



Bruno Alves de Sá Manso

**Avaliação da Confiabilidade de Sistemas de Geração e
Transmissão via Simulação Monte Carlo Otimizada por
Técnicas de Redução de Variância e Aprendizado de Máquina**

Tese de Doutorado

Tese de doutorado apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio.

Orientador: Armando Martins Leite da Silva

Coorientador: André Milhorange de Castro

Coorientador: Fernando Aparecido de Assis

Rio de Janeiro
Abril de 2025



Bruno Alves de Sá Manso

**Avaliação da Confiabilidade de Sistemas de Geração e
Transmissão via Simulação Monte Carlo Otimizada por
Técnicas de Redução de Variância e Aprendizado de Máquina**

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

Prof. Armando Martins Leite da Silva

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. André Milhorange de Castro

Coorientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Fernando Aparecido de Assis

Coorientador

Universidade Federal de São João del-Rei – UFSJ

Prof. Mauro Augusto da Rosa

Universidade Federal de Santa Catarina – UFSC

Prof. Leonidas Chaves de Resende

Universidade Federal de São João del-Rei – UFSJ

Prof. José Filho da Costa Castro

Universidade Federal de Pernambuco – UFPE

Prof. João Guilherme de Carvalho Costa

Universidade Federal de Itajubá – UNIFEI

Rio de Janeiro, 24 de abril de 2025

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

Bruno Alves de Sá Manso

Graduou-se em Engenharia Elétrica na Universidade Federal de São João del-Rei (UFSJ) em 2018. Obteve o título de mestre em Engenharia Elétrica em 2021 pelo Programa de Pós-Graduação em Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro.

Ficha Catalográfica

Manso, Bruno Alves de Sá

Avaliação da confiabilidade de sistemas de geração e transmissão via Simulação Monte Carlo otimizada por técnicas de redução de variância e aprendizado de máquina / Bruno Alves de Sá Manso ; orientador: Armando Martins Leite da Silva ; coorientadores: André Milhorange de Castro, Fernando Aparecido de Assis. – 2025.

144 f. : il. color. ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2025.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Amostragem por importância. 3. Aprendizado de máquina. 4. Confiabilidade composta. 5. Entropia cruzada. 6. Simulação de eventos raros. I. Silva, Armando Martins Leite da. II. Castro, André Milhorange de. III. Assis, Fernando Aparecido de. IV. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. V. Título.

CDD: 621.3

*Dedico esta tese a meus pais e meu irmão:
Luiz, Ana e Lucas.*

Agradecimentos

Primeiramente, agradeço a meu orientador Prof. Armando Martins Leite da Silva por ter me recebido como aluno mais uma vez, por todo suporte e orientação no desenvolvimento da tese, pelos ensinamentos e pelas histórias.

A meus coorientadores, Prof. André Milhorange de Castro e Prof. Fernando Aparecido de Assis, pelo suporte, cada um em uma parte essencial da tese.

A meus pais e meu irmão, Luiz, Ana e Lucas, gostaria de expressar minha enorme gratidão pelo apoio, incentivo e momentos em família.

Agradeço, ainda, a todos os professores, colegas, funcionários e amigos que estiveram presentes durante a minha formação acadêmica, na UFSJ e na PUC-Rio.

Por fim, agradeço à PUC-Rio pela oportunidade, e à CAPES pelo apoio financeiro.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

Manso, Bruno Alves de Sá; Leite da Silva, Armando Martins (Orientador); Castro, André Milhorce de (Coorientador); Assis, Fernando Aparecido de (Coorientador). **Avaliação da Confiabilidade de Sistemas de Geração e Transmissão via Simulação Monte Carlo Otimizada por Técnicas de Redução de Variância e Aprendizado de Máquina**. Rio de Janeiro, 2025. 144p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A simulação Monte Carlo (SMC) é uma ferramenta eficiente na análise da confiabilidade de grandes sistemas compostos de geração e transmissão. Na prática, porém, os planejadores têm maior interesse em avaliar apenas suas regiões de atuação, que geralmente englobam um conjunto reduzido de barras consumidoras. Nesse caso, a falha de suprimento (i.e., corte de carga) é muitas vezes um evento raro, o que eleva substancialmente o custo computacional da SMC. Para lidar com essa situação, o emprego de técnicas de redução de variância foi proposto em inúmeros trabalhos, os quais apresentam bons resultados, mas ainda encontram alguma dificuldade em modelar as restrições físicas da rede de transmissão. Técnicas de inteligência computacional também foram empregadas no intuito de aumentar a eficiência de algoritmos de confiabilidade composta. A abordagem por aprendizado de máquina engloba a maioria dessas técnicas. Em geral, sua tarefa é reduzir diretamente execuções de análise de adequação de estados por meio de classificadores. Tais classificadores precisam ser “treinados” a partir de amostras com boa representação da falha, o que restringe sua aplicação, tendo em vista que a falha de interesse dos agentes de um sistema é, em geral, um evento raro. Neste trabalho, propõe-se uma nova aplicação da técnica amostragem por importância, via método da entropia cruzada, que modela adequadamente as restrições da rede de transmissão. Adicionalmente, buscando melhorar ainda mais o desempenho da ferramenta, propõe-se acoplar um classificador de estados para diferenciá-los em estados de falha ou de sucesso estressado (i.e., estados de sucesso em que o sistema/área/barra está próximo de falhar). A técnica de simulação empregada é a SMC quasi-sequencial com o intuito de devidamente modelar fontes renováveis. O método proposto é avaliado por meio de uma versão modificada do IEEE Reliability Test System – 1996, que apresenta características de falha composta semelhantes às de sistemas reais.

Palavras-Chave

Amostragem por importância; aprendizado de máquina; confiabilidade composta; entropia cruzada; simulação de eventos raros; simulação Monte Carlo; sistemas de geração e transmissão.

Abstract

Manso, Bruno Alves de Sá; Leite da Silva, Armando Martins (Advisor); Castro, André Milhorange de (Co-Advisor); Assis, Fernando Aparecido de (Co-Advisor). **Reliability Assessment of Generation and Transmission Systems via Monte Carlo Simulation Optimized by Variance Reduction Techniques and Machine Learning**. Rio de Janeiro, 2025. 144p. Doctoral Thesis – Department of Electrical Engineering, Pontific Catholic University of Rio de Janeiro.

Monte Carlo simulation (MCS) is an efficient tool in the reliability analysis of large composite generation and transmission systems. In practice, however, planners take greater interest in evaluating only their operating regions, which generally encompasses a reduced set of consumer buses. In this case, supply failure (i.e., load shedding) is often a rare event, which substantially increases the computational cost of the MCS. To deal with this situation, the use of variance reduction techniques was proposed in numerous works, which present good results, but still find some difficulty in modeling the physical restrictions of the transmission network. Computational intelligence techniques have also been employed in order to increase the efficiency of composite reliability algorithms. Machine learning approach encompasses most of these techniques. Generally, its task is to directly reduce the execution of state adequacy analysis through classifiers. Such classifiers need to be “trained” from samples with good representation of the failure, which restricts its application, given that the failure of interest of system agents is, in general, a rare event. In this work, a new application of the importance sampling technique, via cross-entropy method, that adequately models transmission network restrictions, is proposed. Additionally, seeking to further improve the tool’s performance, it is proposed to couple a state classifier in order to differentiate states between failure or strained success (i.e., success states in which system/area/bus is close to failure). The simulation technique employed is the quasi-sequential MCS with the purpose of properly modeling renewable sources. The method is evaluated by the means of a modified version of the IEEE Reliability Test System – 1996, which presents composite failure characteristics as those of real systems.

Keywords

Composite reliability; cross-entropy; generation and transmission systems; importance sampling; machine learning; Monte Carlo simulation; rare event simulation.

Sumário

1 . Introdução	21
1.1. Considerações Iniciais	21
1.2. Desenvolvimento Histórico	24
1.2.1. Aprendizado de Máquina	26
1.2.2. Redução de Variância	28
1.2.3. Técnicas Combinadas	32
1.3. Estrutura do Documento	33
2 . Avaliação da Confiabilidade de Sistemas de Geração e Transmissão	35
2.1. Introdução	35
2.2. Conjunto Geração-Transmissão	35
2.3. Principais Métodos de Avaliação	37
2.4. Simulação Monte Carlo Quasi-Sequencial	39
2.5. Modelagem do Sistema	40
2.5.1. Modelo da Carga	40
2.5.2. Modelo da Energia Renovável	41
2.5.3. Modelo da Disponibilidade de Equipamentos	42
2.6. Índices de Confiabilidade	44
2.7. Análise de Adequação de Estados	47
2.8. Conclusões	50
3 . Confiabilidade Composta Otimizada por Redução de Variância	51
3.1. Introdução	51
3.2. Método da Entropia Cruzada	51
3.2.1. Simulação de Evento Raro	52
3.2.2. Entropia Cruzada	55
3.2.3. Família Exponencial Natural	56
3.2.4. Níveis Múltiplos	59
3.2.5. Algoritmo	60
3.3. Método da Entropia Cruzada em Confiabilidade	61

3.3.1. Confiabilidade de Geração	62
3.3.2. Confiabilidade Composta	63
3.3.3. Proposta	65
3.3.4. Algoritmo	68
3.4. Avaliação do Método Proposto	69
3.4.1. Sistema RTS-96-REN-T	70
3.4.2. Análise de Sistema	73
3.4.3. Análise de Zona	73
3.4.4. Análise de Barra	75
3.4.5. Estudo da Dispersão da Falha	77
3.5. Conclusões	85
 4 . Confiabilidade Composta Otimizada por Aprendizado de Máquina	 87
4.1. Introdução	87
4.2. Regressão Logística Binária	88
4.2.1. Função Logística	88
4.2.2. Ajuste do Classificador	90
4.2.3. Algoritmo	94
4.3. Proposta	95
4.3.1. Variáveis Explicativas	96
4.3.2. Amostra de Treinamento e Limiar de Classificação	97
4.3.3. Algoritmo	99
4.3.4. Complementariedade entre VRT e Classificador	100
4.4. Avaliação do Método Proposto	102
4.4.1. Análise de Sistema	103
4.4.2. Análise de Zona	104
4.4.3. Análise de Barra	105
4.4.4. Estudo da Dispersão da Falha	107
4.5. Redes Neurais Multicamadas	111
4.5.1. Rede GMDH	111
4.5.2. Implementação	113
4.5.3. Resultados	114
4.6. Conclusões	115

5 . Conclusões e Propostas de Trabalhos Futuros	117
5.1. Conclusões	117
5.2. Propostas de Trabalhos Futuros	120
6 . Referências Bibliográficas	121
Apêndice A. Dados do Sistema RTS-96-REN-T	133
A.1. Dados de Barra	133
A.2. Dados de Estações Geradoras	134
A.3. Dados de Circuito	135
A.4. Séries Hídricas	138
Apêndice B. Exemplo de Rede BLR – Sistema RTS-96-REN-T	140

Lista de Figuras

Figura 2.1: Níveis hierárquicos de sistemas em confiabilidade.	35
Figura 2.2: Modelo de Markov multiestados não agregado.	41
Figura 2.3: (a) Modelo de Markov a dois estados; (b) Modelo de Markov multiestados para unidade equivalente.	42
Figura 3.1: Diagrama representativo do sistema RTS-96.	72
Figura 3.2: Diagrama representativo da Área 3 do sistema RTS-96-REN-T.	74
Figura 4.1: Fluxograma do método completo.	99
Figura 4.2: Exemplo ilustrativo de topologia de uma rede GMDH.	113

Lista de Tabelas

Tabela 3.1: Análise do sistema (RTS-96).	71
Tabela 3.2: Substituição de unidades térmicas por renováveis.	72
Tabela 3.3: Análise do sistema (RTS-96-REN-T).	73
Tabela 3.4: Análise da zona A3SUL (RTS-96-REN-T).	74
Tabela 3.5: Análise da Barra 301 (RTS-96-REN-T).	75
Tabela 3.6: Análises da Barra 301 (RTS-96-REN-T e RTS-96-REN-T#1).	78
Tabela 3.7: Comparação de lista ranqueada para a Barra 301.	79
Tabela 3.8: Comparação de lista ranqueada para a Barra 307.	81
Tabela 3.9: Análises da Barra 307 (RTS-96-REN-T e RTS-96-REN-T#2).	82
Tabela 3.10: Comparação de lista ranqueada para a Barra 320.	83
Tabela 3.11: Análises da Barra 320 (RTS-96-REN-T e RTS-96-REN-T#3).	84
Tabela 4.1: Análise do sistema (RTS-96-REN-T).	104
Tabela 4.2: Análise da zona A3SUL (RTS-96-REN-T).	105
Tabela 4.3: Análise da Barra 301 (RTS-96-REN-T).	106
Tabela 4.4: Análise da Barra 301 (RTS-96-REN-T#1).	107
Tabela 4.5: Análise da Barra 307 (RTS-96-REN-T).	108
Tabela 4.6: Análise da Barra 307 (RTS-96-REN-T#2).	109
Tabela 4.7: Análise da Barra 320 (RTS-96-REN-T).	110
Tabela 4.8: Análise da Barra 320 (RTS-96-REN-T#3).	110
Tabela 4.9: Análises do sistema e da zona A3SUL (RTS-96-REN-T).	114

Tabela 4.10: Análises das barras 301, 307 e 320 (RTS-96-REN-T).	115
Tabela A.1: Dados de Barra.	133
Tabela A.2: Dados de Estações Geradoras.	134
Tabela A.3: Dados de Circuito.	135
Tabela A.4: Localização dos Reservatórios.	138
Tabela A.5: Séries Hídricas (aproveitamento mensal em pu).	138
Tabela B.1: Lista ranqueada para o sistema RTS-96-REN-T.	140
Tabela B.2: Variáveis explicativas geradas para o sistema RTS-96-REN-T.	142

Lista de Siglas e Abreviaturas

AC	<i>alternate current</i>
ANN	<i>artificial neural networks</i>
BGC	balanço de geração-carga
BLR	<i>binary logistic regression</i>
CE	<i>cross-entropy</i>
CNN	<i>convolutional neural networks</i>
COPT	<i>capacity outage probability table</i>
<i>CUI</i>	custo unitário médio de interrupção
DC	<i>direct current</i>
EENS	<i>expected energy not supplied</i>
EPNS	<i>expected power not supplied</i>
FACTS	<i>flexible alternating current transmission system</i>
FDCP	fluxo de potência DC com perdas
F&D	frequência e duração
FEN	família exponencial natural
<i>FOR</i>	<i>forced outage rate</i>
FPO	fluxo de potência ótimo
GLPK	GNU Linear Programming Kit
GMDH	<i>group method of data handling</i>
GNN	<i>graph neural networks</i>
GT	geração e transmissão
IS	<i>importance sampling</i>
KNN	<i>k-nearest neighbor</i>
LF	<i>likelihood function</i>
LOLC	<i>loss of load cost</i>
LOLD	<i>loss of load duration</i>
LOLE	<i>loss of load expectation</i>
LOLF	<i>loss of load frequency</i>
LOLP	<i>loss of load probability</i>

LR	<i>logistic regression</i>
LVQ	<i>learning vector quantization</i>
MLE	<i>maximum likelihood estimation</i>
MLP	<i>multilayer perceptron</i>
NH0	nível hierárquico 0
NH1	nível hierárquico 1
NH2	nível hierárquico 2
NH3	nível hierárquico 3
<i>NIT</i>	número de instantes de tempo
<i>NP</i>	<i>number of periods</i>
<i>NS</i>	<i>number of samples</i>
<i>NC</i>	<i>number of components</i>
OSF	<i>one-step forward</i>
pdf	<i>probability density function</i>
pdf-IS	<i>importance sampling probability density function</i>
pu	por unidade
RBF	<i>radial basis function</i>
RTS-79	IEEE Reliability Test System – 1979
RTS-96	IEEE Reliability Test System – 1996
SEP	sistemas elétricos de potência
SMC	simulação Monte Carlo
SMCC	simulação Monte Carlo convencional
SMCIS	simulação Monte Carlo com <i>importance sampling</i>
SMCIS-BLR	simulação Monte Carlo quasi-sequencial com <i>importance sampling</i> via método <i>cross-entropy</i> e classificador <i>binary logistic regression</i> para sistemas de geração e transmissão
SMCIS-G	simulação Monte Carlo quasi-sequencial com <i>importance sampling</i> via método <i>cross-entropy</i> com foco em problemas de geração
SMCIS-GMDH	simulação Monte Carlo quasi-sequencial com <i>importance sampling</i> via método <i>cross-entropy</i> e classificador <i>group method of data handling</i> para sistemas de geração e transmissão

SMCIS-GT	simulação Monte Carlo quasi-sequencial com <i>importance sampling</i> via método <i>cross-entropy</i> para sistemas de geração e transmissão
SMC-NS	simulação Monte Carlo não-sequencial
SMC-QS	simulação Monte Carlo quasi-sequencial
SMC-S	simulação Monte Carlo sequencial
SOM	<i>self-organizing maps</i>
SS	<i>subset simulation</i>
SVM	<i>support vector machine</i>
VM	<i>variance minimization</i>
VRT	<i>variance reduction technique</i>

Lista de Símbolos

U	probabilidade de estar indisponível
$\mathbf{X}_k$	estado k do sistema
$X_{k,j}$	estado k do j -ésimo componente do sistema
$\mathcal{X}$	espaço de estados
$F(\cdot)$	função teste (representação a espaço de estados)
$\hat{E}[\cdot]$	estimativa de valor esperado
$\mathbf{Y}_k$	série temporal que descreve todos os estados resididos pelo sistema em um período k de simulação
$G(\cdot)$	função teste (representação cronológica)
T	período
t	instante de tempo
β	coeficiente de incerteza relativa
$\hat{V}(\cdot)$	variância amostral
Δt	duração de um estado (modelo de Markov multiestados não agregado)
$\lambda_{\Delta t}$	taxa de transição entre estados (modelo de Markov multiestados não agregado)
$L_b(t)$	demanda da barra b no instante de tempo t
$XG_z(t, y)$	fator multiplicativo de capacidade de geradores renováveis da zona z no instante de tempo t do ano y
C_k	capacidade disponível para o estado com k unidades fora de operação (modelo de Markov multiestados para unidade equivalente)
C_{eqv}	capacidade disponível para o estado equivalente (modelo de Markov multiestados para unidade equivalente)
λ	taxa de falha (modelo de Markov a dois estados)
μ	taxa de reparo (modelo de Markov a dois estados)
A	probabilidade de estar disponível
λ_{ij}	taxa de transição entre dois estados i e j quaisquer (modelo de Markov multiestados para unidade equivalente)

$F_{LOLP}(\cdot)$	função teste da LOLP (representação a espaço de estados)
$F_{EPNS}(\cdot)$	função teste da EPNS (representação a espaço de estados)
$F_{LOLF}(\cdot)$	função teste da LOLF (representação a espaço de estados)
ΔP_k	montante de carga cortado no estado k
λ_k	soma das taxas de saída para todos os estados de sucesso que podem ser alcançados mediante uma transição
λ_{out}	soma das taxas de todas as transições de saída
Ω_k	conjunto de estados de sucesso que se comunicam ao estado k pela transição de apenas um componente do sistema
P_{km}	probabilidade de o sistema transitar do estado k para o m
θ_{ik}	abertura angular de tensão entre as barras i e k
$\{f(\cdot; \sigma)\}$	família paramétrica de funções densidade de probabilidade
σ	vetor real de parâmetros de uma função densidade de probabilidade
$\mathbf{u}$	vetor real de parâmetros da distribuição original do problema
$H(\cdot)$	função real
$\mathbb{E}[\cdot]$	valor esperado
ℓ	probabilidade do evento $\{S(\mathbf{X}) \geq \gamma\}$
$S(\mathbf{X})$	função real no espaço $\mathcal{X}$ (função performance do estado $\mathbf{X}$)
γ	número real
$I_{\{\cdot\}}$	avaliador de expressão lógica
$\mathbb{P}(\cdot)$	função de probabilidade
$\hat{\ell}$	estimador de ℓ
$\mathbf{v}$	vetor real de parâmetros de uma distribuição alternativa
$W(\cdot)$	razão de verossimilhança
$g^*(\cdot)$	distribuição amostragem por importância ótima
$\mathcal{D}(\cdot)$	divergência de Kullback-Leibler
$\mathbf{v}^*$	vetor real de parâmetros ótimo (caracteriza a distribuição amostragem por importância ótima)
$h(\cdot)$	função real positiva
$\eta(\cdot)$	função real
$t(\cdot)$	função real

