Theta \rangle$.

5.3 Amostragem dos Estados do Modelo na Simulao Estocstica

Uma vez obtido o modelo, o algoritmo base da Simulao Monte Carlo No Sequencial pode ser modificado para incorporar a amostragem dos elementos variantes no tempo, conforme os passos descritos a seguir:

1. Amostragem multivariada das variveis uniformes discretas;

Uma amostra, i , do conjunto de variveis aleatrias discretas descritas pela rede Bayesiana, $L^{Si} = [L_1^{Si} \ L_2^{Si} \ \dots \ L_w^{Si}]$, onde o ndice S  utilizado para diferenciar as variveis amostradas de suas respectivas observaces histricas,  obtida percorrendo-se os ns do grafo da rede Bayesiana, G , recursivamente, no sentido direto dos ns pais para os ns filhos.

Para um n L_j , que no possui ns pais, $pa(L_j) = 0$, o estado ser amostrado a partir da distribuo de probabilidade acumulada associada ao n, $P(L_j)$, codificada no parmetro Θ da rede Bayesiana.

Para um n L_j , que possui ns pais, $pa(L_j) \neq 0$, o estado ser amostrado a partir da distribuo de probabilidade acumulada associada ao n condicionada aos estados

amostrados para os nós pais, $P(L_j|pa(L_j))$, também codificada no parâmetro Θ da rede Bayesiana.

Se a rede Bayesiana ajustada for suficientemente precisa, de modo que $P(L_1, L_2, \dots, L_w) \approx \prod_{j=1}^n P(L_j|pa(L_j))$, ter-se-á:

$$I(L_i, L_j) \approx I(L_i^S, L_j^S) \forall i, j \in [1, W] \quad (5.8)$$

2. Transformação discreta->contínua;

A partir do conjunto de estados obtidos no passo 1, $L^{Si} = [L_1^{Si} \ L_2^{Si} \ \dots \ L_w^{Si}]$, obtém-se a i -ésima amostra das variáveis transformadas u_j^{Si} , a partir do sorteio de um valor real qualquer no intervalo correspondente ao estado discreto obtido, a partir da seguinte definição:

$$\begin{cases} u_j^{Si} = \text{unif}\left(0, \frac{1}{C}\right), \text{ se } L_j^{Si} = l_1 \\ u_j^{Si} = \text{unif}\left(\frac{1}{C}, \frac{2}{C}\right), \text{ se } L_j^{Si} = l_2 \\ \vdots \\ u_j^{Si} = \text{unif}\left(\frac{C-1}{C}, 1\right), \text{ se } L_j^{Si} = l_C \end{cases} \quad (5.9)$$

Onde, $\text{unif}(a, b)$ representa uma função de geração de número aleatório uniformemente distribuído no intervalo compreendido entre a e b .

Para um número suficientemente grande de amostras, para cada elemento variante no tempo, como os estados discretos l_1, l_2, \dots, l_C são equiprováveis, a densidade de probabilidade das variáveis transformadas amostradas, u_j^S se aproximará de uma densidade uniforme definida no intervalo (0,1).

Sendo ainda que:

$$I(L_i^S, L_j^S) \approx I(u_i^S, u_j^S) \forall i, j \in [1, W] \quad (5.10)$$

Uma vez que a função $\text{unif}(\cdot)$, utilizada em (5.10), é aplicada de maneira independente em cada variável L_j^S , não introduzindo, ou removendo, dependência estatística na transformação $L^S \rightarrow u^S$.

3. Transformação das séries uniformes;

Por fim, pode-se transformar as amostras das variáveis aleatórias uniformes de modo a se obter as amostras das variáveis aleatórias que representam os elementos

variantes no tempo a partir da aplicação da inversa da transformação integral de probabilidade.

Aplicando-se a função \hat{F}_{xj}^{-1} à variável u_j^{Si} , tem-se que:

$$\begin{cases} v_1^{Si} = \hat{F}_{x1}^{-1}(u_1^{Si}) \\ v_2^{Si} = \hat{F}_{x2}^{-1}(u_2^{Si}) \\ \vdots \\ v_w^{Si} = \hat{F}_{xw}^{-1}(u_w^{Si}) \end{cases} \quad (5.11)$$

Se \hat{F}_{xj}^{-1} é uma boa estimativa para a inversa da função de probabilidade acumulada de v_j , como u_j^{Si} se aproxima de uma densidade de probabilidade uniforme, a densidade de probabilidade de v_j^{Si} se aproximará da função de densidade de probabilidade estimada para o histórico:

$$\hat{f}_x(v_j^S) \approx \hat{f}_x(v_j) \quad (5.12)$$

Conforme discutido na seção 2.1, a informação mútua entre variáveis aleatórias é preservada sob qualquer transformação inversível, desta forma, duas novas relações podem ser introduzidas:

$$\begin{cases} I(v_i, v_j) = I(F_x(v_i, v_j)) = I(u_i, u_j) \\ I(u_i^S, u_j^S) = I(F_x^{-1}(u_i^S, u_j^S)) = I(v_i^S, v_j^S) \end{cases} \quad (5.13)$$

Combinando-se (5.8), (5.10) e (5.13), tem-se finalmente que:

$$I(v_i, v_j) \approx I(v_i^S, v_j^S) \forall i, j \in [1, W] \quad (5.14)$$

Desta forma, os principais objetivos do modelo proposto são alcançados: gerar amostras para os elementos variantes no tempo, preservando a densidade de probabilidade não paramétrica estimada, equação (5.12), e a informação mútua das variáveis, equação (5.14).

Para possibilitar uma visão geral, os procedimentos relacionados à estimação dos parâmetros e geração de amostras do modelo proposto são apresentados no fluxograma da Figura 40.

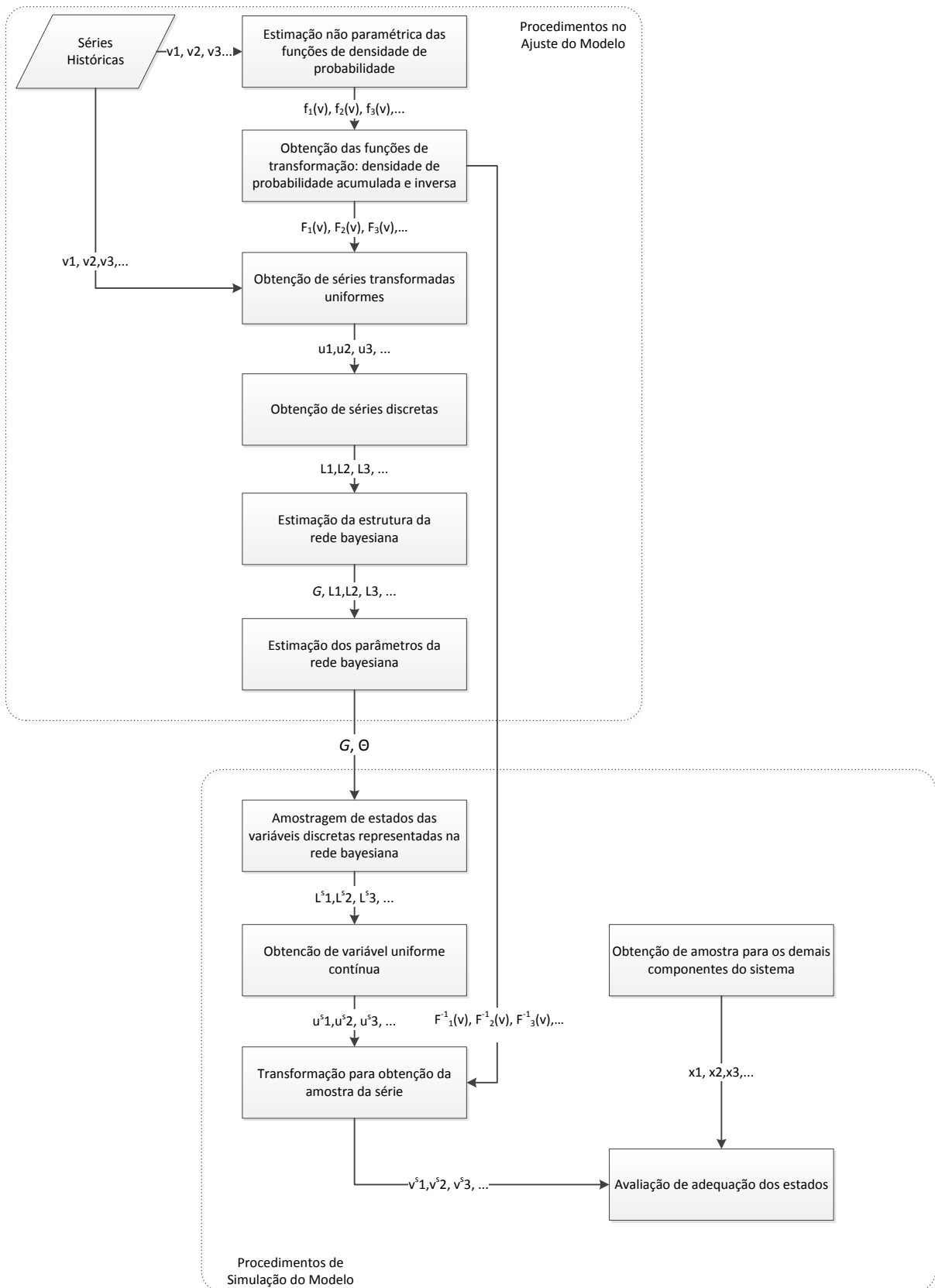


Figura 40– Processo de Estimação e Geração de Amostras a partir do Modelo Proposto

5.4 Representação da Incerteza na Direção do Vento

A importância da representação da incerteza da direção do vento está principalmente relacionada à modelagem das perdas de velocidade das turbinas, que se traduz em perda da energia total gerada pela usina eólica, devido ao efeito esteira. A rigor, a inclusão destas variáveis no modelo estocástico duplica o número de variáveis e faz com que a função de densidade de probabilidade conjunta fique muito mais complexa.

Uma das principais vantagens da representação de dependência estatística por meio de redes Bayesianas é que elas naturalmente evitam a explosão da dimensão do modelo com o aumento do número de variáveis. Desta maneira, a inclusão da representação da incerteza da direção do vento no modelo proposto é relativamente simples, sendo necessária apenas a adição de novos nós ao grafo da rede, que estarão associados a estas novas variáveis representadas.

O *dataset* de observações históricas discretas, $[\overline{L}_1 \ \overline{L}_2 \ \dots \ \overline{L}_w]$, obtido após a etapa de discretização das séries transformadas uniformes, pode ser expandido para considerar as novas observações discretas de direção do vento:

$$D' = [\overline{L}_1 \ \overline{L}_2 \ \dots \ \overline{L}_w \ \overline{D}_1 \ \overline{D}_2 \ \dots \ \overline{D}_w]$$

As estimações da estrutura e dos parâmetros da rede Bayesiana devem seguir o mesmo procedimento descrito para o caso em que se representa apenas os módulos da velocidade do vento. Isto é, a estrutura será estimada novamente por um método *score-based*, com aplicação do algoritmo de busca *hill climbing*, e os parâmetros serão estimados a partir da maximização da verossimilhança dos parâmetros, método *MLE*.

5.5 Representação de Variáveis Exógenas

A representação de informações climáticas exógenas em simulações estocásticas tem recebido destaque na área de modelagem de séries temporais, visando fazer com que a inclusão deste tipo de informação possibilite a obtenção de cenários sintéticos mais aderentes ao que seria efetivamente observado para o horizonte futuro de estudo. Conforme exposto anteriormente, um exemplo deste tipo de abordagem é a consideração de índices ENSO[11] na construção de modelos estocásticos para representação das incertezas nas vazões afluentes e da incerteza do vento.

O modelo proposto neste trabalho é flexível para a incorporação deste tipo de informação, requerendo apenas que as variáveis exógenas sejam adicionadas ao *dataset* da rede Bayesiana, e que sejam observadas restrições de causalidade, de modo a evitar

que o processo de estimação de estrutura da rede Bayesiana convirja para soluções em que as variáveis exógenas sejam representadas por nós descendentes das variáveis aleatórias que representam os elementos variantes no tempo, ou que, o caso de haver mais de uma variável exógena representada, que se crie arcos de dependência entre elas.

1. Restrições de causa/efeito

Estas restrições estabelecem que os nós correspondentes às variáveis exógenas não podem descender das variáveis que representam os elementos variantes no tempo.

2. Restrições de independência das variáveis exógenas

Estas restrições estabelecem que as variáveis exógenas não podem depender explicitamente de outras variáveis exógenas.

Estas duas restrições são adicionadas no problema de inferência da rede a partir da definição do seguinte conjunto *blacklist*, que limita o espaço de busca do problema de otimização do grafo:

$$blacklist = \begin{cases} V_i \rightarrow ex_j & \forall i \in 1, 2, \dots, w, j \in 1, 2, \dots, E \\ ex_j \rightarrow ex_k & \forall j, k \in 1, 2, \dots, E \end{cases} \quad (5.15)$$

Onde ex_j representa a j -ésima variável exógena e E é o conjunto de todas as variáveis exógenas representadas.

Na Figura 41 é apresentado o fluxograma de estimação dos parâmetros e geração de amostras do modelo proposto, anteriormente apresentado na Figura 40, mas incorporando agora à representação das variáveis exógenas.

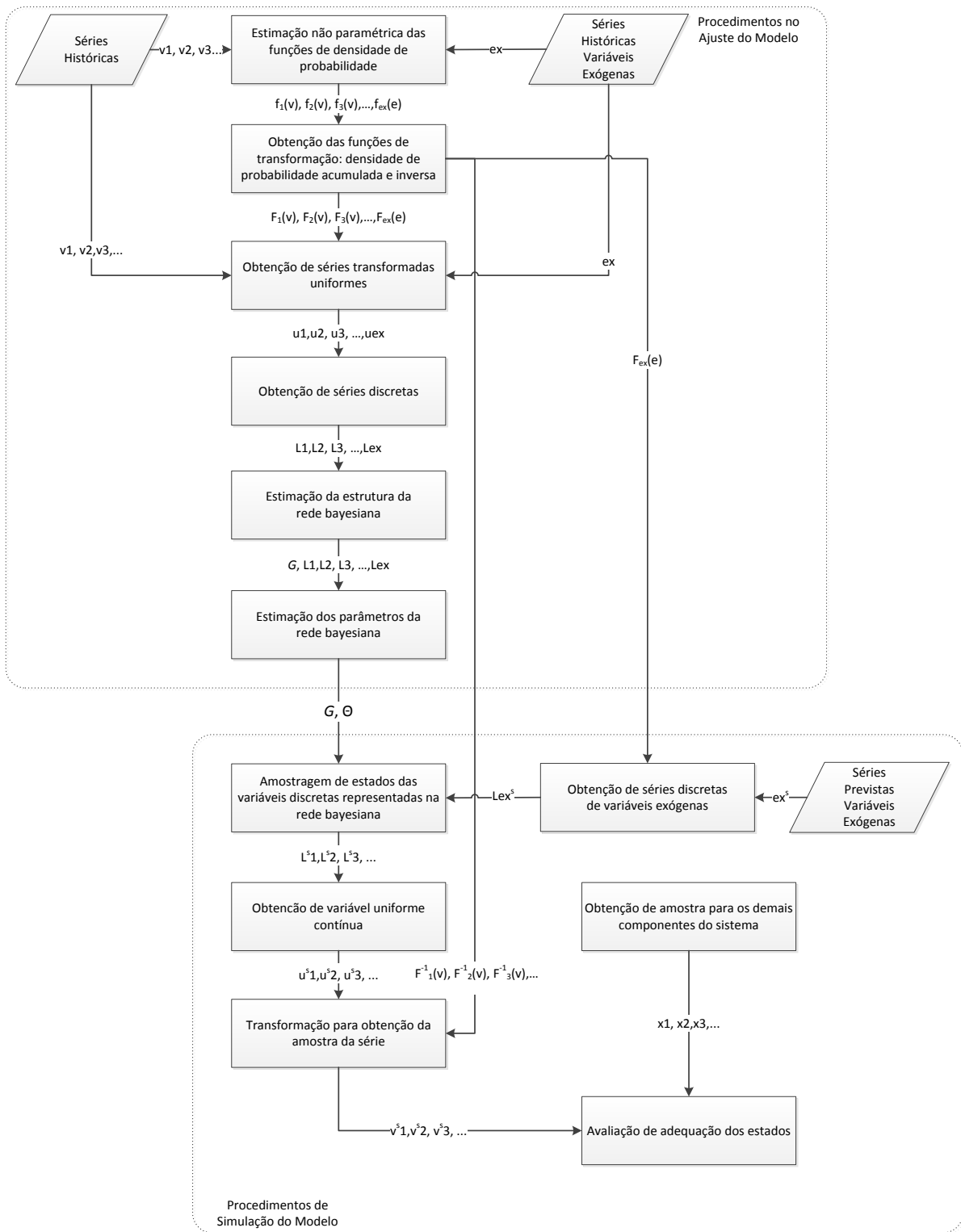


Figura 41– Incorporação da Representação de Variáveis Exógenas a Visão Geral do Processo

É importante frisar que este modelo de representação de variáveis exógenas não é restrito para a representação de variáveis climáticas, sendo adaptável para a representação, por exemplo, da dependência estatística entre séries de vento e séries de vazões afluentes.

Nos estudos de planejamento e operação dos sistemas de potência é usual o emprego de modelos auto-regressivos periódicos cuja discretização é semanal ou mensal, dada a dinâmica lenta das vazões e dos reservatórios das usinas[3]. Trabalhar com níveis de discretizações menores neste caso, pode levar a um aumento considerável da relação sinal/ruído das séries temporais, distorcendo a qualidade dos cenários sintéticos gerados. Esta grande diferença entre o nível de discretização da geração eólica e das vazões afluentes torna mais complicada a geração de cenários que preservem a dependência estatística entre estas variáveis.

Uma alternativa para construir um modelo estocástico que preserve esta dependência é através da utilização desta abordagem de representação de variáveis exógenas no modelo proposto. Desta forma, as vazões afluentes podem ser representadas de forma convencional, através de um modelo auto-regressivo periódico, enquanto que a geração eólica é representada através do modelo proposto, mas empregando as observações das vazões afluentes, semanais ou mensais, como um *dataset* adicional que condiciona a geração eólica.

Na fase de geração das amostras, as séries sintéticas de vazões afluentes são geradas a partir de um processo de amostragem externo, referente ao modelo auto-regressivo periódico adotado, e alimentam o modelo proposto, influenciando, desta forma, as amostras obtidas para a geração eólica.

Este procedimento é apresentado na Figura 42.

Ressalta-se que, por conveniência do estudo, pode-se estimar o modelo proposto de maneira periódica. Isto é, para estudos cuja representação de vazões sejam mensais, pode-se trabalhar com doze modelos temporais independentes, onde cada um representa a dinâmica de um mês do ano e é estimado considerando-se apenas os *datasets* de observações daquele respectivo mês.

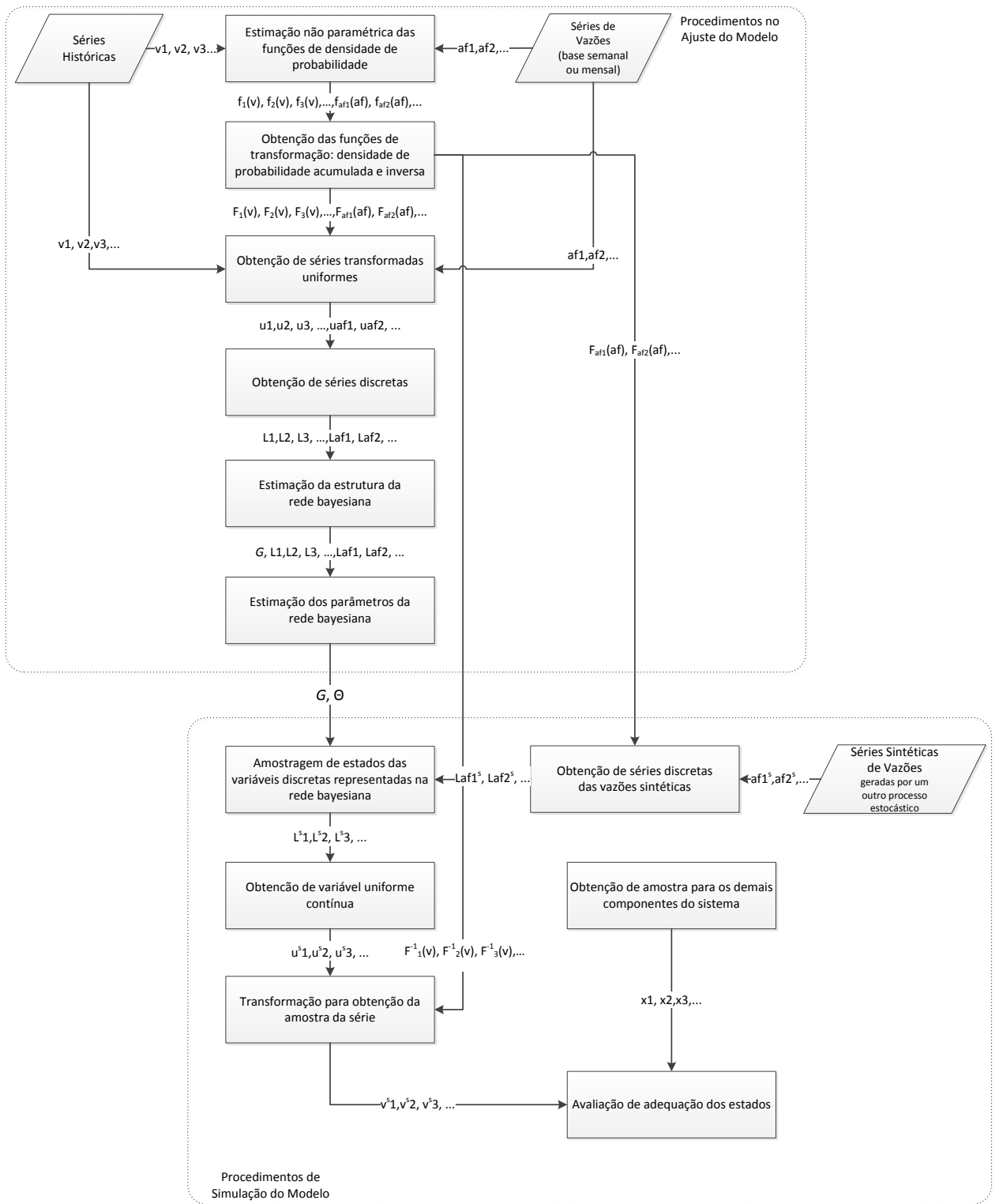


Figura 42– Incorporação da Representação de Dependência Estatística com Vazões Afluentes

Uma forma alternativa de representação da dependência estatística entre as vazões afluentes e a geração eólica, e que pode ser mais conveniente, dependendo do sistema e dos dados disponíveis, é a utilização das séries de energias afluentes, em detrimento de se utilizar diretamente as séries de vazões afluentes.

A energia afluente de um sistema é definida como a energia teórica que é possível de se obter das usinas hidrelétricas em um período, sendo calculada a partir das vazões afluentes deste período, o fator de produção das usinas, e a topologia hidrológica do sistema.

Esta formulação alternativa é apresentada na Figura 43.

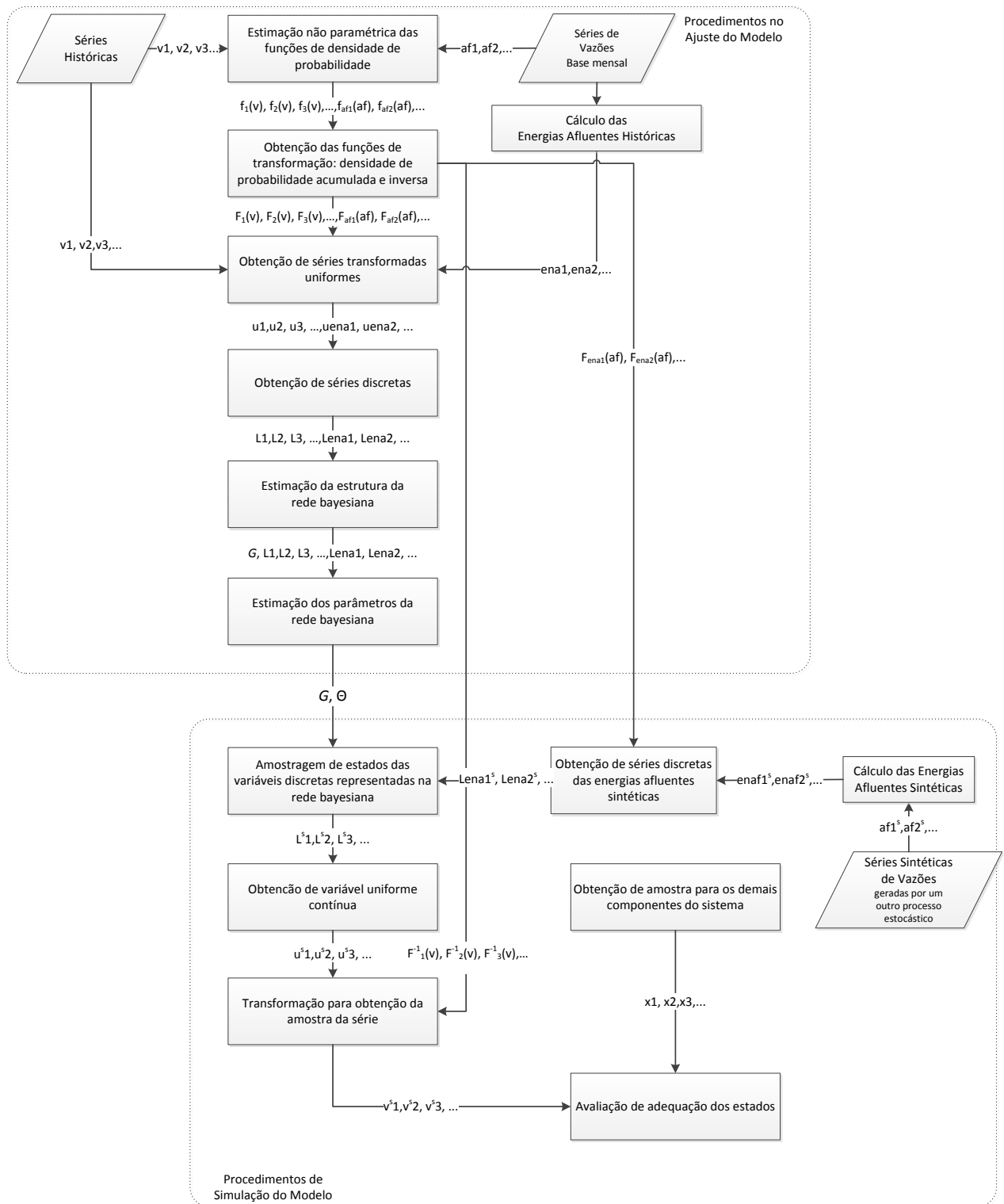


Figura 43– Formulação Alternativa – Representação da Dependência Estatística com as Energias Afluentes

5.6 Estimação de Índices de Frequência e Duração

5.6.1 Modelo de Transição de Estados

Conforme apresentado na seção anterior, o modelo para representação de elementos variantes no tempo através de uma função de densidade de probabilidade conjunta, definida pela combinação das estimativas não paramétricas das densidades de probabilidades marginais das variáveis com uma estrutura de rede Bayesiana discreta, é o suficiente para possibilitar a estimação de índices de confiabilidade baseados em probabilidade e energia. No entanto, para a estimação de índices de confiabilidade baseados em frequência e duração esta representação não é suficiente, uma vez que estes índices dependem da representação do acoplamento temporal de transição dos elementos variantes no tempo.

Conforme anteriormente discutido, uma maneira de representar o processo estocástico que descreve as transições de uma variável ao longo do tempo é através do emprego de cadeias de Markov. Se os elementos variantes no tempo fossem independentes, seria possível construir modelos independentes de transição para cada uma das variáveis, já tendo sido bastante discutida na literatura a estimação destes modelos independentes. Em [52], por exemplo, é apresentado um modelo de transição de estados para a representação de geração eólica. Na prática, assim como existe uma forte correlação espacial entre os elementos variantes no tempo, também é esperado que haja algum acoplamento nas dinâmicas de transição destas variáveis.

Considerando-se duas variáveis aleatórias discretas, que representam dois elementos variantes no tempo no instante t , $V_1^t = V_{1,1}$ e $V_2^t = V_{2,1}$, e que transitam para outros estados no instante posterior, $t+1$, $V_1^{t+1} = V_{1,k}$ e $V_2^{t+1} = V_{2,l}$, se existe dependência estatística nas transições destas variáveis, pela regra da cadeia pode-se dizer que:

$$P(V_1^{t+1} = V_{1,k}, V_2^{t+1} = V_{2,l} | V_1^t = V_{1,1}, V_2^t = V_{2,1}) = P(V_2^{t+1} = V_{2,l} | V_1^t = V_{1,1}, V_2^t = V_{2,1}, V_1^{t+1} = V_{1,k}) \times P(V_1^{t+1} = V_{1,k} | V_1^t = V_{1,1}) \quad (5.16)$$

O termo que define a probabilidade condicional de $V_2^{t+1} = V_{2,l}$ pode ser interpretado como a probabilidade de V_2 transitar do primeiro estado para o estado l , dado que V_1 irá transitar do primeiro estado para o estado k .