$A(\cdot)$	função real
$K_X(\cdot)$	função geradora de cumulantes
$\widehat{\boldsymbol{v}}^*$	estimador de $\boldsymbol{v}^*$
$\widehat{v}_j^*$	estimador do j -ésimo elemento de $\boldsymbol{v}^*$
X_{ij}	realização do j -ésimo elemento do sistema para o i -ésimo estado amostrado
γ_k	k -ésimo nível de γ
ρ	parâmetro de níveis múltiplos ou probabilidade de $\{S(\boldsymbol{X}) \geq \gamma_k\}$
Ω^v	conjunto das variáveis aleatórias que compõem o problema estocástico
$GCAP(\boldsymbol{X})$	capacidade de geração disponível amostrada no estado $\boldsymbol{X}$
L^{pesada}	algum valor de carga pesada
$GCAP^{max}$	capacidade de geração máxima
$L^{corte}(\boldsymbol{X})$	corte de carga do estado $\boldsymbol{X}$
$L(\boldsymbol{X})$	carga amostrada no estado $\boldsymbol{X}$
β^{F1}	valor máximo para coeficientes de incerteza relativa da Fase 1 do método proposto (controla precisão da estimativa da pdf-IS)
β^{F2}	valor máximo para coeficientes de incerteza relativa da Fase 2 do método proposto (controla precisão das estimativas de índices de confiabilidade)
α	parâmetro para amortecimento de $\widehat{\boldsymbol{v}}^*$
β_{LOLP}	β da estimativa do índice LOLP
β_{LOLF}	β da estimativa do índice LOLF
β_{LOLP}^{F1}	β^{F1} para o índice LOLP
β_{LOLF}^{F1}	β^{F1} para o índice LOLF
$\boldsymbol{v}^G$	estimativa do vetor real de parâmetros de estações geradoras
v_j^G	estimativa do j -ésimo elemento de $\boldsymbol{v}^G$ (novo parâmetro probabilidade de estar indisponível dos geradores da estação j)
n_j^{uni}	número de geradores da estação j
X_{ij}^G	número de geradores indisponíveis na estação j
Ω^G	conjunto de estações geradoras
$\boldsymbol{v}^C$	estimativa do vetor real de parâmetros de circuitos

v_j^c	estimativa do j -ésimo elemento de $\mathbf{v}^c$ (novo parâmetro probabilidade de estar indisponível do circuito j)
X_{ij}^c	condição de indisponível do circuito j
Ω^c	conjunto de circuitos
Γ	coeficientes da função logística
$\mathbb{O}(\cdot)$	função chances (<i>odds function</i>)
$\mathbb{P}^{thr}$	limiar de classificação para a probabilidade de um dado evento
$\mathcal{L}(\cdot)$	função verossimilhança (<i>likelihood function</i>)
$\ell(\cdot)$	função log-verossimilhança (<i>log-likelihood function</i>)
v_j^{ref}	indisponibilidade-IS de referência
G^{total}	ganho total
G^{IS}	ganho da simulação-IS
G^{BLR}	ganho do classificador BLR
G^{BLR^*}	ganho real do classificador BLR
FPO^{SMCC}	número de FPO executados em uma SMCC
FPO^{SMCIS}	número de FPO executados em uma SMCIS
FPO^{F1}	número de FPO executados na Fase 1 de uma SMCIS-BLR
$FPO^{F2.1}$	número de FPO executados na Fase 2.1 de uma SMCIS-BLR
$FPO^{F2.2}$	número de FPO executados na Fase 2.2 de uma SMCIS-BLR
ϕ	aproveitamento percentual do ganho do classificador BLR

1. Introdução

1.1. Considerações Iniciais

O surgimento da energia elétrica, ao final do século 19, conduziu a humanidade ao seu período de maior transformação. No século 20, economias cada vez maiores e mais interligadas levaram a demanda por eletricidade a crescer continuamente e com isso a complexidade dos sistemas elétricos. A progressiva dependência da sociedade por energia elétrica lhe deu caráter de recurso básico, criando o imperativo de suprir a crescente demanda com qualidade, economia e sem interrupções. Evidentemente, os componentes de sistemas elétricos de potência (SEP) falham com alguma probabilidade, de modo que a eventual falha de suprimento (i.e., corte de carga), ainda que seja rara, é inevitável. Logo, não é difícil perceber que a confiabilidade é fator fundamental nos planos de operação e expansão de sistemas.

No século 21, a incorporação de novos recursos e tecnologias (e.g., *smart grids*, geração distribuída etc.) trouxe mais qualidade, economia e ecologia à geração e consumo de energia elétrica, mas também elevou ainda mais a complexidade de modelagem e análise de SEP. Os recursos energéticos renováveis intermitentes compõem muito provavelmente o fator de maior transformação dos sistemas modernos, trazendo muitas incertezas e restrições complexas aos modelos empregados por planejadores e operadores. Naturalmente, novas pesquisas em confiabilidade de SEP modelam, invariavelmente, a inserção de fontes intermitentes.

Ao passo que os SEP não apresentam mais uma clara estrutura verticalizada de geração, transmissão e distribuição, ainda é possível desagregar as partes de um sistema sob o ponto de vista de corte de carga. A distribuição, por exemplo, é a partição do sistema onde ocorre a maioria das falhas, as quais, porém, são localizadas, tendo impacto reduzido. Já no conjunto geração-transmissão (GT), as falhas são muito menos frequentes, mas, atingem, simultaneamente, grandes quantidades de consumidores, de variados tipos (residencial, comercial, industrial etc.), apresentando um impacto elevado e abrangente. Mesmo com o justo e necessário

crescente interesse nos sistemas de distribuição, muito em função da geração distribuída, o conjunto GT segue sendo o mais crítico e recebendo mais atenção dos planejadores de SEP.

No intuito de garantir suprimento ininterrupto, a geração e transmissão devem apresentar reserva operativa e estática a todo momento. Para exercer esse papel, operadores e planejadores trabalham mais com critérios determinísticos. No entanto, é reconhecido que os métodos probabilísticos são mais capacitados para quantificar adequadamente o risco de não suprimento de um sistema. A simulação Monte Carlo (SMC) apresenta elevada eficiência computacional e baixa sensibilidade ao tamanho do problema (i.e., espaço de estados operativos de um SEP). Contudo, isso não impede que esforços sejam destinados ao aprimoramento do seu desempenho. Considerando que grande parte do tempo de processamento de uma SMC se concentra na análise de adequação de estados – geralmente na forma de fluxo de potência linear/não linear convencional e ótimo – é fácil notar que técnicas capazes de reduzir a necessidade de tais análises devem, por consequência, melhorar o desempenho computacional da simulação. Na literatura, é possível encontrar inúmeros trabalhos com este intuito, os quais podem ser objetivamente divididos em: (i) por redução de variância do espaço amostral representativo da falha, como forma de reduzir o tamanho da amostra necessária a representar o evento falha; e (ii) por aprendizado de máquina, como forma de evitar diretamente execuções de análise de adequação.

Ambas as abordagens mencionadas para a otimização de algoritmos baseados em SMC foram extensivamente exploradas, tanto na confiabilidade da geração, quanto na composta (i.e., confiabilidade de sistemas compostos por geração e transmissão). Na confiabilidade composta, porém, apesar de bons resultados, as metodologias propostas, em ambas as vertentes, ainda encontram restrições. Na abordagem por aprendizado de máquina, um classificador precisa ser treinado para poder qualificar estados amostrados sem empregar análise de adequação, sendo necessário obter um montante considerável de amostras da falha em estudo. Isso é um problema quando a falha é um evento raro, situação comumente encontrada na prática, em que a análise é voltada para conjuntos reduzidos de barras consumidoras. Em sistemas de grande porte, é muito provável que a quase todo momento seja possível observar falha no suprimento de carga em algum local da rede. O sistema interligado brasileiro, por exemplo, apresentava, em 2014, probabilidade de falha

virtualmente igual a 1,0 pu [1]. Isso demonstra que avaliar a adequação de suprimento global de um sistema real de grande porte não traz benefícios relevantes aos processos decisórios de planejamento da operação ou expansão. Aliás, os agentes dos sistemas estão interessados em conhecer a resiliência apenas de sua região de atuação. Logo, o foco da avaliação da confiabilidade de sistemas GT fica direcionado para conjuntos reduzidos de barras consumidoras. A falha observada por área ou barramento é, invariavelmente, um evento raro, o que leva a uma simulação onerosa, dificultando o treinamento de classificadores.

As técnicas baseadas em redução de variância (VRTs – *variance reduction techniques*) apresentam, por sua vez, maior capacidade de lidar com o problema de raridade no evento de interesse. Elas configuram, portanto, a abordagem com maior potencial e aplicabilidade em confiabilidade de sistemas GT, sendo a amostragem por importância (IS – *importance sampling*) via método da entropia cruzada (CE – *cross-entropy*) a mais proeminente. Contudo, os métodos empregados em confiabilidade composta ainda encontram dificuldade em modelar problemas que envolvem as restrições físicas da rede de transmissão. Um problema presente tanto em classificadores quanto em VRTs.

A motivação desta tese de doutorado é propor um novo método eficiente para avaliação de confiabilidade composta de falha rara, por meio de um plano de duas partes. Primeiro, propõe-se desenvolver uma nova ferramenta de simulação Monte Carlo auxiliada por amostragem por importância, via método da entropia cruzada, que considera adequadamente as restrições da rede de transmissão. Adicionalmente, buscando elevar ainda mais o desempenho da ferramenta, propõe-se acoplar um classificador para diferenciar os estados que apresentam problemas (i.e., violações de limites operativos via fluxo de potência convencional) em estado de falha ou de sucesso estressado (i.e., estados de sucesso em que o sistema/área/barra está próximo de falhar). Assim, é possível evitar a execução de fluxo de potência ótimo para estados de sucesso estressado, para os quais se sabe que a otimização de medidas remediadoras resolverá os problemas. Por fim, destaca-se que recursos energéticos intermitentes são modelados e que, para lidar com aspectos cronológicos sem perda de eficiência computacional, a SMC utilizada é do tipo quasi-sequencial.

1.2. Desenvolvimento Histórico

A aplicação de métodos probabilísticos na avaliação da confiabilidade de SEP está bem documentada em dez amplas bibliografias publicadas em: 1966 [2], 1972 [3], 1978 [4], 1984 [5], 1988 [6], 1989 [7], 1994 [8], 1999 [9], 2001 [10] e 2022 [11]. A maioria dos trabalhos encontrados na literatura é voltada para sistemas de geração e/ou transmissão. Com relação aos sistemas de distribuição, cabe destacar que, em decorrência do impacto localizado e isolado de suas falhas, a sua análise é normalmente realizada separadamente, sem considerar o conjunto GT.

De modo geral, pode-se afirmar que os primeiros métodos probabilísticos desenvolvidos para auxiliar os planejadores eram voltados para a avaliação da confiabilidade de sistemas de geração e permitiam calcular somente índices de perda de carga como a LOLE (*loss of load expectation*) e a EENS (*expected energy not supplied*) [12]. Basicamente, esses métodos combinavam em uma tabela de probabilidade os diferentes níveis de capacidade indisponível (COPT – *capacity outage probability table*) com um modelo apropriado da curva de carga do sistema, para obter os índices de confiabilidade. Esses métodos, denominados analíticos, foram largamente utilizados no passado para o dimensionamento da reserva estática de geração. Posteriormente, os métodos probabilísticos evoluíram de modo a permitir o cálculo de índices de frequência e duração (F&D), sendo [13] a primeira referência encontrada nesse tema. O trabalho [14], que inclui uma revisão de artigos com foco em índices F&D, propôs uma formulação generalizada para os modelos de geração e carga, a qual prescinde da propriedade do equilíbrio de frequência. Dessa forma, foi possível reproduzir com maior precisão o comportamento da carga, utilizando um modelo de Markov com múltiplos estados, desequilibrados em frequência. Ademais, viabilizou a avaliação de índices F&D para sistemas reais de grande porte.

Cabe destacar que até meados do século 20, os sistemas apresentavam dimensões não elevadas, operando de forma isolada. Com a expansão das redes de transmissão e/ou a adoção de operação interligada, foi necessário desenvolver técnicas mais eficientes. Ao mesmo tempo, o surgimento de computadores com maior capacidade de memória e processamento viabilizou o estudo mais detalhado dos SEP. Neste contexto, é na década de 60 que a avaliação da confiabilidade de sistemas passa a incluir a rede de transmissão [7], sendo utilizadas técnicas analíticas, via

processos de Markov. Um trabalho mais conceitual foi apresentado em [15], sendo realizada uma comparação entre os resultados obtidos por este, e os resultados alcançados utilizando as técnicas apresentadas em [16]. No entanto, é apenas a partir da década de 80 que ocorre um aumento mais significativo do interesse nessa área. Questões como o acesso aberto à transmissão por produtores independentes, o aumento do intercâmbio econômico entre as concessionárias e as limitações na construção de novas redes de transmissão, deveriam ser tratadas através de métodos que levassem em consideração os efeitos de geração e transmissão simultaneamente. A referência [17] apresenta uma análise dos principais métodos utilizados na avaliação da confiabilidade de sistemas compostos de geração e transmissão até o início da década de 90.

Tendo em vista a confiabilidade composta de sistemas de grande porte, fortemente conectados, os métodos de simulação se provam notoriamente superiores aos analíticos [1], [17], o que é ainda mais notável se a cronologia da carga e recursos renováveis intermitentes são considerados. Esses métodos são divididos em representação cronológica (SMC sequencial – SMC-S) e a espaço de estados (SMC não-sequencial – SMC-NS). A SMC-S tem maior capacidade de representação, mas custo computacional elevado. Buscando aliar as qualidades da SMC-S com a velocidade da SMC-NS, foi desenvolvida a ferramenta SMC pseudo-sequencial [18], depois aprimorada para a pseudo-cronológica [19]. A fim de obter uma flexibilidade ainda maior, especialmente com a consideração de renováveis intermitentes, a SMC quasi-sequencial (SMC-QS) foi proposta em [20]. Em relação à SMC-S, a SMC-QS só não é capaz de calcular distribuições de probabilidade para os índices de confiabilidade.

O algoritmo geral para avaliação da confiabilidade de sistemas GT é o mesmo para avaliação da geração. Basta selecionar um estado, analisar sua adequação e atualizar a estimativa de índices. No entanto, a confiabilidade composta sofre com um elevado custo computacional na etapa de análise de adequação de estados, a qual é muito mais complexa por depender de estudos de fluxo de potência com otimização de medidas corretivas. O problema fica ainda maior ao considerar as análises adicionais, necessárias ao processo de estimação dos índices F&D. Para essa questão, foram desenvolvidas metodologias: [21-24], entre outras, que visam evitar ou reduzir o volume das análises adicionais. Destaque deve ser dado para a técnica de transição um passo à frente, ou *one-step forward* (OSF) [24], a qual não

depende de hipótese de coerência para o comportamento da falha, o que confere à técnica OSF flexibilidade para lidar com qualquer composição de curva de carga para as barras consumidoras do sistema, bem como com a intermitência de fontes renováveis não despacháveis.

Mesmo atenuando o problema de análises extras para estimação de índices de frequência ao ponto em que isso deixa de ser uma preocupação, a condição de esforço computacional elevado na confiabilidade de sistemas GT permaneceu. Em decorrência disso, basicamente desde a década de 90, pesquisadores buscam otimizar algoritmos de simulação em confiabilidade composta. Majoritariamente, as pesquisas exploraram duas vertentes: (i) aprendizado de máquina; e (ii) técnicas de redução de variância. Outras técnicas isoladas, mas interessantes, são: processamento paralelo [25]; equivalentes para adequação de sistemas de grande porte [26], programação linear multiparamétrica (*multiparametric linear programming*) [27] e algoritmos de busca populacional inteligente (*population-based intelligent search*) [28-30].

1.2.1.

Aprendizado de Máquina

Métodos baseados em aprendizado de máquina para a avaliação de confiabilidade de SEP são empregados, em termos gerais, para classificar estados amostrados como de falha/sucesso sem executar análise de adequação. Nesse âmbito, têm-se trabalhos com redes neurais artificiais (ANN – *artificial neural networks*) tradicionais, e outras menos comuns, bem como algoritmos baseados em outros conceitos. Independentemente do método, os padrões de interesse devem ser identificados por um classificador, o qual deve aprender a associar características selecionadas de entrada a um determinado padrão/classe de saída do fenômeno em estudo, seja o padrão/classe pré-estabelecido ou não. Os classificadores podem ser objetivamente diferenciados em de aprendizado “supervisionado” e “não-supervisionado”.

Classificadores de aprendizado supervisionado aprendem a função que mapeia algum problema com base na relação de pares de entrada-saída já conhecidos/simulados. Em confiabilidade de SEP, o aprendizado supervisionado apresenta uma extensa e variada gama de trabalhos. Dentre as ANN, têm-se as redes *group method of data handling* (GMDH) [31-32], *convolutional neural networks* (CNN) [33-34], *graph neural networks* (GNN) [35-36], *learning vector quantization* (LVQ) [37] e *radial basis function* (RBF) [38]. Em [31], faz-se uso das redes

polinomiais construtivas GMDH para a detecção de estados de sucesso, enquanto em [32], cujo foco de avaliação é a saúde de sistemas GT, a rede necessita também de classificar estados de sucesso como saudáveis e marginais. As redes CNN e GNN, que pertencem à área chamada *deep learning* [39], são mais atuais e compõem o estado da arte de redes neurais artificiais. Em [33], os autores propõem o uso de CNN acompanhada de uma técnica chamada *transfer learning*, de modo a possibilitar um rápido retreino de uma rede previamente utilizada. Essa aplicação é interessante quando se quer avaliar um sistema sob alterações de parâmetros (e.g., análise de sensibilidade). Já em [34] e [36], a CNN e GNN, respectivamente, são empregadas com intuito de não apenas classificar um estado como de sucesso/falha, mas também quantificar o corte de carga, evitando a análise de adequação para todos os estados, inclusive os de falha. No trabalho [35], por sua vez, a rede GNN é incumbida de prever diretamente a estimativa de algum índice de confiabilidade, uma tarefa muito ambiciosa. A motivação, nesse caso, é de se obter rápidas estimativas ao se avaliar um sistema sob mudanças de topologia causadas por manutenção emergencial de linhas de transmissão.

Com semelhante volume e variedade, os classificadores de aprendizado supervisionado, que não são ANN, aparecem em trabalhos na forma de técnicas como *support vector machine* (SVM) [40-43], *k-nearest neighbor* (KNN) [44] e *logistic regression* (LR) [45]. A ferramenta SVM é reconhecida pela sua capacidade de generalização, de modo que a sua aplicação em confiabilidade de sistemas GT complexos é bem-vinda e os trabalhos citados apresentam bons resultados. Em [44], a KNN é aplicada com o objetivo de evitar a análise de adequação de estados por completo, inclusive para estados de falha, mas sem quantificar o corte de carga. Assim, tem-se uma análise restrita ao índice de probabilidade, apresentando, portanto, potencial de aplicação limitado. Já em [45] a regressão logística binária foi empregada, apresentando bons resultados para uma versão equivalente de um subsistema da região sul do Brasil, que configura um sistema real de grande porte e rede de transmissão bem conectada. Vale destacar que a LR, cujo desempenho se mostra promissor, é um classificador de conceito e implementação simples, e, portanto, elevada aplicabilidade.

No aprendizado não-supervisionado, os classificadores devem aprender padrões que diferenciam grupos de elementos de um banco de dados sem classificação pré-estabelecida pelo usuário, que, por sua vez, é o responsável por interpretar os

grupos formados e designar categorias/rótulos a eles. Isso é muito útil quando se quer catalogar um banco de dados em múltiplas classes, mas pouco atraente quando o foco é acentuado em uma classe apenas. Em confiabilidade de SEP, por mais que a falha e o sucesso possam ser fragmentados em níveis de criticidade, o interesse é, em geral, exclusivo pelos estados de falha. Adicionalmente, o aprendizado supervisionado é menos sensível a dados de entrada não expostos ao classificador durante o seu treinamento, sendo o mais indicado para um aprendizado promovido com poucas amostras. Naturalmente, o aprendizado não-supervisionado foi menos explorado. Boa parte dos trabalhos publicados passa pela ANN popularizada como mapas de Kohonen, ou *self-organizing maps* (SOM), como em [46-49], e em menor proporção por outras técnicas como *k-means* e *k-medoids*. Em publicação recente, o artigo [50] traz um estudo dos três métodos citados: SOM, *k-means* e *k-medoids*; com a finalidade de exemplificar a praticidade do emprego de classificadores não-supervisionados simples no aumento de eficiência da SMC para análise de sistemas GT.

1.2.2.

Redução de Variância

Todas as aplicações de aprendizado de máquina citadas são afligidas pelo mesmo problema: a raridade no evento falha. Independente da técnica e objetivo, o uso de classificadores em confiabilidade de SEP implica a necessidade de captar um montante considerável de estados de falha para o aprendizado do padrão de interesse. Se a falha em estudo é rara, o treinamento por uma simulação breve (i.e., com uma amostra pequena) se torna impraticável. Outra vertente de pesquisa que busca otimizar algoritmos em confiabilidade, e tem potencial para lidar com o problema de falha rara, é a de técnicas de redução de variância. As VRTs buscam reduzir o esforço computacional de métodos de simulação por meio da redução da amostra necessária a representar o evento de interesse com a precisão desejada. Na análise de SEP, isso significa estimar índices de confiabilidade via SMC com um número reduzido de amostras. Bem como ocorre com aprendizado de máquina, uma extensa gama de trabalhos que empregam VRTs é encontrada na literatura: variáveis de controle e antitéticas; poda do espaço de estados; amostragem estratificada; amostragem condicional etc. No entanto, a que encontrou terreno de pesquisa mais fértil e entregou melhores resultados é a amostragem por importância, especialmente via método da entropia cruzada.

Duas das técnicas mais utilizadas no início do emprego de VRTs em confiabilidade de SEP são variáveis de controle e antitéticas. Ambas se baseiam na correlação entre variáveis aleatórias para alcançar a desejada redução de variância daquelas que se quer estimar parâmetros. Pela sua afinidade, as técnicas foram apreciadas em conjunto nos trabalhos [51-52], que avaliam confiabilidade composta com SMC sequencial. As variáveis de controle foram aplicadas assumindo forte correlação positiva entre o corte por déficit de geração (variável explicativa, ou seja, de controle) com o corte por efeito composto de geração e transmissão (variável em estudo). Essa abordagem tende a não funcionar em sistemas reais interligados onde a falha por déficit de geração é basicamente inexistente e pouco explica como se dão os “gargalos” da rede. Com isso em mente, os trabalhos [53-54] propõem o uso da geração e da transmissão para configurar variáveis de regressão (i.e., de controle). As variáveis da transmissão estão atreladas às violações de limites operativos observadas em fluxo de potência, dada uma configuração da geração. Essa nova abordagem parece sensata, contudo, em fluxo de potência, eventuais violações estão fortemente associadas à escolha dos fatores de participação dos geradores utilizados na construção de um caso base (e.g., carga pico e sistema intacto), não sendo essas violações necessariamente explicativas das limitações da rede no estado em análise.

As variáveis antitéticas, por sua vez, costumam funcionar melhor, especialmente em aplicações de SMC sequencial. A proposta é estimar parâmetros pela média de um par de realizações antitéticas (i.e., de correlação negativa) da variável aleatória de saída. Em [51-52], a cada ano amostrado na simulação sequencial outro é gerado com os complementos de cada ciclo falha/reparo sorteado para os equipamentos (i.e., se um ciclo é sorteado com probabilidade U , seu complemento é gerado com probabilidade $1 - U$). Dessa forma, os índices de confiabilidade podem ser estimados pela média de dois anos de correlação negativa, o que mitiga o impacto da dispersão das distribuições dos índices na simulação e leva as estimativas a evoluírem mais próximas aos respectivos valores esperados. As médias amostrais dos índices tendem para os mesmos valores que as de uma simulação simples, enquanto as variâncias são reduzidas. Assim, tem-se uma VRT de impacto ainda pequeno, mas eficaz, seja na análise de confiabilidade de geração ou composta. No entanto, são as variáveis de controle que aparecem em trabalhos mais recentes, como em [55-56]. Em [55], a técnica de variáveis de controle é combinada com outra VRT, a IS via método CE, enquanto, em [56], ela é utilizada em conjunto com

aprendizado de máquina para análise de confiabilidade no planejamento da operação um dia à frente.

A técnica de poda do espaço de estados é uma VRT que busca reduzir o espaço de estados de um problema de simulação em um cuja densidade do evento de interesse é maior. Isso é feito por meio da poda de porções do espaço que são compostas objetivamente por estados normais (i.e., estados não pertinentes ao estudo). Em [57], a técnica é aplicada de forma determinística via uma hipótese de coerência que permite caracterizar estados como de sucesso sem avaliar a adequação dos mesmos. Já em [58], são enumeradas as abordagens que buscaram uma alternativa à hipótese de coerência por meio de heurísticas baseadas em busca populacional inteligente. Muitos trabalhos foram publicados com este cunho. Publicações mais recentes, porém, são mais escassas e difíceis de serem encontradas.

Outra abordagem bastante aplicada é a de amostragem estratificada. Nesse caso, a amostragem aleatória simples da simulação Monte Carlo é substituída pela amostragem estratificada em múltiplas dimensões (i.e., para múltiplas variáveis aleatórias). O objetivo é fazer a simulação visitar as diversas regiões do espaço de estados de um problema com um montante reduzido de amostras, e sem repetições. Ou seja, pretende-se representar com mais eficiência todos os segmentos das distribuições de probabilidade dos parâmetros em estimação, inclusive as pontas/caldas das distribuições, cuja probabilidade é menor e pode ser rara. A grande vantagem desse método é ser uma VRT que não altera a função densidade de probabilidade (pdf – *probability density function*) das variáveis aleatórias do problema, sendo capaz de estimar não somente o valor esperado dos parâmetros, mas também suas distribuições. Por outro lado, é importante destacar que a técnica demanda preparação dos dados do problema antes da execução do algoritmo e definição subjetiva de parâmetros (algo que aflige muitos dos trabalhos de otimização de algoritmos de simulação), sendo que o retorno em ganho computacional não é de potencial elevado. O método do Hipercubo Latino é o mais observado na literatura de confiabilidade composta: [59-63]. Outra aplicação de amostragem estratificada é a *cluster-based stratified sampling* [64]. Destaque para os trabalhos [61] e [64], que, diferentemente dos demais, empregam seus respectivos métodos na simulação sequencial, que é mais exigente; e para o trabalho [63], que combina a técnica do Hipercubo Latino com outra VRT, a amostragem por importância.

Mais uma abordagem para o mecanismo de amostragem é a amostragem condicional. Em problemas de simulação de evento raro, a técnica chamada *subset simulation* (SS) é amplamente empregada em problemas de engenharia, ainda que menos comum na confiabilidade de SEP. A SS propõe expressar a probabilidade rara do evento de interesse por meio do produto de probabilidades condicionais de eventos intermediários, que são amostrados com maior frequência. Para isso, é necessário gerar um primeiro conjunto de amostras de forma eficiente, o que é contemplado pelo uso de cadeias de Markov Monte Carlo, e depois gerar, de forma adaptativa, novas amostras cada vez mais próximas do evento de interesse, até que ele seja alcançado. As duas partes desse processo são complexas conceitualmente, e o seu uso é recomendado quando a aplicação de outras VRTs mais eficientes é inviável. De qualquer forma, a SS foi bem adaptada para a confiabilidade de sistemas GT nos trabalhos [65-67]. Em [65] e [67], a SS é empregada individualmente, enquanto, em [66], a técnica dá suporte à de amostragem por importância, ao aprimorar o processo de otimização do método CE, que produz uma estimativa da distribuição ótima a ser utilizada na simulação via IS.

A amostragem por importância é a VRT de premissa mais concreta e certamente a mais pesquisada, com os melhores resultados dentro de confiabilidade de SEP. Sua premissa é a de que existem distribuições, diferentes da original, que levam a simulação de um problema a amostrar os estados de um evento de interesse com uma frequência maior, tão maior quanto for a qualidade da distribuição alternativa. Para manter os estimadores de parâmetros não tendenciosos, basta corrigir a probabilidade das amostras retiradas para o âmbito da distribuição original por meio da razão de verossimilhança (i.e., razão entre a probabilidade de realização da amostra via pdf original pela via pdf alternativa). O grande desafio está em determinar a distribuição alternativa (ou pdf-IS) que melhor atende às expectativas. Dois métodos de otimização propõem resolver o desafio: minimização da variância do estimador dado em função da pdf-IS (método VM – *variance minimization*), ou minimização da entropia cruzada entre a pdf-IS e a original (método CE). Em ambos os casos, *a priori*, a pdf-IS ótima pode ser obtida de forma analítica. Isso não sendo exequível, como ocorre em confiabilidade de SEP, a pdf-IS ótima pode ser estimada por algum mecanismo de simulação. O método VM foi desenvolvido primeiro, depois o método CE surge como uma ferramenta mais poderosa que pode ser empregada em diversos problemas de otimização combinatória além do de

simulação de evento raro. Ambas as metodologias podem ser apreciadas em detalhe em [68] de Rubisntein e Kroese, os proponentes do método CE.

Em confiabilidade de SEP, o método VM foi pouco empregado, apesar de sua primeira publicação ser de 1999 [69]. Recentemente, dois trabalhos foram publicados [70-71]. Nesses artigos, a minimização da variância é solucionada por um programa convexo com o intuito de obter uma solução global. Entretanto, o modelo utilizado demanda relaxamento de restrições para que possa ser tratado como convexo, e, ainda assim, um processo de captação de amostras é necessário para formar o desejado problema de otimização clássico. Desse modo, a solução obtida não é a pdf-IS ótima, pois está condicionada ao modelo aproximado e à amostra utilizada.

O método CE ganhou mais força entre os pesquisadores e apresenta a mais extensa gama de trabalhos entre as VRTs. Sua aplicação é encontrada tanto para SMC não-sequencial quanto para sequencial, em confiabilidade de geração [72-75], e composta [76-86]. No caso da confiabilidade de geração, o método CE apresentou excelente adequação, pois a função performance, que qualifica estados de sucesso quanto ao grau de proximidade à falha e é essencial para uma boa estimativa da pdf-IS, é intuitiva e direta. Isso não se repetiu no problema de confiabilidade de sistemas GT devido à dificuldade de se modelar uma função performance quando a rede de transmissão é considerada. Dos trabalhos citados, inclusive aqueles que combinam a amostragem por importância com outras VRTs [55], [63] e [66], a questão da função performance não recebe a atenção devida e muito esforço é destinado a: melhorar o processo de distorção de parâmetros da pdf original [79], considerar variáveis aleatórias de distribuição contínua, bem como aspectos cronológicos e de correlação entre componentes do sistema [80], [82-83] e [85-86], etc. Outro detalhe pouco discutido é o do critério de parada da pré-simulação do método CE, o qual é subjetivo e sujeito a produzir uma estimativa insuficiente para a pdf-IS ou uma pré-simulação muito onerosa computacionalmente. Tanto a questão da função performance quanto do critério de parada no método CE foram tratadas no trabalho apresentado neste documento.

1.2.3. Técnicas Combinadas

Buscando otimizar os algoritmos de confiabilidade composta em duas frentes, pesquisadores exploraram o emprego de classificadores e VRTs em conjunto: [56],

[87-91]. Em [87], combina-se a IS (sem metodologia clara para obtenção da pdf-IS) com uma rede SOM, com foco no impacto de dispositivos de proteção na confiabilidade. Em [88], o método do Hipercubo Latino é usado com uma rede chamada *generalized regression neural network* em um problema de otimização para obter a confiabilidade de cenários de investimento em um sistema. O artigo [56], anteriormente citado no texto, combina variáveis de controle com técnicas de aprendizado de máquina para a análise de confiabilidade no planejamento da operação um dia à frente. Em [89-90], os mesmos autores combinam a IS via método CE com a rede CNN e RBF, respectivamente. Por fim, em [91], a IS via método CE é aplicada à SMC, enquanto o classificador *k-means* é empregado na modelagem de distribuições de probabilidade no tempo para a disponibilidade energética de fontes eólicas – diferentemente dos demais trabalhos citados, os quais usam classificadores para evitar execuções de fluxo de potência ótimo.

Nenhum dos trabalhos enumerados fazem proveito da inteligência produzida pela VRT empregada para aperfeiçoar o treinamento dos classificadores escolhidos, as ferramentas são apenas utilizadas em conjunto. É imprescindível destacar que isso não apenas representa uma oportunidade perdida, mas também expõe a insistência em usar classificadores para aprenderem um padrão que é, em geral, raro. Para identificar a falha de sistemas GT e, mais ainda, inferir a qualidade do estado pelo corte de carga associado, é necessária uma boa amostra, de certo tamanho e representatividade. Se as VRTs são capazes de aumentar a densidade de estados de falha em uma dada amostra, nada mais coerente do que acoplar uma VRT ao treinamento de um classificador. Essa é a essência da parte final desta tese de doutorado, um adequado acoplamento entre um método de amostragem por importância via minimização da entropia cruzada e um classificador de estados em falha/sucesso.

1.3. Estrutura do Documento

Este documento de tese de doutorado é composto por cinco capítulos, brevemente descritos a seguir.

No presente capítulo de introdução, linhas gerais da análise de confiabilidade de sistemas GT com participação de energia renovável intermitente são apresentadas para estabelecer a área base de atuação deste trabalho. Deste ponto, a motivação

e proposta da pesquisa são comentadas, o que pode ser resumido em otimização de algoritmos de simulação para confiabilidade composta via técnicas de redução de variância e aprendizado de máquina. Uma revisão bibliográfica é detalhada para situar e justificar a pesquisa proposta em meio à literatura existente.

No Capítulo 2, o problema da confiabilidade de sistemas GT é extensivamente detalhado em termos técnicos. A simulação Monte Carlo quasi-sequencial é descrita, bem como a modelagem de cargas e equipamentos, especialmente os geradores de fonte renovável. Adicionalmente, os índices de confiabilidade pertinentes ao trabalho e a metodologia para análise de adequação de estados são devidamente apresentados.

No Capítulo 3, a ferramenta de amostragem por importância via método CE é descrita em detalhes para que a sugerida necessidade de uma nova abordagem em confiabilidade composta possa ser justificada. Na sequência, a metodologia da nova abordagem proposta é apresentada. Ao final do capítulo tem-se a apreciação do método proposto com análises para o sistema, área e barra a partir de um sistema teste de características afins a de sistemas reais que são interligados e de grande porte.

O Capítulo 4 traz a parte final da pesquisa, o acoplamento da nova ferramenta IS via método CE a um classificador de estados baseado em aprendizado de máquina. O classificador utilizado é a regressão logística binária, uma ferramenta eficiente e de simples implementação. Com a metodologia completa descrita em detalhes, os estudos de caso apresentados nos Capítulo 3 são reavaliados e seus resultados extensivamente discutidos. No final do capítulo, o uso de classificadores de maior complexidade é discutido.

Por fim, o Capítulo 5 traz as principais conclusões acerca do trabalho, assim como suas contribuições, e estabelece um prognóstico para a sequência da pesquisa.

2. Avaliação da Confiabilidade de Sistemas de Geração e Transmissão

2.1. Introdução

Neste capítulo, o problema de avaliação da confiabilidade de sistemas de geração e transmissão é apresentado, com foco no método de avaliação empregado na pesquisa da tese: a simulação Monte Carlo quasi-sequencial (SMC-QS). Primeiro, o estudo do conjunto geração-transmissão (GT) é contextualizado, incluindo a caracterização do evento falha de suprimento. Na sequência, um breve resumo dos principais métodos de avaliação de confiabilidade pode ser apreciado, sendo o método escolhido, a SMC-QS, devidamente apresentado na seção seguinte. Por fim, tem-se a descrição da modelagem de sistemas (carga, recursos energéticos renováveis e equipamentos), dos índices de confiabilidade e do procedimento de análise de adequação de estados.

2.2. Conjunto Geração-Transmissão

Historicamente, os sistemas elétricos de potência foram observados sob a ótica de níveis hierárquicos, conforme representação da Figura 2.1.

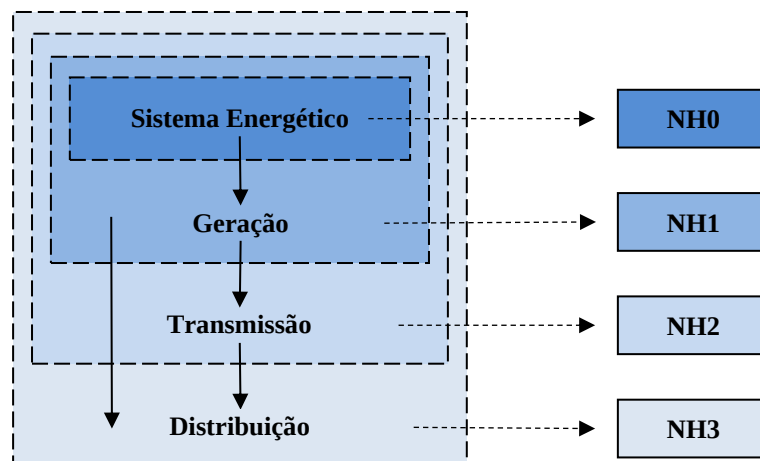


Figura 2.1: Níveis hierárquicos de sistemas em confiabilidade.

O nível hierárquico “0” (NH0) foi introduzido pela referência [92] para representar o sistema energético, cujos problemas são sazonais. Neste nível, é avaliado o balanço entre produção de energia elétrica e disponibilidade de recursos energéticos primários, com maior foco em recursos administráveis, como água e combustíveis fósseis, e menor foco nos não administráveis, como vento e radiação solar. Já os níveis hierárquicos 1, 2 e 3 (NH1, NH2 e NH3) estão atrelados aos problemas de potência, que são momentâneos. No NH1, é realizada a avaliação da confiabilidade da geração. A rede de transmissão é considerada 100% confiável e de capacidade térmica ilimitada, de modo que a carga e geração do sistema são representadas em barra única. No NH2, o conjunto GT é avaliado, tornando-se necessário representar as limitações impostas pela rede de transmissão. Portanto, são utilizados algoritmos de fluxo de potência, incluindo rotinas de otimização de medidas corretivas (redespacho de potência, ajustes nos taps de transformadores etc., e, em último caso, corte de carga). Por fim, no NH3, tem-se a avaliação integrada da confiabilidade de um sistema, incluindo geração, transmissão e distribuição, a qual é pouco investigada. Em geral, os estudos de confiabilidade da distribuição são realizados de forma independente, sendo o conjunto GT considerado 100% confiável e representado por potências injetadas equivalentes na fonte das redes de distribuição.