Analogamente ao problema da maldição da dimensionalidade na representação das probabilidades estacionárias, a complexidade desta definição tende a crescer exponencialmente à medida que se aumenta o número de elementos variantes no tempo

representados. Porém, também de maneira análoga, é possível a adoção de um modelo baseado em rede Bayesiana para representar estas transições de estados destas variáveis, evitando-se o crescimento exponencial dos parâmetros do modelo.

Neste caso, os nós da rede Bayesiana devem representar os estados dos elementos variantes no tempo no instante atual e no instante imediatamente posterior, conforme o grafo conceitual apresentado na Figura 44, no qual é ilustrada a possibilidade de dependência dos estados das variáveis no instante em $t+1$ com os respectivos estados em t , e dos estados das demais variáveis também em $t+1$.

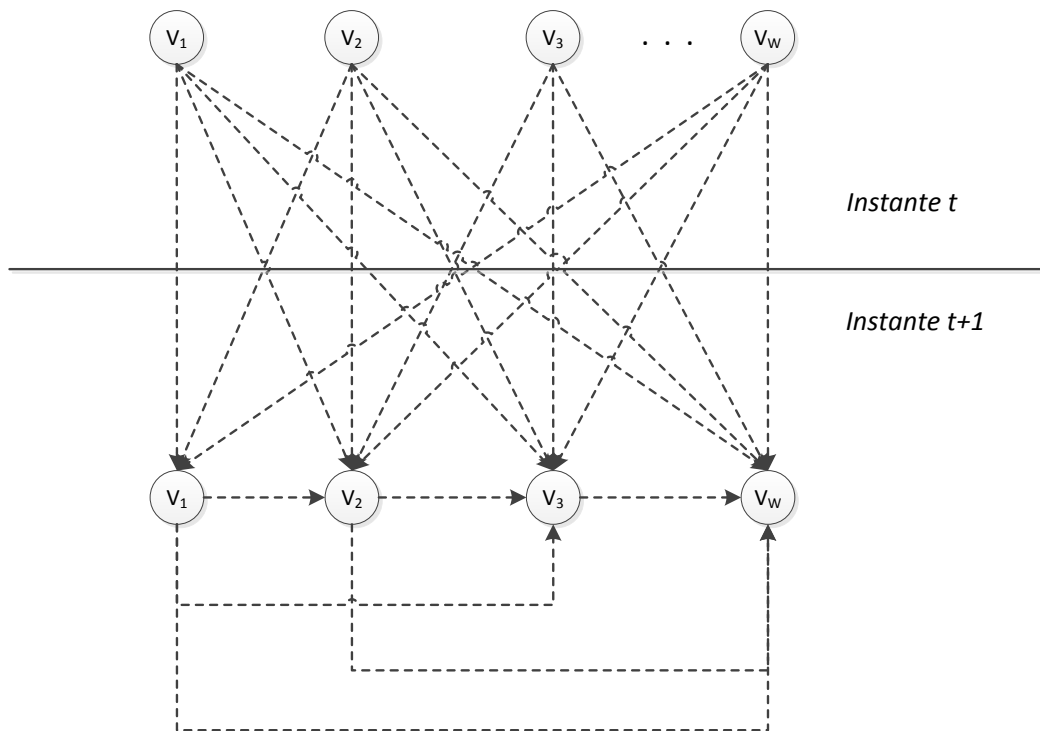


Figura 44– Rede Bayesiana Conceitual para a Transição de Estados

É importante ressaltar que nesta rede não existe necessidade de se explicitar a dependência entre as variáveis em t , pois os estados das variáveis neste instante já são conhecidos no momento da amostragem de uma transição, conforme será discutido nos algoritmos de amostragem de transição utilizados nos cálculos dos índices de frequência e duração.

Desta maneira, para a estimação da rede Bayesiana que descreve as transições dos estados dos elementos variantes no tempo, define-se um conjunto de dados, D_{trans} , onde cada registro corresponde à concatenação das observações históricas dos estados

discretos dos elementos variantes no tempo no instante t e no instante imediatamente posterior a t .

$$D_{trans} = \begin{bmatrix} V_1^1 & V_2^1 & \dots & V_w^1 & V_1^2 & \dots & V_w^2 \\ V_1^2 & V_2^2 & \dots & V_w^2 & V_1^3 & \dots & V_w^3 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ V_1^{T-1} & V_2^{T-1} & \dots & V_w^{T-1} & V_1^T & \dots & V_w^T \end{bmatrix} \quad (5.17)$$

A representação proposta exige que algumas restrições de topologia sejam adicionadas ao problema de estimação da topologia desta rede Bayesiana:

1. Restrições de cronologia (causa/efeito)

Estas restrições estabelecem que os nós correspondentes ao instante t , não podem depender dos nós correspondentes ao instante posterior, $t+1$.

Esta restrição é necessária para que o processo de ajuste da rede não produza inversões na relação de causa/efeito entre os nós, convergindo para soluções na qual a amostragem do estado de alguma variável em um instante t dependa do prévio conhecimento do estado de alguma variável no instante posterior, sendo a mesma ou qualquer outra variável. Ao contrário da representação das probabilidades estacionárias, na qual a relação causa/efeito não era relevante, neste caso a preservação desta relação é fundamental.

2. Restrições de independência das variáveis no instante atual

Estas restrições estabelecem que as variáveis no instante atual não são condicionadas a nenhuma outra variável da rede Bayesiana.

Na realidade, como já mencionado, no processo de transição de estados o estado atual é sempre conhecido, não sendo nunca obtido por esta rede de representação das transições. Os nós correspondentes a estas variáveis são incluídos na estrutura da rede Bayesiana por uma questão de atendimento aos princípios e aos algoritmos de estimação da rede.

Estas duas restrições são adicionadas no problema de estimação da rede a partir da definição do seguinte conjunto *blacklist*, que limita o espaço de busca do problema de otimização do grafo:

$$blacklist = \begin{cases} V_i^{t+1} \rightarrow V_j^t & \forall i, j \in 1, 2, \dots, w \\ V_i^t \rightarrow V_j^t & \forall i, j \in 1, 2, \dots, w \end{cases} \quad (5.18)$$

5.6.2 Aplicação do Modelo para o Cálculo dos Índices de Frequência e Duração

No Capítulo 2 foi apresentado o algoritmo conceitual para a Simulação Monte Carlo Pseudo-Sequencial, que é uma alternativa eficiente para o cálculo dos índices de frequência e duração na avaliação da confiabilidade. A partir do modelo definido para a representação das transições dos estados é possível propor um algoritmo de estimação dos índices de frequência e duração, que assim como o algoritmo da Simulação Monte Carlo Pseudo-Sequencial, também baseados em simulações *forward* e *backward* a partir do ponto de falha.

Uma diferença conceitual do algoritmo proposto em relação a Simulação Monte Carlo Pseudo-Sequencial é que esta exige que as séries sintéticas de operação do sistema sejam previamente geradas, enquanto que na proposição apresentada não existe esta necessidade. Para que isto seja possível, no entanto, é necessária a construção de um segundo modelo de transição que descreva a dinâmica inversa das transições das variáveis. Este modelo é aplicado na fase *backward* das análises das subsequências de falhas.

Dessa forma, o procedimento para aplicação do modelo para o cálculo dos índices de frequência e duração podem ser resumidos como:

- a) obtenha uma amostra para os estados dos componentes convencionais;
- b) obtenha uma amostra para os estados dos elementos variantes no tempo a partir do modelo proposto
- c) avalie o estado x . Se este estado é de falha, prossiga para o passo (d); caso contrário, retorne ao passo (a);
- d) obtenha uma subsequência de falha (interrupção de energia) através das simulações *forward* e *backward*:
 - execute uma simulação cronológica no sentido direto do tempo começando do estado selecionado até que seja encontrado um estado de sucesso:
 - I. os estados subsequentes dos componentes convencionais são amostrados a partir das suas respectivas taxas de transição;
 - II. os estados subsequentes dos elementos variantes no tempo são amostrados a partir do modelo de transição apresentado.
 - execute uma simulação cronológica no sentido inverso do tempo começando do estado selecionado até que seja encontrado um estado de sucesso

- I. os estados reversos dos componentes convencionais são amostrados a partir das suas taxas de transição *backward* (μ_{back} e λ_{back})
 - II. os estados reversos dos elementos variantes no tempo são amostrados a partir do modelo de transição *backward*
- e) calcule as estimativas dos índices de confiabilidade. Se a precisão destas estimativas for aceitável, pare; caso contrário, retorne ao passo (a).

De maneira análoga a Simulação Pseudo-Sequencial, a função teste do índice de frequência e duração é definida pela equação (2.31).

Na Figura 45 são apresentados novamente os procedimentos relacionados à estimação dos parâmetros e geração de amostras do modelo proposto, mas incorporando também os modelos de transição de estados.

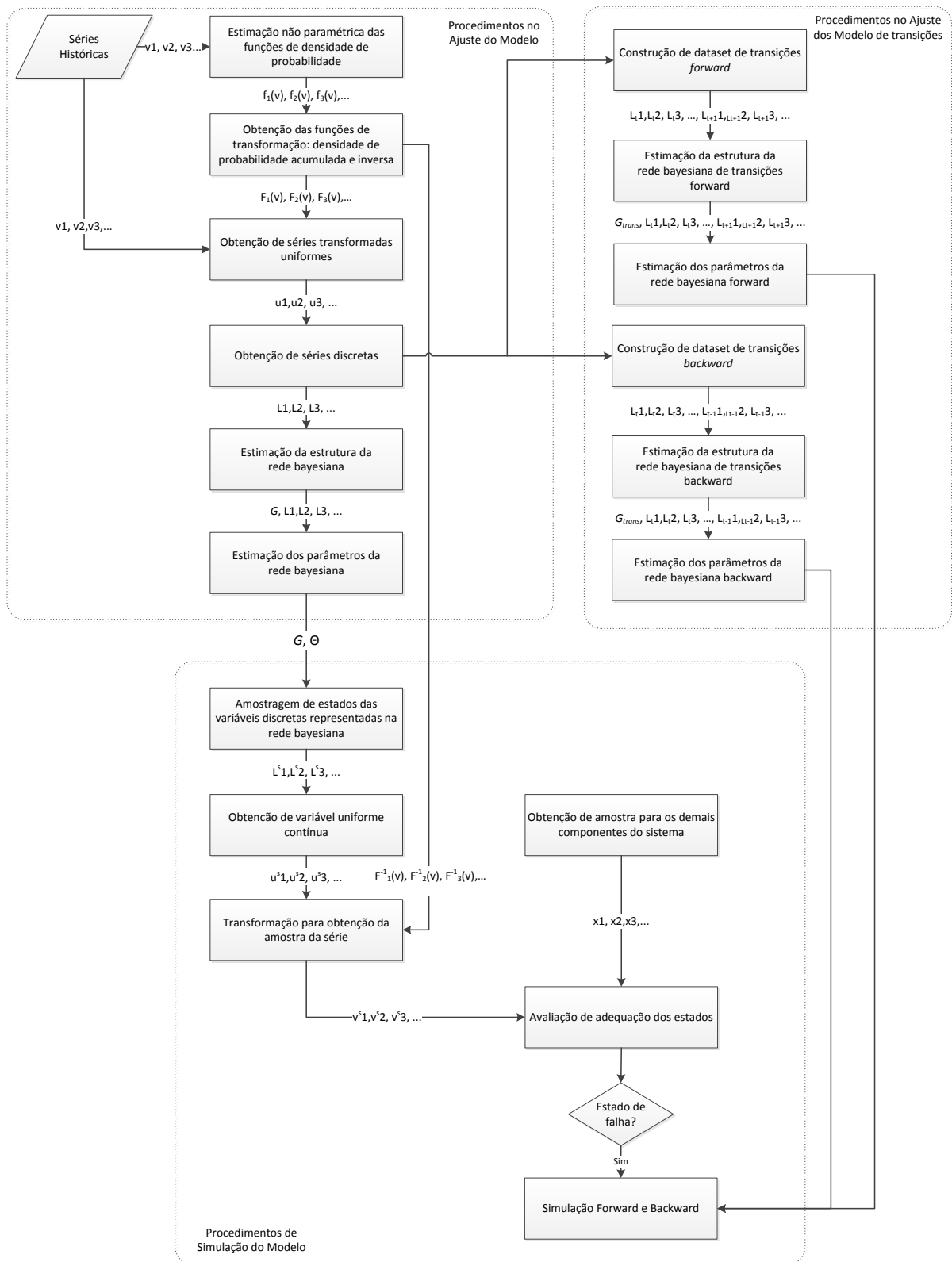


Figura 45– Incorporação do Modelo de Transição de Estados a Visão Geral do Processo

Capítulo 6

Resultados

6.1 Introdução

Neste capítulo serão apresentados os resultados obtidos pelo modelo desenvolvido neste trabalho, tendo sido avaliados quatro casos distintos.

Os primeiros três casos apresentados são baseados no sistema IEEE-RTS[62], com a introdução de algumas modificações, referentes à inclusão de vinte e sete usinas eólicas e seus cenários próprios de velocidade do vento.

No primeiro caso apenas a representação do módulo da velocidade do vento foi considerada, sendo desconsiderada a representação da direção do vento. O modelo proposto foi estimado a partir dos dados históricos da velocidade de vento e os resultados da simulação foram comparados com os resultados obtidos a partir das simulações com os dados históricos, utilizando modelos tradicionais da literatura.

O segundo caso é bastante similar ao primeiro, no entanto, neste caso se representou não apenas a incerteza dos módulos da velocidade dos ventos, mas também a incerteza nas direções do vento. Mais uma vez os resultados da simulação foram comparados com os resultados obtidos a partir das simulações com os dados históricos.

O objetivo dos dois primeiros casos foi comprovar a eficácia e eficiência do modelo desenvolvido para representar a dinâmica dos elementos variantes no tempo, tendo sido importante por este motivo a comparação dos resultados com os resultados de referência, obtidos diretamente a partir da simulação dos dados históricos.

O objetivo do terceiro caso é demonstrar a flexibilidade do modelo para considerar variáveis exógenas no processo estocástico representado. Desta forma, neste caso o modelo é estimado e simulado a partir das mesmas séries de velocidade de vento anteriormente representadas, no entanto foi incorporada ao modelo uma variável climática adicional, associada à medida de intensidade do fenômeno El niño. Para este caso não é possível estabelecer uma referência a partir dos dados históricos, sendo as comparações dos resultados realizadas contra os resultados anteriormente obtidos, no

intuito de destacar o quanto a introdução de informação no modelo, por meio da representação de variáveis exógenas, pode modificar os resultados esperados.

Por fim, no último caso foi simulado o sistema brasileiro, representando séries de geração eólica reais, e cuja informação disponibilizada corresponde já às séries de geração, e não as séries de vento, sendo também explorada neste caso a representação da dependência estatística entre as séries de geração eólica com as séries de vazões afluentes.

Todas as simulações foram realizadas com um simulador desenvolvido em C++, denominado *RelSim*, sob o enfoque da modelagem orientada a objetos, e que corresponde a uma evolução do simulador inicialmente desenvolvido em [63].

O simulador possui integração, provida por bibliotecas de linkedição em tempo de execução, com o ambiente da linguagem R[64]. Este ambiente, distribuído em código aberto, provê uma ampla variedade de funcionalidades estatísticas para modelagem linear, não linear, testes estatísticos, análise de séries temporais, classificação, regressão, classificação, entre outros, além de um ferramental avançado para análise gráfica dos resultados.

O ambiente da linguagem R é constituído por um conjunto de pacotes desenvolvidos de forma colaborativa, tendo sido utilizados neste trabalho, em especial, os pacotes estatísticos: *stats*, que é o pacote de estatísticas básicas do R, *minerva*[65] que disponibiliza funcionalidades para o cálculo do coeficiente de máxima informação e outros índices associados à exploração não paramétrica de informação, e o *bnlearn*, que implementa as funcionalidades para estimação de estrutura e parâmetros de redes Bayesianas.

6.2 Dados de Vento

Neste trabalho foi utilizada a base de dados do Projeto KNMI HYDRA, provido pelo *Royal Netherlands Meteorological Institute (KNMI)*[57].

A preferência pela utilização desta base de dados se justifica por ela agregar características que ainda são muito difíceis de serem encontradas em bases de dados de séries de vento, principalmente em relação ao sistema elétrico brasileiro, para o qual são poucos os dados de geração eólica disponibilizados.

Primeiramente, esta é uma base aberta, na qual todas as informações utilizadas estão disponíveis *online*, sendo possível, desta maneira, que qualquer um possa trabalhar com o mesmo conjunto de informações, podendo reproduzir e comparar análises apresentadas neste trabalho. Em comparação com outras bases de dados de

séries de vento, principalmente em comparação com muitas das séries medidas no sistema brasileiro, esta é uma base que apresenta um horizonte de medições bastante extenso, havendo muitas séries com mais de 20 anos de medições reais. Esta base de dados reúne mais de 30 estações de medições, das quais 27 são efetivamente utilizadas nas análises do modelo desenvolvido, e o que é uma excelente característica para estudar a representação de dependência estatística de alta dimensionalidade, principalmente pelo fato do relacionamento entre as diferentes estações ser bastante heterogêneo, sendo que existem tantos postos *offshore*, quanto em terra.

Além disso, a base de dados não apenas contem os registros do módulo da velocidade do vento, como também a direção do vento, que é uma informação necessária para a simulação do efeito do sombreamento, abordado nas análises realizadas.

A lista de estações de medidas de velocidades de vento utilizadas nas análises é apresentada na Tabela 4, contendo, respectivamente, o código e a identificação da estação.

<i>Código</i>	<i>Identificação</i>
s-210	Valkenburg
s-225	IJmuiden
s-229	Texelhors
s-235	De Kooy
s-240	Schiphol
s-242	Vlieland
s-248	Wijdenes
s-249	Berkhout
s-251	Hoorn (Tersch.)
s-252	K13
s-254	Meetpost Noordwijk
s-265	Soesterberg
s-269	Lelystad
s-275	Deelen
s-277	Lauwersoog
s-280	Eelde
s-283	Hupsel
s-320	L.E. Goeree
s-321	Europlatform
s-323	Wilhelminadorp
s-324	Stavenisse
s-330	Hoek van Holland
s-340	Woensdrecht
s-356	Herwijnen
s-370	Eindhoven
s-380	Beek
s-391	Arcen

Tabela 4– Estações de Medição de Velocidade de Vento

A seleção deste subconjunto de postos de medição se deu a partir de uma primeira filtragem dos dados históricos disponibilizados, de maneira a eliminar os postos com muitos pontos de medições faltantes, ou com muitos pontos classificados como medidas de baixa qualidade pelo instituto responsável.

Na Figura 46 é apresentado um mapa contendo a distribuição geográfica dos postos de medição, destacando-se os vinte e sete postos que foram efetivamente utilizados.



Figura 46- Mapa Contendo a Distribuição Geográfica dos Postos de Medição[57]

6.3 Sistema IEEE-RTS

O sistema IEEE-RTS foi originalmente proposto em [62], se consolidando como o sistema base para testes de novas metodologias relacionadas ao estudo de confiabilidade. Para este sistema são definidos, além dos dados estocásticos típicos de confiabilidade, dados relacionados à capacidade de geração, limites da transmissão e carga. Não é definido o ponto de operação do sistema, isto é, o módulo e ângulo das tensões, valores de injeção de ativo e reativo, etc. No entanto, em [66] estes dados de operação são definidos, sendo adotados para os sistemas analisados neste trabalho.

As características gerais do sistema IEEE-RTS são apresentadas na Tabela 5, os dados das usinas convencionais são apresentados na Tabela 6 e os dados de confiabilidade das linhas de transmissão e transformadores são apresentados na Tabela 7.

O diagrama unifilar do sistema IEEE-RTS base é apresentado na Figura 47.

<i>Característica</i>	
Número de Barras	24 unidades
Número de Circuitos	38 unidades
Número de Usinas	15 unidades
Geração Total (capacidade)	3405 MW
Carga Total (pico)	3135 MW
Carga Média	1926 MW

Tabela 5– Características Básicas do Sistema IEEE-RTS

Barra de Conexão	Unidades Geradoras	Capacidade por Unidade (MW)	Taxa de Falha (oc/hora)	Taxa de Retorno (oc/hora)
1	2	20	0,002222	0,020000
1	2	76	0,000510	0,025000
2	2	20	0,002222	0,025000
2	2	76	0,000510	0,025000
7	3	100	0,000833	0,020000
13	3	197	0,001053	0,020000
15	5	12	0,000340	0,016667
15	1	155	0,001042	0,025000
16	1	155	0,001042	0,025000
18	1	400	0,000909	0,006667
21	1	400	0,000909	0,006667
22	6	50	0,000505	0,050000
23	2	155	0,001042	0,025000
23	1	350	0,000870	0,010000

Tabela 6– Dados das Usinas do Sistema IEEE-RTS

Barra 01	Barra 02	Taxa de Falha (oc/hora)	Taxa de Retorno (oc/hora)
1	2	0,000027	0,062500
1	3	0,000058	0,100000
1	5	0,000038	0,100000
2	4	0,000045	0,100000
2	6	0,000055	0,100000
3	9	0,000043	0,100000
3	24	0,000002	0,001302
4	9	0,000041	0,100000
5	10	0,000039	0,100000
6	10	0,000038	0,028571
7	8	0,000034	0,100000
8	9	0,000050	0,100000
8	10	0,000050	0,100000
9	11	0,000002	0,001302
9	12	0,000002	0,001302
10	11	0,000002	0,001302
10	12	0,000002	0,001302
11	13	0,000046	0,090909
11	14	0,000045	0,090909
12	13	0,000046	0,090909
12	23	0,000059	0,090909
13	23	0,000056	0,090909
14	16	0,000043	0,090909
15	16	0,000038	0,090909
15	21	0,000047	0,090909
15	21	0,000047	0,090909
15	24	0,000047	0,090909
16	17	0,000040	0,090909
16	19	0,000039	0,090909
17	18	0,000037	0,090909
17	22	0,000062	0,090909
18	21	0,000040	0,090909
18	21	0,000040	0,090909
19	20	0,000043	0,090909
19	20	0,000043	0,090909
20	23	0,000039	0,090909
20	23	0,000039	0,090909
21	22	0,000051	0,090909

Tabela 7– Dados das Linhas de Transmissão e Transformadores do Sistema IEEE-RTS

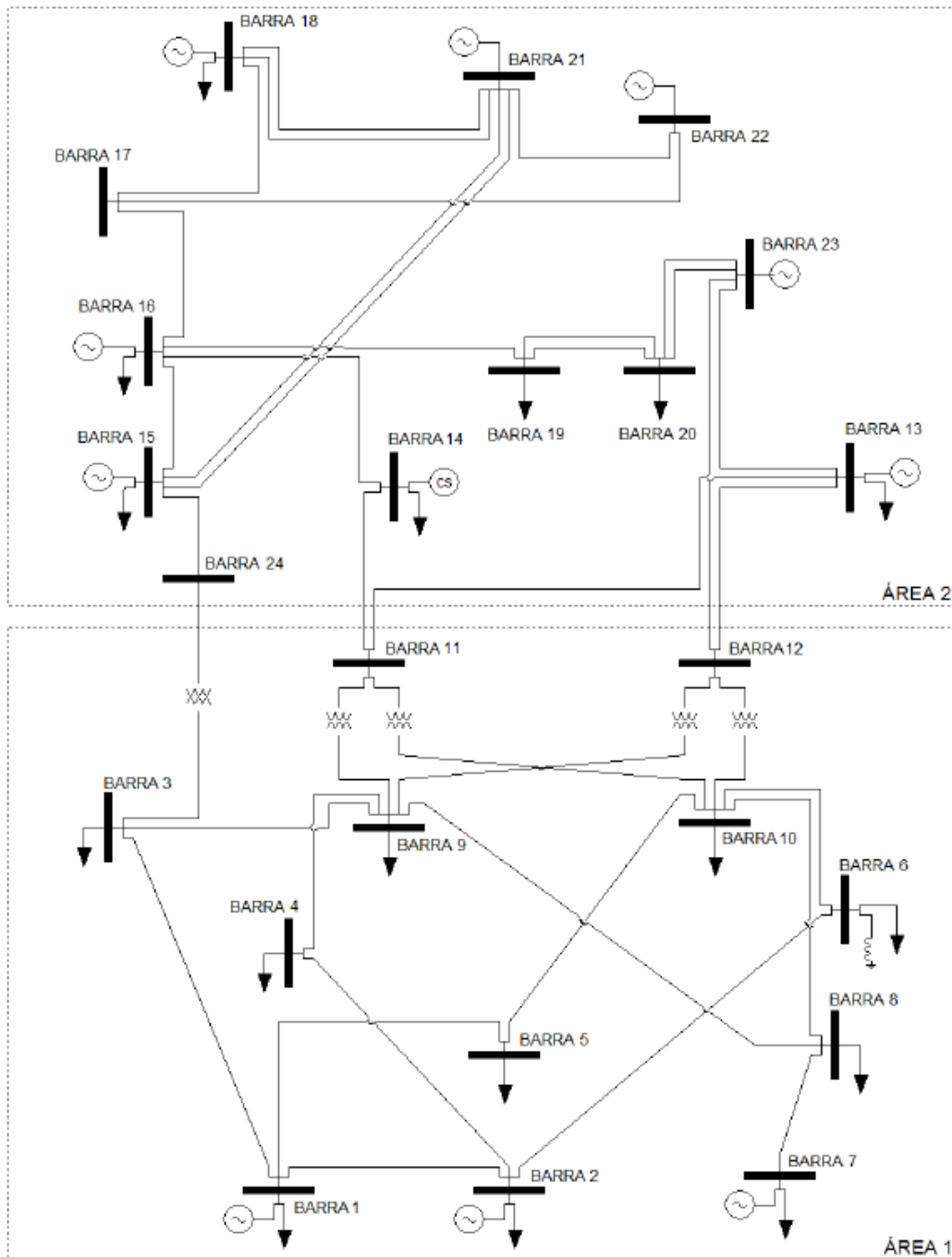


Figura 47- Representação Unifilar do Sistema IEEE-RTS[62]

Para a criação do sistema IEEE-RTS modificado utilizado nas análises deste trabalho, foram definidas 27 usinas eólicas fictícias, cada uma correspondendo a um posto de medição da base de dados de vento utilizada.

A distribuição destas usinas eólicas fictícias ao longo do sistema IEEE-RTS é apresentada na Tabela 8, na qual se observa que foi adotado o mesmo código de identificação dos postos de medição para cada uma das respectivas usinas.

<i>Barra</i>	<i>Usinas Eólicas</i>		
1	s-210	s-240	s-330
2	s-225	s-249	s-356
7	s-254	s-269	s-323
15	s-235	s-324	s-340
16	s-242	s-251	s-280
18	s-283	s-320	s-391
21	s-265	s-275	s-370
22	s-229	s-321	s-380
23	s-248	s-252	s-277

Tabela 8– Distribuição das Usinas Eólicas ao Longo do RTS.

Foi estabelecido o mesmo número de turbinas instaladas para cada uma das usinas eólicas, 25 unidades de 2,5 MW cada, sendo adotada também uma mesma característica de geração para todas as turbinas representadas no sistema, extraída de [67], apresentada na Figura 48.

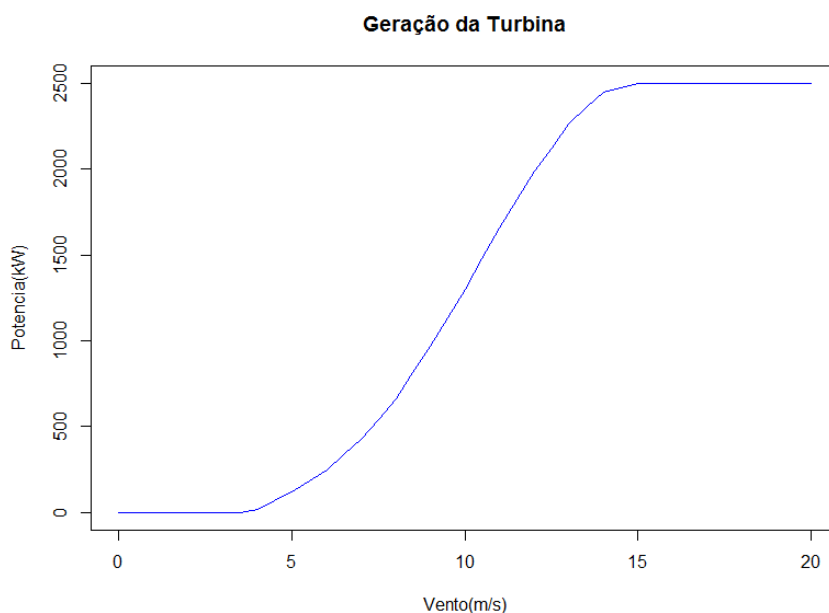


Figura 48- Geração da Turbina em Função da Velocidade do Vento

Desta maneira, a capacidade instalada total adicionada pelas usinas eólicas é de aproximadamente 1687,50 MW (27 usinas x 25 unidades x 2,5 MW), sendo 187,50 MW (3 usinas x 25 unidades x 2,5 MW) acrescentado a cada barra com geração eólica.

6.3.1 Caso 1 - RTS + 27 Usinas Eólicas – Velocidade do Vento

Neste caso, a base de dados utilizada correspondeu a um período de oito anos das observações das séries históricas de velocidades de vento, abrangendo os registros horários de janeiro de 1997 a dezembro de 2003.

De modo a estabelecer uma referência para o impacto da incorporação das usinas eólicas no sistema, o caso foi primeiramente simulado sem se considerar a geração eólica, apenas com a representação de um aumento de carga em 10%, resultando em um novo pico de carga para o sistema de 3448,5 MW, sendo obtidos os índices de confiabilidade apresentados na Tabela 9.

<i>Índice de Confiabilidade</i>	<i>Caso Referência</i>
LOLP	0,0238 p.u
EPNS	4,182 MW
LOLF	36,36 oc/ano

Tabela 9– Índices de Confiabilidade para o Sistema – Sem Eólicas

Após o estabelecimento dos valores de referência, foi simulado o caso proposto, com a incorporação das vinte e sete usinas eólicas, tendo sido aplicados, além do modelo proposto, os modelos de Simulação Monte Carlo Sequencial, Pseudo-Sequencial, Pseudo-Cronológico e a Simulação Monte Carlo Não Sequencial, sem representação de dependência estatística entre os elementos variantes no tempo. O objetivo da aplicação da Simulação Sequencial foi estabelecer uma referência para os índices de confiabilidade esperados para o caso, enquanto que as simulações Pseudo-Sequencial e Pseudo-Cronológica foram inseridas principalmente para estabelecer uma referência para o tempo computacional. A Simulação Monte Carlo Não Sequencial foi aplicado para prover uma sensibilidade da ordem de grandeza dos erros obtidos ao se desconsiderar a representação da dependência estatística das variáveis neste caso.

Os índices de confiabilidade obtidos para este caso são apresentados na Tabela 10. Os intervalos de confiança para os índices de referência (Simulação Monte Carlo Sequencial), com 95% de certeza, são apresentados na Tabela 11.

<i>Modelo SMC</i>	<i>LOLP (pu)</i>	<i>EPNS (MW)</i>	<i>LOLF (oc/ano)</i>	<i>Tempo</i>
Sequencial	0,0095 (0,0198)	1,6028 (0,0300)	16,49 (0,0151)	8hs:38min (8745306 casos)
Modelo Independente	0,0060 (0,0219)	0,8927 (0,0300)	24,64 (0,0221)	31min:39sec (345156 casos)
Pseudo-Sequencial	0,0098 (0,0223)	1,7030 (0,0300)	15,79 (0,0277)	22min:31sec (202263 casos)
Pseudo-Cronológico	0,0097 (0,0127)	1,6897 (0,0173)	16,50 (0,0300)	65min:53sec (632399 casos)
Modelo Proposto	0,0098 (0,0223)	1,6275 (0,0300)	16,88 (0,0264)	16min:46sec (203165casos)

Tabela 10– Índices de Confiabilidade para o Sistema – Caso 1

	<i>LOLP (pu)</i>	<i>EPNS (MW)</i>	<i>LOLF (oc/ano)</i>
Limite Inferior	0,0091	1,509	16,00
Limite Superior	0,0099	1,697	16,99

Tabela 11– Intervalos de Confiança para os Índices da SMC Sequencial – Caso 1

De modo a tornar ainda mais clara a comparação entre a precisão dos métodos de simulação, na Tabela 12 são apresentados os erros percentuais absolutos entre os índices obtidos pelos modelos e a simulação de referência, a Simulação Monte Carlo Sequencial.

<i>Modelo SMC</i>	<i>ERRO LOLP (%)</i>	<i>ERRO EPNS (%)</i>	<i>ERRO LOLF (%)</i>
Convencional	36,8%	44,3%	4,2%
Pseudo-Sequencial	3,1%	6,3%	5,4%
Pseudo-Cronológico	2,1%	5,4%	< 0,1%
Modelo Proposto	3,1%	1,5%	2,4%

Tabela 12– Erros Percentuais para os Índices de Confiabilidade do Sistema – Caso 1

6.3.1.1 Análise do Desempenho Computacional

O primeiro aspecto a se observar é o que os resultados da Simulação Monte Carlo Não Sequencial convencional são bem imprecisos, estando todos os índices bem distantes dos intervalos de confiança estabelecidos pela simulação de referência. Este fato pode ser observado também a partir dos erros percentuais absolutos obtidos para este caso, que são bem altos, estando entre 36% e 45%. A observação desta grande imprecisão é interessante para a análise da precisão do modelo proposto, pois permite dizer que a geração eólica representada é relevante para os índices de confiabilidade obtidos. Do contrário, não se verificaria um impacto tão grande nos índices da Simulação Monte Carlo Não Sequencial, e, como consequência, qualquer inadequação

do modelo proposto poderia ser ocultada pelo fato da ordem de grandeza dos efeitos representados não serem significativos para o sistema.

Em comparação com a Simulação Monte Carlo Sequencial, observa-se que os índices de confiabilidade obtidos pelo modelo proposto foram bem próximos dos valores esperados, estando todos os índices dentro do intervalo de confiança estabelecido. Verifica-se que para a Simulação Pseudo-Sequencial o índice LOLF está um pouco fora do intervalo de confiança.

Quanto a análise do esforço computacional do modelo, o ganho em relação à Simulação Monte Carlo Sequencial foi bastante significativo, como já esperado, tendo sido o modelo proposto aproximadamente 31 vezes mais rápido do que a Simulação Monte Carlo Sequencial. Em comparação com o desempenho da Simulação Monte Carlo Pseudo-Sequencial, o modelo proposto foi aproximadamente 1,3 vezes mais rápido, não sendo possível, no entanto, generalizar esta conclusão a respeito da comparação do custo computacional dos dois modelos. A princípio, espera-se que os dois modelos de simulação tenham custos computacionais similares.

Um aspecto adicional a ser considerado a respeito do desempenho computacional destes dois modelos é que o modelo proposto precisou avaliar 902 estados a mais de adequação do que a Simulação Pseudo-Sequencial, valor correspondente a 0,4% do total de estados avaliados. A Simulação Pseudo-Sequencial, no entanto, precisou avaliar mais estados subsequentes aos estados de falhas do que o modelo proposto, o que justifica o tempo mais elevado.

O modelo proposto foi 3,9 vezes mais rápido do que a Simulação Pseudo-Cronológica. Este fato pode ser explicado pela maior demora na convergência da LOLF na Simulação Pseudo-Cronológica, em comparação tanto com o modelo proposto, quanto com a Simulação Pseudo-Sequencial. Embora este método tenha como principal vantagem a necessidade de se avaliar um único estado adicional para cada estado de falha, diferentemente da Simulação Pseudo-Sequencial, que precisa avaliar um maior número de estados em torno do estado de falha, o número de estados adicionais que são avaliados devido a mais lenta convergência da LOLF acaba comprometendo o desempenho do método.

Na Figura 49 é apresentada a evolução da convergência da LOLF ao longo do número de estados de adequação avaliados para os três modelos. Observa-se que existe uma grande similaridade entre a convergência do modelo proposto e a Simulação Pseudo-Sequencial, enquanto que a Simulação Pseudo-Cronológica converge bem mais lentamente neste caso.

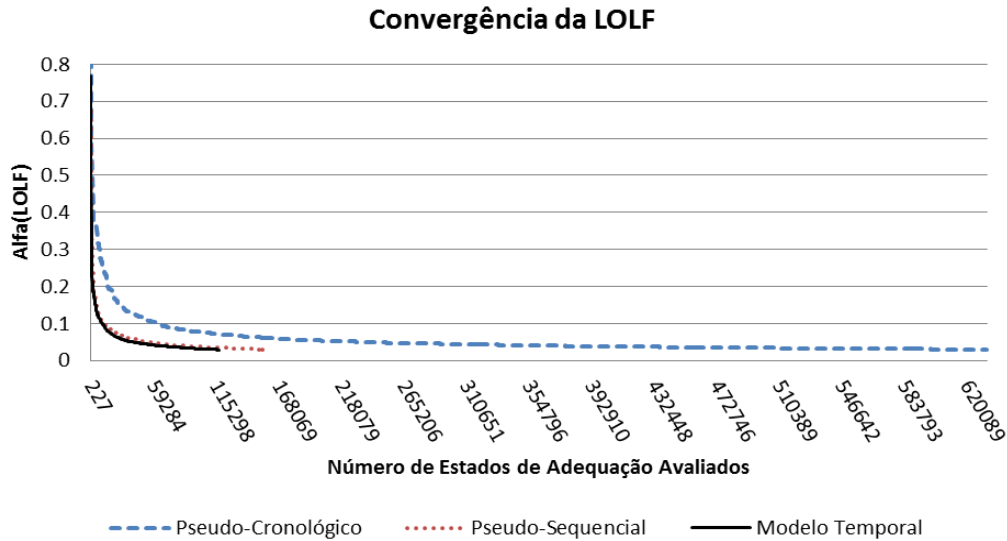


Figura 49– Comparação das Convergências da LOLF para o Caso 1

Para uma contabilização mais justa do tempo total despendido durante todo o processo, na Tabela 13 são apresentados os tempos necessários para efetuar as operações adicionais realizadas na estimação do modelo proposto.

O tempo total para as etapas adicionais acrescentam aproximadamente um tempo de 86 segundos ao processo total, um pouco menos que um minuto e meio, no qual se percebe que os maiores tempos se devem principalmente a estimação da estrutura das redes Bayesianas.

Embora o modelo proposto tenha este processamento adicional, é importante frisar que ele dispensa a necessidade de se gerar a priori quaisquer conjuntos de séries sintéticas, que pode ser uma etapa bastante dispendiosa, considerando-se o número de séries sintéticas desejadas e o horizonte. No modelo proposto todas as amostras são obtidas durante a própria simulação, com custo computacional similar a amostragem dos estados dos componentes convencionais, que é irrelevante para o tempo da simulação, como pode ser visto na Tabela 14, na qual são apresentados os tempos despendidos para as principais etapas do algoritmo de simulação. Observa-se que o tempo total de amostragem dos estados para este caso (4s) corresponde a aproximadamente 0,1% do tempo total do processo.

<i>Atividade</i>	<i>Tempo</i>
Estimação das funções de densidade de probabilidade marginais e funções de transformação	20s
Transformação das variáveis aleatórias e discretização	5s
Estimação da estrutura da rede Bayesiana para representação das probabilidades dos estados	22s
Estimação dos parâmetros da rede Bayesiana para representação das probabilidades dos estados	3s
Estimação da estrutura das redes Bayesianas para representação das transições forward e backward	32s
Estimação dos parâmetros das redes Bayesianas para representação das transições forward e backward	4s

Tabela 13– Tempo Despendido para as Atividades de Estimação do Modelo Proposto

<i>Etapas</i>	<i>Tempo</i>
Geração de amostras dos estados	4s
Montagem dos problemas de rede	3s
Resolução dos problemas de rede	935s
Escrita de resultados	14s

Tabela 14– Tempo Despendido para as Etapas da Simulação do Modelo Proposto

6.3.1.2 Análise da Confiabilidade

Para a análise dos índices de confiabilidade deste sistema, foram utilizados apenas os resultados produzidos pelo modelo proposto de simulação.

Inicialmente, destaca-se que os valores médios dos índices de confiabilidade diminuíram com a introdução da geração eólica no sistema. A *LOLP* passou de 2,38% para 0,98%, a *EPNS* passou de 4,182 MW para 1,628 MW e a *LOLF* passou de 36,36 oc/ano para 16,88 oc/ano.

A inclusão da capacidade adicional de geração eólica de 1687 MW resultou na diminuição do corte de carga médio em aproximadamente 2,55 MW, equivalente a uma redução de aproximadamente 61% do total de corte de carga que havia no sistema.

Como a maior parte dos estados avaliados não possui corte de carga, pois apenas 0,98% dos estados apresentam corte de carga, tanto o valor médio do corte de carga, dado pela *EPNS*, como a própria distribuição de probabilidade do corte de carga são muito influenciados pelos estados sem corte de carga. Desta maneira, para possibilitar uma visão mais representativa de como se distribuem estes cortes de carga, são apresentadas duas densidades de probabilidade de corte de carga, onde apenas os estados do sistema com corte de carga foram considerados para a estimação destas

densidades de probabilidade. A primeira densidade, apresentada na Figura 50, corresponde ao caso sem geração eólica, e a segunda densidade, apresentada na Figura 51, corresponde ao caso com geração eólica.

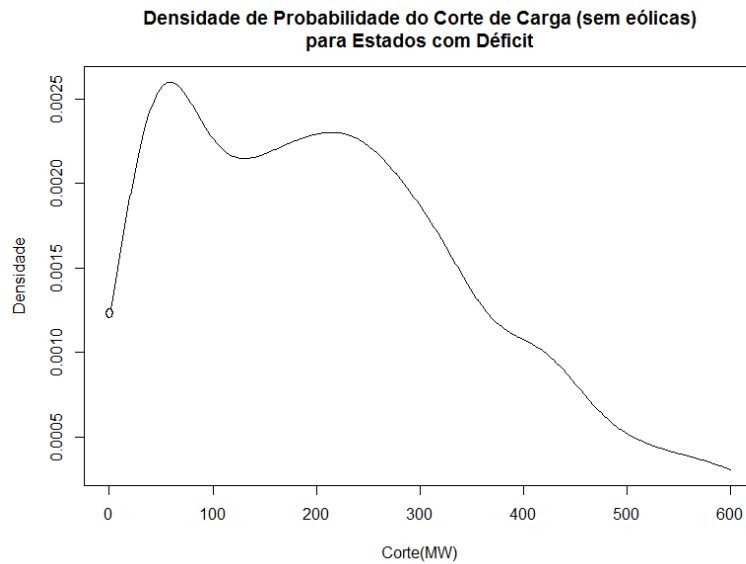


Figura 50- Densidade de Probabilidade do Corte de Carga para Estados com Déficit – Sem Geração

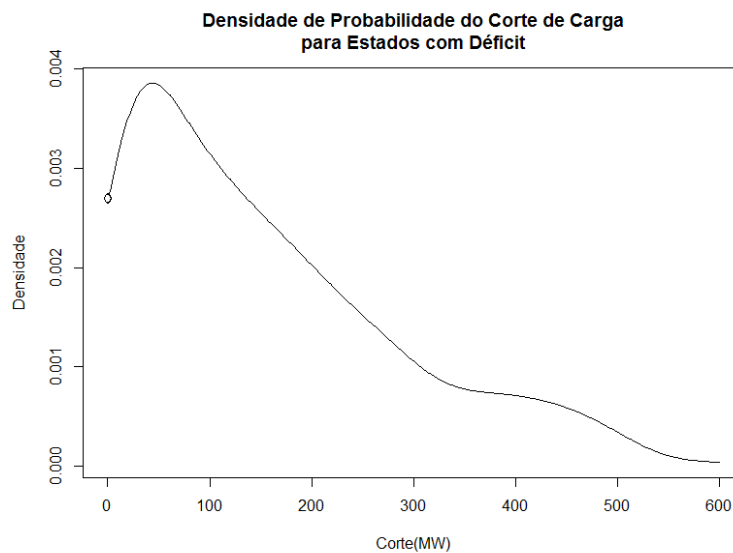


Figura 51- Densidade de Probabilidade do Corte de Carga para Estados com Déficit – Caso 1

Observa-se que a densidade de probabilidade do caso de referência, sem geração eólica, apresenta duas modas bem definidas, indicando a existência de uma maior

concentração de cortes de carga inferiores a 100 MW, mas com uma segunda grande concentração de cortes entre 100 MW e 300 MW. O efeito da introdução da geração eólica é principalmente o de atenuação desta segunda concentração, havendo um aumento na concentração dos cortes de valores inferiores a 100 MW.

Quantitativamente, este efeito pode ser medido obtendo-se a distribuição de probabilidade acumulada, a partir da integração destas funções. Com isto, tem-se que as probabilidades dos cortes de carga serem inferiores a 100 MW, são de aproximadamente 25%, no caso sem geração eólica, e de 44%, no caso com geração eólica, sugerindo que a introdução das usinas eólicas resultou na atenuação de grande parte dos cortes de carga entre 100 MW e 300 MW, fazendo com que estes cortes passassem a estar na faixa dos cortes de carga inferiores a 100 MW.

Na Figura 52 são apresentados os cortes de carga médios por barra para o caso de referência, gráfico da esquerda, e para o caso com a introdução da geração eólica, gráfico da direita.

Observa-se que as maiores parcelas de corte estão concentrados nas quatro barras com maior corte médio de carga, as barras 15, 16, 18 e 19. O total de corte de carga médio destas barras é de aproximadamente 1,18 MW, correspondente a quase 25% do corte médio total. A comparação dos gráficos mostra que a introdução das eólicas teve um efeito bastante significativo para estas barras, e em especial para a barra 19, a qual apresenta o maior corte médio. O corte de carga desta barra diminuiu de 0,832 MW para 0,257 MW, correspondendo a uma queda de aproximadamente 69% do valor de corte médio.

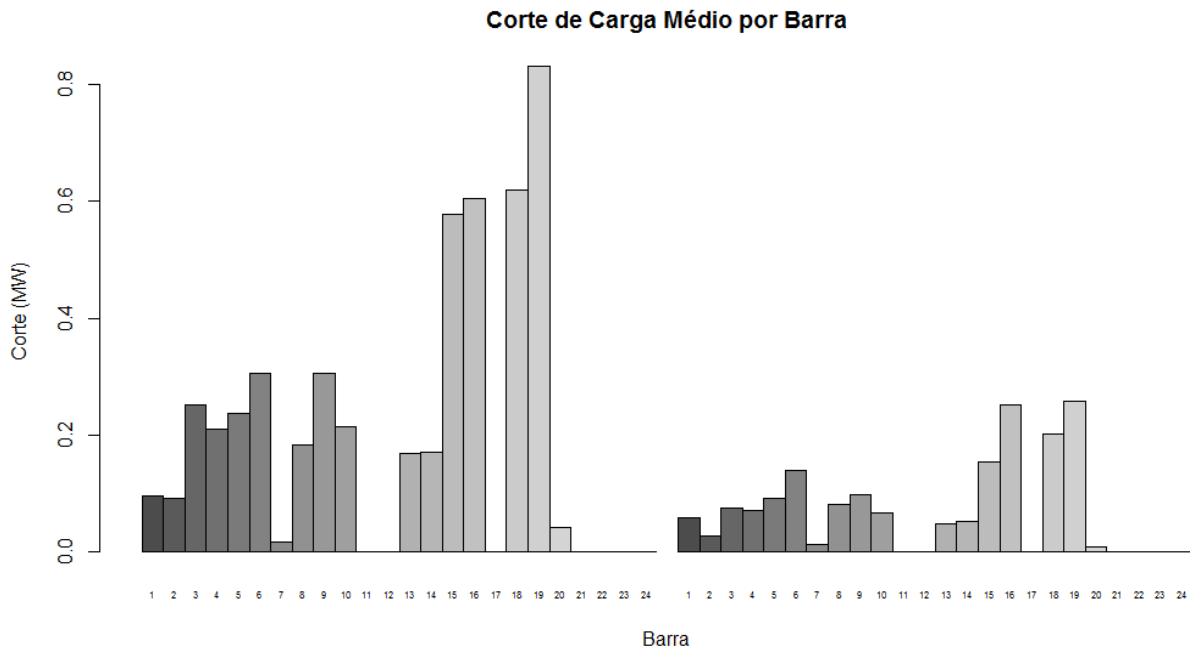


Figura 52- Corte de Carga Médio por Barra – Caso 1

Na Figura 53 é apresentada a redução de corte de carga médio por barra.

Observa-se que este percentual de redução é bastante desigual entre as barras. O maior valor observado, de aproximadamente 80%, está associado à barra 20, cujo corte de carga inicial já era bastante reduzido em relação às outras barras, e com isso qualquer redução absoluta em termos absolutos, resulta em um percentual relativo bastante significativo.

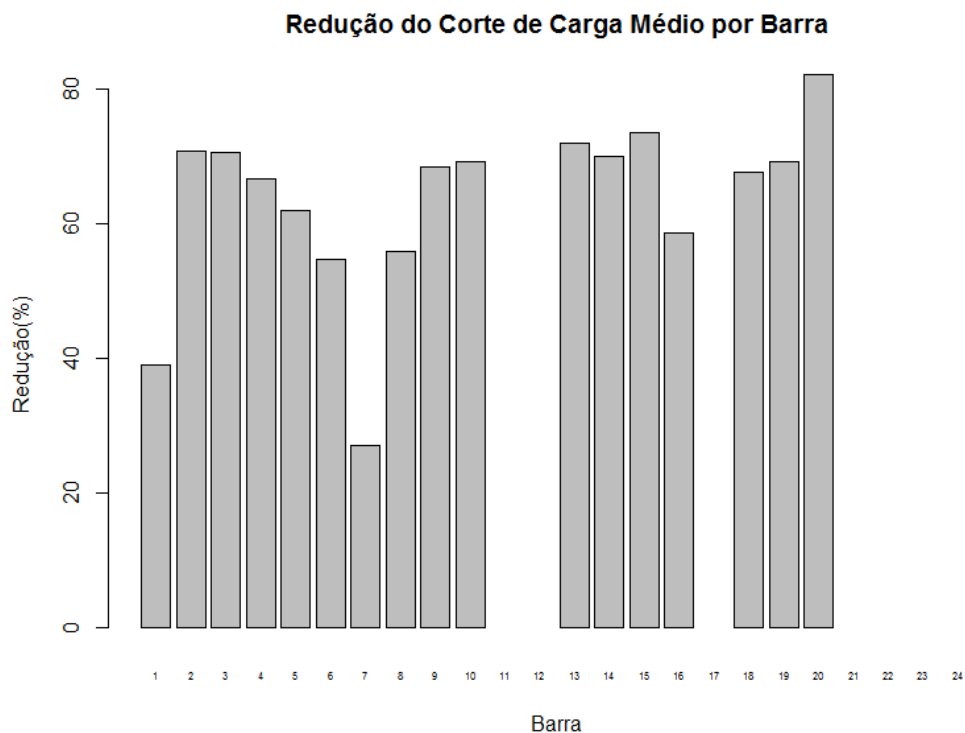


Figura 53- Redução Percentual de Corte de Carga Médio por Barra – Caso 1

6.3.1.3 Análise das Amostras Sintéticas Geradas

Outro aspecto importante de ser analisado é a qualidade das amostras das variáveis aleatórias geradas, no que tange à preservação das estatísticas do histórico.

Para uma verificação qualitativa da eficácia do modelo proposto de preservar as densidades de probabilidade marginais, na Figura 54 são apresentadas as funções de densidade de probabilidade para as variáveis aleatórias correspondentes as séries de velocidade de vento: s-210, s-229, s-240 e s-248, comparando-se também a estimativa realizada sobre os dados históricos e sobre as séries sintéticas.

Observa-se que visualmente, para os quatro casos apresentados, as densidades de probabilidade estimadas para os dados sintéticos são bem próximas àquelas estimadas para os dados históricos.

Para uma análise quantitativa mais abrangente, as densidades obtidas para todas as séries podem ser comparadas a partir da aplicação do teste estatístico de *Mann-Whitney-Wilcoxon Test* [68], utilizado para verificar a hipótese de dois conjuntos de amostras pertencerem a mesma distribuição, com um certo nível de significância.

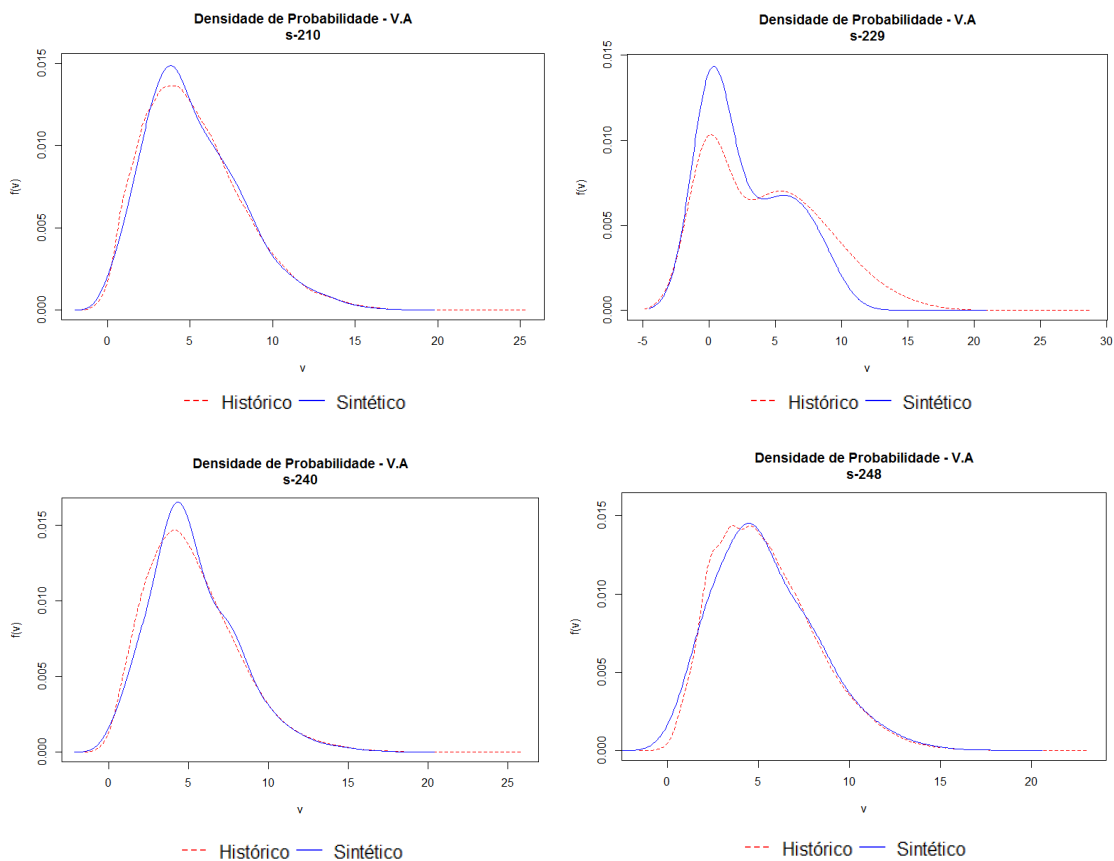


Figura 54– Densidade de Probabilidade para as Variáveis Aleatórias de Vento

Os valores de *p-value* do teste estatístico para as quatro séries destacadas nos gráficos anteriores são apresentados na Tabela 15.

Observa-se que os *p-value* obtidos estão acima do nível de significância de 0,01 e apenas uma série, a s-240, está um pouco abaixo do nível de significância de 0,05, indicando quantitativamente que as densidades obtidas são compatíveis entre si, de acordo com o teste estatístico realizado.

Série	<i>p-value</i>
s-210	0,23
s-229	0,07
s-240	0,03
s-248	0,17

Tabela 15– Referência para os Valores de *p-value*.

A preservação da dependência estatística entre as diferentes séries de vento pode ser inicialmente avaliada a partir da comparação das densidades de probabilidade conjunta (densidade 2D) estimadas a partir dos dados históricos e das séries sintéticas amostradas. Esta comparação é apresentada para algumas das séries nos gráficos de níveis da Figura 55.

Das séries comparadas, destaca-se que para a combinação dada pelas variáveis s-210 e s-225, o formato obtido para a densidade de probabilidade estimada sugere que o relacionamento entre estas séries poderia ser bem caracterizado através somente da correlação linear. Isto é, é possível visualizar a densidade conjunta obtida como sendo uma dispersão bem comportada em torno de uma linha reta.

Numericamente, um indicador que sinaliza que a correlação linear é suficientemente precisa para caracterizar a dependência entre estas variáveis é o **MIC-R2**, que é negativo para este caso: **MIC-R2** = -0,12. Isto é, o coeficiente de máxima informação não descreve nenhuma relação adicional àquela capturada pela correlação.

Para as duas densidades posteriores, (s-229, s-210) e (s-240, s-249), isto já não é mais verdade. É possível ver que existem basicamente duas grandes áreas de densidade distintas, e que seria bastante impreciso tentar caracterizar a probabilidade conjunta destas variáveis por meio de uma dispersão em torno de uma única reta.

Analogamente ao caso anterior, pode-se mensurar o quanto a correlação linear pode ser imprecisa para estes casos a partir do indicador, **MIC-R2**, que vale respectivamente para os dois conjuntos de variáveis analisados: 0,08 e 0,16,

Para a última densidade analisada, que descreve a probabilidade conjunta de s-240 e s-248, o formato novamente se aproxima de uma densidade normal bivariada, sendo o relacionamento entre estas variáveis também bem caracterizado pela correlação linear. Este fato também é numericamente observável, uma vez que o índice **MIC-R2** para este caso também é negativo, **MIC-R2** = -0,07.

Pode-se dizer que independentemente da complexidade do formato da densidade dos dados históricos, o modelo proposto foi capaz de reproduzir bem a probabilidade conjunta das variáveis avaliadas. Nota-se que existe uma grande similaridade visual entre as densidades dos dados históricos e das amostras sintéticas.