Evidentemente, os sistemas modernos não são mais verticalizados. A distribuição está sendo transformada pela geração distribuída e desenvolvimento de *smart-grids*, o que muda completamente a dinâmica da sua operação e estrutura, que tende a se tornar menos radial com o tempo (na Figura 2.1, esse fenômeno é representado pela seta que conecta diretamente os conjuntos geração e distribuição). No entanto, do ponto de vista da falha de suprimento de energia, a análise segregada dos sistemas ainda é válida e desejada. Observe que na distribuição as falhas são bem mais frequentes, mas são localizadas e isoladas, tendo impacto socioeconômico reduzido. Por outro lado, as falhas no conjunto GT são muito menos frequentes, mas de impacto socioeconômico elevado e abrangente, pois atingem, simultaneamente, grandes contingentes de consumidores, de variados tipos (residencial, comercial, industrial etc.). Com características tão distintas, é claro que a falha da distribuição e da geração-transmissão devem ser estudadas separadamente.

Ainda que o estudo da confiabilidade da distribuição seja cada vez mais importante, o conjunto GT segue sendo o mais crítico e o mais pesquisado, sendo também o objeto de análise deste trabalho.

Antes de apresentar os principais métodos de avaliação para sistemas GT, é importante também caracterizar a falha de suprimento. No NH1, a interpretação é bem direta: se a capacidade de geração disponível é menor que a carga momentânea, então há falha. Já a análise no NH2 é chamada de análise composta, pois existe um efeito combinado de dois tipos de falha: a por insuficiência de recursos de geração (falha G), e a por insuficiência de recursos de transmissão (falha T).

Visando eliminar eventuais dúvidas acerca das falhas G, T e composta, tem-se as seguintes definições. A falha G leva em conta apenas os equipamentos de geração e observa a ocorrência de déficit. Apenas após computar o eventual déficit de geração, a falha T pode ser avaliada. Nesse caso, consideram-se todos os equipamentos do sistema (geradores e circuitos) e procura-se saber, por meio de fluxo de potência, se o sistema tem a flexibilidade operativa necessária para atender toda a carga da região de interesse (sistema, área/zona ou barra). Como todo déficit observado na falha G é aplicado na forma de corte de carga antes da análise da falha T, têm-se dois eventos independentes, que podem ocorrer tanto concomitantemente quanto exclusivamente. A falha composta, por sua vez, é a devida combinação das falhas G e T. Ela é igual à soma das falhas em qualquer estado avaliado, o que não significa que os índices da falha composta sejam dados pela soma dos índices das demais falhas. Para índices de energia, a soma pode ser realizada, mas para os índices de probabilidade e de frequência e duração, isso não é aplicável. Um desacoplamento do evento falha ainda mais detalhado pode ser verificado em [93]. Por fim, conclui-se que a adequada distinção das falhas G e T bem como dos índices associados é muito importante para evitar interpretações e, conseqüentemente, metodologias inadequadas.

2.3. Principais Métodos de Avaliação

A confiabilidade de sistemas elétricos de potência é tratada na operação como um problema determinístico. No entanto, sabe-se que a adequada qualificação do risco de suprimento de um sistema só pode ser obtida por métodos probabilísticos. Tais métodos se baseiam na estimação de índices a partir de um conjunto de estados operativos. A forma como o conjunto de estados é gerado distingue as metodologias em do tipo analítico e de simulação.

A metodologia analítica é marcada pela enumeração do espaço de estados do problema, de modo que os índices estimados são de fato valores esperados. Sua implementação é simples, porém, sua aplicabilidade em problemas combinatórios de elevada dimensão é limitada, como é o caso de sistemas elétricos interligados de grande porte. Esforços interessantes foram feitos para tornar a enumeração de estados competitiva com os métodos de simulação [93-95]. Entretanto, se levados em consideração aspectos cronológicos como curva de carga e recursos renováveis, a enumeração de estados se torna inexequível até mesmo para sistemas de pequeno porte. Por outro lado, as técnicas de simulação fornecem boas estimativas para os índices. Contudo, o controle sobre a incerteza das estimativas, aliado à flexibilidade de representação e robustez frente ao tamanho do espaço amostral do problema, faz das simulações uma poderosa ferramenta na análise de sistemas GT.

Os métodos de simulação, por sua vez, são diferenciados pela representação do problema estocástico em espaço de estados (SMC não-sequencial – SMC-NS) e cronológica (SMC sequencial – SMC-S). A SMC-S é capaz de representar qualquer elemento de um sistema, de qualquer comportamento cronológico. Essa habilidade, porém, vem com custo computacional elevado, em contraposto à SMC-NS, cuja flexibilidade é menor, mas eficiência é bem maior. Conforme descrito no capítulo de introdução (na Seção 1.2), tentativas de aliar as qualidades da SMC-S com a velocidade da SMC-NS foram implementadas, começando pela SMC pseudo-sequencial [18], depois aprimorada para a pseudo-cronológica [19], até culminar na ferramenta SMC quasi-sequencial [20], a qual supera a pseudo-cronológica em eficiência, mas apresenta menor precisão para a estimativa do índice de custo de interrupção [96].

Na representação a espaço de estados, cada estado k de um sistema é dado por um vetor $\mathbf{X}_k = (X_{k,1}, \dots, X_{k,n})$, onde $X_{k,j}$ é o estado k do j -ésimo componente. Cada componente de $\mathbf{X}_k$ está associado a um elemento do sistema (gerador, transformador, linha de transmissão, carga ou energia renovável). O espaço de estados, ou seja, todas as possíveis realizações de $\mathbf{X}_k$ é denotado por $\mathcal{X}$. Já os índices de confiabilidade são obtidos a partir do cálculo do valor esperado de suas respectivas funções teste, denotadas por $F(\mathbf{X})$. O objetivo destas funções teste é verificar se, e/ou de que forma, uma determinada configuração de geradores e circuitos é capaz de suprir certo patamar de carga, dada a capacidade energética momentânea dos

geradores de fonte renovável. A estimativa do valor esperado de $F(\mathbf{X})$ é obtida pela média de $F(\mathbf{X})$ sobre os NS sorteios realizados:

$$\hat{E}[F(\mathbf{X})] = \frac{1}{NS} \sum_{k=1}^{NS} F(\mathbf{X}_k) \quad (2.1)$$

Na representação cronológica, os estados são amostrados sequencialmente no tempo, o que permite reproduzir com precisão a evolução temporal do sistema. Assim, dois estados sucessivos do sistema diferem-se pela alteração do status de apenas um de seus componentes, seja ele o patamar da carga, a condição operativa de um gerador ou circuito, ou a capacidade energética de fontes renováveis. Isso implica em um procedimento de estimação de índices razoavelmente distinto daquele empregado pela representação a espaço de estados. Uma função teste não é mais aplicada ao estado $\mathbf{X}_k$ do sistema, mas à série temporal $\mathbf{Y}_k$, variável que descreve todos os estados resididos pelo sistema em um período de simulação. A notação de uma função teste é $G(\mathbf{Y}_k)$ e a estimativa de seu valor esperado é dada pela média de $G(\mathbf{Y}_k)$ sobre os NP períodos simulados:

$$\hat{E}[G(\mathbf{Y})] = \frac{1}{NP} \sum_{k=1}^{NP} G(\mathbf{Y}_k) \quad (2.2)$$

Importante destacar que enquanto na representação a espaço de estados os índices são estimados sob a perspectiva momentânea, na cronológica a perspectiva é periódica. Por exemplo, enquanto se estima a probabilidade de falhar a qualquer momento na SMC-NS, na SMC-S estima-se por quantos instantes de tempo o sistema deverá estar em falha, em média, a cada período. As estimativas são evidentemente equivalentes, estando apenas em bases distintas.

2.4. Simulação Monte Carlo Quasi-Sequencial

A SMC-QS é uma ferramenta que “amarra” a cronologia das amostras na representação a espaço de estados da SMC-NS [20], pois associa cada amostra a um instante de tempo. Para tanto, um modelo de Markov multiestados não agregado é aplicado para todos os elementos variantes no tempo do sistema. Tais elementos podem ser selecionados do espaço de estados do problema com a correlação temporal desejada. Em geral, séries anuais históricas com informação horária de

demanda ou de disponibilidade energética de renováveis são obtidas sem muita dificuldade e produzem uma excelente representação cronológica desses elementos. Com essas séries em mãos, a simulação quasi-sequencial pode proceder de duas formas: (i) selecionar sequencialmente o instante de tempo associado ao estado amostrado (basta seguir a ordem das séries históricas); ou (ii) sortear o instante de tempo a partir de uma distribuição uniforme no intervalo $[1, T]$, onde T é o período da série. Assim, a SMC-QS pode ser interpretada como uma SMC-NS que considera o instante de tempo em que cada estado amostrado reside.

Neste trabalho, foi optado por sortear o instante de tempo t via distribuição uniforme no intervalo $[1, T]$. Um algoritmo básico de SMC-QS é descrito a seguir.

- Passo 1) Inicie o contador de número de sorteios: $NS = 0$;
- Passo 2) Sorteie o instante de tempo t e o vetor $\mathbf{X}_{NS} \in \mathcal{X}$;
- Passo 3) Avalie a adequação do estado $\mathbf{X}_{NS}$, estime índices, conforme eq. (2.1), e calcule as incertezas das estimativas via coeficiente de incerteza relativa β , conforme eq. (2.3);
- Passo 4) Se a incerteza da estimativa de qualquer índice for maior que uma dada tolerância, faça $NS = NS + 1$ e retorne ao Passo 2); senão, pare.

$$\beta = \frac{\sqrt{\hat{V}(\hat{E}[F(\mathbf{X})])}}{\hat{E}[F(\mathbf{X})]} \times 100\% \quad (2.3)$$

onde $\hat{V}(\hat{E}[F(\mathbf{X})])$, variância amostral do estimador $\hat{E}$, é obtido conforme eq. (2.4).

$$\hat{V}(\hat{E}[F(\mathbf{X})]) = \frac{1}{NS} \times \hat{V}(F(\mathbf{X})) = \frac{1}{NS} \times \frac{\sum_{k=1}^{NS} (F(\mathbf{X}_k) - \hat{E}[F(\mathbf{X})])^2}{(NS - 1)} \quad (2.4)$$

2.5.

Modelagem do Sistema

2.5.1.

Modelo da Carga

Para a carga, um modelo de Markov multiestados não agregado é utilizado [20] e [78]. Tal modelo corresponde a uma sequência de estados não balanceada em frequência, com taxa de transição unidirecional e constante, vide Figura 2.2. Nesse caso, a duração média de todo estado é $\Delta t = 1\text{h}$, o período de análise é de

um ano, ou $T = 8760h$, e a taxa de transição entre dois instantes de tempo é $\lambda_{\Delta t} = 1 \text{ oc./h}$. Assim, com um modelo simples, cada barra de carga $b \in [1, B]$, com demanda $L_b(t)$, pode seguir suas próprias séries históricas, representando basicamente qualquer comportamento cronológico.

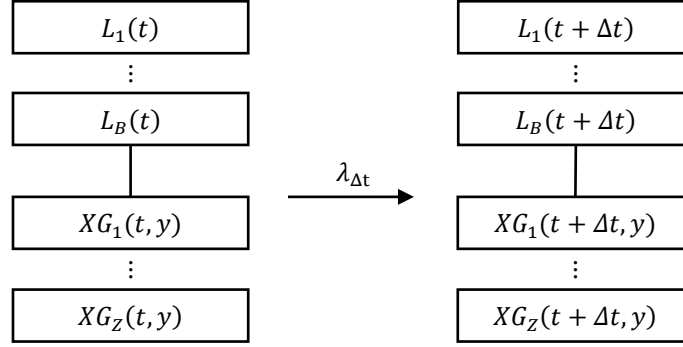


Figura 2.2: Modelo de Markov multiestados não agregado.

2.5.2. Modelo da Energia Renovável

Para geradores de energia renovável, séries históricas do recurso natural (e.g., água e vento) são convertidas em séries de capacidade de geração, dada em “por unidade” (pu) do valor nominal. Cada barra com geração renovável pertence a alguma zona, e para cada zona um certo número de anos deve ser utilizado (e.g., 10 anos), de modo que períodos de características climáticas diferentes sejam propriamente representados [97]. A cada estado amostrado, a capacidade nominal de cada gerador renovável é multiplicada por um fator $XG_z(t, y)$, o qual está atrelado ao instante de tempo $t \in [1, T]$, da série/ano $y \in [1, Y]$, da zona $z \in [1, Z]$ a que o gerador pertence (vide Figura 2.2). Neste trabalho, a série/ano de cada zona é sorteada com base em probabilidades associadas aos anos contemplados, os quais são considerados equiprováveis, aplicando total correlação entre zonas do mesmo tipo de fonte renovável e nenhuma correlação entre zonas de fontes distintas. Desse modo, para cada estado amostrado, um instante de tempo é sorteado, comum a todos os elementos do sistema, enquanto um ano é sorteado para as zonas de cada tipo de fonte de forma independente. Feito isso, a disponibilidade de cada equipamento é amostrada.

2.5.3. Modelo da Disponibilidade de Equipamentos

Em termos de disponibilidade, os equipamentos são separados em dois grupos. Os geradores térmicos e hidráulicos e os circuitos seguem um modelo de Markov a dois estados, como o representado pela Figura 2.3(a) [12], enquanto os geradores eólicos seguem um modelo de Markov multiestados para unidade equivalente, como o representado pela Figura 2.3(b) [19-20].

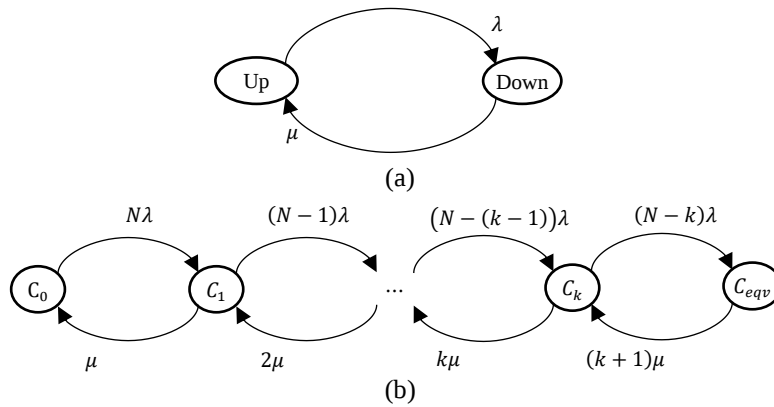


Figura 2.3: (a) Modelo de Markov a dois estados; (b) Modelo de Markov multiestados para unidade equivalente.

No modelo equivalente, C_k representa a capacidade disponível para o estado com k unidades fora de operação e C_{eqv} a capacidade do estado equivalente, o qual absorve a probabilidade acumulada de todos os estados cujo nível de contingência é superior ao do estado equivalente (evidentemente, a probabilidade acumulada desses estados precisa ser menor que uma dada tolerância). Em geral, múltiplas fazendas eólicas ocupam a mesma região geográfica e, por isso, são arranjadas no mesmo barramento na representação do sistema, o que resulta em centenas de geradores por barra com presença de fonte eólica. Se esses geradores são representados individualmente por modelo de Markov a dois estados, um número exorbitante de sorteios será realizado para cada estado amostrado. Portanto, o modelo equivalente é essencial para aliviar o custo computacional. Para que o modelo funcione corretamente, é necessário considerar que todos os geradores eólicos representados em uma barra não sejam apenas sujeitos à mesma intermitência do vento, mas também à mesma capacidade nominal e taxas de falha-reparo; i.e., sejam tratados como unidades idênticas.

Independentemente do modelo, todos os equipamentos seguem uma distribuição Bernoulli. Portanto, individualmente, todos podem residir em exatamente dois estados operacionais: disponível (*Up*) e indisponível (*Down*), os quais se comunicam por meio da taxa de falha (λ) e de reparo (μ), usualmente fornecidas em ocorrências por ano. Com base no histórico do equipamento, a taxa λ é dada pela razão entre o total de ocorrências de falha e o tempo total de disponibilidade, enquanto a taxa μ é dada, analogamente, pela razão entre o total de reparos e o tempo total de indisponibilidade. A probabilidade de um equipamento qualquer ser encontrado em condição disponível (A) ou indisponível (U) pode, então, ser obtida pelas equações (2.5) e (2.6), respectivamente. O acrônimo *FOR* refere-se à taxa de saída forçada, ou *forced outage rate*.

$$A = 1 - FOR = \frac{\mu}{\lambda + \mu} \quad (2.5)$$

$$U = FOR = \frac{\lambda}{\lambda + \mu} \quad (2.6)$$

Retomando o modelo equivalente, tem-se que a probabilidade de um estado C_k é dada pelo produto entre a disponibilidade A dos geradores disponíveis e a indisponibilidade U dos geradores indisponíveis, vezes a combinação de N unidades geradoras tomadas k a k (i.e., via distribuição binomial). A taxa de saída para o estado com uma contingência a menos é a soma da taxa de reparo dos k geradores em falha, enquanto para o estado com uma contingência a mais, a taxa é a soma da taxa de falha dos $N - k$ geradores em operação.

Vale ressaltar que o funcionamento de uma unidade geradora depende de diversos componentes auxiliares. A falha de um desses componentes pode não resultar na perda total de capacidade, de forma que a unidade continua funcionando em um estado de capacidade intermediária. Assim, as unidades geradoras podem ser representadas com maior precisão por modelos de Markov multiestados, bem como o modelo para unidade equivalente representado na Figura 2.3(b). Nesse caso, a taxa de transição λ_{ij} entre dois estados i e j quaisquer é definida como a razão entre o número de vezes que o gerador transita do estado i para o estado j e o tempo total de permanência no estado i .

2.6. Índices de Confiabilidade

Os índices usualmente calculados na avaliação da confiabilidade de sistemas GT se referem à falha de suprimento, a qual se caracteriza pela insuficiência de recursos de geração e/ou de transmissão. O objetivo desses índices é estimar a probabilidade de ocorrência, a frequência e a duração média das falhas do sistema, assim como a expectativa de energia não suprida e os custos associados. A seguir, apresentam-se os índices de confiabilidade tradicionalmente calculados.

- LOLP – *loss of load probability* (probabilidade da perda de carga)
- LOLE – *loss of load expectation* (expectativa da perda de carga)
- EPNS – *expected power not supplied* (potência esperada não suprida)
- EENS – *expected energy not supplied* (energia esperada não suprida)
- LOLF – *loss of load frequency* (frequência da perda de carga)
- LOLD – *loss of load duration* (duração da perda de carga)
- LOLC – *loss of load cost* (custo da perda de carga)

Outro índice bastante utilizado é o de severidade [98], de uso destacado no Brasil [1]. Seu resultado é dado em minutos (unidade conhecida como *system minutes*) e denota a expectativa do montante de tempo, para um período anual, em que o sistema, área ou barra deixaria de suprir sua carga totalmente se a mesma demandasse seu valor de pico durante toda a falha simulada. Para tanto, a expectativa de energia não suprida que foi de fato calculada para o ano em questão é dividida pelo valor de pico de carga no período. Com o advento da análise de bem-estar, também conhecida como confiabilidade preventiva [99-101] e [24], índices adicionais foram introduzidos com o objetivo de avaliar o grau de conforto do sistema em estados de sucesso.