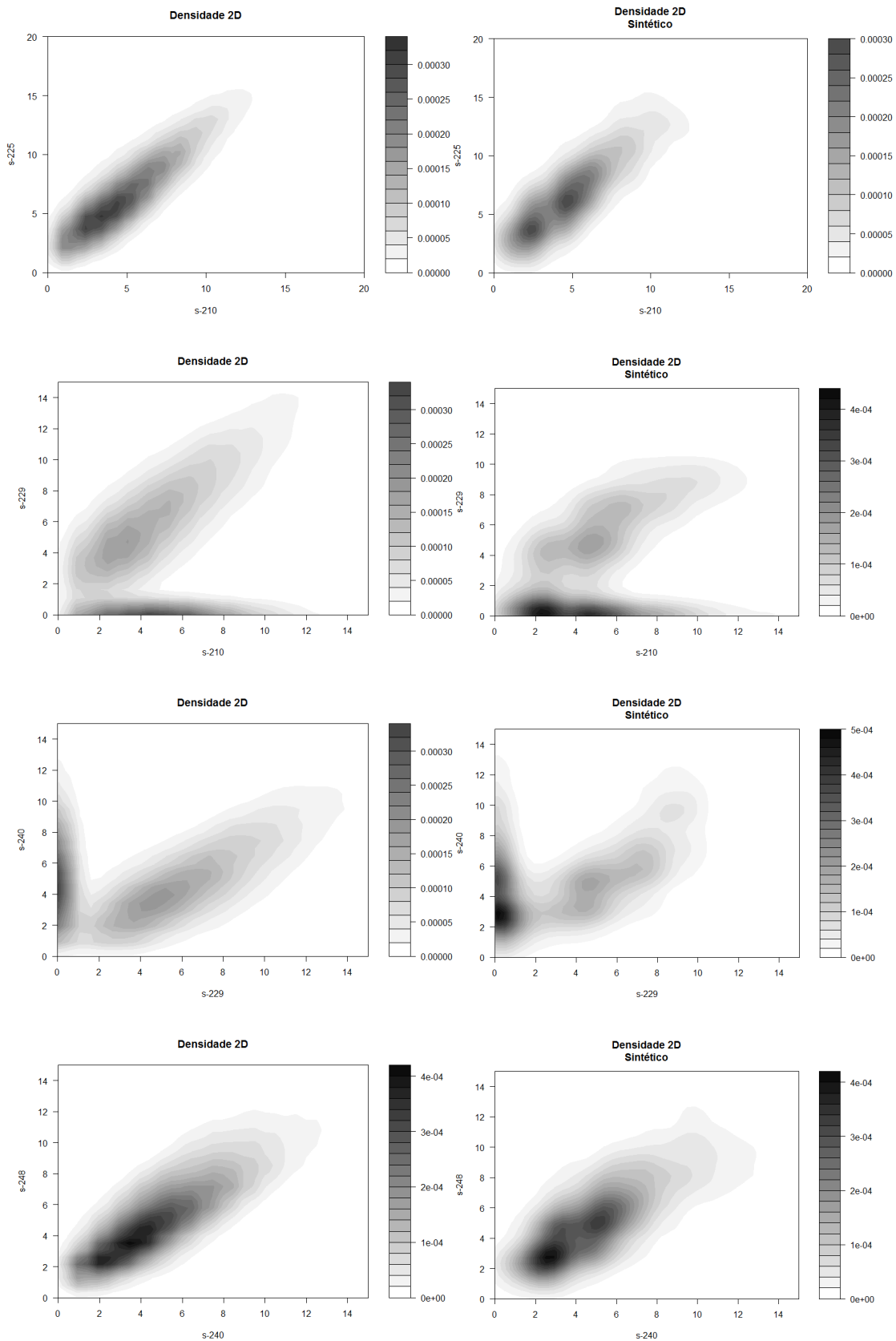


Figura 55– Densidades de Probabilidade Conjunta (Densidades 2D) para Algumas da Séries

Para uma análise mais abrangente, pode-se obter o coeficiente de informação máxima para todas as combinações 2 a 2 de variáveis aleatórias, de modo a se verificar a diferença entre os indicadores medidos a partir dos dados históricos e a partir das amostras sintéticas. Calculando-se as diferenças absolutas ($|MIC_{hist}-MIC_{sint}|$) para cada combinação e construindo um vetor com estas diferenças, tem-se um conjunto de medidas que pode ser usado na obtenção de um histograma, apresentado na Figura 56, cujo objetivo é possibilitar a visualização de como se distribuem estas diferenças.

Observa-se que a grande maioria das diferenças está concentrada no intervalo entre 0,00 e 0,05, estando ainda a moda desta distribuição bem perto da origem. Mesmo para os casos em que as diferenças são superiores a 0,05, elas ainda são na sua grande maioria inferiores a 0,1, restando muito poucas medidas que chegam, ou superam a diferença de 0,15.

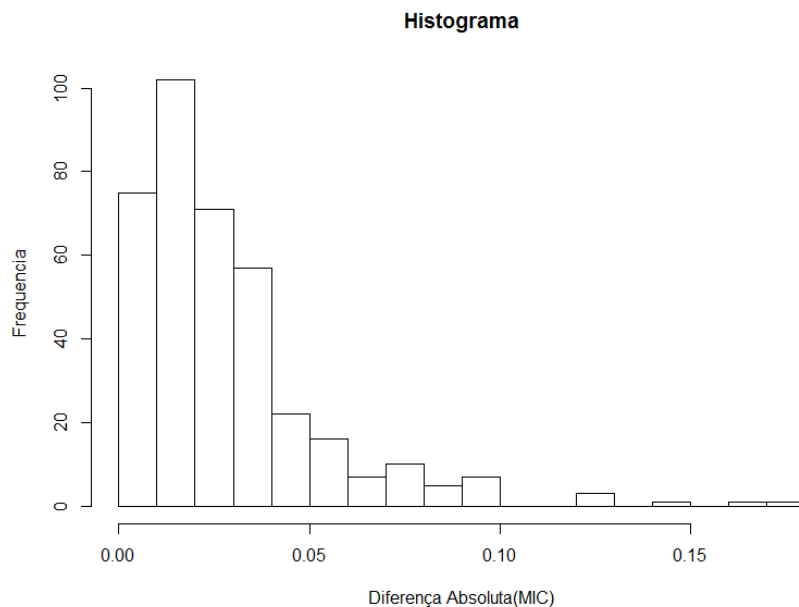


Figura 56– Histograma da Diferença Absoluta dos MICs (Histórico e Sintética)

6.3.2 Caso 2 - RTS + 27 Usinas Eólicas – Velocidade e Direção do Vento

Este segundo caso corresponde a uma extensão do primeiro caso avaliado, incluindo a consideração das séries de direção dos ventos, além do módulo da velocidade. O objetivo é incorporar ao modelo incertezas importantes associadas a representação das perdas pelo efeito do sombreamento das usinas. Desta maneira, procura-se avaliar a capacidade do modelo proposto de representar o processo estocástico conjunto das velocidades e direções do vento, representando de forma adequada, em especial, tanto as densidades de probabilidade das séries de direção, que

são de difícil caracterização paramétrica, como as dependências estatísticas introduzidas por estas novas variáveis aleatórias representadas.

A incorporação das séries de direção resulta na duplicação do número de variáveis representadas, e, desta forma, este caso também permite a avaliação da capacidade do modelo de representar uma função de densidade de probabilidade conjunta ainda mais complexa do que a do primeiro caso, uma vez que agora existem 54 variáveis representadas associadas aos elementos variantes no tempo.

Conforme discutido na seção 5.4, o objetivo da introdução do efeito do sombreamento neste trabalho, restringe-se ao aspecto da representação da dependência estatística entre a série de módulo de velocidade e da direção da velocidade, sem abranger a modelagem detalhada das perdas aerodinâmicas, devido ao efeito esteira. Desta maneira, a abordagem dada para este caso foi a utilização de uma tabela de eficiência da usina eólica em função da direção do vento, apresentada no estudo [54], estimada a partir de um período de medidas da variação da eficiência da usina de Middelgrunden em relação à variação da direção do vento para um ponto de referência.

Os valores utilizados para compor a curva utilizada neste caso são apresentados na Tabela 16. A função eficiência construída a partir destes dados é apresentada na Figura 57.

<i>Direção (Ângulo)</i>	<i>Eficiência</i>
0	40%
30	95%
60	100%
120	100%
150	90%
180	30%
210	95%
240	100%
300	100%
330	90%
360	40%

Tabela 16– Valores Adotados de Eficiência por Direção

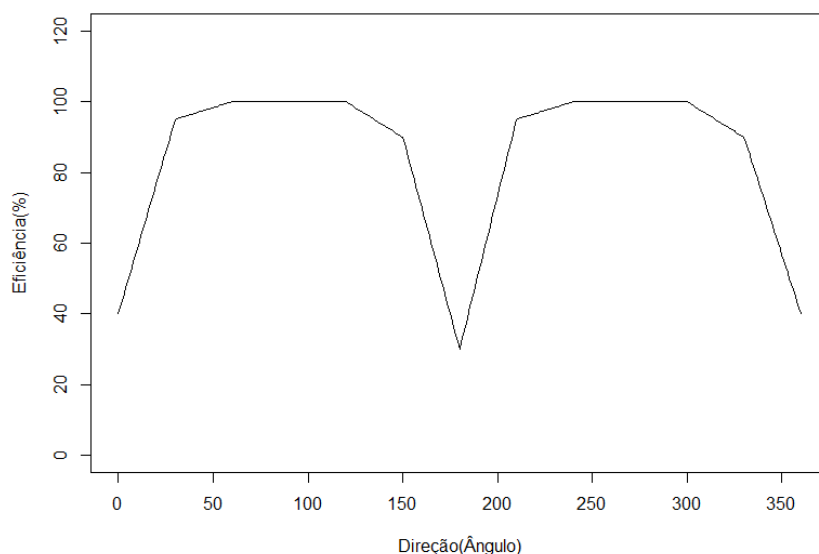


Figura 57– Função de Eficiência das Usinas pela Direção do Vento

Os índices de confiabilidade obtidos para este caso são apresentados na Tabela 17. Os intervalos de confiança para os índices de referência (SMC Sequencial), com 95% de certeza, são apresentados na Tabela 18. Os erros percentuais absolutos entre os índices obtidos pelos modelos e a simulação de referência são apresentados na Tabela 19.

<i>Modelo SMC</i>	<i>LOLP</i> (<i>pu</i>)	<i>EPNS</i> (<i>MW</i>)	<i>LOLF</i> (<i>oc/ano</i>)	<i>Tempo</i>
Sequencial	0,0103 (0,0200)	1,7505 (0,0300)	18,13 (0,0154)	8hs:03min (7979862 casos)
Pseudo-Sequencial	0,0104 (0,0223)	1,8027 (0,0300)	17,59 (0,0276)	26min:32sec (191423 casos)
Pseudo-Cronológico	0,0103 (0,0129)	1,7949 (0,0175)	18,00 (0,0300)	59min:12sec (578826 casos)
Modelo Proposto	0,0101 (0,0224)	1,7647 (0,0300)	17,70 (0,0272)	21min:05sec (196128 casos)

Tabela 17– Índices de Confiabilidade para o Sistema – Caso 2

	<i>LOLP</i> (<i>pu</i>)	<i>EPNS</i> (<i>MW</i>)	<i>LOLF</i> (<i>oc/ano</i>)
Limite Inferior	0,0099	1,650	17,58
Limite Superior	0,0107	1,853	18,68

Tabela 18– Intervalos de Confiança para os Índices da SMC Sequencial – Caso 2

<i>Modelo SMC</i>	<i>LOLP (pu)</i>	<i>EPNS (MW)</i>	<i>LOLF (oc/ano)</i>
Pseudo-Sequencial	< 0,1%	2,9%	3,0%
Pseudo-Cronológico	0,0%	2,5%	< 0,1%
Modelo Proposto	2,0%	0,1%	2,4%

Tabela 19– Erros Percentuais para os Índices de Confiabilidade do Sistema – Caso 2

6.3.2.1 Análise do Desempenho Computacional

Assim como no primeiro caso, os índices de confiabilidade obtidos pelo modelo proposto estão bem próximos dos valores esperados, estando todos os índices dentro dos intervalos de confiança estabelecidos a partir da Simulação Monte Carlo Sequencial. A precisão do modelo pode ser ainda mais bem observada a partir dos erros percentuais apresentados na Tabela 19, que são ainda menores do que aqueles observados para o Caso 1.

Em relação ao esforço computacional do modelo, observa-se a mesma ordem de grandeza para o obtido para o Caso 1. O modelo proposto foi aproximadamente 23 vezes mais rápido que a Simulação Monte Carlo Sequencial, se manteve 1,3 vezes mais rápido do que a Simulação Pseudo-Sequencial, e 2,8 vezes mais rápido do que a Simulação Pseudo-Cronológica.

6.3.2.2 Análise da Confiabilidade

A comparação dos índices de confiabilidade obtidos pelo modelo proposto para este caso e para o Caso 1, no qual não se representou a variação da eficiência das usinas com a variação da direção do vento, é apresentada na Tabela 20.

<i>Caso</i>	<i>LOLP (pu)</i>	<i>EPNS (MW)</i>	<i>LOLF (oc/ano)</i>
Caso 1 – Sem Representação da Direção	0,0098	1,6275	16,88
Caso 2 – Com Representação da Direção	0,0101	1,7647	17,70

Tabela 20– Comparação dos Índices de Confiabilidade entre os Casos 1 e 2

Observa-se que houve uma ligeira piora na confiabilidade do sistema, com um aumento aproximado de 8% no índice EPNS, que passou de 1,6275 MW para 1,7647 MW, e de 5% no índice LOLF, que passou de 16,88 oc/ano para 17,70 oc/ano.

Este aumento nos índices já era esperado, uma vez que no Caso 2 a eficiência das usinas passou a ser representada, diminuindo-se a disponibilidade de geração das usinas eólicas.

Na Figura 58 é apresentada uma comparação entre as densidades de probabilidade do corte de carga para os Casos 1 e 2, considerando apenas os estados de corte de carga.

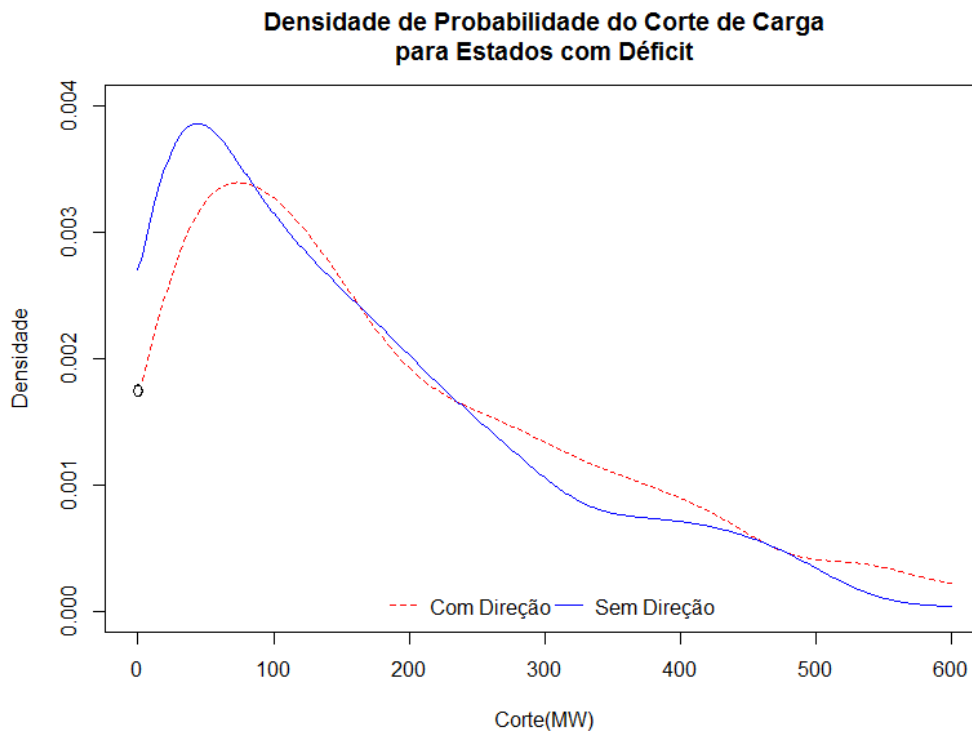


Figura 58- Densidade de Probabilidade do Corte de Carga para Estados com Corte de Carga – Casos 1 e 2

Observa-se que o efeito da inclusão da representação da variação da eficiência das usinas com a variação da direção do vento provocou um pequeno deslocamento da moda da distribuição para a direita, o que significa dizer que os valores mais prováveis de corte de carga agora são um pouco mais altos. Enquanto o valor mais provável de corte de carga era de aproximadamente 45 MW no Caso 1, passou a ser de aproximadamente 75 MW no caso 2, correspondente a um aumento de quase 66%.

Tem-se ainda, que a probabilidade do corte de carga ser inferior a 100MW é de aproximadamente 35% neste caso, contra a probabilidade de 44% que havia sido verificada no Caso 1.

Com isto, pode-se concluir que, embora em termos de valores médios os índices de confiabilidade tenham variado pouco em relação ao caso sem representação da direção do vento, o impacto no aumento da severidade dos cortes não é desprezível.

6.3.2.3 Análise das Amostras Sintéticas Geradas

Na Figura 59 são apresentadas as densidades de probabilidade marginais estimadas para algumas das séries de direção, sobrepondo-se a estimativa obtida a partir dos dados históricos e das amostras sintéticas. Pode-se observar que as densidades de probabilidade marginais obtidas são bem mais complexas do que as anteriormente obtidas para as variáveis de velocidade do vento. Para as variáveis de direção as densidades são multimodais, não sendo possível caracteriza-las por uma típica família de distribuição de probabilidade.

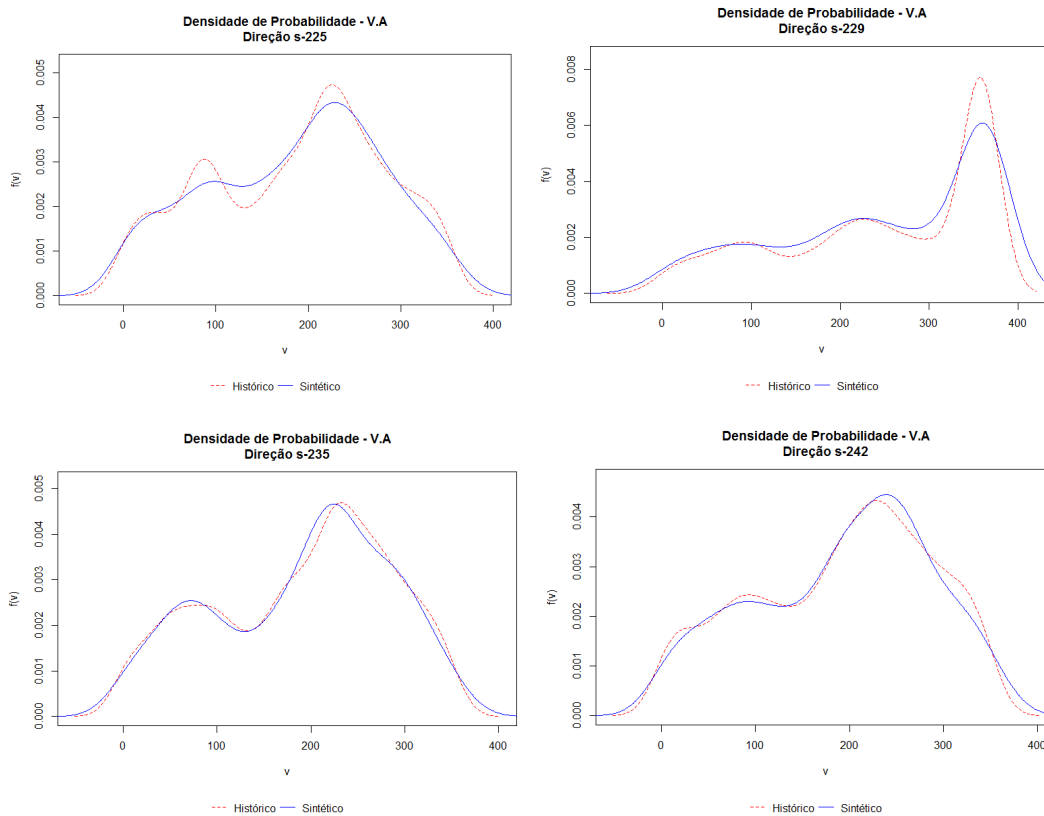


Figura 59– Densidades de Probabilidade para as Variáveis Aleatórias de Direção do Vento

Na Figura 60, é apresentada a comparação entre a densidade 2D estimada para o par de variáveis associadas a série de vento s-225, módulo e direção.

Observa-se que o aspecto da densidade conjunta das duas variáveis foi bem reproduzido pelo modelo proposto, mesmo a relação entre elas sendo bastante complexa.

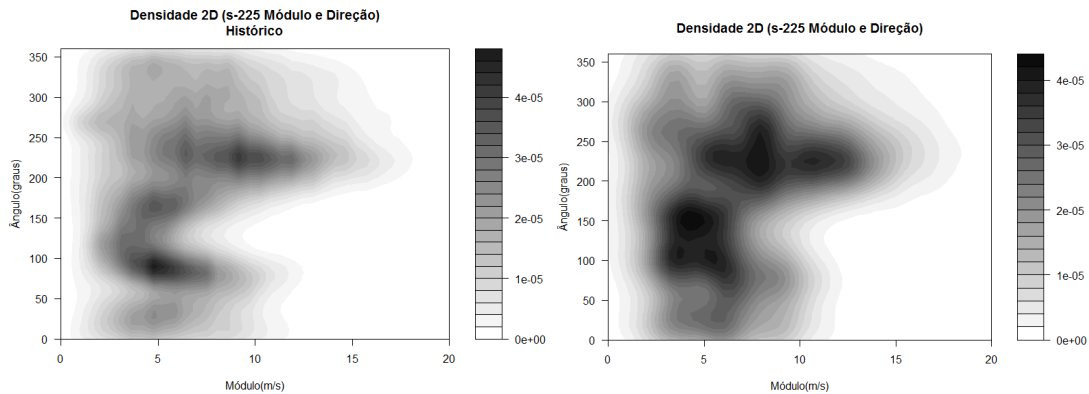


Figura 60– Densidade de Probabilidade Conjunta para a Série de Vento s-225 (Módulo e Direção)

Para ilustrar a dependência estatística entre as séries de direção de vento entre postos distintos, na Figura 61 é apresentada uma comparação entre a densidade de probabilidade conjunta estimada a partir dos dados históricos e das respectivas amostras sintéticas de direção do vento para os pares de séries: (s-210 e s-225), (s-225 e s-249) e (s-229 e s-254).

Nota-se que cada combinação descreve um padrão distinto de dependência. Na primeira densidade, entre as séries s-210 e s-225, nota-se um relacionamento bastante linear entre as direções de vento, enquanto que para a segunda densidade, entre as séries s-225 e s-249, não existe mais esta caracterização. Na última densidade, para as séries s-229 e s-254, pode-se observar basicamente duas regiões distintas. Uma região em que se tem um relacionamento bem linear entre estas variáveis, com um aspecto similar ao da primeira densidade apresentada, definido praticamente para a região do gráfico na qual s-229 é inferior a 300°, e uma segunda região na qual praticamente s-229 e s-254 são independentes, na região para qual s-229 é superior a 300°.

Independente do formato definido pela densidade de probabilidade conjunta das variáveis, qualitativamente se observa que o modelo proposto conseguiu capturar e reproduzir bem o comportamento das séries históricas.

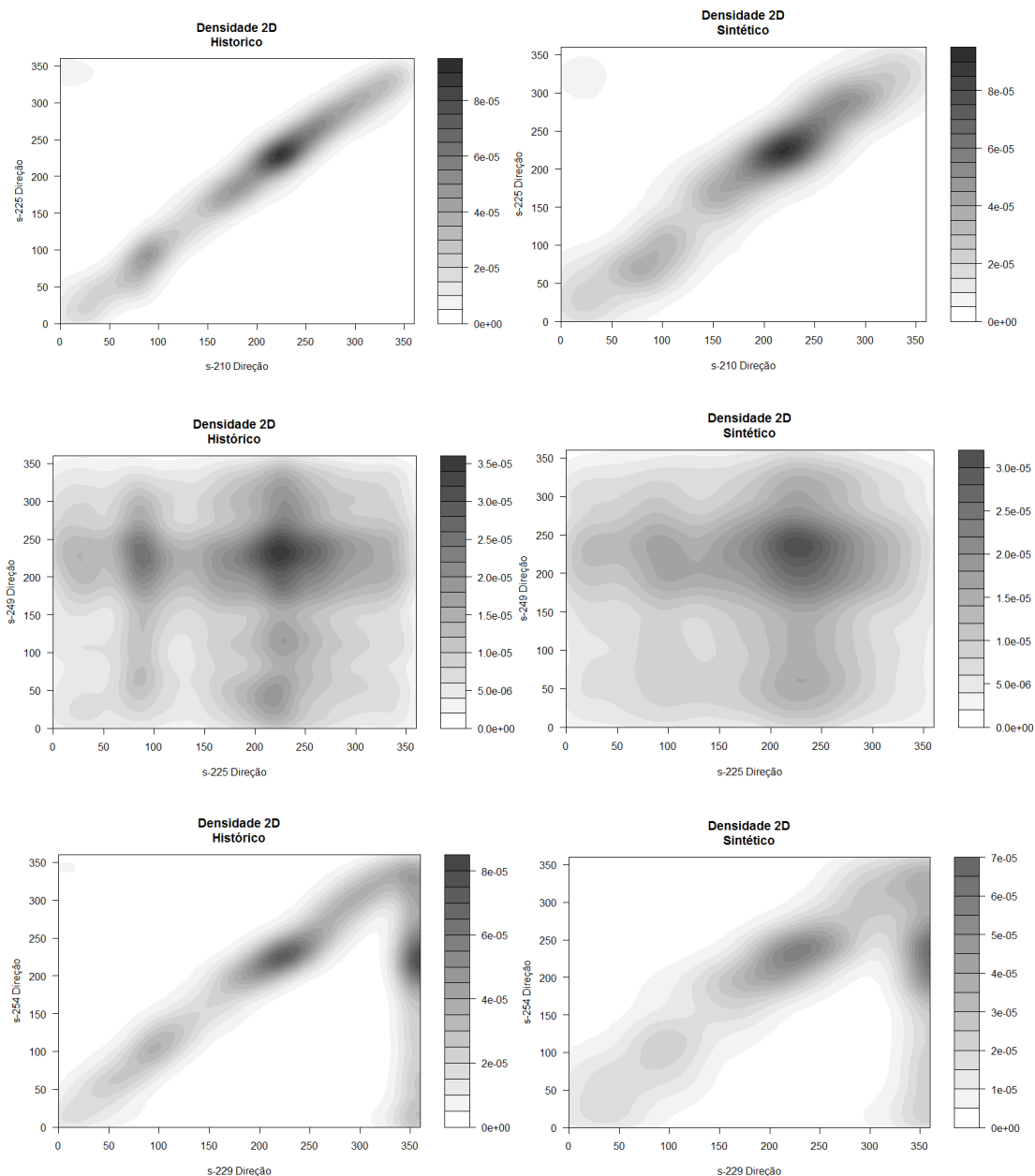


Figura 61– Densidade de Probabilidade Conjunta para Séries de Direção

6.3.3 Caso 3 - RTS + 27 Usinas Eólicas – Variável Climática Exógena

Este terceiro caso também corresponde a uma extensão do primeiro caso avaliado, onde, ao invés de considerar as séries de direção de vento, é considerada a representação de uma variável climática no modelo, que é exógena aos dados de vento, informando tanto os registros históricos da variável quanto a previsão do índice para o ano corrente, de modo a se verificar o impacto da inclusão desta informação nas amostras geradas e, consequentemente, nos resultados do modelo.

O objetivo deste caso é apresentar a flexibilidade do modelo proposto para incluir informações exógenas aos dados das séries históricas, uma vez que este tipo de abordagem não é possível através das abordagens mais tradicionais.

O índice arbitrado é o *Oceanic Niño Index* (ONI)[69], que é um indicador que mensura a intensidade do efeito do El Niño a partir da diferença entre a temperatura média em um mês do Oceano Pacífico, medida em um ponto de referência, e a temperatura esperada, dada pela média mensal histórica. A escolha particular por este índice foi motivada pelo fato de que os centros de monitoramento de meteorologia tem apontado indícios de manifestação do El Niño para o período compreendido entre o biênio, 2014 e 2015. Desta maneira, neste estudo de caso foi proposta a avaliação da hipótese do efeito esperado para o El Niño possuir o mesmo grau de intensidade observado para os anos de 1997 e 1998, período em que se mediu a maior intensidade do fenômeno, desde que este começou a ser monitorado.

Na Tabela 21 são apresentados os índices mensais medidos para o *ONI*[69] a partir do ano de 1990, destacando-se o período citado, 1997 e 1998, no qual se vê o quanto o índice se mantém bem superior aos valores normais.

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1990	0.1	0.2	0.3	0.3	0.2	0.2	0.3	0.3	0.4	0.3	0.4	0.4
1991	0.3	0.2	0.2	0.3	0.5	0.7	0.8	0.7	0.7	0.8	1.2	1.4
1992	1.6	1.5	1.4	1.2	1.0	0.7	0.3	0.0	-0.2	-0.3	-0.2	0.0
1993	0.2	0.3	0.5	0.6	0.6	0.5	0.3	0.2	0.2	0.2	0.1	0.1
1994	0.1	0.1	0.2	0.3	0.4	0.4	0.4	0.4	0.5	0.7	1.0	1.2
1995	1.0	0.8	0.6	0.3	0.2	0.0	-0.2	-0.4	-0.7	-0.8	-0.9	-0.9
1996	-0.9	-0.8	-0.6	-0.4	-0.3	-0.2	-0.2	-0.3	-0.3	-0.3	-0.4	-0.5
1997	-0.5	-0.4	-0.1	0.2	0.7	1.2	1.5	1.8	2.1	2.3	2.4	2.3
1998	2.2	1.8	1.4	0.9	0.4	-0.2	-0.7	-1.0	-1.2	-1.3	-1.4	-1.5
1999	-1.5	-1.3	-1.0	-0.9	-0.9	-1.0	-1.0	-1.1	-1.1	-1.3	-1.5	-1.7
2000	-1.7	-1.5	-1.2	-0.9	-0.8	-0.7	-0.6	-0.5	-0.6	-0.6	-0.8	-0.8
2001	-0.7	-0.6	-0.5	-0.4	-0.2	-0.1	0.0	0.0	-0.1	-0.2	-0.3	-0.3
2002	-0.2	0.0	0.1	0.3	0.5	0.7	0.8	0.8	0.9	1.2	1.3	1.3
2003	1.1	0.8	0.4	0.0	-0.2	-0.1	0.2	0.4	0.4	0.4	0.4	0.3
2004	0.3	0.2	0.1	0.1	0.2	0.3	0.5	0.7	0.8	0.7	0.7	0.7
2005	0.6	0.4	0.3	0.3	0.3	0.3	0.2	0.1	0.0	-0.2	-0.5	-0.8
2006	-0.9	-0.7	-0.5	-0.3	0.0	0.1	0.2	0.3	0.5	0.8	1.0	1.0
2007	0.7	0.3	-0.1	-0.2	-0.3	-0.3	-0.4	-0.6	-0.8	-1.1	-1.2	-1.4
2008	-1.5	-1.5	-1.2	-0.9	-0.7	-0.5	-0.3	-0.2	-0.1	-0.2	-0.5	-0.7
2009	-0.8	-0.7	-0.5	-0.2	0.2	0.4	0.5	0.6	0.8	1.1	1.4	1.6
2010	1.6	1.3	1.0	0.6	0.1	-0.4	-0.9	-1.2	-1.4	-1.5	-1.5	-1.5
2011	-1.4	-1.2	-0.9	-0.6	-0.3	-0.2	-0.2	-0.4	-0.6	-0.8	-1.0	-1.0
2012	-0.9	-0.6	-0.5	-0.3	-0.2	0.0	0.1	0.4	0.5	0.6	0.2	-0.3
2013	-0.6	-0.6	-0.4	-0.2	-0.2	-0.3	-0.3	-0.3	-0.3	-0.2	-0.3	-0.4

Tabela 21– Medições para o *Oceanic Niño Index* (ONI)

Para a modelagem da hipótese sugerida, considera-se que os índices mensais serão exatamente os medidos para este período, conforme apresentado na Tabela 22.

<i>Jan</i>	<i>Fev</i>	<i>Mar</i>	<i>Mai</i>	<i>Abr</i>	<i>Jun</i>	<i>Jul</i>	<i>Ago</i>	<i>Set</i>	<i>Out</i>	<i>Nov</i>	<i>Dez</i>
2.2	1.8	1.4	0.9	0.7	1.2	1.5	1.8	2.1	2.3	2.4	2.3

Tabela 22– Índices Considerados para o *Oceanic Niño Index (ONI)*

Desta forma, o índice ONI é modelado como uma variável discreta de doze estados, correspondente a cada um dos meses do ano, e cada estado tem probabilidade de ocorrência dada pela razão entre a duração do respectivo mês e a duração total do ano.

Como o tempo necessário para esta variável transitar de um estado para o outro, que corresponde a mudança de mês, é muito maior do que o tempo que leva para qualquer um dos elementos variantes no tempo transitar, o efeito da transição desta variável foi desconsiderada no modelo de transição de estados.

Os índices de confiabilidade deste caso são apresentados na Tabela 23.

<i>Modelo SMC</i>	<i>LOLP</i> (<i>pu</i>)	<i>EPNS</i> (<i>MW</i>)	<i>LOLF</i> (<i>oc/ano</i>)	<i>Tempo</i>
Modelo Proposto	0,0084 (0,0224)	1,3712 (0,0300)	15,51 (0,0272)	39min:27sec (378617 casos)

Tabela 23– Índices de Confiabilidade para o Sistema

Observa-se, a partir da comparação entre o resultado deste caso e o resultado do primeiro caso, que a introdução da representação da variável climática no modelo modificou substancialmente os índices de confiabilidade obtidos. A *LOLP* caiu de 0,98% para 0,84%, a *EPNS* caiu de 1,6275 MW para 1,3712 MW e a *LOLF* de 17,22 oc/ano para 15,51. Isto sugere que de fato o processo estocástico dos ventos representados são afetados pelo fenômeno El Niño, do contrário, não teriam sido observadas diferenças tão significativas nos índices obtidos.

Verifica-se ainda que a modificação na dinâmica dos ventos foi positiva para a confiabilidade do sistema, o que impactou inclusive aumentando bastante o número de estados analisados até a convergência do processo.

Na Figura 62 são apresentadas comparações entre as funções de densidades de probabilidade estimadas para os casos sem representação da variável climática, linha tracejada, e com representação da variável climática, linha cheia, para algumas das séries de velocidade de vento. O principal efeito observado é que a introdução da

variável climática ocasionou uma pequena modificação nas distribuições, deslocando a moda um pouco mais para a direita. Na prática, significa que neste caso existe uma probabilidade um pouco maior de se ter ventos maiores do que no caso anterior. O que justifica a queda nos índices de confiabilidade.

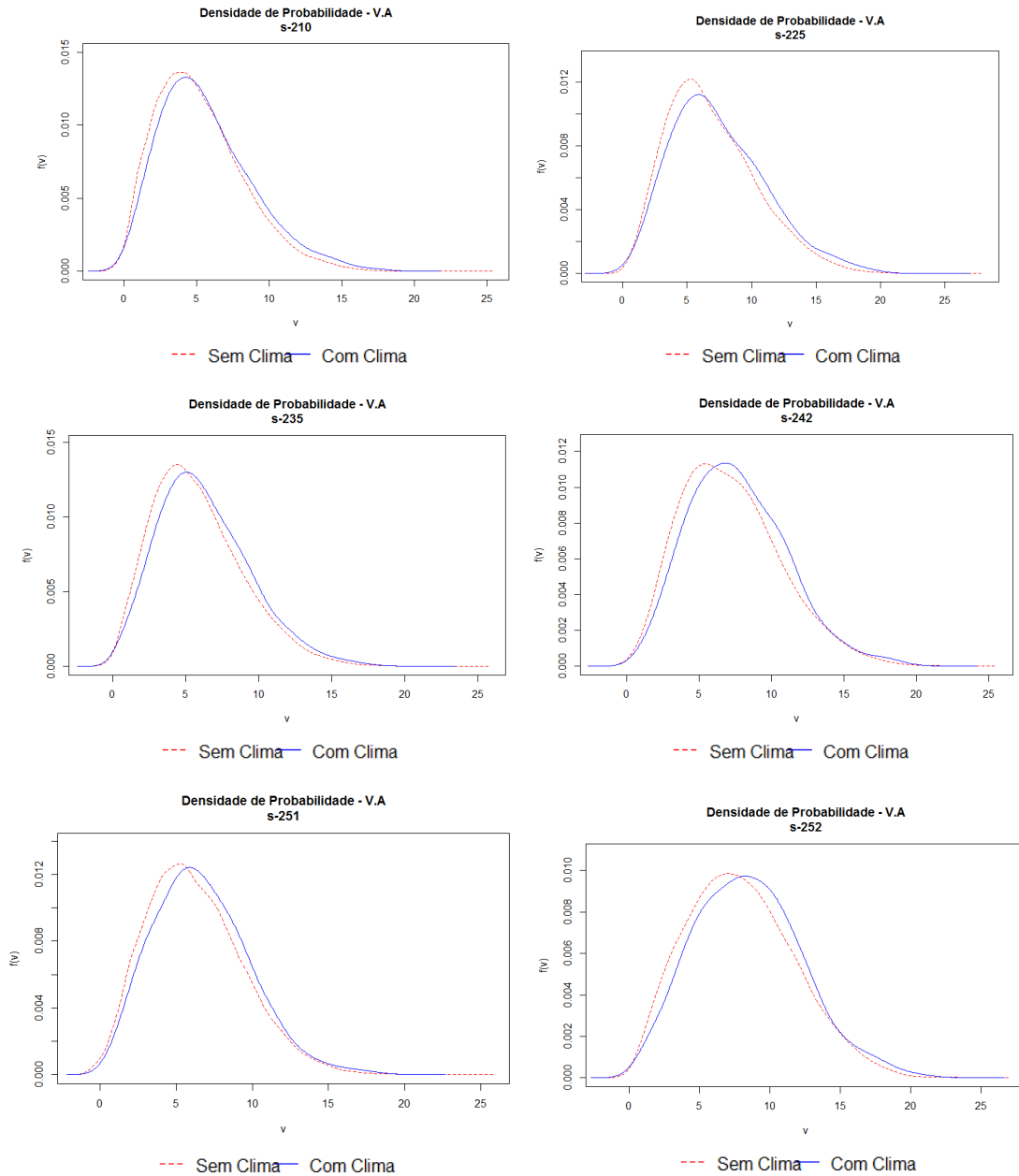


Figura 62– Densidade de Probabilidade para as Variáveis Aleatórias de Vento

6.3.4 Caso 4 – Sistema Brasileiro

Este último caso corresponde à representação do sistema real brasileiro, incluindo a representação completa da rede de transmissão e das usinas geradoras individualizadas. Os objetivos da análise deste caso são avaliar a aplicabilidade do modelo em um caso de grande porte, e principalmente, explorar a representação da correlação entre a geração eólica e as vazões afluentes em um sistema que é predominantemente hidrelétrico.

Na Figura 63, extraído de [70], é apresentado um mapa da visão geral do sistema brasileiro.

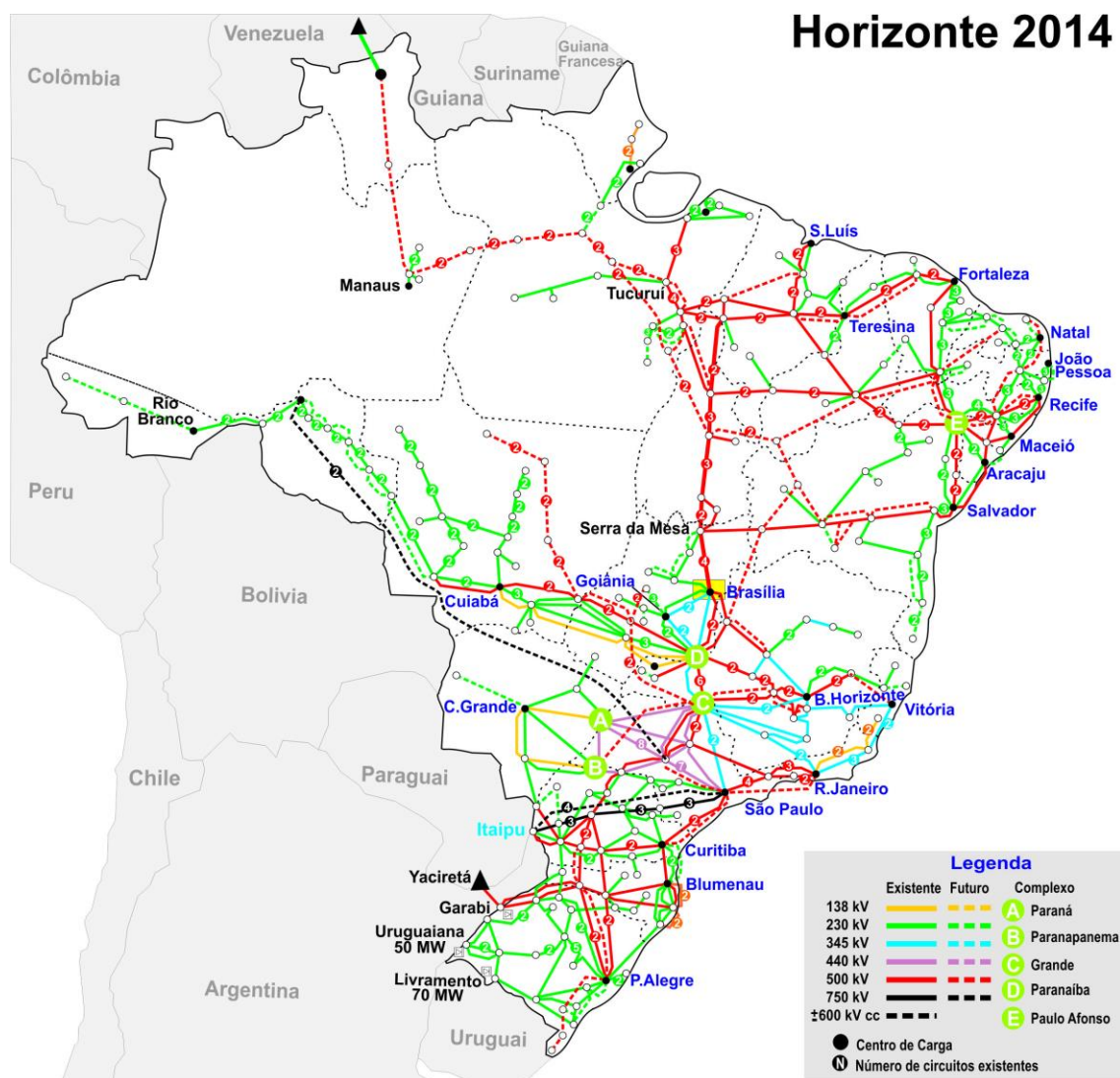


Figura 63– Visão Geral do Sistema de Transmissão do Sistema Brasileiro

A base de dados utilizada foi consolidada a partir da combinação de uma base de dados energética do sistema brasileiro com uma base de dados elétrica. A base de dados energética corresponde a uma modificação da base original no formato do modelo NEWAVE[71], disponibilizada pelo Operador Nacional do Sistema brasileiro, ONS[70], de modo a representar maiores detalhes da topologia das usinas hidrelétricas no sistema brasileiro. A base elétrica corresponde ao caso de referência para estudos de regime permanente, no formato do modelo ANAREDE[71], disponibilizado também pelo ONS. A integração destas duas bases resultou em uma nova base no formato do modelo SDDP[72], sendo esta diretamente processada pelo simulador desenvolvido.

As características básicas do sistema brasileiro representadas neste estudo são apresentadas na Tabela 24.

É importante ressaltar que os números apresentados, principalmente para as usinas, não corresponde exatamente ao número de usinas físicas do sistema brasileiro, pois algumas usinas são desmembradas em várias usinas fictícias para permitir uma melhor representação de aspectos elétricos, como a conexão das usinas a diferentes barras no sistema, e aspectos energéticos, como a topologia hidrológica das usinas.

<i>Característica</i>	
Número de Subsistemas	5
Número de Barras	6177
Número de Circuitos	9192
Número de Usinas Térmicas	144
Número de Usinas Hidrelétricas	160
Número de Usinas Eólicas Representadas	63
Número de Séries Hidrológicas	155
Número de Séries Eólicas Horárias	5

Tabela 24– Características Básicas da Representação do Sistema Brasileiro

Na Tabela 25 são apresentadas as usinas eólicas do sistema brasileiro, com suas respectivas localizações e potências nominais. A capacidade total instalada destas usinas é de 3106,4 MW.

<i>Estado</i>	<i>Usina</i>	<i>Potência Nominal (MW)</i>	<i>Estado</i>	<i>Usina</i>	<i>Potência Nominal (MW)</i>
BA	IGAPORÃ ALVORADA	38,4	CE	TRAIRÍ	25,4
BA	IGAPORÃ GUIRAPÁ	52,8	CE	VOLTA DO RIO	42,0
BA	IGAPORÃ LICÍCIO DE ALMEIDA	73,6	RN	ALEGRIA I	51,0
BA	IGAPORÃ NOSSA SENHORA DA CONCEIÇÃO	76,8	RN	ALEGRIA II	100,7
BA	IGAPORÃ PLANALTINA	52,8	RN	AREIA BRANCA	27,3
BA	MACAÚBAS	35,1	RN	MANGUE SECO 1	26,0
BA	NOVO HORIZONTE	30,1	RN	MANGUE SECO 2	26,0
BA	PEDRA BRANCA	30,1	RN	MANGUE SECO 3	26,0
BA	SÃO PEDRO DO LAGO	30,0	RN	MANGUE SECO 5	26,0
BA	SEABRA	30,1	RN	MAR E TERRA	23,1
BA	SETE GAMELEIRAS	30,0	RN	MIASSABA 3	68,5
CE	BONS VENTOS	50,0	RN	MORRO DOS VENTOS	145,2
CE	CANOA QUEBRADA	57,0	RN	REI DOS VENTOS 1	58,5
CE	CEARÁ II	60,0	RN	REI DOS VENTOS 3	60,1
CE	COLÔNIA	18,9	RN	RIO DO FOGO	49,0
CE	EMBUACA	27,3	RN	SANTA CLARA	187,2
CE	ENACEL	31,5	RN	UNIÃO DOS VENTOS	169,6
CE	FAÍSA I	29,4	RS	ATLÂNTICA	120,0
CE	FAÍSA II	27,3	RS	CERRO CHATO I	30,0
CE	FAÍSA III	25,2	RS	CERRO CHATO II	30,0
CE	FAÍSA IV	25,2	RS	CERRO CHATO III	30,0
CE	FLEIXEIRAS 1	30,0	RS	CIDREIRA I	70,0
CE	GUAJIRÚ	30,0	RS	ÍNDIOS	50,0
CE	IACARAÍ I	27,3	RS	LIVRAMENTO 2	30,0
CE	IACARAÍ II	37,8	RS	OSÓRIO	50,0
CE	ICARAZINHO	54,6	RS	OSÓRIO 2	24,0
CE	MUNDAÚ	30,0	RS	OSÓRIO 3	26,0
CE	PRAIA DO MORGADO	28,8	RS	SANGRADOURO	50,0
CE	PRAIA FORMOSA	105,0	RS	SANGRADOURO 2	26,0
CE	TAÍBA ÁGUA	23,1	RS	SANGRADOURO 3	24,0
CE	TAÍBA ANDORINHA	14,7	SC	ÁGUA DOCE	129,0

Tabela 25– Usinas Eólicas do Sistema Brasileiro

Para este caso não foi considerada uma curva horária de carga, e sim a representação da demanda proveniente da base energética. Esta corresponde ao agrupamento por mês dos três patamares de carga mais significativos, incluindo em especial a representação da demanda de ponta do sistema, com a consequente distribuição desta demanda, através de fatores de participação pelas barras do sistema.

A demanda em energia do sistema para o ano de 2015 é apresentada, de forma agregada por mês, na Figura 64.

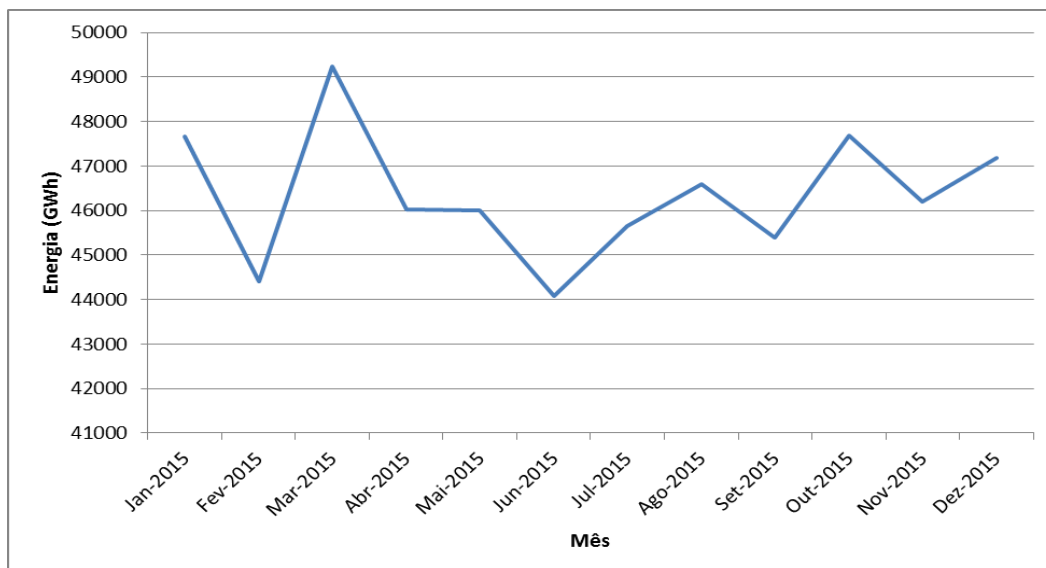


Figura 64– Demanda do Sistema Brasileiro para o Ano de 2015

Na Figura 65 são apresentadas as barras do sistema brasileiro com os maiores fatores de participação em relação à demanda total do sistema. Destaca-se que todas as barras pertencem ao subsistema sudeste.

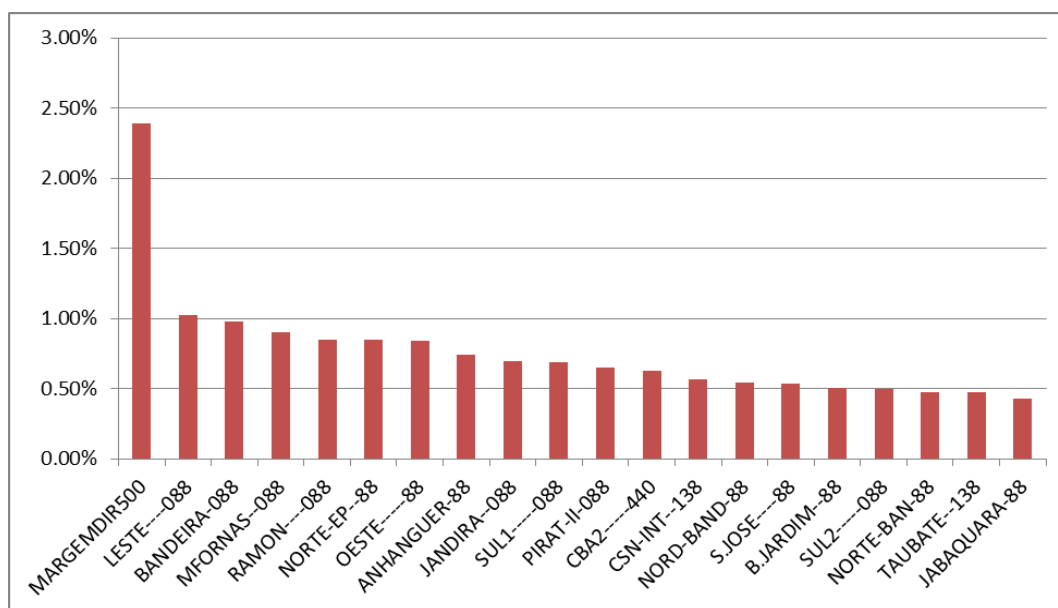


Figura 65– Barras do Sistema Brasileiro com Maiores Fatores de Participação

Ressalta-se que por indisponibilidade dos dados de taxa de falha e duração média das falhas das unidades geradoras, uma vez que a base original do NEWAVE, de onde foram extraídos os dados de geração, contempla apenas a informação consolidada de

probabilidade de falha dos equipamentos, não foi possível a obtenção dos índices baseados em frequência e duração do sistema.

Na Figura 66 são apresentadas as séries de gerações eólicas mensais para ao sistema do nordeste e as séries de energia afluyente (ENA) dos sistemas sudeste e sul para um período contínuo de cinco anos, tendo sido as séries normalizadas para permitir a compatibilização das escalas. Estas duas séries de energia afluyente foram escolhidas porque apresentam respectivamente as maiores correlações negativas e positivas com a série de geração eólica.

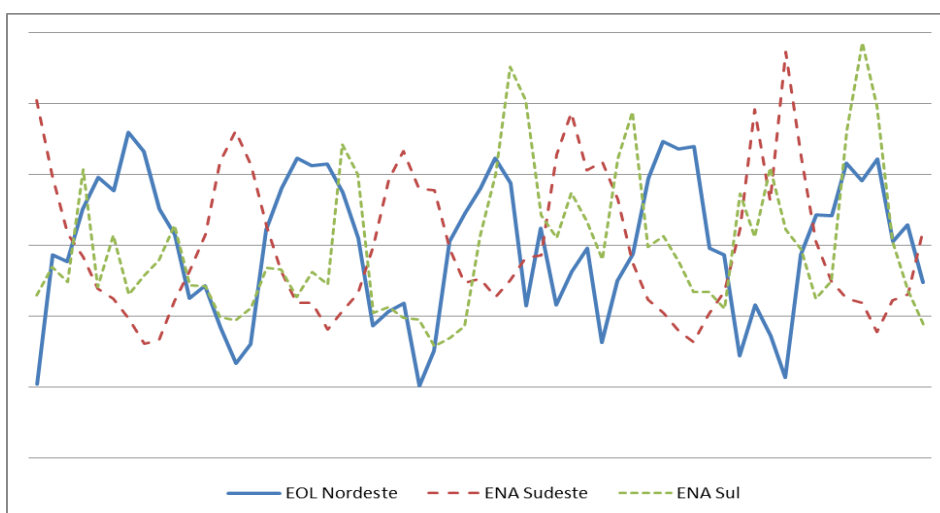


Figura 66– Séries Normalizadas da Eólica do Nordeste e das ENAs do Sudeste e do Sul

O efeito de complementariedade que existe entre a geração eólica do nordeste e a energia afluyente do sudeste fica bem evidente neste gráfico. Observa-se que os períodos úmidos do sudeste, relacionados às maiores energias afluentes, correspondem aos períodos de menor geração eólica. Assim como os períodos secos, relacionados às menores energias afluentes, correspondem aos períodos de maior geração eólica.

Já a comparação da geração eólica do nordeste com a energia afluyente do sul revela um padrão mais similar entre as séries, coincidindo mais os períodos de maior geração eólica com os períodos de maior energia afluyente, embora esta relação não seja constante ao longo do horizonte.

Na Tabela 26 são apresentadas as correlações entre as séries de gerações eólicas mensais e as energias afluentes para os subsistemas do sistema brasileiro, com o objetivo de possibilitar uma ideia geral do relacionamento entre elas.

É possível observar algumas relações bem significativas entre as energias afluentes e a geração eólica das regiões. Em relação às observações anteriormente

realizadas, a complementariedade entre a geração eólica do nordeste e a energia afluenta do sudeste é confirmada pela correlação observada de -0,81, enquanto que a correlação positiva de 0,237 também respalda a observação a respeito do relacionamento entre as séries de geração eólica do nordeste e da energia afluenta do sul.

Também podem ser observados outros sistemas com alta complementariedade entre a geração eólica e a energia afluenta, como entre a geração eólica do nordeste e as energias afluentes do próprio nordeste.

<i>Correlações Geração Eólica x Energia Afluenta</i>				
	ENA Sudeste	ENA Sul	ENA Nordeste	ENA Norte
EOL Sul	-0,173	0,225	-0,068	-0,351
EOL Nordeste	-0,810	0,237	-0,787	-0,443

Tabela 26– Correlações entre a Geração Eólica e a Energia Afluenta por Sistema

Preliminarmente à avaliação da confiabilidade, foi obtida uma política de operação para o sistema para o ano de 2015 para 300 cenários de vazões afluentes, a partir do modelo SDDP, com o objetivo de estabelecer as restrições de operação para as usinas hidrelétricas para o horizonte do estudo.

A simulação da operação do sistema resultou nos riscos de déficit apresentados na Tabela 27, indicando haver um indicio de problema de suprimento de energia no sistema brasileiro, independentemente dos fatores associados a confiabilidade do sistema.

<i>Riscos de Déficit</i>				
	Sistema Sudeste	Sistema Sul	Sistema Nordeste	Sistema Norte
Risco (%)	11%	13%	3%	3%

Tabela 27– Riscos de Déficit para o Sistema Brasileiro

Para expurgar o efeito do problema de suprimento de energia da avaliação da confiabilidade, foram eliminadas do estudo as séries de operação para as quais foram verificados os déficits de energia, sendo as demais séries de afluências e operação do sistema consideradas na Simulação Monte Carlo para avaliação da confiabilidade.

A partir da política obtida, foi simulado inicialmente um caso de referência para o sistema brasileiro, sem a representação da geração eólica. Foi utilizada a configuração do sistema prevista para o ano de 2015.

Estes índices de confiabilidade de referência são apresentados na Tabela 28.

<i>Modelo SMC</i>	<i>LOLP (pu)</i>	<i>EPNS (MW)</i>	<i>EENS (GWh)</i>	<i>Tempo</i>
Modelo Proposto	0,0357	4,9875	43,69	2h: 44min (16783 casos)

Tabela 28– Índices de Confiabilidade para o Sistema Brasileiro, sem Eólicas

Na Figura 67 são apresentadas as barras com maiores probabilidades de corte de carga, sendo a primeira da lista a barra IVAIP---I-69, pertencente ao sistema Sul, que apresenta 0,28% de probabilidade de corte.