Na representação a espaço de estados, os índices LOLP, EPNS e LOLF são obtidos por meio das seguintes funções teste:

$$F_{LOLP}(\mathbf{X}_k) = \begin{cases} 0, & \text{se } \mathbf{X}_k \text{ é estado de sucesso} \\ 1, & \text{se } \mathbf{X}_k \text{ é estado de falha} \end{cases} \quad (2.7)$$

$$F_{EPNS}(\mathbf{X}_k) = \begin{cases} 0, & \text{se } \mathbf{X}_k \text{ é estado de sucesso} \\ \Delta P_k, & \text{se } \mathbf{X}_k \text{ é estado de falha} \end{cases} \quad (2.8)$$

$$F_{LOLF}(\mathbf{X}_k) = \begin{cases} 0, & \text{se } \mathbf{X}_k \text{ é estado de sucesso} \\ \lambda_k, & \text{se } \mathbf{X}_k \text{ é estado de falha} \end{cases} \quad (2.9)$$

onde ΔP_k (usualmente dado em MW) representa o montante de carga cortado no estado $\mathbf{X}_k$, e λ_k (usualmente dado em oc./ano) representa a soma das taxas de saída de $\mathbf{X}_k$ para todos os estados de sucesso que podem ser alcançados mediante uma transição (i.e., transição de apenas um componente do sistema). Os demais índices podem ser calculados ao final da simulação, tal como:

$$LOLE = LOLP \times NIT \quad (2.10)$$

$$EENS = EPNS \times NIT \quad (2.11)$$

$$LOLD = LOLE / LOLP \quad (2.12)$$

$$LOLC = EENS \times CUI \quad (2.13)$$

onde NIT é o número de instantes de tempo que compõem o período considerado (usualmente 8760 h/ano) e CUI é o custo unitário médio de interrupção (usualmente dado em unidade monetária por MWh).

Para o índice $LOLD$, tal como na eq. (2.12), deve-se ter em mente que $LOLE$ é utilizado no lugar de $LOLP$ para que a duração média da falha seja dada na unidade do instante de tempo, ao invés de uma fração da unidade do período. Também é importante destacar que a eq. (2.13), utilizada para o cálculo da $LOLC$, apresenta fortes simplificações [102], sendo que vários métodos foram propostos visando prover estimativas mais consistentes via técnicas voltadas para as simulações não sequenciais. No entanto, somente as técnicas capazes de reproduzir a evolução cronológica do sistema [103], ao menos para os processos de falha [104], podem fornecer boas estimativas para o índice $LOLC$ [18] e [102].

Dos índices apresentados, somente $LOLC$ não integra este trabalho. Sobre os demais, cabe comentários adicionais para a $LOLF$, pois o cálculo de sua função teste via (2.9) é computacionalmente ineficiente e foi preterido por outra técnica. Em princípio, utilizando a eq. (2.9), para cada estado de falha selecionado de um sistema com NC componentes, seriam necessárias NC análises de adequação adicionais para atualizar a estimativa da frequência. Note que, para cada análise extra, um novo estudo de adequação de estados se faz necessário, incluindo a execução de fluxo de potência com otimização de medidas corretivas no caso de sistemas GT,

de modo a verificar se o novo estado será de sucesso. Na década de 90, duas técnicas eficientes foram desenvolvidas para evitar este esforço computacional proibitivo. Uma utiliza filtros baseados nos multiplicadores de Lagrange, produzidos pelo algoritmo de otimização das medidas corretivas [21], a outra é baseada no conceito de probabilidade condicionada, ou frequência incremental [22]. Em ambos os métodos é necessário adotar a hipótese de coerência: uma transição de falha de um equipamento ou de elevação da carga nunca melhora a condição operativa do sistema, enquanto a transição de reparo de um equipamento ou de redução da carga nunca piora a condição operativa do sistema. Posteriormente, no início dos anos 2000, foi proposto um método denominado processo de transição de estados um passo à frente [23-24]. Este método implica em um número reduzido de análises adicionais e, por prescindir da hipótese de coerência, traz mais flexibilidade às simulações não sequenciais, sendo o escolhido para estimar o índice de frequência da falha neste trabalho.

Com o método de transição um passo à frente (OSF – *one-step forward*), a função teste da LOLF passa a ser obtida por meio da eq. (2.14). A taxa λ_{out} corresponde à soma das taxas de todas as transições de saída do estado $\mathbf{X}_k$, não apenas para estados de sucesso como ocorre com λ_k , e $\mathbf{X}_m$ corresponde a um estado qualquer que se comunica com $\mathbf{X}_k$ pela transição de apenas um componente do sistema.

$$F_{LOLF}(\mathbf{X}_k) = \begin{cases} \lambda_{out}, & \text{se } \mathbf{X}_k \text{ é estado de falha e } \mathbf{X}_m \text{ é de sucesso} \\ 0, & \text{caso contrário} \end{cases} \quad (2.14)$$

Observe que a transição de um estado $\mathbf{X}_k$ a um $\mathbf{X}_m$ ocorre com uma probabilidade, que é a proporção da taxa da referida transição para com todas as possíveis incluindo a mesma, vide eq. (2.15). Observe também que λ_k pode ser obtida pela soma das taxas de toda transição que leva $\mathbf{X}_k$ para o sucesso, tal como eq. (2.16). Ω_k é o conjunto de estados de sucesso que se comunicam com $\mathbf{X}_k$ pela transição de apenas um componente do sistema.

$$P_{km} = \frac{\lambda_{km}}{\lambda_{out}} \quad (2.15)$$

$$\lambda_k = \sum_{m \in \Omega_k} \lambda_{km} \quad (2.16)$$

De (2.15), tem-se que λ_{km} é igual ao produto entre P_{km} e λ_{out} . Então, λ_k pode ser obtida como:

$$\lambda_k = \sum_{m \in \Omega_k} P_{km} \times \lambda_{out} = \left(\sum_{m \in \Omega_k} P_{km} \right) \times \lambda_{out} \quad (2.17)$$

Enfim, pode-se constatar pela eq. (2.14) que a probabilidade de cada transição de $\mathbf{X}_k$ a um $\mathbf{X}_m \in \Omega_k$ será capturada (i.e., estimada) através da combinação dos processos de amostragem da simulação principal e do sorteio de $\mathbf{X}_m$, e será multiplicada por λ_{out} . Logo, as funções testes em (2.9) e (2.14) são de fato equivalentes. Certamente, o processo de OSF demandará um montante de amostras bem maior para que a incerteza da estimativa da LOLF atinja o valor limite. Ainda assim, o número de análises adicionais, e, portanto, custo computacional, será muito menor.

2.7.

Análise de Adequação de Estados

A análise de adequação de estados para sistemas GT exige a execução de fluxo de potência. Idealmente, o algoritmo convencional de fluxo de potência não linear AC [105], deve ser empregado, de modo a capturar toda a complexidade das leis físicas que governam as redes elétricas de potência em regime estático e representar os principais componentes que integram a operação de um sistema de potência, de compensadores síncronos a tap de transformadores. Apesar de ser solucionável por meio de métodos numéricos, o fluxo AC traz um custo computacional elevado para os estudos de operadores e planejadores (além de problemas com divergência da solução para análises com contingência de equipamentos). Em confiabilidade de sistemas GT, onde milhares a milhões de fluxos de potência precisam ser resolvidos, o custo se torna proibitivo. Como alternativa, um modelo de fluxo de potência linear DC [106] pode ser utilizado mediante três simplificações:

- i) resistência de circuitos é desprezada;
- ii) módulo de tensão das barras é fixado no seu valor nominal (ou 1,0 pu);
- iii) $\theta_{ik} \rightarrow 0 \Rightarrow \sin \theta_{ik} = \theta_{ik}$ e $\cos \theta_{ik} = 1$;

onde θ_{ik} é a abertura angular de tensão entre as barras i e k .

As redes de transmissão são compostas por transformadores, que são naturalmente equipamentos de baixa resistência elétrica, e linhas, que, por estarem

submetidas a altos valores de tensão, apresentam baixa razão resistência por reatância (atende à primeira simplificação). Adicionalmente, sabe-se que recursos são empregados (e.g., tap de transformadores e dispositivos FACTS – *flexible alternating current transmission system*) para controlar o fluxo de reativo na rede e, por consequência, o módulo de tensão de barras (atende à segunda simplificação). Por fim, tem-se que sistemas interligados apresentam rede robusta e boa distribuição de fluxo, de modo que as aberturas angulares de tensão entre barramentos diretamente conectados por circuito são, em geral, reduzidas (atende à terceira simplificação).

De fato, sistemas GT são compatíveis com o fluxo DC e suas simplificações. Esses sistemas, inclusive, estão cada vez mais conectados e com geração cada vez mais distribuída (em função da crescente exploração de energia renovável), o que reduz a ocorrência de fluxos unidirecionais por grupos de circuitos em série e a consequente extensa abertura angular entre os barramentos terminais desses fluxos. No entanto, um sistema de médio a grande porte apresenta magnitude de potência gerada-consumida elevada e grande quantidade de circuitos, de modo que as perdas, que individualmente são negligenciáveis, acumulam para um valor significativo, ainda que represente um percentual pequeno da capacidade do sistema. Para aproximar a solução do fluxo DC ao do fluxo AC, sobretudo quando se quer avaliar sistemas de médio a grande porte, pode-se empregar o fluxo DC com perdas [107], método escolhido para integrar a análise de adequação de estados neste trabalho.

Em uma simulação, um fluxo de potência convencional tem o papel de definir a distribuição do fluxo para o estado amostrado, sendo necessário especificar um despacho para os geradores disponíveis no sistema, o que é feito a partir de alguma heurística. Certamente, nenhuma heurística será tão bem elaborada a ponto de refletir adequadamente o comportamento da operação frente a todo e qualquer cenário amostrado na simulação. Na prática, os geradores são comissionados, geração e reserva, com um horizonte de 24 horas, enquanto a operação em tempo real otimiza os recursos disponíveis e redespacha o sistema. Portanto, um fluxo de potência com otimização de medidas corretivas (redespacho de geração ativa e reativa, controle de tap de transformadores, uso de FACTS e, em último caso, aplicação de corte de carga) é o procedimento empregado em virtualmente todo estudo de confiabilidade composta. Neste trabalho, um fluxo DC ótimo é utilizado, o qual redespacha a geração e, se necessário, corta carga [17].

É importante destacar que um fluxo de potência ótimo (FPO) é oneroso computacionalmente e deve ser evitado. Como a análise da confiabilidade está interessada no evento de falha de suprimento, a execução de FPO para a adequação de estados não é necessária quando o estado é claramente de sucesso, de modo que boa parte das amostras de uma simulação pode receber uma análise simplificada. Uma maneira bem simples de fazer isso é definir um estado claramente de sucesso como um estado sem problemas, isto é, que não apresenta violações de limites operativos na solução de um fluxo de potência convencional. Para tanto, uma heurística pode ser implementada para equilibrar, sem considerar a rede, a geração disponível com a carga momentânea do estado amostrado, e, então, o despacho de geração resultante é aplicado em um fluxo DC. Não existindo nenhuma violação de capacidade de fluxo em circuitos, o estado em questão é nitidamente de sucesso e a execução de um FPO pode ser evitada.

Embasado pelo que foi discutido nesta seção, o procedimento empregado neste trabalho para a análise de adequação de estados é definido em três partes:

- i) Balanço de Geração-Carga (BGC): Geração e carga são balanceadas em um modelo “barra única”, produzindo um novo despacho para os geradores, obedecendo aos fatores de participação de um despacho base [108], o qual é dado pela solução de um FPO que minimiza o fluxo de potência na rede para a condição de carga pico sem contingências. Eventuais cortes de carga por déficit de geração são observados (i.e., falha G).
- ii) Fluxo de potência DC com Perdas (FDCP): Um fluxo de potência linearizado com perdas é executado para determinar se o despacho obtido no BGC para o estado amostrado apresenta violação de capacidade de circuitos. Em caso positivo, um FPO é executado na próxima etapa para tentar resolver os problemas da rede. Não há observação de corte de carga nesta etapa.
- iii) Fluxo de Potência Ótimo: Um fluxo de potência ótimo linearizado é aplicado com mínimo desvio do ponto de operação encontrado no FDCP. Eventuais cortes de carga por inflexibilidade no transporte de energia são observados (i.e., falha T).

2.8. Conclusões

Este capítulo apresentou os conceitos básicos da avaliação da confiabilidade de sistemas de geração e transmissão, tendo como foco a simulação Monte Carlo quasi-sequencial. Foram apresentados a modelagem dos componentes de sistemas, incluindo a representação de recursos renováveis, os índices de confiabilidade usualmente avaliados, com destaque para o indicador de frequência da falha, assim como o procedimento adotado para avaliar a adequação de estados amostrados na simulação.

O propósito deste capítulo é estabelecer o *framework* sobre o qual a metodologia de avaliação de confiabilidade, a ser apresentada nos próximos dois capítulos, foi desenvolvida. Todos os elementos de sistemas elétricos de potência e demais aspectos do problema de confiabilidade composta, que são considerados no trabalho, foram descritos. Todavia, o destaque fica para a interpretação da falha de suprimento no contexto do conjunto geração-transmissão. Foi evidenciado que o desacoplamento da falha composta em falha por insuficiência de recursos de geração e por insuficiência de recursos de transmissão é de suma importância para a correta interpretação do problema objetivo da tese e a efetividade da metodologia proposta.

3. Confiabilidade Composta Otimizada por Redução de Vari- ância

3.1. Introdução

Conforme estabelecido no capítulo de introdução, o objetivo desta tese de doutorado é otimizar algoritmos de simulação, em confiabilidade de sistemas de geração e transmissão (GT), por meio de um adequado acoplamento entre um método de amostragem por importância (IS – *importance sampling*), via minimização da entropia cruzada (CE – *cross-entropy*), e um classificador de estados. Destaca-se, porém, que boa parte da inovação deste trabalho reside na nova abordagem para o emprego do método CE em sistemas GT apresentada neste capítulo.

Na próxima seção, o método CE original é descrito em detalhe, enquanto a seção seguinte traz um resumo da aplicação mais bem sucedida em confiabilidade de geração e composta, com o intuito de explicitar a necessidade da nova abordagem para a confiabilidade composta. Ainda nesta seção, é apresentada a proposta para o novo método e seu algoritmo. Na quarta seção, são apresentados estudos de caso para demonstrar o comportamento da ferramenta desenvolvida em diferentes cenários de análise de um sistema com características semelhantes às de sistemas interligados de médio-grande porte. A quinta e última seção traz as conclusões do capítulo.

3.2. Método da Entropia Cruzada

O método da entropia cruzada é conceitualmente simples, eficiente e de grande capacidade de generalização, capaz de resolver variados problemas de otimização combinatória [68]. Sua origem, porém, reside mais especificamente na busca por aprimorar algoritmos de simulação de evento raro que são baseados em técnicas de redução de variância (VRTs – *variance reduction techniques*).

O método CE pode ser resumido como um processo de otimização iterativo de dois estágios: (i) simulação Monte Carlo auxiliada por IS; (ii) dimensionamento de nova distribuição de probabilidade. Seu propósito é endereçar o desafio da IS: encontrar uma distribuição de probabilidade, diferente da original, capaz de levar uma simulação a amostrar um determinado evento com maior frequência, reduzindo a variância da estimativa dos parâmetros em estudo. O método assume que existe uma distribuição de probabilidade ótima capaz de fazer uma simulação a representar o evento de interesse com variância zero, a partir de qualquer amostra simples (i.e., amostra de apenas um estado). Desse ponto, é destacado que é possível determinar tal distribuição ótima analiticamente por meio da minimização da entropia cruzada entre uma distribuição alternativa (partindo da original) e a ótima. Como o formato da função densidade de probabilidade das distribuições (original, alternativa e ótima) é invariavelmente difícil de ser definida, a determinação analítica da distribuição ótima é quase sempre inviável. Logo, o método recorre a uma solução mais prática: estimar a distribuição ótima, o que é feito a partir do programa estocástico equivalente do problema de minimização da entropia cruzada. Por fim, não sendo ainda exequível estimar a distribuição em apenas uma simulação, propõe-se “caminhar” iterativamente em direção a ela com o suporte da ferramenta IS.

A seguir, uma descrição matemática detalhada do método CE, baseada no livro [68], é apresentada.

3.2.1. Simulação de Evento Raro

Seja $\mathbf{X} = (X_1, \dots, X_n)$ um vetor aleatório que assume valores no espaço $\mathcal{X}$ e $\{f(\cdot; \boldsymbol{\sigma})\}$ uma família paramétrica de funções densidade de probabilidade (pdf – *probability density function*) no mesmo espaço $\mathcal{X}$, onde $\boldsymbol{\sigma}$ é um vetor real de parâmetros. Então, o valor esperado de uma função mensurável $H(\mathbf{X})$ qualquer é dado por (3.1) para distribuições contínuas e (3.2) para distribuições discretas. O subscrito do operador $\mathbb{E}$ indica a partir de qual vetor de parâmetros o valor esperado é obtido.

$$\mathbb{E}_{\boldsymbol{\sigma}}[H(\mathbf{X})] = \int_{\mathcal{X}} \dots \int_{\mathcal{X}} H(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (3.1)$$

$$\mathbb{E}_{\sigma}[H(\mathbf{X})] = \sum_{X_{k,n} \in \mathcal{X}} \dots \sum_{X_{k,1} \in \mathcal{X}} H(X_{k,1}, \dots, X_{k,n}) f(X_{k,1}, \dots, X_{k,n}) \quad (3.2)$$

Considere que é desejado estimar a probabilidade ℓ de um certo evento $\{S(\mathbf{X}) \geq \gamma\}$ em $f(\mathbf{X}; \mathbf{u})$, conforme eq. (3.3). Ou seja, quer-se saber o valor esperado de $H(\mathbf{X}) = I_{\{S(\mathbf{X}) \geq \gamma\}}$, onde $I_{\{\cdot\}}$ representa um avaliador de expressão lógica, $S(\mathbf{X})$ é uma função real no espaço $\mathcal{X}$ e γ um número real.

$$\ell = \mathbb{P}_{\mathbf{u}}(S(\mathbf{X}) \geq \gamma) = \mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) \geq \gamma\}}] \quad (3.3)$$

A maneira mais óbvia de estimar ℓ é simplesmente retirar uma amostra $\mathbf{X}_1, \dots, \mathbf{X}_N$ de $f(\mathbf{X}; \mathbf{u})$ e calcular a média amostral de $I_{\{S(\mathbf{X}) \geq \gamma\}}$, tal como:

$$\hat{\ell} = \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}} \quad (3.4)$$

onde $\hat{\ell}$ é um estimador não tendencioso de ℓ . Tal estimador é insensível ao tamanho do espaço de estados gerado por $f(\mathbf{X}; \mathbf{u})$, sendo bastante eficiente, exceto no caso em que ℓ é uma probabilidade muito pequena, ou seja, o evento $\{S(\mathbf{X}) \geq \gamma\}$ é raro. Isso pode ser constatado pela expressão do erro relativo do estimador $\hat{\ell}$ (eq. (3.5)), ou coeficiente de incerteza relativa β , dado pela razão entre o desvio padrão amostral do estimador e o valor estimado.

$$\beta = \frac{\sqrt{\hat{V}(\hat{\ell})}}{\hat{\ell}} = \frac{\sqrt{\hat{V}(I_{\{S(\mathbf{X}) \geq \gamma\}})/N}}{\hat{\ell}} \quad (3.5)$$

$$N = \frac{\hat{V}(I_{\{S(\mathbf{X}) \geq \gamma\}})}{(\beta \times \hat{\ell})^2} \quad (3.6)$$

Rearranjando (3.5), é possível obter o tamanho N da amostra necessária a alcançar um determinado nível de incerteza β . Isso é expresso pela eq. (3.6), da qual se tiram duas conclusões:

- i) N é inversamente proporcional ao quadrado do tamanho de ℓ , pois $\hat{\ell} \rightarrow \ell$.
- ii) N é diretamente proporcional à variância da função cuja média quer-se estimar.