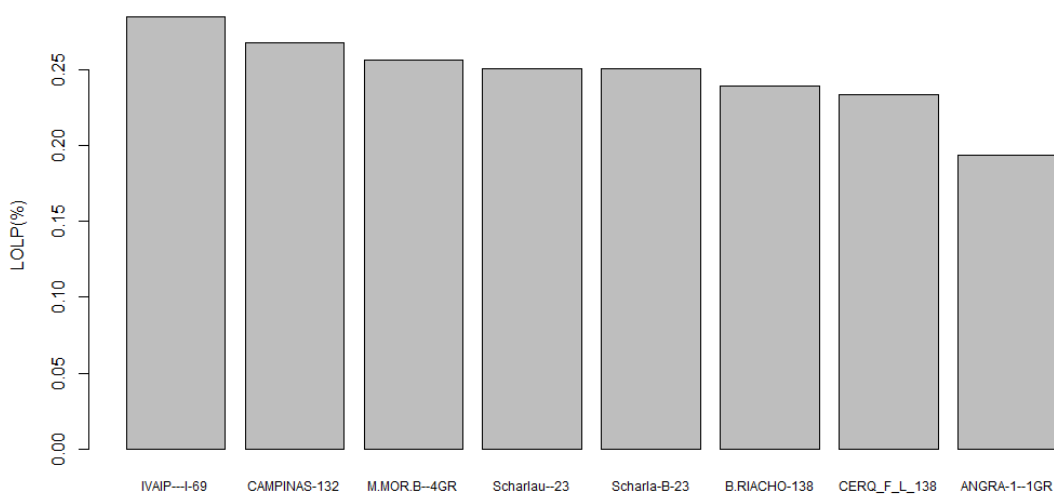


Figura 67– Barras com Maiores Probabilidades de Perda de Carga.

Na Figura 68 são apresentadas as barras com maiores valores esperados de potência não suprida, sendo a barra NORTE-BAN-88, pertencente ao sistema Sudeste, a que apresenta o maior valor esperado, de 0,341 MW, correspondente a aproximadamente 7% do total de potência não suprida.

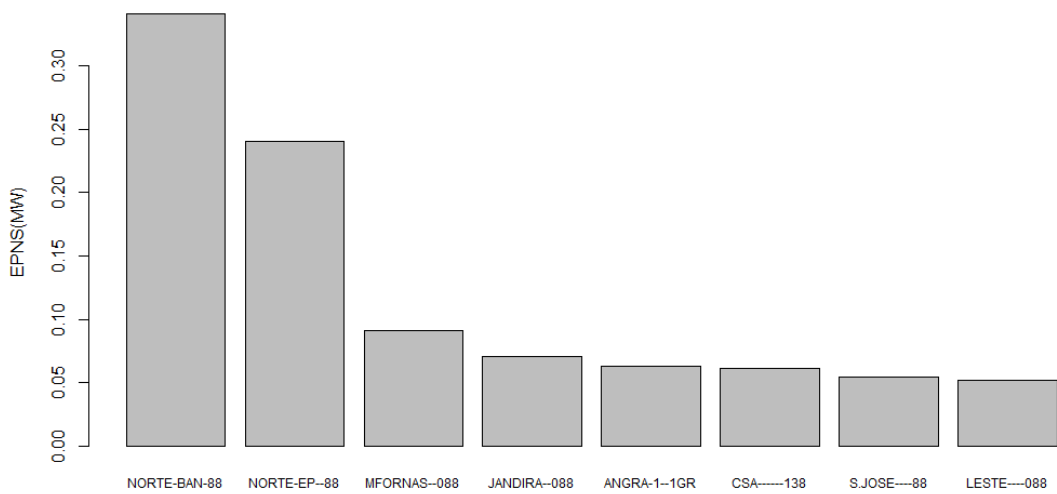


Figura 68– Barras com Maiores EPNS.

Na Figura 69 é apresentada a densidade de probabilidade do corte de carga do sistema brasileiro, considerando-se apenas os estados de déficit do sistema.

Observa-se que o sistema brasileiro possui uma maior concentração de cortes de carga até o valor de 200 MW, porém apresentando também outros níveis de corte bem mais severos, que são, entretanto, muito raros.

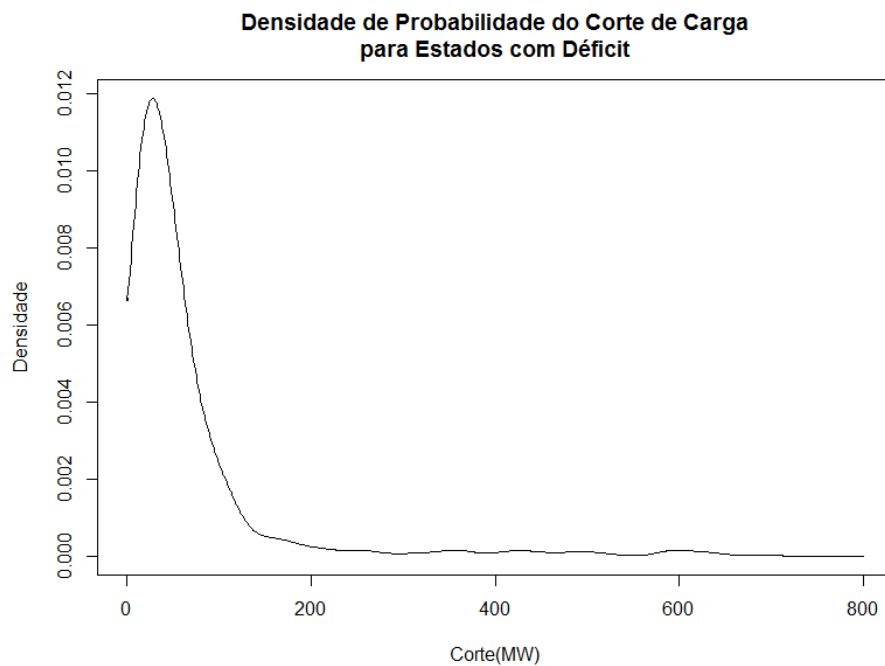


Figura 69– Densidade de Probabilidade do Corte de Carga para o Sistema Brasileiro

Este fato pode ser mais bem verificado a partir da distribuição de probabilidade acumulada do corte de carga, apresentada na Figura 70, e com um maior detalhamento da escala de interesse na Figura 71.

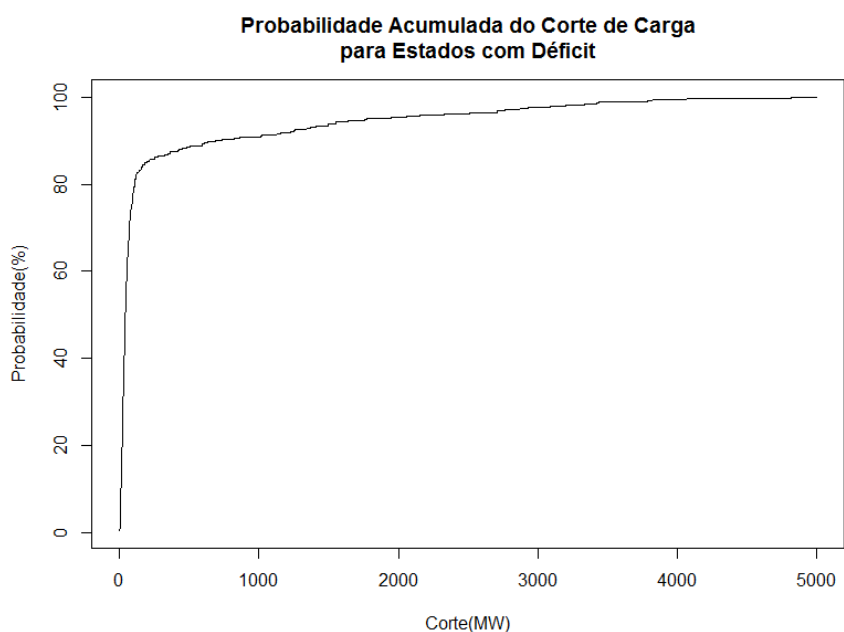


Figura 70– Probabilidade Acumulada do Corte de Carga para Estados com Déficit.

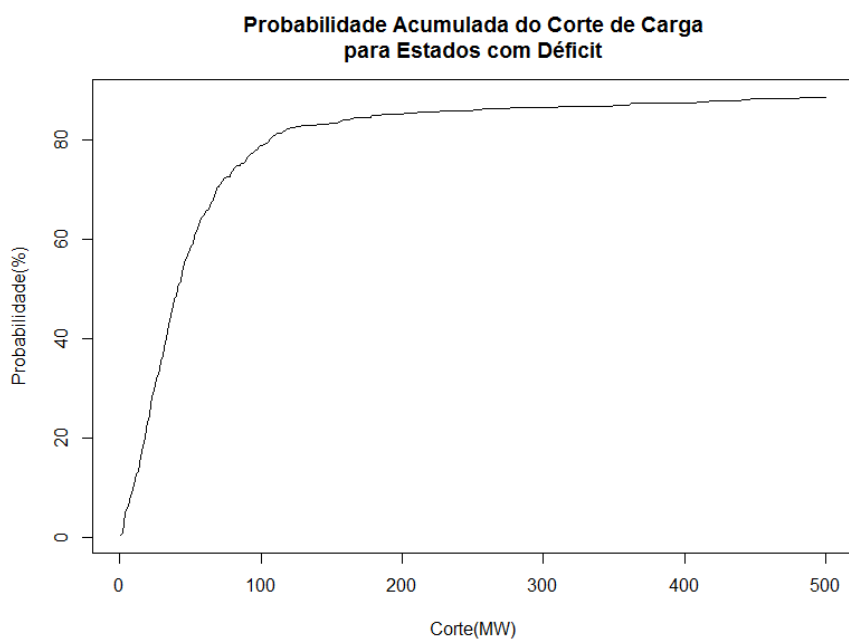


Figura 71– Probabilidade Acumulada do Corte de Carga para Estados com Déficit – Escala até 500 MW

A função de probabilidade acumulada estimada mostra que a probabilidade de haver um corte de até 100 MW é de quase 80%, sendo que o crescimento lento da função a partir deste ponto indica baixa probabilidade de se obter qualquer um dos estados de corte de carga superiores a este valor.

Para complementar esta análise, na Tabela 29 são apresentadas as probabilidades de se obter os níveis de corte de carga mais severos para o sistema brasileiro, sendo apresentadas tanto as probabilidades condicionais, isto é, considerando-se apenas o subconjunto de amostras dos estados de corte do sistema, como as probabilidades dos eventos de forma não condicionada, isto é, considerando-se todos os estados do sistema.

<i>Corte de Carga</i>	<i>Probabilidade Condicional</i>	<i>Probabilidade do Evento</i>
Superior a 1 GW	9,0%	0,003%
Superior a 2 GW	4,6%	0,002%
Superior a 3 GW	2,4%	Inferior a 0,001%
Superior a 4 GW	0,5%	Inferior a 0,001%
Superior a 5 GW	Inferior a 0,01%	Inferior a 0,00001%

Tabela 29– Probabilidades dos Estados Mais Severos de Corte de Carga do Sistema Brasileiro

Observa-se que embora a maior parte dos estados de falha do sistema seja caracterizada por cortes de carga de até 200 MW, existem alguns cenários bem mais severos, com cortes de carga na faixa de 1GW até 5GW. Estes cenários, no entanto, são bem raros, sendo a probabilidade de se ter um corte superior a 1GW de aproximadamente 0,003% apenas. Esta probabilidade decai à medida que se aumenta a severidade do corte de carga, sendo a probabilidade de se ter um corte superior a 5 GW praticamente desprezível.

Na Figura 72 é apresentada a distribuição dos cortes de carga médios por sistema. Observa-se que a maior parte dos cortes de carga está no sistema Sudeste, correspondente a 75% do total do valor médio do corte, o que é justificado pelo fato deste sistema concentrar a maior parte da carga do sistema brasileiro, seguido respectivamente pelos sistemas Sul, com 22%, Nordeste, com 2% e Norte, com 1%.

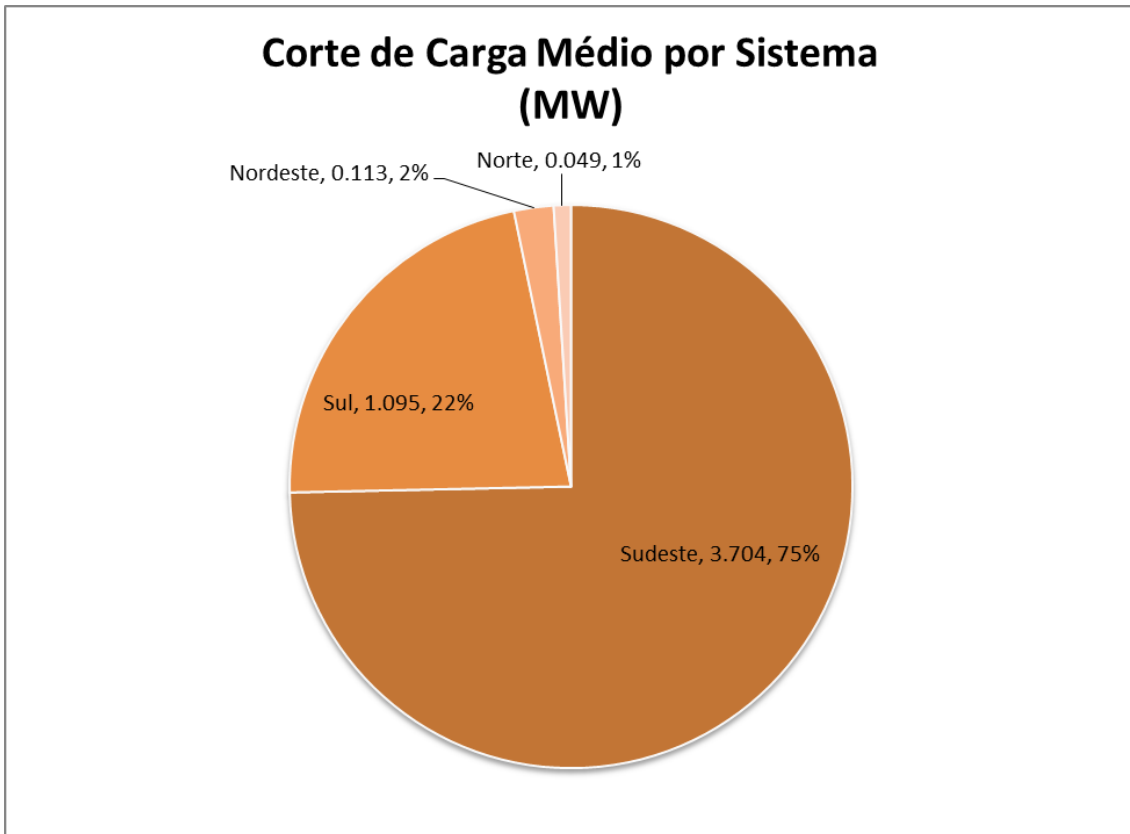


Figura 72– Distribuição de Corte de Carga Médio por Sistema – Sem Eólicas

Em um segundo passo, foram incorporadas ao sistema as usinas eólicas.

A base de dados disponível não contém dados históricos de geração individualizados para cada usina eólica, sendo a geração total representada de maneira agregada para os estados aos quais as usinas pertencem. Desta maneira, para o modelo de simulação a geração total de cada estado foi desagregada pelos respectivos geradores pertencentes, de modo a permitir a representação de geração eólica horária para de cada uma das usinas.

Os índices de confiabilidade foram obtidos a partir do modelo proposto de SMC Não Sequencial, sendo estes apresentados na Tabela 30.

<i>Modelo SMC</i>	<i>LOLP (pu)</i>	<i>EPNS (MW)</i>	<i>EENS (GWh)</i>	<i>Tempo</i>
Modelo Proposto	0,0313 (0,0378)	3,6585 (0,100)	32,04 (0,100)	2h: 58min (18566 casos)

Tabela 30– Índices de Confiabilidade para o Sistema Brasileiro, com Eólicas

Os índices obtidos mostram a melhora da confiabilidade do sistema com a entrada da geração eólica, tendo a LOLP caído de 3,57% para 3,13% e a EPNS caído de 4.9875 MW para 3,6585 MW, resultado na queda da EENS de 43,69 GWh para 32,04 GWh.

Na Figura 73 é apresentada a densidade de probabilidade do corte de carga do sistema brasileiro, considerando-se apenas os estados de corte de carga. Para permitir uma melhor comparação com o caso de referência, a densidade de probabilidade anterior foi inserida no mesmo gráfico, sendo a densidade dos cortes de carga para o caso com representação da geração eólica dada pela linha cheia, enquanto que a densidade do caso de referência é dada pela linha tracejada.

Observa-se que o efeito da inclusão da geração eólica é um deslocamento para o lado da esquerda bem pequeno da função, indicando que há uma pequena redução nos valores de corte de carga. A moda da função de densidade do caso de referência indica que o corte de carga mais provável é de aproximadamente 28 MW, enquanto que a moda da densidade do caso com representação da geração eólica indica um corte mais provável de aproximadamente 25 MW, equivalente a uma redução de aproximadamente 11% do valor, como pode ser visto na Figura 74, na qual estas funções de densidade são rerepresentadas em uma escala mais conveniente.

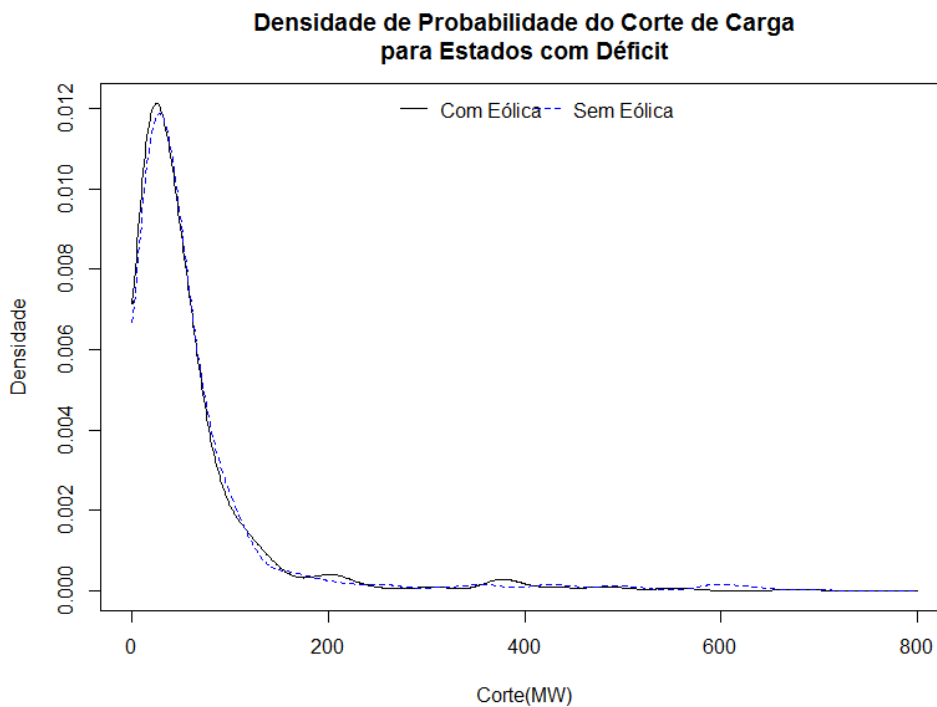


Figura 73– Densidade de Probabilidade do Corte de Carga para o Sistema Brasileiro – Com Eólica x Sem Eólica

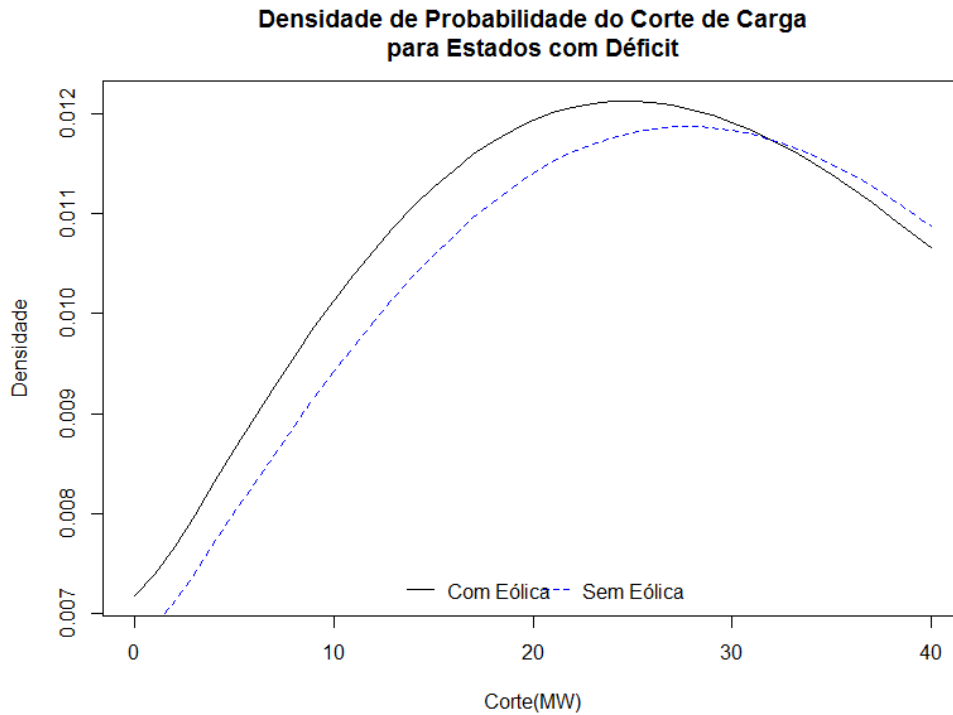


Figura 74– Densidade de Probabilidade do Corte de Carga para o Sistema Brasileiro – Com Eólica x Sem Eólica – Escala de Corte até 40 MW

O efeito do deslocamento das probabilidades de corte de carga pode ser mais bem observado a partir da comparação das funções de probabilidade acumulada dos dois casos, apresentadas na Figura 75, e com maior destaque na escala de interesse na Figura 76.

O deslocamento citado fica bem visível na escala que compreende os cortes de carga até 100 MW.

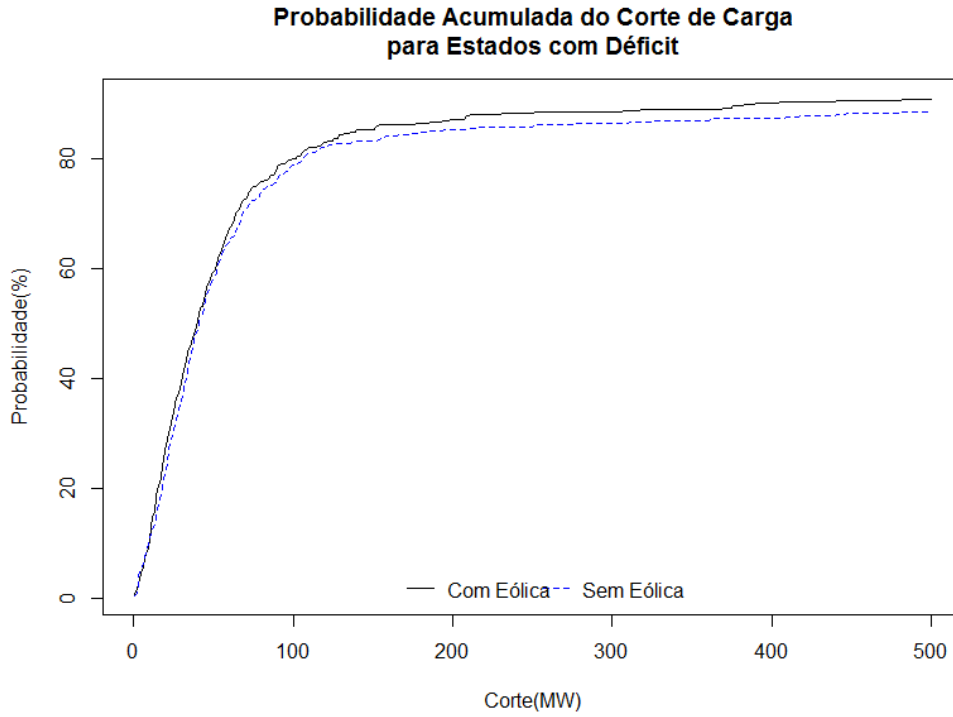


Figura 75– Probabilidade Acumulada do Corte de Carga para Estados com Déficit – Escala até 500 MW – Com Eólica x Sem Eólica

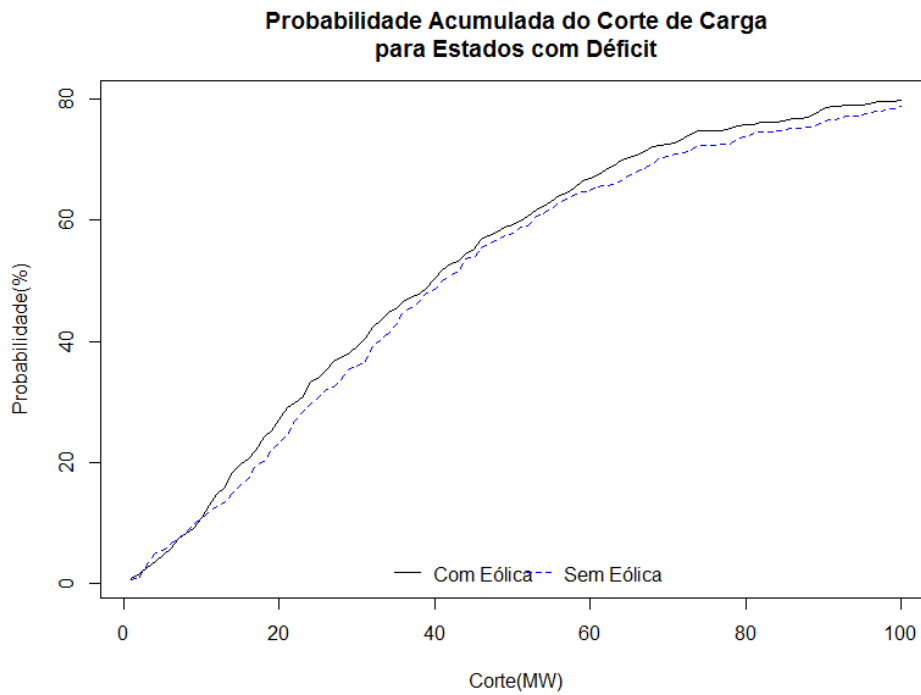


Figura 76– Probabilidade Acumulada do Corte de Carga para Estados com Déficit – Escala até 100 MW – Com Eólica x Sem Eólica

A severidade dos cortes dos dois casos pode ser também comparada, sendo apresentadas na Tabela 31 as probabilidades de se obter os níveis de corte de carga mais severos para o sistema brasileiro com a representação da geração eólica.

Os valores anteriormente apresentados para o caso sem representação da geração eólica são reapresentados para permitir uma melhor comparação entre estes resultados obtidos.

<i>Corte de Carga</i>	<i>Probabilidade Condicional (com eólica)</i>	<i>Probabilidade do Evento (com eólica)</i>	<i>Probabilidade Condicional (sem eólica)</i>	<i>Probabilidade do Evento (sem eólica)</i>
Superior a 1 GW	7,5%	0,002%	9,0%	0,003%
Superior a 2 GW	3,4%	0,001%	4,6%	0,002%
Superior a 3 GW	1,5%	Inferior a 0,001%	2,4%	Inferior a 0,001%
Superior a 4 GW	0,5%	Inferior a 0,001%	0,5%	Inferior a 0,001%
Superior a 5 GW	Inferior a 0,01%	Inferior a 0,00001%	Inferior a 0,01%	Inferior a 0,00001%

Tabela 31– Probabilidades dos Estados Mais Severos de Corte de Carga do Sistema Brasileiro – Com Eólica x Sem Eólica

Observa-se que houve uma queda nas probabilidades de déficit com os níveis de corte de carga mais severos. Esta queda, no entanto, não é tão significativa, principalmente quando se considera a probabilidade não condicional dos cortes de carga.

Na Figura 77 é apresentada a distribuição dos valores médios de corte de carga para os subsistemas do sistema brasileiro após a representação da geração eólica. Observa-se que em termos percentuais a distribuição obtida é bastante similar àquela obtida sem a representação da geração eólica, havendo apenas uma pequena diminuição da participação do sistema Sul, que passou de 22% para 21%, com o consequente aumento da participação do sistema Sudeste, que passou de 75% para 76%. Esta mudança pode ser explicada pelo fato do sistema Sul ser o segundo maior sistema em termos de capacidade instalada de geração eólica, e desta forma, tende a ser mais beneficiado com esta modalidade de geração que o Sudeste.