A primeira conclusão aponta o problema: o elevado esforço da simulação de evento raro; enquanto a segunda conclusão aponta a solução: redução de variância.

Naturalmente, os algoritmos propostos para lidar com simulação de eventos raros fazem uso de VRTs. A técnica amostragem por importância é considerada a VRT fundamental e é componente essencial do método CE [68].

Retomando o problema da eq. (3.4), considere, agora, que é possível obter uma nova pdf $f(\mathbf{X}; \mathbf{v})$, da mesma família da pdf $f(\mathbf{X}; \mathbf{u})$, mas com vetor de parâmetros $\boldsymbol{\sigma} = \mathbf{v}$. Para fins de maior clareza no equacionamento, $f(\mathbf{X}; \mathbf{v})$ será representado por $g(\mathbf{X}; \mathbf{v})$. Então, a probabilidade ℓ pode ser estimada a partir de uma amostra em $g(\mathbf{X}; \mathbf{v})$ (vide eq. (3.7)). Observe que o estimador $\hat{\ell}$ permanece não tendencioso, pois a razão de verossimilhança $W(\mathbf{X}; \mathbf{u}, \mathbf{v}) = \frac{f(\mathbf{X}; \mathbf{u})}{g(\mathbf{X}; \mathbf{v})}$ corrige a probabilidade de cada estado $\mathbf{X}_i$ amostrado para o âmbito da distribuição original.

$$\hat{\ell} = \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}} \frac{f(\mathbf{X}_i; \mathbf{u})}{g(\mathbf{X}_i; \mathbf{v})} \quad (3.7)$$

O grande desafio é determinar a distribuição amostragem por importância $g(\mathbf{X}; \mathbf{v})$, aquela que reduzirá a variância do evento $\{S(\mathbf{X}) \geq \gamma\}$ de modo que uma amostra menor seja suficiente para estimar ℓ a um nível de incerteza β . Idealmente, a melhor distribuição amostragem por importância, ou pdf-IS (pdf - *importance sampling*), é dada por:

$$g^*(\mathbf{x}) = \frac{I_{\{S(\mathbf{x}) \geq \gamma\}} f(\mathbf{x}; \mathbf{u})}{\ell} \quad (3.8)$$

Note que, ao se considerar $g(\mathbf{X}; \mathbf{v}) = g^*(\mathbf{x})$, qualquer amostra simples $\mathbf{X} \in \mathcal{X}$, seja $\{S(\mathbf{X}) \geq \gamma\}$ verdadeiro ou falso, renderá à expressão de (3.7) o valor exato de ℓ . Em outras palavras, tem-se variância nula, e $g^*(\mathbf{x})$ é a pdf-IS ótima. Evidentemente, tal distribuição é difícil de ser obtida, tendo em vista que ℓ é justamente o parâmetro não conhecido que se quer estimar. É nesse ponto que a minimização da entropia cruzada entra como uma alternativa para obter $g^*(\mathbf{x})$ a partir de um processo que não depende do conhecimento do valor de ℓ . A seguir, o conceito de entropia cruzada é descrito para denotar como o método CE se propõe a determinar a pdf-IS $g(\mathbf{X}; \mathbf{v})$.

3.2.2. Entropia Cruzada

A entropia cruzada entre duas pdfs g e h , também conhecida como a divergência de Kullback-Leibler, é dada pela eq. (3.9) e mede a distância estatística entre distribuições. Observe que a divergência é um valor esperado, obtido a partir da distribuição que se tem como alvo. Logo, $\mathcal{D}$ não pode ser vista como uma distância métrica, pois ela é uma medida referenciada sem simetria. A divergência que h tem para com g é diferente daquela que g tem para com h ; i.e., $\mathcal{D}(g, h) \neq \mathcal{D}(h, g)$.

$$\mathcal{D}(g, h) = \mathbb{E}_g \left[\ln \frac{g(\mathbf{X})}{h(\mathbf{X})} \right] = \int g(\mathbf{x}) \ln g(\mathbf{x}) d\mathbf{x} - \int g(\mathbf{x}) \ln h(\mathbf{x}) d\mathbf{x} \quad (3.9)$$

Visando esclarecer a conceptualização da expressão de entropia cruzada, pode-se defini-la em três níveis de abstração: (i) expectativa da divergência entre as pdfs; (ii) expectativa da divergência logarítmica entre as pdfs; (iii) expectativa do logaritmo da razão de verossimilhança entre as pdfs.

Retomando o problema da simulação com amostragem por importância, agora é possível determinar a pdf-IS ótima por meio da minimização da divergência que a pdf-IS $g(\mathbf{X}; \mathbf{v})$ tem para com a ótima $g^*(\mathbf{x})$. Isso significa encontrar o vetor de parâmetros $\mathbf{v}$ que minimiza a divergência de $g(\mathbf{X}; \mathbf{v})$ para $g^*(\mathbf{x})$. Como $g^*(\mathbf{x})$ é a referência e, portanto, seu vetor de parâmetros não varia, a minimização de:

$$\mathcal{D}(g^*, g(\mathbf{x}; \mathbf{v})) = \int g^*(\mathbf{x}) \ln g^*(\mathbf{x}) d\mathbf{x} - \int g^*(\mathbf{x}) \ln g(\mathbf{x}; \mathbf{v}) d\mathbf{x}$$

depende apenas de $g(\mathbf{x}; \mathbf{v})$. Ou seja, basta minimizar, em $\mathbf{v}$, o termo $(-\int g^*(\mathbf{x}) \ln g(\mathbf{x}; \mathbf{v}) d\mathbf{x})$, ou maximizar o seu oposto:

$$\min \mathcal{D}(g^*, g(\mathbf{x}; \mathbf{v})) \equiv \max_{\mathbf{v}} \int g^*(\mathbf{x}) \ln g(\mathbf{x}; \mathbf{v}) d\mathbf{x}$$

Substituindo $g^*(\mathbf{x})$ por (3.8) e observando que ℓ é um número escalar, tem-se que:

$$\min \mathcal{D}(g^*, g(\mathbf{x}; \mathbf{v})) \equiv \max_{\mathbf{v}} \int I_{\{S(\mathbf{x}) \geq \gamma\}} \ln g(\mathbf{x}; \mathbf{v}) f(\mathbf{x}; \mathbf{u}) d\mathbf{x} \quad (3.10)$$

O vetor $\mathbf{v}$ que resolve a otimização em (3.10) corresponde ao vetor de parâmetros ótimo $\mathbf{v}^*$, com o qual, enfim, é possível obter a pdf-IS ótima sem precisar conhecer o valor de ℓ . Observando que o argumento da maximização em (3.10) é o

valor esperado de $(I_{\{S(\mathbf{x}) \geq \gamma\}} \ln g(\mathbf{x}; \mathbf{v}))$ em $f(\mathbf{X}; \mathbf{u})$, a solução ótima de $\mathbf{v}$ pode ser obtida como:

$$\mathbf{v}^* = \underset{\mathbf{v}}{\operatorname{argmax}} \mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{x}) \geq \gamma\}} \ln g(\mathbf{X}; \mathbf{v})] \quad (3.11)$$

E se o argumento da maximização é uma função convexa e diferenciável em $\mathbf{v}$, $\mathbf{v}^*$ é obtido como a solução do sistema de equações:

$$\mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{x}) \geq \gamma\}} \nabla(\ln g(\mathbf{v}))] = \mathbf{0} \quad (3.12)$$

Essa abordagem é interessante, pois a solução de (3.12) pode ser determinada de forma analítica [68], especialmente no caso de distribuições que pertencem a uma família exponencial natural (FEN). A seguir, o conceito de FENs é apresentado para denotar como a solução de (3.12) pode ser de fato obtida de forma prática.

3.2.3. Família Exponencial Natural

Em probabilidade e estatística, uma família exponencial é um conjunto paramétrico de distribuições com pdf no seguinte formato:

$$f(\mathbf{x}; \boldsymbol{\theta}) = h(\mathbf{x}) e^{(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{t}(\mathbf{x}) - A(\boldsymbol{\theta}))} \quad (3.13)$$

onde $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathbb{R}^q$, $\boldsymbol{\eta}(\boldsymbol{\theta}), \mathbf{t}(\mathbf{x}) \in \mathbb{R}^r$, $A(\boldsymbol{\theta}) \in \mathbb{R}$ e $h(\mathbf{x}) > 0$.

Já uma família exponencial natural é um subconjunto de uma família exponencial, para a qual $\boldsymbol{\eta}(\boldsymbol{\theta})$ e $\mathbf{t}(\mathbf{x})$ são funções identidade, tal que:

$$f(\mathbf{x}; \boldsymbol{\theta}) = h(\mathbf{x}) e^{(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta}))} \quad (3.14)$$

onde $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p$ e $\boldsymbol{\theta} \in \mathbb{R}^p$.

Muitos problemas podem ser modelados por FENs, as quais apresentam elevada capacidade de representação. A maior evidência disso é que uma família multivariada e multiparamétrica $f(\mathbf{x}; \boldsymbol{\theta})$ é equivalente a uma distribuição conjunta de FENs, independentes entre si, de apenas uma variável aleatória e um parâmetro. Neste caso, tem-se que $h(\mathbf{x}) = \prod h_i(x_i)$ e $A(\boldsymbol{\theta}) = \sum A_i(\theta_i)$, tal que:

$$f(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i \in [1, p]} h_i(x_i) e^{(\theta_i x_i - A_i(\theta_i))} \quad (3.15)$$

Definido o conceito e formato da pdf de uma FEN, é possível resolver, sem grande dificuldade, o gradiente $\nabla(\ln f(\boldsymbol{\theta}))$:

$$\begin{aligned}\nabla(\ln f(\boldsymbol{\theta})) &= \nabla \left(\ln \left[h(\mathbf{x}) e^{(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta}))} \right] \right) \\ \nabla(\ln f(\boldsymbol{\theta})) &= \frac{1}{h(\mathbf{x}) e^{(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta}))}} \times \nabla \left(h(\mathbf{x}) e^{(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta}))} \right) \\ \nabla(\ln f(\boldsymbol{\theta})) &= \frac{1}{h(\mathbf{x}) e^{(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta}))}} \times \left(h(\mathbf{x}) e^{(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta}))} \nabla(\boldsymbol{\theta}^T \mathbf{x} - A(\boldsymbol{\theta})) \right)\end{aligned}$$

tal que:

$$\nabla(\ln f(\boldsymbol{\theta})) = \mathbf{x} - \nabla(A(\boldsymbol{\theta})) \quad (3.16)$$

Observe, agora, que um membro de uma FEN tem a propriedade de ter a sua função geradora de cumulantes (*cumulant generating function*), a qual é dada pelo logaritmo neperiano da função geradora de momentos (*moment generating function*), expressa pela seguinte função de $A(\boldsymbol{\theta})$:

$$K_X(\mathbf{t}) = A(\boldsymbol{\theta} + \mathbf{t}) - A(\boldsymbol{\theta}) \quad (3.17)$$

Sabe-se que, para uma função univariada, o primeiro cumulante (i.e., a primeira derivada de $K_X(t)$ em t para $t = 0$) é igual à média da distribuição. No caso multivariado, o primeiro cumulante (i.e., o gradiente de $K_X(\mathbf{t})$ em $\mathbf{t}$ para $\mathbf{t} = \mathbf{0}$) é igual ao vetor de médias $\boldsymbol{\mu}$. Então:

$$\begin{aligned}\boldsymbol{\mu} &= \nabla(K_X(\mathbf{t})) \big|_{\mathbf{t}=\mathbf{0}} \\ \boldsymbol{\mu} &= \frac{\partial}{\partial \mathbf{t}} (A(\boldsymbol{\theta} + \mathbf{t})) \big|_{\mathbf{t}=\mathbf{0}} - \frac{\partial}{\partial \mathbf{t}} (A(\boldsymbol{\theta})) \big|_{\mathbf{t}=\mathbf{0}} \\ \boldsymbol{\mu} &= \frac{\partial}{\partial \mathbf{t}} (A(\boldsymbol{\theta} + \mathbf{t})) \big|_{\mathbf{t}=\mathbf{0}} \\ \boldsymbol{\mu} &= \frac{\partial}{\partial(\boldsymbol{\theta} + \mathbf{t})} (A(\boldsymbol{\theta} + \mathbf{t})) \big|_{\mathbf{t}=\mathbf{0}} \times \frac{\partial}{\partial \mathbf{t}} (\boldsymbol{\theta} + \mathbf{t}) \big|_{\mathbf{t}=\mathbf{0}} \\ \boldsymbol{\mu} &= \frac{\partial}{\partial(\boldsymbol{\theta})} (A(\boldsymbol{\theta})) \times 1 = \nabla(A(\boldsymbol{\theta}))\end{aligned}$$

ou seja:

$$\nabla(A(\boldsymbol{\theta})) = \boldsymbol{\mu} \quad (3.18)$$

Substituindo (3.18) em (3.16):

$$\nabla(\ln f(\boldsymbol{\theta})) = \mathbf{x} - \boldsymbol{\mu} \quad (3.19)$$

Retomando o sistema de equações em (3.12), e considerando que $g(\mathbf{X}; \mathbf{v})$ pertence a uma FEN e é parametrizado pelo vetor de médias $\mathbf{v}$, é possível atualizar o sistema de equações para:

$$\mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) \geq \gamma\}}(\mathbf{X} - \mathbf{v}^*)] = \mathbf{0} \quad (3.20)$$

Manipulando (3.20) tem-se que $\mathbf{v}^*$ pode ser dado por:

$$\mathbf{v}^* = \frac{\mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) \geq \gamma\}}\mathbf{X}]}{\mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) \geq \gamma\}}]} \quad (3.21)$$

No entanto, aqui ainda há um problema: a distribuição original $f(\mathbf{X}; \mathbf{u})$ pode não ser conhecida, ou no mínimo apresentar uma função de difícil modelagem, especialmente no caso de distribuições discretas. Assim, recorre-se à simulação novamente, de modo que o problema de (3.10) pode ser aproximado pelo seu programa estocástico equivalente [68], e, por conseguinte, $\mathbf{v}^*$ pode ser estimado por:

$$\widehat{\mathbf{v}}^* = \frac{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}} \mathbf{X}_i}{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}}} \quad (3.22)$$

E para cada parâmetro v_j^* :

$$\widehat{v}_j^* = \frac{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}} X_{ij}}{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}}} \quad (3.23)$$

Imprescindível observar que as amostras são tomadas em respeito à distribuição original $f(\mathbf{X}; \mathbf{u})$ nos estimadores das equações (3.22) e (3.23).

Apesar de se ter, enfim, um mecanismo direto e prático para estimar os parâmetros da pdf-IS ótima de alguma FEN, este procedimento permanece vulnerável à raridade do evento $\{S(\mathbf{X}) \geq \gamma\}$. Repare que para quase toda a amostra, a função $I_{\{S(\mathbf{X}_i) \geq \gamma\}}$ será nula, o que leva a estimativa de $\mathbf{v}^*$ a ser insuficiente até mesmo para elevados montantes de sorteios. A solução para isto passa pelo emprego da ferramenta IS na forma de um procedimento iterativo de níveis múltiplos. A seguir, tal procedimento é explicado, e, assim, a descrição do método CE é finalizada.

3.2.4. Níveis Múltiplos

Em uma simulação, o evento a ser estimado corresponde a uma função performance $S(\mathbf{X})$ estar acima de um nível γ . Logo, se o nível for rebaixado a um valor $\gamma_1 < \gamma$, o tamanho da amostra, necessária para estimar $\widehat{\mathbf{v}}^*$, será reduzido, pois a probabilidade do evento de interesse será artificialmente elevada de $\mathbb{P}(\{S(\mathbf{X}) \geq \gamma\})$ para $\mathbb{P}(\{S(\mathbf{X}) \geq \gamma_1\})$. Obviamente, levar em consideração estados que não pertencem ao evento de interesse produzirá uma estimativa menos consistente. Contudo, como a função performance, em tese, traduz corretamente a proximidade de um estado para com a ocorrência de $\{S(\mathbf{X}) \geq \gamma\}$, uma escolha adequada de γ_1 induzirá uma estimativa razoável de $\mathbf{v}^*$, de modo que uma nova simulação via IS, com vetor de parâmetros $\widehat{\mathbf{v}}^*$, poderá amostrar o evento $\{S(\mathbf{X}) \geq \gamma\}$ com maior probabilidade. Fica fácil perceber, porém, que, a depender da raridade do evento, mais de uma realização do procedimento descrito será necessária para que uma eventual simulação alcance, para uma amostra reduzida, a proporção desejada de estados com performance $S(\mathbf{X}) \geq \gamma$. Portanto, um processo iterativo de múltiplos níveis $\gamma_k \leq \gamma$, onde k representa o k -ésimo nível, deve ser empregado [68].

Quando se fala em simulação com vetor de parâmetros diferente do vetor original, deve-se ter em mente que a probabilidade dos estados amostrados precisa ser corrigida. Logo, para que a ideia de níveis múltiplos funcione, é necessário adaptar a expressão da estimativa dada em (3.23) para a da eq. (3.24), onde $W(\mathbf{X}_i; \mathbf{u}, \mathbf{w}) = \frac{f(\mathbf{X}_i; \mathbf{u})}{g(\mathbf{X}_i; \mathbf{w})}$ é a razão de verossimilhança entre $f(\mathbf{X}_i; \mathbf{u})$ e $g(\mathbf{X}_i; \mathbf{w})$ e as amostras são tomadas em respeito à distribuição $g(\mathbf{X}_i; \mathbf{w})$. O vetor de parâmetros $\mathbf{w}$ recebe a estimativa de $\mathbf{v}^*$ da iteração anterior (i.e., $\widehat{\mathbf{v}}^*(k-1)$), sendo que na primeira iteração $k = 1$, $\mathbf{w}$ é o vetor original $\mathbf{u}$ e $W(\mathbf{X}_i; \mathbf{u}, \mathbf{w})$ é igual a um para todo $\mathbf{X}_i$.

$$\widehat{v}_j^*(k) = \frac{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma_k\}} W(\mathbf{X}_i; \mathbf{u}, \mathbf{w}) X_{ij}}{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma_k\}} W(\mathbf{X}_i; \mathbf{u}, \mathbf{w})} \quad (3.24)$$

Para a escolha do nível de cada iteração, estipula-se que γ_k deve receber o valor de $S(\mathbf{X})$ do estado amostrado com a pior performance dentre aqueles que compõem o ρ -quantil com as melhores performances. O parâmetro ρ deve receber um valor de probabilidade pequeno, mas não raro, de forma que seja maior que a

probabilidade rara a ser estimada (i.e., a probabilidade ℓ). A ideia é, por meio de $\{S(\mathbf{X}) \geq \gamma_k\}$, classificar, artificialmente, $\rho\%$ da amostra como pertencente ao evento de interesse. Desse modo, o novo vetor de parâmetros levará a próxima simulação a apresentar para $\rho\%$ dos estados amostrados uma performance superior à dos $\rho\%$ estados amostrados na iteração anterior, tal que $\gamma_k > \gamma_{k-1}$. Logicamente, quando γ_k for maior ou igual a γ deve-se interromper o processo, o qual é considerado convergido. Isso ocorre, pois nesse caso tem-se que $\mathbb{P}(\{S(\mathbf{X}) \geq \gamma\}) \geq \mathbb{P}(\{S(\mathbf{X}) \geq \gamma_k\})$, ou seja, sob a distribuição $g(\mathbf{X}; \widehat{\mathbf{v}}^*(k-1))$, a probabilidade do evento original $\{S(\mathbf{X}) \geq \gamma\}$ é no mínimo igual à do evento artificial $\{S(\mathbf{X}) \geq \gamma_k\}$, a qual deve ser aproximadamente igual ao valor de ρ . Logo, a simulação final, feita com a pdf-IS $g(\mathbf{X}; \widehat{\mathbf{v}}^*(k))$, irá amostrar o evento de interesse com probabilidade maior do que ρ , ao invés da probabilidade ℓ , que é rara.