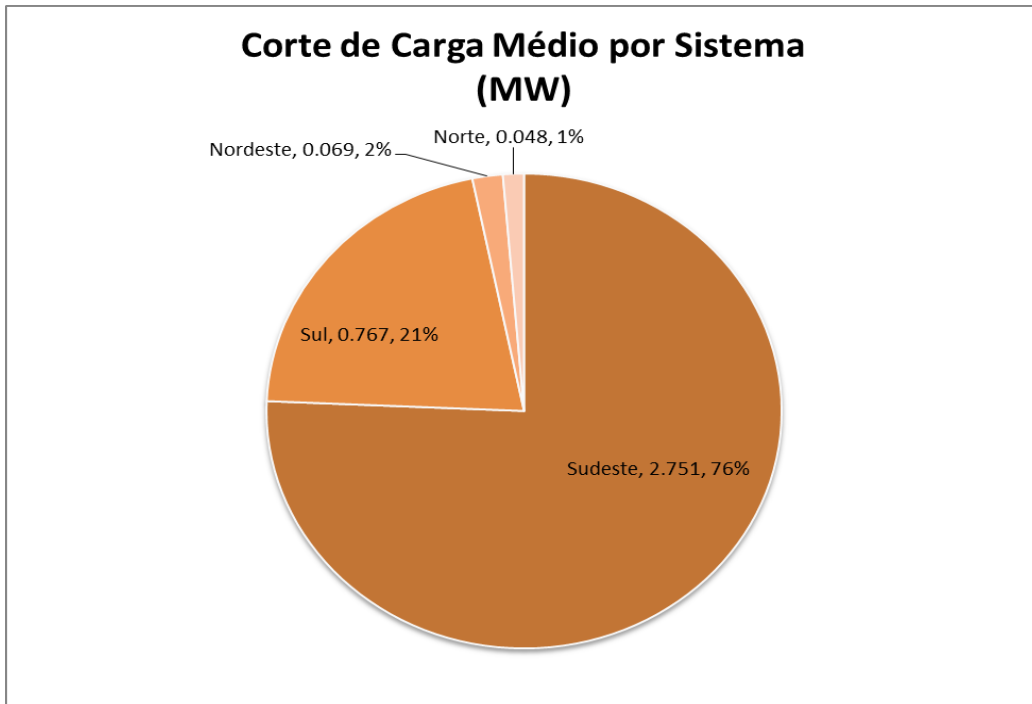


Figura 77– Distribuição de Corte de Carga Médio por Sistema – Com Eólicas

Na Figura 78 é apresentada uma comparação entre os valores médios de corte de carga para os subsistemas do sistema brasileiro, sem a representação da geração eólica e com a representação da geração eólica, sendo apresentadas na Figura 79 as reduções percentuais correspondentes a estes valores.

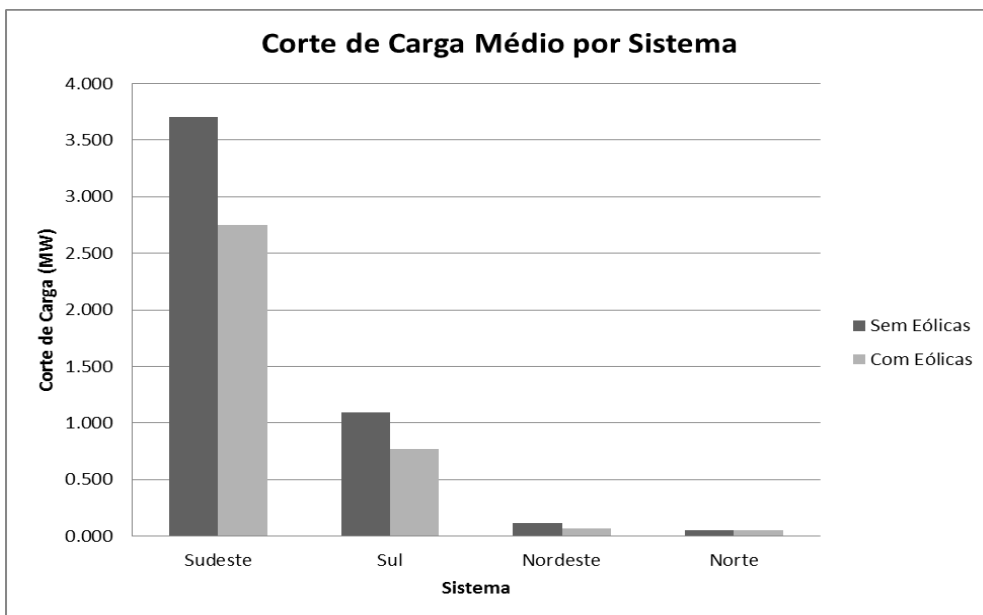


Figura 78– Comparação de Valores Médios de Corte de Carga

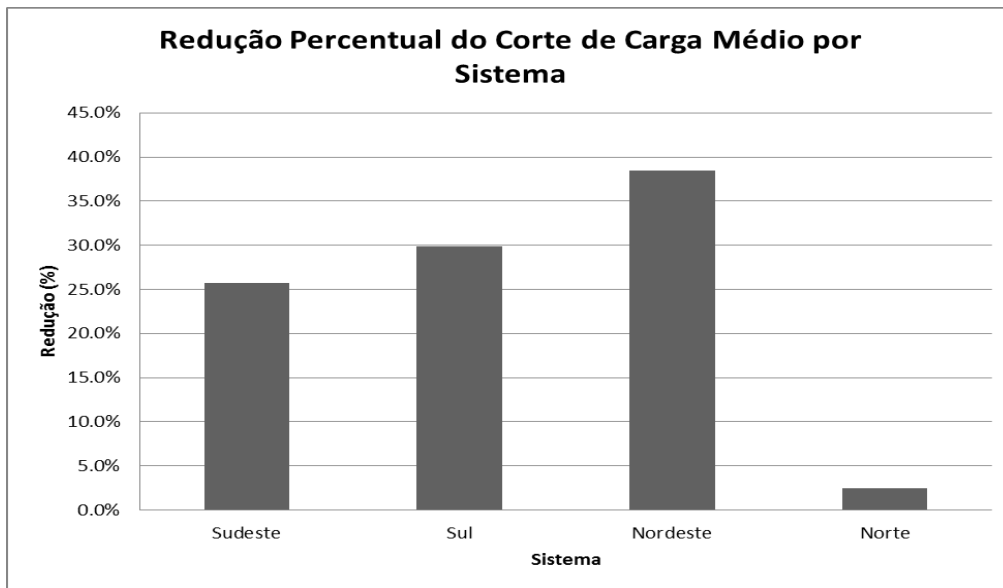


Figura 79– Redução Percentual do Corte de Carga Médio por Sistema

Estes gráficos mostram que a redução percentual dos valores médios de corte de carga não se dá de forma homogênea para todos os sistemas, sendo interessante de se observar que os maiores percentuais correspondem exatamente aos subsistemas com maiores percentuais de participação de geração eólica em relação ao total do sistema brasileiro.

Capítulo 7

Conclusões e Continuação do Trabalho

7.1 Considerações Gerais

Este trabalho apresentou o desenvolvimento de um modelo para representação de elementos variantes no tempo estatisticamente dependentes na Simulação Monte Carlo Não Sequencial.

O primeiro objetivo do desenvolvimento deste modelo foi prover flexibilidade para a caracterização das variáveis aleatórias de forma não paramétrica, de maneira a se evitar o procedimento de arbítrio de famílias de distribuição de probabilidade para as variáveis aleatórias, que pode levar a adoção de ajustes pouco precisos. Este fator é importante principalmente quando o sistema possui muitos elementos variantes no tempo, pois aumentam as chances de não haver uma única família de distribuição que seja bem ajustável a todas as variáveis.

No caso da representação de séries de direção de vento este ponto se torna crítico, pois estas variáveis não são tão facilmente enquadradas em uma família de distribuição.

O segundo objetivo foi possibilitar a representação de relacionamentos estatísticos não lineares entre as variáveis, o que em geral não é possível pelos modelos convencionais, que usualmente adotam a correlação de *Pearson* como medida de dependência estatística, e que tem como principal limitação medir apenas relacionamentos lineares e ser, a rigor, aplicável apenas em condições específicas que em geral não são satisfeitas pelos elementos variantes no tempo. No que tange a representação da incerteza da geração eólica, esta questão se torna fundamental, mais uma vez, quando se representam as séries de direção dos ventos, por estas introduzirem relacionamentos não lineares bastante significativos, tanto entre as diversas variáveis que representam a direção, como entre as variáveis de direção e as de módulo da velocidade.

O terceiro objetivo foi o tratamento da representação de alta dimensionalidade, isto é, para quando existe a necessidade de se representar muitos elementos variantes no tempo, o que implicaria no aumento exponencial dos parâmetros da função de densidade de probabilidade conjunta das variáveis, caso esta fosse representada

explicitamente. Este fator também é fundamental na análise da confiabilidade composta de sistema de potências que possuem um elevado percentual de geração eólica nas suas matrizes elétricas.

Para alcançar estes objetivos, o modelo proposto foi desenvolvido combinando-se um método de estimação não paramétrico de funções de densidade de probabilidade de variáveis contínuas, o método *kernel density estimation*, a adoção de uma medida de dependência estatística não linear, a informação mútua, complementada pelo coeficiente de informação máxima, e uma abordagem de representação de probabilidades condicionais em grafo, dada por uma estrutura de redes Bayesianas. A partir da aplicação destes conceitos, foi possível uma proposição para se representar as funções de densidade de probabilidade dos elementos variantes no tempo de forma indireta, a partir de duas componentes principais. A primeira componente descreve as variáveis aleatórias de maneira independente, por meio das estimativas não paramétricas das suas funções de densidade de probabilidade marginais, e a segunda componente codifica a dependência estatística entre as variáveis, por meio de uma rede Bayesiana discreta.

A partir desta definição inicial, foram desenvolvidas extensões no modelo para tratar aspectos adicionais bastante relevantes na representação da geração eólica em estudos de confiabilidade composta de sistema de potências. Entre estes aspectos, destaca-se a possibilidade de se representar variáveis exógenas ao processo estocástico dos elementos variantes no tempo, como, por exemplo, a representação de variáveis climáticas que podem afetar o comportamento do vento, possibilitando a geração de cenários mais plausíveis para o estudo.

Adicionalmente, foi tratada também a possibilidade de representação da correlação da geração eólica com as vazões afluentes, uma vez que esta relação é de fundamental importância na avaliação da confiabilidade composta de sistemas hidrotérmicos e esta representação não é trivial, dado que as vazões afluentes possuem uma dinâmica muito mais lenta do que a geração eólica, possuindo níveis de discretização diferentes.

Para possibilitar a estimação dos índices baseados em frequência e duração das falhas dos sistemas, é necessária a representação da auto-correlação cruzada entre as variáveis. Desta forma, esta é caracterizada por meio de um modelo de transição de estados, também construído a partir da aplicação de redes Bayesianas.

A partir da definição do modelo proposto e do algoritmo de estimação dos seus parâmetros, foi proposta uma variação no algoritmo conceitual da Simulação Monte Carlo Não Sequencial para incorporar a amostragem dos estados deste modelo, visando à estimação dos índices baseados em probabilidade e energia, bem como incorporar a

simulação do modelo de transição dos elementos, visando à estimação dos índices de frequência e duração.

Para avaliar a precisão, a eficiência e a flexibilidade de representação do modelo, foram realizadas simulações com o sistema RTS, modificado para incorporar usinas eólicas fictícias, que foram construídas a partir de séries de ventos reais.

As primeiras simulações realizadas tiveram como objetivo validar o modelo desenvolvido, tendo sido os resultados comparados com os resultados das simulações de referência, realizadas com os modelos tradicionais de confiabilidade. Esta comparação mostrou que o modelo conseguiu reproduzir bem o comportamento das séries temporais dos elementos variantes no tempo, sendo bastante preciso na obtenção dos mesmos índices de confiabilidade que a Simulação Sequencial, e bastante eficiente ao despender custos computacionais similares, e para os casos avaliados, até menores que os métodos de simulação empregados: Simulação Monte Carlo Pseudo-Sequencial e Simulação Monte Carlo Pseudo-Cronológica.

A análise das amostras sintéticas geradas mostrou que o modelo conseguiu reproduzir bem as estatísticas das séries históricas, sendo bem reproduzidas tanto as densidades de probabilidade marginais das variáveis aleatórias, quanto à dependência estatística existente entre elas.

As simulações posteriores tiveram como objetivo avaliar aspectos que a princípio não são triviais de serem representados nos modelos estocásticos tradicionais utilizados na confiabilidade, mas que podem ser facilmente representados pelo modelo proposto, dado a sua flexibilidade.

Foi avaliada a inclusão no processo de uma variável exógena referente a um indicador do fenômeno El Niño, e os resultados da simulação mostraram que a confiabilidade pode ser bastante afetada ao se incluir esta representação, sendo importante que o modelo que representa o processo estocástico dos elementos seja flexível para permitir tal informação.

Por fim, foi avaliado o caso do sistema real brasileiro com a inclusão da geração eólica. Entre os principais aspectos de interesse na avaliação deste caso, destaca-se o fato deste ser um sistema de grande porte, o que permite avaliar a robustez do modelo de simulação frente a uma representação bem mais complexa. Destaca-se também que foi necessário se representar a geração eólica a partir das séries de geração total das usinas, por indisponibilidade de dados históricos de vento. A caracterização da geração total de cada usina por meio da sua função de densidade de probabilidade tende a ser mais complicada do que a caracterização do vento, uma vez que as variáveis aleatórias que representam a geração total já correspondem a transformações não lineares das

variáveis aleatórias que representam o vento, dada a característica não linear de geração das turbinas.

Por fim, outro aspecto fundamental do sistema brasileiro é que este é constituído predominantemente por usinas hidrelétricas, o que torna a questão da correta representação da dependência estatística entre as séries temporais de vazões afluentes e geração eólica extremamente crítica para os resultados do estudo.

As simulações foram realizadas considerando e desconsiderando a representação das usinas eólicas, e, desta maneira, os resultados obtidos permitiram que fosse avaliado o impacto da geração eólica na confiabilidade do sistema brasileiro.

De maneira geral, observou-se que a probabilidade de corte de carga não foi tão impactado pela inclusão da geração eólica, mas sendo obtido uma redução de aproximadamente 26% do valor médio de corte de carga para o sistema, que, entretanto, já é baixo para o caso sem representação da geração eólica. Verificou-se que a redução no corte de carga não é homogêneo por todos os subsistemas do sistema brasileiro, sendo os maiores percentuais de redução obtidos exatamente para os subsistemas com maior percentual de capacidade instalada de geração eólica em relação ao total do sistema brasileiro: o subsistema Nordeste, para o qual se observou uma redução de quase 40% no valor médio de corte de carga, e o subsistema Sul, para o qual se observou uma redução de quase 30% do total.

Adicionalmente, foi identificado que a incorporação das usinas eólicas quase não teve efeito sobre os eventos mais severos de corte de carga, sendo observadas apenas pequenas reduções na probabilidade destes eventos mais severos.

7.2 Desenvolvimentos Futuros

Algumas linhas de pesquisa promissoras para a continuidade do trabalho desenvolvido podem ser enumeradas.

1. Adoção de abordagem periódica para o Modelo.

Este aspecto chegou a ser brevemente citado na seção 5.5. Em estudos de operação e planejamento de sistemas de potência é comum se trabalhar com resultados discretizados semanalmente ou mensalmente. Para possibilitar este nível de representação, pode-se adotar, ao invés de um único modelo de representação, um conjunto de modelos, de modo que cada modelo corresponda a um período do ano.

Esta abordagem parece ser ainda mais promissora ao se considerar a representação de dependência estatística entre a geração eólica e as vazões afluentes, uma vez que os

modelos estocásticos de vazões são representados de forma periódica na maior parte dos estudos realizados, o que é justificável pelo tipo de sazonalidade observada nas séries de vazões, e espera-se que com isso seja possível melhorar ainda mais a representação da complementariedade da geração hidroelétrica com a geração eólica.

2. Representação de outros elementos variantes no tempo

Embora, por construção, o modelo seja flexível para a representação de qualquer elemento variante no tempo, neste trabalho o enfoque foi apenas a representação da geração eólica. Desta maneira, seria importante avaliar variáveis de outras naturezas, igualmente importantes para a avaliação da confiabilidade.

Dentre estas variáveis, seria interessante se avaliar a introdução de variáveis representadas por processos estocásticos discretos no modelo e outras fontes alternativas de energia, como a energia solar, que por depender de uma fonte comum, o sol, espera-se que pode apresentar algum grau de dependência estatística com a energia eólica.

3. Exploração do aspecto *datamining* das redes Bayesianas

As redes Bayesianas são estruturas com uso bastante difundido em aplicações de *datamining*, tendo alcançado bastante sucesso nesta área. Este aspecto poderia ser aproveitado para minimizar o esforço computacional da avaliação da confiabilidade por Simulação Monte Carlo, no sentido de se dotar a rede Bayesiana com a capacidade de evitar a avaliação de estados similares do sistema a outros estados anteriormente avaliados. O trabalho consistiria em identificar as adaptações na estrutura de rede Bayesiana desenvolvida neste trabalho, de modo a torna-la parte desta função de identificação de similaridade, objetivo básico do *datamining*.

4. Adaptação para abordagem cronológica

Embora o enfoque deste trabalho, e por consequência a natureza do modelo desenvolvido, seja voltado para a representação não cronológica, a extensão deste modelo para a abordagem cronológica é um aspecto importante a ser considerado, uma vez que a exigência de cronologia do processo pode ser uma restrição do estudo que se deseja realizar, e desta maneira, se incorporaria a flexibilidade provida pelo modelo também à representação cronológica.

Apêndice I

Kernel Density Estimation

I.1 Medidas de Desempenho

Dentre as principais medidas de desempenho para as estimativas obtidas por *kernel smoothing*, destaca-se o erro médio quadrático, (*MSE* - *mean squared error*) e o erro médio quadrático integrado (*MISE* - *mean integrated squared error*).

A medida *MSE* é dada por:

$$MSE(\hat{f}(x)) = E(\hat{f}(x) - f(x))^2 \quad (1.1)$$

Que pode ser escrita como:

$$MSE(\hat{f}(x)) = Bias^2(\hat{f}(x)) + var(\hat{f}(x)) \quad (1.2)$$

Onde,

$$Bias^2(\hat{f}(x)) = (E[\hat{f}(x)] - f(x))^2 ;$$

$$var(\hat{f}(x)) = E(\hat{f}(x) - E[\hat{f}(x)])^2$$

Esta medida é relativa a um dado argumento x . A eficiência da estimativa de $\hat{f}(x)$ ao longo de toda função é obtida pela medida *MISE*, dada por:

$$MISE(\hat{f}) = E \int_{-\infty}^{\infty} MSE(\hat{f}(x)) dx \quad (1.3)$$

Que pode ser escrita como:

$$MISE(\hat{f}) = \int_{-\infty}^{\infty} Bias^2(\hat{f}(x)) dx + \int_{-\infty}^{\infty} var(\hat{f}(x)) dx \quad (1.4)$$

Em [73] é apresentado um desenvolvimento para obtenção dos termos $E(\hat{f}(x))$, $var(\hat{f}(x))$ e $Bias(\hat{f}(x))$, produzindo-se as seguintes aproximações para as medidas de desempenho:

$$MSE(\hat{f}(x)) \approx \frac{1}{4} h^4 k_2^2 f''(x)^2 + \frac{1}{nh} f(x) j_2 \quad (1.5)$$

$$MISE(\hat{f}) \approx \frac{1}{4} h^4 k_2^2 \beta(f) + \frac{1}{nh} j_2 \quad (1.6)$$

Onde,

$$k_2 = \int z^2 K(z) dz ;$$

$$j_2 = \int K(z)^2 dz ;$$

$$\beta(f) = \int f''(x)^2 dx ;$$

I.2 Estimação do Parâmetro de Alisamento

A determinação do parâmetro de alisamento ótimo tem sido formulada como um problema de maximização da eficiência da estimativa por minimização dos índices dos erros desta estimativa. Uma abordagem simples e bastante utilizada é apresentada a seguir.

Da expressão (1.6), tem-se que o parâmetro de alisamento ótimo que minimiza o *MISE* é obtido por:

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5} \quad (1.7)$$

Onde,

$$\gamma(K) = j_2 k_2^{-2}$$

Desta expressão, observa-se que $\gamma(K)$ depende apenas da função *kernel* escolhida, no entanto o termo $\beta(f)$ depende da densidade real da função, mais precisamente do termo $f''(x)$, que é obviamente desconhecido. Uma aproximação bastante utilizada é assumir que a função de densidade segue uma distribuição normal, para a qual se pode mostrar que:

$$\beta(f) = \frac{3\sigma^{-5}}{8\sqrt{\pi}} \quad (1.8)$$

Adotando-se ainda a *kernel* gaussiana, tem-se que:

$$h_{opt} = \left(\frac{4}{3n} \right)^{1/5} \sigma \quad (1.9)$$

A aplicação direta desta expressão apresenta diversos problemas para a estimação, principalmente devido à consideração de que a densidade possui distribuição normal, cuja característica unimodal implica em um efeito indesejado para estimativas de funções de densidade multimodais[43].

Em [73], é proposto um outro estimador para o qual:

$$\hat{h}_{opt} = \frac{0.9\hat{\sigma}}{n^5} \quad (1.10)$$

Onde,

$$\hat{\sigma} = \min\left(s, \frac{IQR_x}{1.34}\right);$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

IQR_x é a distância interquartílica dos dados.

I.3 Estimação de Densidade com Suporte para Variáveis com Limite

A forma mais simples de garantir que o estimador respeite os limites pré-definidos é introduzir uma transformação específica para tal. A ideia básica é estimar a função de densidade de probabilidade de uma variável aleatória transformada $Y = t(X)$, com estes limites bem definidos. Dado que $y \sim g(y)$ é a função de densidade de probabilidade de Y , a relação entre f e g será dada por[43]:

$$f(x) = g(t(x))t'(x) \quad (1.11)$$

Desta maneira, o procedimento para se obter a estimativa com a definição dos limites é constituído pelos seguintes passos:

1. Obter o vetor de observações transformadas: $x_i \rightarrow y_i = t(x_i), i = 1, 2, \dots, n$
2. Aplicar normalmente o método *kernel density estimation* para obter uma estimativa para $g(y)$
3. Obter uma estimativa para $f(x)$ usando $\hat{f}(x) = \hat{g}(t(x))t'(x)$

Para exemplificar, para uma variável que seja definida em $X \in [L, U]$, conforme será visto mais a frente, pode-se aplicar a seguinte transformação:

$$t(x) = \log\left(\frac{x-L}{U-x}\right) \quad (1.12)$$

A derivada da função de transformação será dada por:

$$t'(x) = \frac{1}{x-L} + \frac{1}{U-x} \quad (1.13)$$

Resultando na seguinte expressão para a função de densidade desejada:

$$\hat{f}(x) = \begin{cases} 0 & x \leq L \\ \hat{g}\left(\log\left(\frac{x-L}{U-x}\right)\right) \cdot \left(\frac{1}{x-L} + \frac{1}{U-x}\right) & L < x < U \\ 0 & x \geq U \end{cases} \quad (1.14)$$

Na Figura 80 é apresentada uma comparação entre a função de densidade de probabilidade estimada sem definição e com definição de limite, para uma variável aleatória associada a uma série real de vento. A função de densidade obtida através da estimação pelo método tradicional corresponde a linha tracejada, enquanto que a função obtida pelo método com definição do suporte corresponde a linha cheia.

Para a estimação da função de densidade utilizando o suporte, foi definida a função de transformação dada por: $t(x) = \exp\left(\frac{x}{19.5}\right)$, onde 19.5 m/s corresponde ao valor máximo observado no histórico.

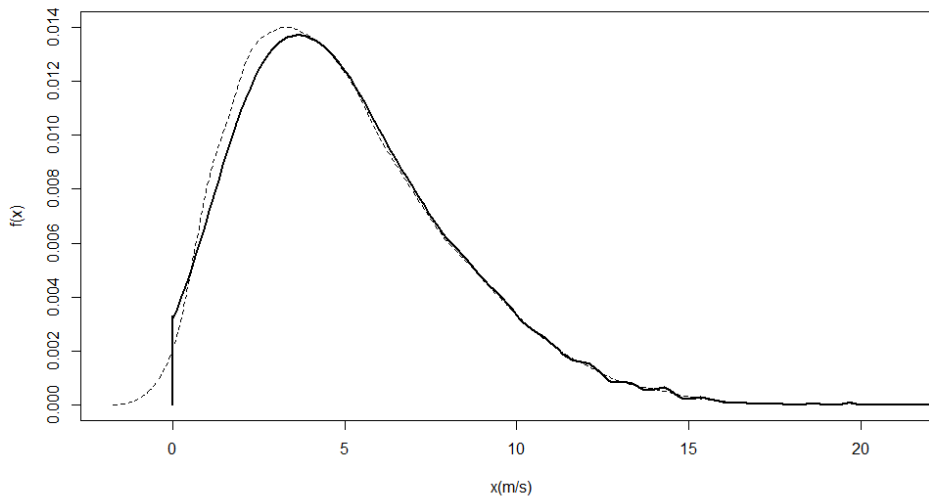


Figura 80– Comparação de Estimativa sem Definição de Suporte e Com Definição de Suporte

Observa-se que as principais diferenças entre as duas estimativas estão no lado esquerdo das funções, existindo na estimativa pelo método tradicional, valores negativos com densidade de probabilidade diferentes de zero, embora estas densidades sejam muito pequenas. Para a estimativa pelo método com suporte, a função de densidade possui um degrau na origem, sendo nula para qualquer valor inferior no eixo horizontal.

A importância desta consideração se deve ao fato das funções de densidades de probabilidade serem diretamente utilizadas na geração de cenários, pelo modelo proposto neste trabalho, e desta maneira se garante que não serão amostradas velocidades negativas para as séries de ventos representadas.

I.4 Proposta de Transformação para Estimação de Densidade com Suporte para Variáveis com Limite

Embora a aplicação da transformação logarítmica garanta o atendimento da restrição do domínio de não negatividade das variáveis aleatórias, ela tende a produzir algumas distorções indesejadas na função de densidade de probabilidade estimada, em relação a função de densidade de probabilidade original, não transformada, como pode ser visto na comparação entre as estimativas apresentadas na Figura 80.

Para contornar esta limitação, neste trabalho é proposta a seguinte função de transformação:

$$T(x) = \frac{\ln(1 + e^{kx})}{k} \quad (1.15)$$

Cuja derivada produz a função sigmoide, expressa como:

$$T'(x) = \frac{1}{1 + e^{-kx}} \quad (1.16)$$

O interesse pelo arbítrio desta função advém do fato da função sigmoide se aproximar o tanto quanto se queira da função degrau, a partir do ajuste da constante k , conforme observado na Figura 81. Nesta figura, a linha preta corresponde a função sigmoide para $k=1$, a linha vermelha corresponde a função para $k=2$ e a linha azul corresponde a função para $k=4$.

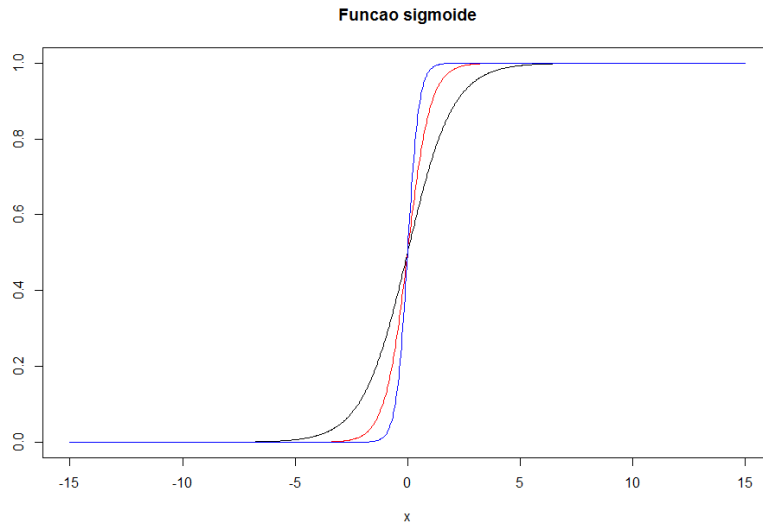


Figura 81– Função sigmoide para diferentes constantes

Desta maneira, a partir da equação (1.11), tem-se que:

$$f(v) = g\left(\frac{\ln(1 + e^{kv})}{k}\right) \cdot \frac{1}{1 + e^{-kv}} \quad (1.17)$$

Onde, $g(y)$ corresponde a função de densidade de probabilidade da variável transformada, $y = T(v)$, conforme apresentado na seção 3.2.2

A parcela da direita da equação tenderá a fazer com que $f(v)$ seja nulo para $v < 0$, para um valor de k suficientemente grande.

Por outro lado, o crescimento de v faz com que e^{-kv} tenda a zero e $\ln(1 + e^{kv})$ tenda a kv , desta maneira o termo a direita tende a 1, e a variável transformada, y , passa a seguir exatamente o valor de v . Como consequência, a função de densidade de probabilidade estimada possui o aspecto apresentada na Figura 82, na qual a estimativa obtida com suporte é a representada pela linha cheia, enquanto que uma estimativa de referência sem uso da definição de suporte é representada pela linha tracejada.

É possível observar que a diferença entre as duas estimativas está distribuída na parte mais a esquerda da função, na qual se nota, principalmente, que a estimação sem suporte indica que existe uma probabilidade de ocorrência de velocidades de vento negativas, mesmo que esta probabilidade seja pequena, enquanto que a estimativa com suporte indica que não existe probabilidade de ocorrência de velocidade de vento negativa.