A seguir, o algoritmo do método CE é formalizado.

3.2.5. Algoritmo

O método CE, conforme já apontado, é um processo iterativo a dois estágios de simulação e atualização de parâmetros de pdf-IS. Os conceitos apresentados de entropia cruzada, família exponencial natural e níveis múltiplos combinam para formar o método cujo algoritmo pode ser apreciado pelo passo a passo abaixo [68].

Passo 1) Carregar dados do problema e definir os parâmetros:

- i. N – número de amostras por iteração;
 - ii. ρ – parâmetro de níveis múltiplos ou probabilidade de $\{S(\mathbf{X}) \geq \gamma_k\}$.
- Fazer $\mathbf{v}(0) = \mathbf{u}$ e $k = 1$;

Passo 2) Gerar amostra $\mathbf{X}_1, \dots, \mathbf{X}_N$ a partir da distribuição $f(\mathbf{X}; \mathbf{v}(k-1))$, calcular a performance $S(\mathbf{X})$ de cada estado e arranjar os resultados em ordem crescente: $S_{(1)} \leq \dots \leq S_{(N)}$. Faça γ_k igual ao quantil $(1 - \rho)$ do vetor ordenado de performances ou a γ , tal como:

$$\gamma_k = \begin{cases} S_{([1-\rho]N)}, & \text{se } S_{([1-\rho]N)} < \gamma \\ \gamma, & \text{se } S_{([1-\rho]N)} \geq \gamma \end{cases} \quad (3.25)$$

Passo 3) Estimar cada elemento do novo vetor de parâmetros $\mathbf{v}(k)$, a partir da amostra obtida no Passo 2), conforme:

$$v_j(k) = \frac{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma_k\}} W(\mathbf{X}_i; \mathbf{u}, \mathbf{v}(k-1)) X_{ij}}{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma_k\}} W(\mathbf{X}_i; \mathbf{u}, \mathbf{v}(k-1))}, \quad j \in \Omega^v \quad (3.26)$$

onde Ω^v é o conjunto das variáveis aleatórias que compõem o problema;

Passo 4) Se $\gamma_k < \gamma$, faça $k = k + 1$ e retorne ao Passo 2); senão, prossiga;

Passo 5) Faça $\widehat{\mathbf{v}}^* = \mathbf{v}(k)$, gere a amostra final $\mathbf{X}_1, \dots, \mathbf{X}_N$, a partir de $g(\mathbf{X}; \widehat{\mathbf{v}}^*)$, e estime a probabilidade rara ℓ por meio do estimador não tendencioso:

$$\hat{\ell} = \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \gamma\}} \frac{f(\mathbf{X}_i; \mathbf{u})}{g(\mathbf{X}_i; \widehat{\mathbf{v}}^*)} \quad (3.27)$$

onde N é redimensionado para acomodar o nível desejado de incerteza.

3.3.

Método da Entropia Cruzada em Confiabilidade

O método CE já integrou inúmeros trabalhos, tanto em confiabilidade da geração quanto composta (vide Subseção 1.2.2). A adaptação do método para a análise da geração foi bem-sucedida, pois o problema é facilmente modelado nesse caso. O problema composto GT, por sua vez, é de complexidade elevada, sobretudo na qualificação dos estados que não pertencem ao evento de interesse (i.e., estados de sucesso), aspecto fundamental para a efetividade do processo de níveis múltiplos. Consequentemente, os trabalhos existentes optaram por fazer simplificações na modelagem do problema, as quais, como é demonstrado nessa seção, não são adequadas e limitam a aplicabilidade do método. Como alternativa, uma nova metodologia é proposta, trazendo as seguintes contribuições:

- i) emprego efetivo do método CE na avaliação de sistemas GT;
- ii) flexibilidade no foco de análise, seja ele para sistema/área/barra;
- iii) procedimento de estimação da pdf-IS com convergência automática;
- iv) menos parâmetros para definição do usuário;
- v) potencial de ganho computacional (i.e., *speed-up*) previamente dimensionável.

3.3.1. Confiabilidade de Geração

O estudo do NH1 (Seção 2.2) busca conhecer a falha de suprimento por déficit de geração (i.e., a falha G), cujo nível de raridade é elevado, principalmente em sistemas interligados. Essa raridade é coerente com a exigência da sociedade por suprimento ininterrupto do curto ao longo prazo. Portanto, verificar tal raridade tem sua importância, e vai de encontro com o objetivo do método CE. A seguir, a metodologia geralmente empregada em confiabilidade da geração é descrita [72].

Na maioria dos trabalhos, a disponibilidade de geradores segue um modelo de Markov a dois estados (conforme Capítulo 2), ou seja, segue uma distribuição Bernoulli, a qual é uma FEN. As variáveis aleatórias do problema, porém, correspondem ao número de geradores indisponíveis na respectiva estação geradora. Como é considerado que os geradores de uma estação são idênticos, as variáveis aleatórias seguem distribuição binomial, também uma FEN. A carga e capacidade energética de geradores renováveis também podem ser modeladas como variável aleatória, contudo, a distorção de suas pdfs dificilmente é implementada.

A falha G é observada pela comparação simples entre a carga momentânea e a capacidade de geração disponível $GCAP(\mathbf{X})$ no estado $\mathbf{X}$. A função performance $S(\mathbf{X})$ é costumeiramente definida como $GCAP(\mathbf{X})$, enquanto o nível γ recebe algum valor de carga pesada L^{pesada} , tendo em vista que a falha ocorre com maior probabilidade na situação de carga pesada. O evento raro, que reflete o comportamento da falha G , é dado por $\{S(\mathbf{X}) < \gamma\}$ (note a alteração do sinal de desigualdade em relação ao método genérico da Seção 3.2). Dada a escolha de um nível de carga pesada e adotando a ordenação decrescente de $S(\mathbf{X})$, uma amostra apresenta a seguinte disposição: a primeira posição é reservada ao caso de sucesso mais confortável ($S(\mathbf{X}) \rightarrow GCAP^{max}$); daí em diante, todos os estados serão de sucesso até chegar no pior caso de sucesso ($S(\mathbf{X}) = L^{pesada}$); na sequência, têm-se os estados de falha, cada vez mais severos, até chegar no pior caso global ($S(\mathbf{X}) \rightarrow 0$). Evidentemente, para que a modelagem seja coerente, as simulações do processo de estimação da pdf-IS precisam ser realizadas com a carga pesada que foi atribuída a γ .

Diferentemente do que será observado no problema da confiabilidade composta, é fácil notar que os estados de sucesso são adequadamente classificados, pois a interpretação é bem direta: quanto maior a folga entre a geração disponível e a

carga pesada escolhida, maior é, em média, a distância do sistema para com a ocorrência de déficit de geração. Talvez $S(\mathbf{X})$ pudesse ser mais bem definida como a razão entre a geração disponível e a carga momentânea. Desse modo, a classificação de estados como falha/sucesso seria mais fiel ao problema de carga variável, tal que $S(\mathbf{X}) \geq 1,0$ representaria os estados de sucesso e $S(\mathbf{X}) < 1,0$ os de falha. Não obstante, a classificação proposta em [72] é suficientemente coerente e eficiente, para esse caso de confiabilidade da geração, o qual é de simples modelagem.

3.3.2. Confiabilidade Composta

Apesar de ser relevante confirmar que um sistema opera com um nível de confiabilidade de geração muito próximo de 1 pu, sabe-se que a confiabilidade do NH2 (Seção 2.2) é mais importante. O conjunto GT apresenta falha bem menos rara e é muito mais complexo, o que justifica maior atenção e maiores esforços para otimizar os algoritmos que o avaliam. A seguir, a metodologia geralmente empregada em confiabilidade composta [76] é descrita e suas limitações comentadas, enquanto, na próxima subseção, a metodologia proposta é enfim apresentada.

Adicionalmente ao que foi descrito para confiabilidade de geração, os circuitos também são representados por modelo de Markov a dois estados (conforme Capítulo 2), sendo sua indisponibilidade também uma variável aleatória com distribuição Bernoulli, uma FEN. A falha composta, conforme é explicado na Seção 2.2, engloba a falha G, compondo-a com a falha por “gargalos” da transmissão (i.e., a falha T). Por isso é uma falha bem menos rara e de complexidade bem maior. A modelagem da função performance $S(\mathbf{X})$ se torna algo desafiador, pois estados de sucesso apresentam sempre a mesma saída na análise de adequação de estados: não há corte de carga. Na tentativa de contornar esse desafio sem fazer um apropriado desacoplamento da falha composta em falhas G e T, foi assumido que a capacidade de geração disponível poderia ser mais uma vez um qualificador de estados de sucesso. Semelhante à confiabilidade de geração, as simulações para estimação da pdf-IS são realizadas com a carga pesada L^{pesada} atribuída ao nível γ . O evento raro também é dado por $\{S(\mathbf{X}) < \gamma\}$, enquanto a função $S(\mathbf{X})$ é definida como:

$$S(\mathbf{X}) = \begin{cases} L^{pesada} - L^{corte}(\mathbf{X}), & \text{se } L^{corte}(\mathbf{X}) > 0 \\ GCAP(\mathbf{X}), & \text{se } L^{corte}(\mathbf{X}) = 0 \end{cases} \quad (3.28)$$

onde $L^{corte}(\mathbf{X})$ é o corte de carga do estado $\mathbf{X}$.

Observe que, em comparação ao problema de geração, a função $S(\mathbf{X})$ é a mesma quando o estado é de sucesso, mas é diferente em estados de falha, pois estes podem ocorrer mesmo quando $GCAP(\mathbf{X})$ é maior do que L^{pesada} .

Assumindo que $GCAP(\mathbf{X})$ é de fato uma variável explicativa da falha composta, a metodologia já apresenta um problema. A consideração de carga constante e pesada nas simulações aumenta a ocorrência da falha e proximidade geral dos estados de sucesso para com a ocorrência da falha. Assim, é bem possível que $\rho\%$ dos estados amostrados sejam de falha já na primeira iteração do processo de níveis múltiplos, o que atende ao critério de parada. Isso pode parecer bom, mas, na verdade, o benefício do processo de níveis múltiplos, o qual eleva a probabilidade da falha paulatinamente, está sendo anulado. Nesse caso, não importa que a probabilidade do evento em estudo esteja sendo elevada, pois o evento falha ocorre de múltiplas maneiras (i.e., para diferentes patamares de carga), as quais não estão sendo contempladas. O método distorce a indisponibilidade de equipamentos para uma situação de carga constante, enquanto a simulação final estima índices de confiabilidade composta para uma situação de carga variável. Em outras palavras, existe incompatibilidade entre o problema que dá origem à pdf-IS e o problema que de fato será simulado com a pdf-IS. Consequentemente, ainda que uma parcela maior da amostra final seja de estados de falha, a incerteza das estimativas de índices permanece elevada, pois a representação da falha codificada pela pdf-IS é pobre.

O problema descrito no parágrafo anterior pode ser resolvido com a consideração da curva de carga nas simulações do processo de níveis múltiplos e nova função performance:

$$S(\mathbf{X}) = \begin{cases} \frac{L(\mathbf{X}) - L^{corte}(\mathbf{X})}{L(\mathbf{X})}, & \text{se } L^{corte}(\mathbf{X}) > 0 \\ \frac{GCAP(\mathbf{X})}{L(\mathbf{X})}, & \text{se } L^{corte}(\mathbf{X}) = 0 \end{cases} \quad (3.29)$$

onde $L(\mathbf{X})$ é a carga amostrada no estado $\mathbf{X}$.

Desse modo, a lógica de níveis múltiplos é atendida, sendo os estados de sucesso responsáveis por conduzir o processo. Entretanto, a função performance para estados de sucesso permanece inadequada. Apesar da falha G apresentar relação direta com capacidade de geração disponível, é a falha T que domina a composição

da falha de sistemas GT, e ela certamente não apresenta relação direta com $GCAP(X)$. Para a falha T, é a disposição geográfica da geração disponível combinada com as restrições físicas da rede de transmissão (i.e., capacidade de fluxo e leis de Kirchhoff de circuitos) que governam o suprimento da carga ao longo da rede. Infelizmente, conforme já destacado, a análise de adequação de sistemas GT só é capaz de qualificar o corte de carga, de modo que os estados de sucesso são todos iguais na perspectiva da ferramenta de análise. Por este motivo, propõe-se, nesta tese, substituir o mecanismo de níveis múltiplos por uma lógica mais simples. A nova lógica é predicada no uso de apenas estados de falha na estimação da pdf-IS e no controle da incerteza da estimativa da pdf-IS. Importante destacar também, que apenas a falha T é considerada devido à sua total dominância sobre a falha G, a qual é desprezível em sistemas interligados, tais como a grande maioria dos sistemas modernos. Ademais, se a análise da falha G é desejada, a mesma deve ser realizada separadamente da análise da falha T, tendo em vista as características bastante distintas entre os tipos de falha. As duas próximas subseções apresentam a proposta e o seu algoritmo.

3.3.3. Proposta

Em vista da dificuldade de se definir uma função performance afeita ao problema da falha T e consistente para os estados de sucesso, o processo de níveis múltiplos foi preterido por uma lógica mais simples. A nova lógica vem da observação prática de que em qualquer simulação é possível monitorar a incerteza da estimativa de cada parâmetro em estudo via coeficiente β . Isso significa que β pode funcionar como um elemento qualificante para a estimativa do vetor de parâmetros ótimo da pdf-IS. Portanto, a entropia cruzada entre a pdf-IS e a ótima pode ser minimizada por meio de uma única simulação convencional (i.e., feita com a pdf original) tomando o próprio β como critério de parada. Evidentemente, a tolerância adotada para a incerteza nesta fase de distorção de parâmetros deve ser maior que a desejada para a convergência das estimativas dos índices de confiabilidade. Uma forma interessante de determinar essa tolerância é por meio de (3.6), da Subseção 3.2.1. Antes de explicitar tal sugestão, porém, é necessário destacar que o método proposto é composto por duas fases:

Fase 1: Simulação Monte Carlo convencional para estimação de pdf-IS, com parada em β^{F1} .

Fase 2: Simulação Monte Carlo com IS (SMCIS) para estimação de índices, com parada em β^{F2} .

Observe que associado à escolha de β^{F1} existe um ganho máximo que o método proposto pode fornecer dado pela razão entre o tamanho da amostra necessária a estimar índices via uma SMC convencional, com parada em β^{F2} , e o daquela necessária a estimar a pdf-IS na Fase 1, também via SMC convencional, mas com parada em β^{F1} . Claramente, este ganho máximo ocorre sob o pressuposto de que a Fase 2 tem custo computacional desprezível. Desse modo, é possível obter conhecimento prévio sobre o ganho máximo dividindo a equação em (3.6) tomada para β^{F2} pela tomada para β^{F1} :

$$MAXGAIN = \left[\frac{\hat{V}(I_{\{S(X) \geq \gamma\}})}{(\beta \times \hat{\ell})^2} \right]_{\beta=\beta^{F2}} \bigg/ \left[\frac{\hat{V}(I_{\{S(X) \geq \gamma\}})}{(\beta \times \hat{\ell})^2} \right]_{\beta=\beta^{F1}} \quad (3.30)$$

Como ambas as amostras são simuladas via SMC convencional, sob o mesmo espaço probabilístico da pdf original, a variância amostral e média amostral da função $I_{\{S(X) \geq \gamma\}}$ não são as mesmas, mas tendem para os mesmos valores (variância e valor esperado). Logo, é possível aproximar o ganho máximo por:

$$MAXGAIN \cong \frac{(\beta^{F1})^2}{(\beta^{F2})^2} \quad (3.31)$$

Neste trabalho, índices de confiabilidade composta são estimados a $\beta^{F2} = 5\%$. Como exemplo, se β^{F1} for dimensionado em 50%, o ganho máximo seria próximo de 100 vezes. Contudo, para uma incerteza tão elevada, os valores da variância e média amostrais de $I_{\{S(X) \geq \gamma\}}$ tomadas para β^{F1} devem ser bem diferentes dos daquelas tomadas para β^{F2} , invalidando a relação em (3.31). Desse modo, a estimativa de ν^* será pobre, o que deve levar a Fase 2 a ser mais onerosa ao invés de menos. Sendo mais conservador e aplicando um β^{F1} de 10%, mais próximo de β^{F2} , já é possível obter um ganho máximo razoável ($\cong 4$ vezes) com um bom grau de certeza. Em confiabilidade composta, β^{F1} e β^{F2} caracterizam a probabilidade da falha T, sendo $\beta^{F1} = 10\%$ um valor bem seguro, que invariavelmente levará a Fase 1 a produzir uma estimativa consistente para ν^* e, assim, uma Fase 2 rápida. No entanto,

a estimação do índice de frequência via o método de transição um passo à frente (OSF – *one-step forward*) é o que define a convergência da simulação, de modo que sua incerteza ainda é bem elevada quando a da estimativa de probabilidade atinge o nível de 10%. Assim, foi estipulado que a incerteza do índice que converge por último deve ser no máximo de 20%, dado que a do índice LOLP (*loss of load probability*) atingiu 10%. Na Fase 2, a estimativa de todos os índices está sujeita ao mesmo critério de $\beta^{F2} = 5\%$.

O ganho máximo deve, então, ser aproximado em função do índice LOLF (*loss of load frequency*), que invariavelmente converge por último. Logo, a depender dos valores de variância e média amostrais tomadas para β^{F1} e β^{F2} , o ganho máximo estará em torno de $20^2/5^2 = 16$ vezes. Vale destacar que a LOLF é estimada por meio da técnica OSF, a qual possui processo amostral próprio, tornando o valor de ganho máximo um pouco menos preciso.

Antes de apresentar o algoritmo, é necessário detalhar mais dois importantes aspectos do método. Primeiro, tem-se que o evento raro $\{S(\mathbf{X}) > \gamma\}$ corresponde à falha T, de modo que a função $S(\mathbf{X})$ é o tamanho do corte de carga no estado $\mathbf{X}$ e o nível que qualifica a falha é $\gamma = 0$. O segundo aspecto é o conceito de condicionamento da pdf-IS. Ainda que β^{F1} de 10% para a LOLP e 20% para a LOLF sejam considerados adequados, deve-se ter em mente que a pdf-IS levará a SMCIS a atuar adentro da região do espaço de estados em que a falha T reside. Isso significa que uma quantidade demasiada de fluxo de potência ótimo (FPO) será executada sem necessidade, pois muitos estados serão de falha profunda (i.e., o estado está a múltiplas transições do sucesso). Como a região da falha de maior probabilidade é a localizada na fronteira sucesso-falha, uma pdf-IS menos “agressiva” produzirá uma SMCIS mais eficiente, com número reduzido de execuções de FPO (lembrando que um FPO só é acionado caso um fluxo DC com perdas (vide Seção 2.7) apresente violação de capacidade em circuitos). Portanto, um amortecimento α é aplicado ao vetor $\widehat{\mathbf{v}}^*$, vide eq. (3.32). A ideia é compor o novo vetor com uma parcela da estimativa $\widehat{\mathbf{v}}^*$ e uma de $\mathbf{u}$, que é o vetor de parâmetros da pdf original do problema.

$$\widehat{\mathbf{v}}^* = \mathbf{u} \times (1 - \alpha) + \widehat{\mathbf{v}}^* \times \alpha \quad (3.32)$$

A ideia de amortecimento da estimativa da pdf-IS vem da proposta original do método CE [68], mas é pouco explorada nos trabalhos existentes em

confiabilidade de sistemas elétricos de potência, para os quais α costuma ser dimensionado com valores próximos de um, ou seja, quase nenhum amortecimento. Para este trabalho, buscou-se dimensionar α da maneira mais abrangente possível. Assim, foi determinado que o amortecimento deverá ser muito forte para as falhas que não apresentam raridade ($LOLP \geq 1\%$), forte para as de baixo nível de raridade ($LOLP < 1\%$) e fraco para as de elevado nível de raridade ($LOLP < 0,01\%$). Essa configuração é coerente, pois o amortecimento reduz (i.e., α cresce) ao passo que a raridade, e, portanto, a distância da amostra para a falha, aumenta. A raridade, por sua vez, pode ser facilmente averiguada pelo valor da estimativa da LOLP a β^{F1} , ao final da Fase 1.