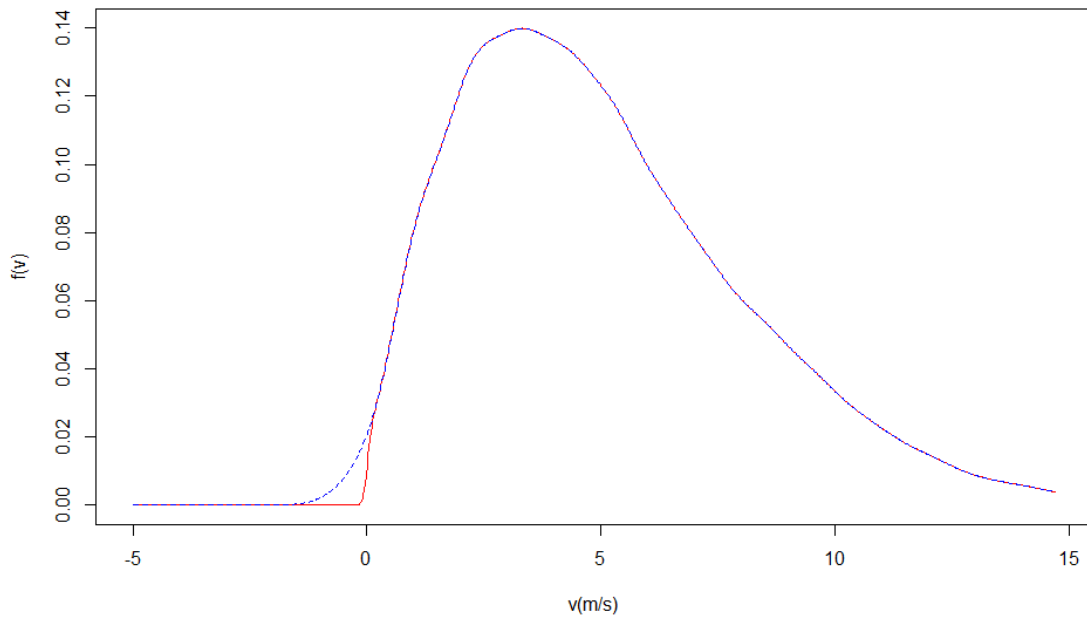


Figura 82– Função de densidade de probabilidade estimada com suporte em $[0, \infty)$

Apêndice II

Rede Bayesiana

II.1 Estimação de Estrutura de Redes Bayesianas pelo algoritmo Hill Climbing

Este método pertence a classe de métodos de estimação de estrutura de redes Bayesianas baseados em pontuação (métodos *score-based*).

Basicamente o processo atribui a cada possível configuração da rede uma medida que quantifica o quão bem a rede Bayesiana descreve o *dataset D*. Esta pontuação pode ser definida como [49]:

$$score(G|D) = P(G|D) \quad (2.1)$$

Onde $P(G|D)$ é probabilidade a posteriori de G , dado um *dataset D*.

O objetivo dos métodos *score-based* é maximizar esta pontuação.

Por conveniência, a equação pode ser transformada a partir da aplicação do teorema de *Bayes*, resultando em:

$$score(G|D) = P(G|D) = \frac{P(D|G)P(G)}{P(D)} \quad (2.2)$$

A estimação do grafo que maximiza esta expressão depende apenas da maximização do numerador da expressão, uma vez que o denominador não depende de G . Mais que isso, usualmente se considera que todas as possíveis estruturas de grafos são equiprováveis, isto é, não há informação a priori a ser considerada, o que elimina também o termo $P(G)$, reduzindo o problema a maximização de $P(D|G)$.

O cálculo de $P(D|G)$ pode ser efetuado a partir de um enfoque Bayesiano [49], em que todos os possíveis parâmetros são ponderados por suas probabilidades a posteriori:

$$P(D|G) = \int P(D|G, p)P(p|G)dp \quad (2.3)$$

Em [48] foi demonstrado que para funções de densidades locais multinomiais tem-se que:

$$P(D|G) = \prod_{i=1}^n \prod_{j=1}^{q_{ij}} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})} \quad (2.4)$$

Onde, α_{ij} e N_{ij} são respectivamente os hyper-parâmetros e os números de funções de densidade de probabilidade de X_i para a j -ésima configuração do pai. No limite, para uma grande quantidade de dados, o termo $P(D|G, p)P(p|G)$ pode ser razoavelmente bem aproximado por uma gaussiana multivariável[49]. A partir desta consideração, aproximando ainda a média da gaussiana com o valor de \hat{p} de máxima verossimilhança e ignorando os termos que não dependem do tamanho do *dataset* N , pode-se mostrar que a pontuação pode ser aproximada por:

$$BICscore(G, D) = \log P(D|\hat{p}, G) - \frac{d}{2} \log N \quad (2.5)$$

Onde o termo \hat{p} é o conjunto de estimações de máxima verossimilhança dos parâmetros p da rede Bayesiana, d é o número de graus de liberdade da gaussiana multivariável, isto é, a sua dimensão, que coincide com o número de graus de liberdade das funções de densidade de probabilidade locais, resultando em: $d = \sum_{i=1}^n q_i (r_i - 1)$.

A principal vantagem da pontuação pelo índice *BIC* advém do fato de que ele não depende da priori dos parâmetros, uma vez que a informação desta priori muitas vezes não está disponível ou é difícil de calcular. Este índice também possui uma interpretação intuitiva de ser uma ponderação da verossimilhança dos dados menos um termo, $\frac{d}{2} \log N$, que penaliza as topologias que possuem estruturas mais complicadas, atuando no sentido de evitar um super-ajuste da estrutura da rede aos dados (*overfitting*).

O grande problema prático da estimação de estruturas de redes Bayesianas pelos métodos *score-based* é, que a rigor, este é um problema *NP-Completo*, pois o processo de maximização da verossimilhança proposto pressupõe que todas as diferentes estruturas sejam avaliadas. O espaço de estados é no mínimo dado por uma função exponencial do número de variáveis n . Existem $n(n-1)/2$ possíveis arcos não direcionados e $2^{n(n-1)/2}$ possíveis estruturas para cada conjunto destes arcos. Mais que isso, cada arco deste pode assumir 2 diferentes orientações para cada diferente escolha.

Embora o método da força-bruta, na qual todas as possíveis combinações são avaliadas, seja de fato utilizado para casos pequenos, para a maior parte dos problemas, e em particular para o problema de representação de dependência estatística de sistemas de alta dimensionalidade abordado neste trabalho, o seu uso é proibitivo. Desta maneira,

uma abordagem bastante utilizada na prática são os métodos heurísticos, que buscam a diminuição drástica deste espaço de busca.

Um dos algoritmos heurísticos mais populares para a resolução do problema de estimação da estrutura de redes Bayesianas é o algoritmo de *hill climbing*. Este corresponde a uma técnica de otimização que pertence à família das técnicas de buscas locais, sendo frequentemente utilizado para resolver problemas que possuem muitas soluções, sendo algumas soluções melhores que outras. Basicamente, o algoritmo inicia de uma solução aleatória do problema (potencialmente ruim) e iterativamente faz ajustes na solução, melhorando um pouco mais a cada iteração. Quando o algoritmo não pode prover mais nenhuma melhora na solução atual, o algoritmo é dito convergido.

Idealmente, a solução final encontrada estará bem próxima do que seria a solução ótima real do problema, mas na prática não há nada que formalmente garanta que o algoritmo irá sempre encontrar um ponto próximo do ótimo real.

No caso do problema da determinação da estrutura da rede Bayesiana, o algoritmo atua avaliando configurações diferentes, mas próximas, a configuração atual, formadas a partir da aplicação de alguma transformação elementar na topologia da rede.

As principais transformações elementares aplicadas a uma determinada configuração de rede no processo de estimação de estrutura de rede Bayesiana são: adição de arestas entre nós, remoção de arestas já existentes entre nós e inversão do sentido das arestas já existentes entre nós.

Os passos do algoritmo conceitual são apresentados a seguir:

- 1) *Arbitrar uma estrutura inicial para a rede* $G \leftarrow G_0$;
- 2) *Avaliar o score da rede* $f_G \leftarrow f(G : D)$ *a partir do dataset* D ;
- 3) *Faça* $improvement \leftarrow false$;
- 4) *Gerar candidatos vizinhos por adição de arestas à rede:*
 - Para cada par de nós X_i e $X_j \notin pa(X_i)$, tal que a adição da aresta $X_j \rightarrow X_i$ não introduza um ciclo fechado em G , calcular a diferença: $diff = f(G + \{X_j \rightarrow X_i\} : D) - f_G$ e armazenar a configuração que maximiza $diff$ em $\langle change_a \quad diff_a \rangle$
- 5) *Gerar candidatos vizinhos por remoção de arestas da rede:*
 - Para cada par de nós X_i e X_j , calcular a diferença: $diff = f(G - \{X_j \rightarrow X_i\} : D) - f_G$ e armazenar a configuração que maximiza $diff$ em $\langle change_d \quad diff_d \rangle$
- 6) *Gerar candidatos vizinhos por inversão dos sentidos das arestas da rede:*

- Para cada par de nós X_i e $X_j \notin pa(X_i)$, tal que a inversão do sentido da aresta $X_j \rightarrow X_i$ para $X_j \leftarrow X_i$ não introduza um ciclo fechado em G , calcular a diferença: $diff = d_1 + d_2$, onde $d_1 = f(G - \{X_j \rightarrow X_i\}; D) - f_G$ e $d_2 = f(G + \{X_j \rightarrow X_i\}; D) - f_G$, armazenar a configuração que maximiza $diff$ em $\langle change_r, diff_r \rangle$

7) Checar se houve melhoria na configuração da rede:

- Fazendo $d^* = \max_{k=a,d,r} diff_k$ e $move^*$ a modificação correspondente na rede,

8) Se $d^* > 0$, faça:

$improvement \leftarrow true$;

$G \leftarrow G + move^*$;

9) Se $improvement = true$, retornar ao passo 2, caso contrário, o algoritmo é finalizado.

Para um melhor entendimento do processo, uma ilustração dos passos de avaliação da vizinhança da estrutura atual da rede é apresentada na Figura 3.

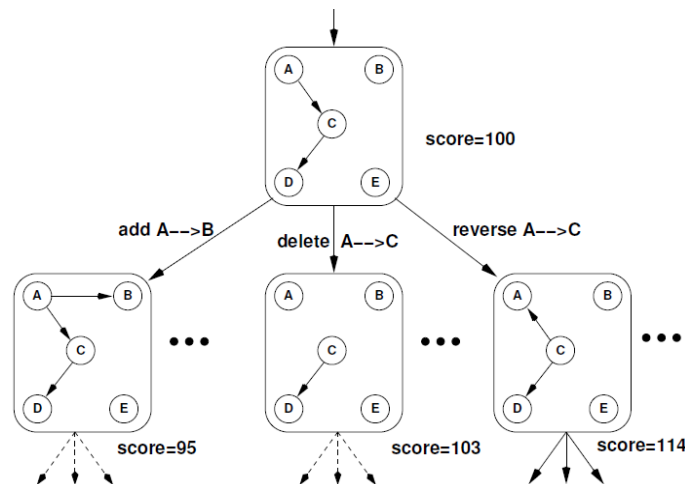


Figura 83– Estimação de rede Bayesiana por *hill climbing*

II.2 Estimação dos Parâmetros de Redes Bayesianas pelo Método MLE (Maximum Likelihood Estimation – Estimação da Máxima Verossimilhança)

Estimar os parâmetros de redes Bayesianas a partir de uma topologia conhecida é um problema bem mais simples do que o problema da estimação da topologia. Um dos métodos mais empregados para este fim é o método da estimação da máxima verossimilhança (MLE – Maximum Likelihood Estimation), utilizado no modelo desenvolvido neste trabalho.

Conforme apresentado na seção 3.4, tem-se que:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i)) \quad (2.6)$$

Definindo-se os parâmetros:

$$\theta_{ijk} = P(X_i = j | pa(X_i) = k), i = 1, \dots, n; j = 1, \dots, r_i; k = 1, \dots, q_i \quad (2.7)$$

Que constituem o conjunto de parâmetros:

$$\Theta = \{\theta_{ijk} | i = 1, \dots, n; j = 1, \dots, r_i; k = 1, \dots, q_i\} \quad (2.8)$$

Onde,

$1, 2, \dots, r_i = |\Omega_{X_i}|$ os números dos estados de X_i ;

$1, 2, \dots, q_i = |\Omega_{pa(X_i)}|$ o número de configurações que os nós pais de X_i podem

assumir;

Definindo-se uma função característica, tal que:

$$\chi(i, j, k : D_i) = \begin{cases} 1 & \text{se } X_i = j, pa(X_i) = k \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

E definindo-se:

$$m_{ijk} = \sum_I \chi(i, j, k : D_i) \quad (2.10)$$

Tem-se que:

$$\begin{aligned} \log L(\Theta | D) &= \sum \log P(D_i | \Theta) \\ &= \sum_I \sum_{i,j,k} \chi(i, j, k : D_i) \log \theta_{ijk} \\ &= \sum_{i,j,k} \sum_I \chi(i, j, k : D_i) \log \theta_{ijk} \\ &= \sum_{i,j,k} m_{ijk} \log \theta_{ijk} \\ &= \sum_{i,k} \sum_j m_{ijk} \log \theta_{ijk} \end{aligned} \quad (2.11)$$

Desta forma, a estimação de máxima verossimilhança pode ser formulada como:

$$\arg \max_{\Theta} \log L(\Theta | D) = \arg \max_{\theta_{ijk}} \sum_{i,k} \sum_j m_{ijk} \log \theta_{ijk} \quad (2.12)$$

Observa-se que $\theta_{ijk} = P(X_i = j | pa(X_i) = k)$ e $\theta_{i'j'k'} = P(X_{i'} = j' | pa(X_{i'}) = k')$ não estão relacionados, se $i \neq i'$ ou $k \neq k'$, o que permite que cada termo que aparece em $\sum_{i,k} [\dots]$ seja maximizado independentemente, por:

$$\arg \max_{\theta_{ijk}} \sum_j m_{ijk} \log \theta_{ijk} \quad (2.13)$$

De [49], tem-se que o parâmetro que maximiza a expressão é dado por:

$$\theta_{ijk}^* = \frac{m_{ijk}}{\sum_j m_{ijk}} \quad (2.14)$$

O que diz que a estimação de máxima verossimilhança para $\theta_{ijk} = P(X_i = j | pa(X_i) = k)$ é dada por:

$$\theta_{ijk}^* = \frac{\text{número de casos em que } X_i = j \text{ e } pa(X_i) = k}{\text{número de casos em que } pa(X_i) = k} \quad (2.15)$$

Referências Bibliográficas

- [1] Leite da Silva, A. M., Endrenyi, J., Wang, L., “Integrated Treatment of Adequacy and Security in Bulk Power system Reliability Evaluations”, *IEEE Transactions on Power Systems*, v. 8, n. 1, pp. 275-285, Fev. 1993.
- [2] Borges, C.L.T, “Avaliação da Confiabilidade Composta de Sistemas de Potência em Ambientes Computacionais Paralelos e Distribuídos”, Tese de Doutorado PESC-COPPE/UFRJ, 1998.
- [3] Matalas, N. C., “Mathematical Assessment of Synthetic Hydrology”, *Water Resources*, v. 3, n. 4, pp. 937-946, 1967.
- [4] Hipel, K.W., McLeod, A.I., **Time Series Modelling of Water Resources and Environmental System**, 1 ed, Amsterdam, Elsevier , 1994.
- [5] Himanshu, A. A., Lester, C. H., **Electricity Demand for Sri Lanka: a Time Series Analysis**. University of Surrey, 2007.
- [6] Saab, S., Badr, E., Nasr, G., “Univariate Modeling and Forecasting of Energy Consumption: the Case of Electricity in Lebanon”. *Energy*, v. 26, n. 1, pp. 1-14, Jan. 2001.
- [7] Chen, P., Pedersen T., Bak-Jensen, B., Chen, Z., “ARIMA-Based Time Series Model of Stochastic Wind Power Generation”, *IEEE Transactions on Power Systems*, v. 25, n. 2, pp. 667-676, Mai. 2010.
- [8] Brown, B. G., Katz, R. W., Murphy, A. H., “Time Series Models to Simulate and Forecast Wind Speed and Wind Power”, *J. Climate and Applied Meteorology*, v. 23, n. 8, pp. 1184-1195, Mai. 1984.
- [9] Welch G., Bishop B., "An Introduction to the Kalman Filter," University of North Carolina, Tech. Rep. TR 95-041, 1995.
- [10] Lingli, S., Changhong, D., Qiushi, X., Chu, Z., “Load Forecasting Considering the Regional Economic Environment”, **In: International Conference on Automatic Control and Artificial Intelligence**, pp. 1310-1314, Xiamen, Mar. 2012.
- [11] Rasmusson, E. M., Wallace, J. M., “Meteorological Aspects of the El Niño/Southern Oscillation”, *Science*, v. 222, n. 4629, pp. 1195-1202, Dez. 1983.

- [12] Menezes Neto, O. L., Oliveira, J. L., Costa, A. A., Sombra, S. S., “Impactos da Circulação Geral em Casos de El Niño e La Niña no Potencial Eólico no Nordeste Brasileiro”. **In: Congresso Brasileiro de Meteorologia**, 14. 2006, Florianópolis, Anais.
- [13] Araújo Junior, L. M., da Silva, E. M., Costa, A. A., Sales, D. C., Vasconcelos Junior, F. C., Guimarães, S. O., “Avaliação dos Recursos Eólicos Simulado no Estado do Ceará: um Estudo de caso para o El Niño 97/98 e para La Niña 98/99”, *Revista Brasileira de Meteorologia*, v. 29 no. 1, pp. 139-152, Mar. 2014.
- [14] Hardle, W., Lutkepohl, H., Chen, R., “A Review of Nonparametric Time Series Analysis”, *International Statistical Review*, v. 65, n. 1, pp. 49–72, Abr. 1997.
- [15] Tuller, S. E., Brett, A. C., "The Characteristics of Wind Velocity that Favor the Fitting of a Weibull Distribution in Wind Speed Analysis", *J. Climate Appl. Meteor*, v. 23, n. 1, pp. 124–134, Jan. 1984.
- [16] Borges, C. L. T., Dias, J. A. S., “A Non Parametric Stochastic Model for River Inflows Based on Kernel Density Estimation”, **In: Proceedings of 13th International Conference on Probabilistic Methods Applied to Power Systems**, pp. 100-107, Durham, Jun. 2014.
- [17] Reinsel, G. C., **Elements of Multivariate Time Series Analysis**, 2ed, Springer, 2013.
- [18] Tiao, G. C., Box, G. E. P., “Modeling Multiple Time Series with Applications”, *Journal of the American Statistical Association*, v. 76, n. 376, pp. 802-816, Dez. 1981.
- [19] ENEL, “The Monte Carlo Method in Planning Electrical Systems”, *Studie Ricerche*, n. 22, Ago. 1971.
- [20] Nofery, P. L., Paris, L. Salvaderi, L., “Monte Carlo Method for Power System Reliability Evaluation in Transmission and Generation Planning”, **In: Proceedings Annual Reliability and Maintainability Symp.**, IEEE , Washington D.C., , 1975.
- [21] Clancy, U. P., Gross, G., Wu, F. F., “A decomposition Monte Carlo Approach for Multi-Area Reliability Calculations”, **In: Proc. 7th PSCC** , pp. 261-269, Lausanne, July 1981.

- [22] Kady, M. A., Sobki, M., Sinha, M. K., "Evaluation of the Reliability of Large-Scale Power Systems Using a Monte Carlo Method", **In: IEEE Electronicom 85**, pp. 338-342, Toronto, 1985.
- [23] Bertoldi, U., Salvaderi, L., Scalcino, S., "Monte Carlo Approach in Planning Studies: An Application to IEEE RTS", *IEEE Transactions On Power Systems*, v. PWRS-3, n.3, pp. 1146-1154, Ago. 1988.
- [24] Giosetto, P., Utsorogi, K. F., "Development of a New Procedure for Reliability Modeling of Wind Turbine Generators," *IEEE Transactions on Power Apparatus and Systems*, v. PAS-102, n. 1, Jan. 1983.
- [25] Wang, X., Dai, H. Z., Thomas, R. J., "Reliability Modeling of Large Wind Farms and Associated Electric Utility Interface System," *IEEE Transactions on Power Apparatus and Systems*, v. PAS-102, n. 3, Mar. 1984.
- [26] Billinton, R., Wenyuan, L., "A Monte Carlo Method for Multi Area Generation System Reliability Assessment," *IEEE Transactions on Power Systems*, v. 7, n. 4, pp.1487-1492, Nov 1992.
- [27] Sankar Krishnam, A., Billinton, R., "Sequential Monte Carlo Simulation for Composite Power System Reliability Analysis with Time Varying Loads", *IEEE Transactions on Power Systems*, v. 10, n. 3, pp.1540-1545, Ago. 1995.
- [28] Mello, A.C.G., Pereira, M.V.F., Leite da Silva, A.M., "Evaluation of Reliability Worth in Composite Systems Based on Pseudo-Sequential Monte Carlo Simulation", *IEEE Transactions on Power Systems*, v. 9, n. 3, pp. 1318-1326, Ago. 1994.
- [29] Leite da Silva, A. M., Manso, L. A. F., Mello, J. C. O., Billinton, R., "Pseudo-Chronological Simulation for Composite Reliability Analysis with Time Varying Loads", *IEEE Transactions On Power Systems*, v. 15, n. 1, pp. 73-80, Fev. 2000.
- [30] Billinton, R., Chowdhury, A. A., "Incorporation of Wind Energy Conversion System in Conventional Generation Capacity Adequacy Assessment" *IEEE Proceedings-C*, v. 139, n. 1, Jan. 1992.
- [31] Billinton, R., Chen, H., Ghajar, R., " A Sequential Simulation Technique for Adequacy Evaluation of Generation System Including Wind Energy," *IEEE Transactions on Energy Conversions*, v. 11, n. 4, Dez. 1996.

- [32] Sayas, F. C., Allan, R. N., "Generation Availability Assessment of Wind Farms," *IEEE Proc-Gener. Transm. distrib.* v. 143, n. 5, Set. 1996.
- [33] Leite, A.P., Borges, C.L.T., Falcao, D.M., "Probabilistic Wind Farms Generation Model for Reliability Studies Applied to Brazilian Sites", *IEEE Transactions on Power Systems*, v. 21, n. 1, pp.1487-1492, Nov 2006.
- [34] Leite da Silva, A.M.,Gonzalez-Fernandez, R.A.,Sales, W.S.,Manso, L.A.F., "Reliability Assessment of Time-Dependent Systems via Quasi-Sequential Monte Carlo Simulation", **In: Proceedings of 11th International Conference on Probabilistic Methods Applied to Power Systems**, pp. 697 - 702, Singapura, Jun. 2010
- [35] Leite da Silva, A.M., Fernandez, R.A.G., Singh, C. "Generating Capacity Reliability Evaluation Based on Monte Carlo Simulation and Cross-Entropy Methods", *IEEE Transactions on Power Systems*, v. 25, n. 1, pp. 129 - 137, Fev. 2010.
- [36] Gonzalez-Fernandez, R.A., Leite da Silva, A.M., "Reliability Assessment of Time-Dependent Systems via Sequential Cross-Entropy Monte Carlo Simulation", *IEEE Transactions on Power Systems*, v. 26, n. 4, pp. 2381-2389, Nov. 2011.
- [37] Billinton, R., Allan, R. N., "Reliability Evaluation of Engineering Systems: Concepts and Techniques", 1 ed., New York, Plenum Press, 1983.
- [38] Pereira, M.V.F, Balu, N.J, "Composite Generation / Transmission Reliability Evaluation", *Proceedings of the IEEE*, v. 80 n. 4, pp. 470-491, Abr. 1992.
- [39] Justino, T. C., "Cálculo de Índices de Frequência e Duração na Avaliação da Confiabilidade de Sistemas Interligados Considerando Curvas Variantes no Tempo", Dissertação de Mestrado, Programa de Engenharia Elétrica – COPPE, Universidade Federal do Rio de Janeiro, Agosto de 2009.
- [40] Mello, J.C.O., Pereira,M.V.F., , A.M.L., "Evaluation of Reliability Worth in Comosie Systems Based on Pseudo-Sequential Monte Carlo Simulation". *IEEE Transactions on Power Systems*, v. 9, n. 3, pp. 1318-1326, Ago. 1993.
- [41] Scott, D. W., Tapia, R. A., Thompson, J. R., Tuller, S. E., Brett, A. C., "Kernel Density Estimation Revisited", *Nonlinear Analysis: Theory, Methods & Applications*, v. 1, n. 4, pp. 339–372, 1977

- [42] Castañeda, L. B., Arunachalam, V., Dharmanja, S., **Introduction to Probability and Stochastic Processes with Applications**, John Wiley & Sons , 2012.
- [43] Zucchini, W., “Applied Smoothing Techniques – Kernel Density Estimation”, Outubro 2003.
- [44] Ferreira, M.R.P., “Método Kernel: Estimação de Densidades e Classificação de Padrões”, Departamento de Estatística, UFPB, Abril de 2009.
- [45] Kraskov, A., Stogbauer, H., Grassberger, P., "Estimating Mutual Information", *Physical review E*, v. 69, n. 6, p. 066138, 2004.
- [46] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turabaugh, P. J., Lander, E. S., Mitzenmacher, M., Sabeti, P. C., “Detecting novel associations in large data sets”, *Science*, v. 334, n. 6062, pp. 1518-24, 2011.
- [47] Kinney, J. B., Atwal, G. S., “Equitability, Mutual Information, and the Maximal Information Coefficient”, *PNAS*, v. 111, n. 9, pp. 3354–3359, Jan. 2014.
- [48] Cooper, G., F., Herskovits, E., H., “A Bayesian Method for the Induction of Probabilistic Networks from Data”, *Machine Learning*, v. 9, n. 4, pp. 309-347, 1992.
- [49] Margaritis, D., “Learning Bayesian Network Model Structure from Data “, Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, 2003.
- [50] Afanasyeva, S., “Optimization of Wind Farm Installations”, Master’s Thesis, Lappeenranta University of Technology, 2011.
- [51] CCEE, Boletim das Usinas Eólicas, No 3, Abril de 2014
- [52] Leite, A.P. “Modelagem de Fazendas Eólicas para Estudo de Confiabilidade”. Tese de Mestrado, Programa de Engenharia Elétrica - COPPE, Universidade Federal do Rio de Janeiro, Abril 2005.
- [53] EWEA, Wind Energy – The Facts, v.1 – Technology – EWEA, 2004.
- [54] Barthelmie, R. J., Frandsen, S. T., Rethore, P. E., Mechali, M., Pryor, S.C., Jensen, L., Sorensen, P., “Modelling and Measurements of Offshore Wakes”, *Owemes* Abr. 2006.

- [55] Réthoré, P. E, “Wind Turbine Wake in Atmospheric Turbulence”, Tese de Doutorado, Department of Civil Engineering, Aalborg University, Outubro de 2009.
- [56] Wand, M. P, “Fast Computation of Multivariate Kernel Estimators”. *Journal of Computational and Graphical Statistics*, v. 3, n. 4, pp. 433-445, Dez. 1994.
- [57] Site do Royal Netherlands Meteorological Institute. Disponível em: <http://www.knmi.nl/index_en.html>. Acessado em 08 de julho de 2014.
- [58] Villanueva, D., Feijóo, A., Pazos, J. L., “Simulation of Correlated Wind Speed Data for Economic Dispatch Evaluation”, *IEEE Transactions on Sustainable Energy*, v.3, n. 1, pp. 142-149, Jan. 2012.
- [59] Box, G. E. P., Jenkins, G. M., Reinsel, G. C, “Time Series Analysis, Forecasting and Control”. Hoboken, NJ: Wiley, 2008.
- [60] Dias, J. A., Borges, C. L. T., “Representation of Wind and Load Correlation in Non-Sequential Monte Carlo Reliability Evaluation.” **In: Proceedings of 12th International Conference on Probabilistic Methods Applied to Power Systems**, pp. 100-107, Istanbul, Jun. 2012.
- [61] Genz, A., “Numerical Computation of Multivariate Normal Probabilities”, *Journal of Computational and Graphical Statistics*, v. 1, n. 2, pp. 141-149, 1992.
- [62] IEEE RTS Task Force of APM Subcommittee, “IEEE Reliability Test System”, IEEE PAS, vol. 98, no.6, pp. 2047-20654, Nov/Dez 1979.
- [63] Dias, J.A. “Avaliação da Confiabilidade Composta Baseada em Simulação Monte Carlo com Representação da Geração Eólica”, Dissertação de Mestrado, Programa de Engenharia Elétrica – COPPE/UFRJ, Junho de 2008.
- [64] A Brief History R: Past and Future History, Ross Ihaka, Statistics Department, The University of Auckland, Auckland, New Zealand;
- [65] Albanese, D., Filosi, M., Visintainer, R., Riccadonna, S., Jurman, G., Furlanello, C., “Cmine Minerva & Minepy: A C Engine for the MINE Suite an its R and Python Wrappers”.
- [66] IEEE RTS Task Force of APM Subcommittee, “IEEE Reliability Test System - 1996”, IEEE PAS, vol. 14, no.3, pp. 1010-1020, August 1999.

- [67] Boehme, T., Taylor, J., Wallace, R., Bialek, J., “Academic Study: Matching Renewable Electricity Generation with Demand: Full Report”, Abr. 2006.
- [68] Hollander, M., Wolfe., D. A., “Nonparametric Statistical Methods”, New York: John Wiley & Sons. Pages 27–33 (one-sample), 68–75 (two-sample).
- [69] Site do National Oceanic and Atmospheric Administration , NOAA. Disponível em: <<http://www.cpc.ncep.noaa.gov/>>. Acessado em 25 de Julho de 2014.
- [70] Site do Operador Nacional do Sistema. Disponível em: <<http://www.ons.org.br/>>. Acessado em 1 de Julho de 2014.
- [71] Site do Centro de Pesquisas de Energia Elétrica. Disponível em: <<http://www.cepel.br/>>. Acessado em 1 de Julho de 2014.
- [72] Site da PSR. Disponível em: <<http://www.psr-inc.com.br/>>. Acessado em 1 de Julho de 2014.
- [73] Siegel S., Castellan N. J., “Nonparametric Statistics for the Behavioural Sciences”, 2nd Edition, McGraw-Hill, 1988.