3.3.4. Algoritmo

O método proposto pode ser resumido pelo seguinte passo a passo:

Passo 1) Carregar dados do problema, definir os parâmetros: (i) $\beta_{LOLP}^{F1} = 10\%$ e (ii)

$\beta_{LOLF}^{F1} = 20\%$, e fazer $N = 1$;

Passo 2) Gerar $\mathbf{X}_N$ a partir da pdf original $f(\mathbf{X}; \mathbf{u})$ e calcular sua performance $S(\mathbf{X}_N)$. Estimar índices para a amostra $\mathbf{X}_1, \dots, \mathbf{X}_N$ e calcular as incertezas das estimativas;

Passo 3) Se $\beta_{LOLP} > \beta_{LOLP}^{F1}$ ou $\beta_{LOLF} > \beta_{LOLF}^{F1}$, faça $N = N + 1$ e retorne ao Passo 2); senão, prossiga;

Passo 4) Estimar o novo vetor de parâmetros $\mathbf{v} = \begin{bmatrix} \mathbf{v}^G \\ \mathbf{v}^C \end{bmatrix}$.

- Para estações geradoras (G):

$$v_j^G = \frac{1}{n_j^{uni}} \left[\frac{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) > \gamma\}} X_{ij}^G}{\sum_{i=1}^N I_{\{S(\mathbf{X}_i) > \gamma\}}} \right], \quad j \in \Omega^G \quad (3.33)$$

onde Ω^G é o conjunto de estações geradoras; n_j^{uni} o número de unidades geradoras da estação j ; X_{ij}^G o número de geradores indisponíveis na estação j ; e v_j^G o novo parâmetro probabilidade de estar indisponível dos geradores da estação j .

- Para circuitos (C):

$$v_j^C = \frac{\sum_{i=1}^N I_{\{S(X_i) > \gamma\}} X_{ij}^C}{\sum_{i=1}^N I_{\{S(X_i) > \gamma\}}} , \quad j \in \Omega^C \quad (3.34)$$

onde Ω^C é o conjunto de circuitos; X_{ij}^C a condição de indisponível do circuito j ; e v_j^C o novo parâmetro probabilidade de estar indisponível do circuito j ;

Passo 5) Defina α como:

$$\alpha = \begin{cases} 0,01, & \text{se } \mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) > \gamma\}}] \geq 1 \times 10^{-2} \\ 0,10, & \text{se } \mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) > \gamma\}}] < 1 \times 10^{-2} \\ 0,50, & \text{se } \mathbb{E}_{\mathbf{u}}[I_{\{S(\mathbf{X}) > \gamma\}}] < 1 \times 10^{-4} \end{cases} \quad (3.35)$$

Passo 6) Faça $\widehat{\mathbf{v}}^* = \mathbf{u} \times (1 - \alpha) + \mathbf{v} \times \alpha$ e estime os índices de confiabilidade a partir de $g(\mathbf{X}; \widehat{\mathbf{v}}^*)$ por meio do estimador não tendencioso:

$$\hat{\rho} = \frac{1}{N} \sum_{i=1}^N H(\mathbf{X}) \frac{f(\mathbf{X}_i; \mathbf{u})}{g(\mathbf{X}_i; \widehat{\mathbf{v}}^*)} \quad (3.36)$$

onde N é redimensionado para acomodar o nível de incerteza $\beta^{F2} = 5\%$ e $H(\mathbf{X})$ é uma função mensurável associada ao índice a ser estimado.

Importante destacar que para o modelo de unidade equivalente (Subseção 2.5.3) a probabilidade dos estados deve ser atualizada pelo novo parâmetro v_j^G .

3.4. Avaliação do Método Proposto

Nesta seção, a metodologia proposta é avaliada por meio de uma série de análises sobre um sistema teste com características semelhantes às de sistemas modernos interligados. Primeiro, são apresentadas as análises de sistema, área/zona e barra para ilustrar a típica relação entre o nível de raridade da falha e a abrangência do foco de análise, e como a ferramenta proposta se comporta nessas situações. Várias observações importantes são feitas sobre desempenho e precisão de resultados. Na sequência, vários casos com determinadas barras são apresentados para avaliar o aspecto de dispersão da falha, o mais influente na performance do método além da raridade. Nesses casos, novas configurações do sistema teste são utilizadas,

de modo que suas análises possam corroborar as suposições feitas acerca do método e do evento falha de suprimento de sistemas elétricos de potência. Em [109], um estudo preliminar da metodologia proposta, sem análise de dispersão da falha, pode ser contemplado. Os índices exibidos nos resultados das análises são: LOLP (*loss of load probability*), EENS (*expected energy not supplied*) e LOLF (*loss of load frequency*). Em termos de performance, são exibidos: número de estados amostrados, número de execuções de fluxo de potência ótimo e tempo de processamento.

O método proposto neste capítulo corresponde, em resumo, a uma simulação Monte Carlo quasi-sequencial com amostragem por importância via método da entropia cruzada para avaliação da confiabilidade de sistemas de geração e transmissão, que considera adequadamente as restrições da rede de transmissão. Portanto, a ferramenta que incorpora o método é referida como SMCIS-GT (SMC quasi-sequencial com IS via método CE para sistemas GT), enquanto a ferramenta usada como *benchmark* é referida como SMCC (SMC convencional). A SMCC é nada mais que uma SMC realizada com a distribuição original do problema. Por fim, sobre a implementação computacional, tem-se que todas as rotinas foram desenvolvidas em linguagem FORTRAN e as simulações executadas em computador de CPU i7-12650H 2,30-4,70 GHz. Da rotina de análise de adequação de estados, destaca-se que o *solver* de programação linear é o pacote GLPK (GNU Linear Programming Kit) [110].

3.4.1. Sistema RTS-96-REN-T

Os sistemas modernos são, em geral, bem interligados e são planejados com o imperativo de disponibilidade energética ininterrupta. Portanto, eles estão sujeitos à conjectura de falha G desprezível e falha T rara, mas mensurável. Ou seja, a falha composta é basicamente a falha T. Logo, todos os índices exibidos nesta seção são referentes à falha composta, ainda que a amostragem por importância da ferramenta SMCIS-GT seja orientada apenas pela falha T.

Buscando um cenário apropriado para validar a metodologia proposta, em acordo ao que foi discutido no último parágrafo, recorreu-se ao sistema IEEE Reliability Test System – 1996 (RTS-96) [111], o qual corresponde a três sistemas IEEE Reliability Test System – 1979 (RTS-79) [112] interligados. O RTS-96 é fruto de um esforço para aumentar o escopo do seu antecessor, o RTS-79, com uma

configuração e representação de diferentes tecnologias mais abrangente. Esse sistema é, portanto, bastante utilizado para apreciar novas metodologias de confiabilidade, pois, além de suas características abrangentes, pode ser facilmente modificado para adequar o contexto de estudo. Na Tabela 3.1, a avaliação convencional (i.e., via SMCC) da confiabilidade do sistema RTS-96 pode ser observada. Importante destacar que neste e em todos os testes apresentados neste documento, a carga segue a curva horária anual do sistema RTS-96, com total correlação entre barras.

Tabela 3.1: Análise do sistema (RTS-96).

Índices		
LOLP [%]	EENS [MWh/ano]	LOLF [oc./ano]
$3,38 \times 10^{-3}$ ($\beta=2,67\%$)	50,1 ($\beta=3,79\%$)	0,104 ($\beta=5,00\%$)
Performance		
Sorteios	FPO	Tempo
41.626.700	316.703	22,55 min

Apesar do aspecto de interligação entre subsistemas, importante para caracterizar um sistema moderno, o RTS-96 apresenta uma falha G considerável, com participação em quase metade da composição dos índices, e falha composta excessivamente rara ($LOLP < 0,01\%$) para um sistema de médio porte. Isso ocorre, pois esse sistema foi projetado com reserva de geração um pouco aquém do ideal e capacidade de fluxo para circuitos demasiadamente elevada, características herdadas do RTS-79. Outro aspecto relevante para sistemas modernos é a crescente participação de energia renovável. No entanto, os sistemas RTS-79 e RTS-96 apresentam como fontes renováveis apenas alguns geradores hidráulicos, os quais, no RTS-96, estão localizados nas barras 122, 222 e 322 (vide Figura 3.1).

Com o intuito de tornar as características do sistema RTS-96 mais realistas, uma nova versão é confeccionada a partir de duas modificações:

Modificação 1: Visando uma configuração com participação renovável maior, alguns geradores térmicos são substituídos por hidráulicos e eólicos (vide Tabela 3.2). O RTS-96 possui 10.215 MW de capacidade, sendo 900 MW de geração renovável (hidráulica). Com a modificação, o sistema passa a ter 13.996 MW de capacidade, sendo $6931 + 900 = 7831$ MW de geração renovável (55,95% da matriz).

Modificação 2: Visando uma configuração com característica de confiabilidade composta mais coerente, um fator de 0,90 é aplicado à carga de todos os consumidores do sistema para reduzir os índices da falha G a patamares desprezíveis, enquanto um fator de 0,45 é aplicado à capacidade de fluxo de todo circuito para acentuar a falha T.

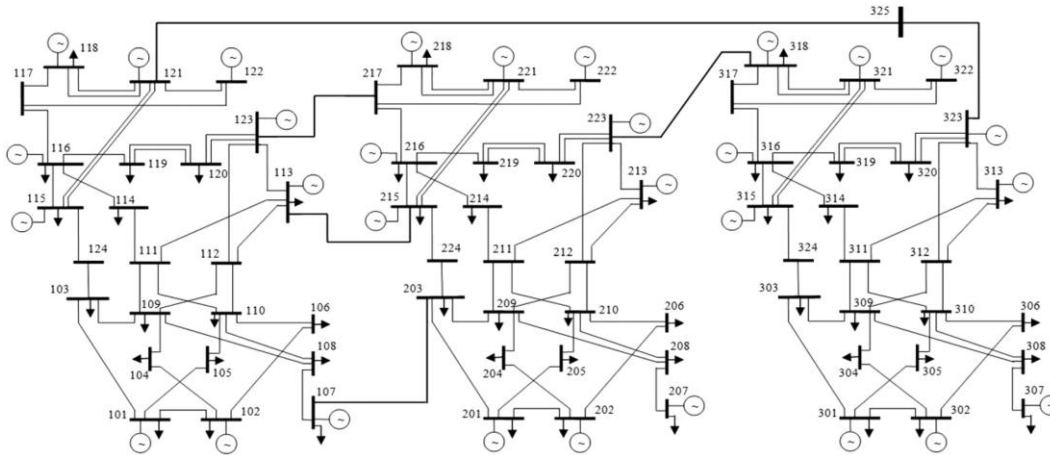


Figura 3.1: Diagrama representativo do sistema RTS-96.

Tabela 3.2: Substituição de unidades térmicas por renováveis.

Barra	RTS-96		RTS-96-REN-T	
	Tipo	Cap. [MW]	Tipo	Cap. [MW]
101	Térmica	$2 \times 76 = 152$	Eólica	$319 \times 2,5 = 797,5$
102	Térmica	$2 \times 76 = 152$	Hidro	$2 \times 84,0 = 168,0$
113	Térmica	$3 \times 197 = 591$	Hidro	$3 \times 212,0 = 636,0$
115	Térmica	$1 \times 155 = 155$	Eólica	$325 \times 2,5 = 812,5$
201	Térmica	$2 \times 76 = 152$	Eólica	$181 \times 2,5 = 452,5$
202	Térmica	$2 \times 76 = 152$	Hidro	$2 \times 80,0 = 160,0$
213	Térmica	$3 \times 197 = 591$	Hidro	$3 \times 253,0 = 759,0$
215	Térmica	$1 \times 155 = 155$	Eólica	$184 \times 2,5 = 460,0$
301	Térmica	$2 \times 76 = 152$	Eólica	$285 \times 2,5 = 712,5$
302	Térmica	$2 \times 76 = 152$	Hidro	$2 \times 96,0 = 192,0$
313	Térmica	$3 \times 197 = 591$	Hidro	$3 \times 352,0 = 1056,0$
315	Térmica	$1 \times 155 = 155$	Eólica	$290 \times 2,5 = 725,0$
Total	Térmica	3150	Renovável	6931

Assim, tem-se o novo sistema RTS-96-REN-T, de participação renovável elevada, falha composta que é basicamente a falha T e uma adequada disposição de níveis de raridade para sistema, área/zona e barramento (i.e., a falha do sistema é não rara, de área/zona é pouco rara e de barra é bastante rara). Os dados do RTS-

96-REN-T podem ser observados no Apêndice A, exceto pelas séries das fontes eólicas, as quais apresentam quantidade de dados muito elevada (5 séries/anos de 8760 entradas/horas para 3 zonas de vento) e podem ser observadas em [113].

3.4.2. Análise de Sistema

Fazendo uma análise com foco na falha de todo o sistema RTS-96-REN-T, têm-se os resultados da Tabela 3.3. A relevância desse teste é demonstrar como a ferramenta SMCIS-GT se comporta em situação de falha não rara ($LOLP \geq 1\%$), circunstância que reduz a eficácia e necessidade da amostragem por importância. Como esperado, nenhum ganho real foi constatado, mas sem prejuízo à performance computacional e estimativa de índices. Isso comprova que o método proposto pode ser utilizado indiscriminadamente, o que é importante, tendo em vista que a raridade da falha não é conhecida a priori.

Tabela 3.3: Análise do sistema (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio
LOLP [%]	1,23 ($\beta=1,67\%$)	1,23 ($\beta=1,68\%$)	0,00%
EENS [MWh/ano]	1580 ($\beta=2,97\%$)	1610 ($\beta=3,01\%$)	1,90%
LOLF [oc./ano]	31,1 ($\beta=5,00\%$)	32,6 ($\beta=5,00\%$)	4,82%
Performance	SMCC	SMCIS-GT	Ganho
Sorteios	289.893	277.333	1,05 pu
FPO	66.004	63.717	1,04 pu
Tempo	4,27 min	4,12 min	1,04 pu

3.4.3. Análise de Zona

Suponha, agora, que um agente de transmissão do sistema RTS-96-REN-T atue na região de tensão de 138 kV da Área 3. Nesse caso, não é do interesse do agente avaliar a confiabilidade do sistema, talvez nem mesmo da sua área, mas somente das barras consumidoras ao sul da sua região. Mais especificamente, quer-se conhecer a confiabilidade da zona que envolve as barras 301, 302, 304, 305, 307 e

308, conjunto destacado pelo retângulo laranja na Figura 3.2. O fato de que há inserção de renováveis nas barras 301 e 302, torna a análise da zona ainda mais crítica.

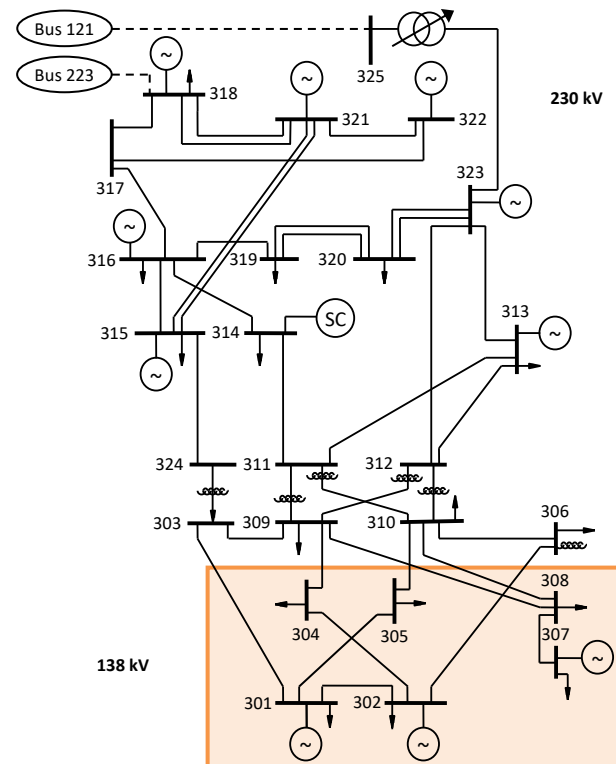


Figura 3.2: Diagrama representativo da Área 3 do sistema RTS-96-REN-T.

Tabela 3.4: Análise da zona A3SUL (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio
LOLP [%]	$2,42 \times 10^{-2}$ ($\beta=1,65\%$)	$2,40 \times 10^{-2}$ ($\beta=1,45\%$)	0,83%
EENS [MWh/ano]	39,6 ($\beta=2,46\%$)	38,1 ($\beta=2,06\%$)	3,79%
LOLF [oc./ano]	0,595 ($\beta=5,00\%$)	0,603 ($\beta=5,00\%$)	1,34%
Performance	SMCC	SMCIS-GT	Ganho
Sorteios	15.213.052	6.448.294	2,36 pu
FPO	3.456.715	1.508.261	2,29 pu
Tempo	3,84 h	1,84 h	2,10 pu

A Tabela 3.4 apresenta os resultados da análise da zona em questão, chamada de zona A3SUL. Observe que, nesse caso, a falha já pode ser considerada rara, ou pouco rara ($0,01\% < \text{LOLP} < 1\%$), o que leva a ferramenta SMCIS-GT a atuar ainda com algum conservadorismo (i.e., amortecimento de $\alpha = 0,10$) e, assim, a apresentar ganho ainda pequeno. Não há como escapar desse princípio, pois, conforme

discutido na Subseção 3.3.3, se o amortecimento fosse mais leve (i.e., α maior), uma amostra menor seria gerada por SMCIS-GT, mas com uma carga de FPO maior, o que significa custo computacional maior.

Com relação às estimativas dos índices, observa-se que o desvio para a EENS é maior que a incerteza alcançada na SMCC. Contudo, deve-se constatar também que pela sua faixa de incerteza ($\pm 2,06\%$) a estimativa em SMCIS-GT pode chegar a $38,1 \times 1,0206 = 38,9$, valor cujo desvio para a estimativa em SMCC é de $1,77\%$ e inferior a $\beta = 2,46\%$. Ou seja, há interseção entre as faixas de incerteza das estimativas e, portanto, a estimativa de todos os índices apresenta precisão satisfatória.

Por fim, destaca-se que uma ferramenta baseada na metodologia em [76] (primeiro trabalho a adaptar o método CE ao problema de confiabilidade composta) foi executada. Tal ferramenta, que pode ser chamada de SMCIS-G devido à sua afinidade com o problema de geração, apesar de não apresentar prejuízo à performance e estimativa de índices, não apresentou ganho computacional.

3.4.4. Análise de Barra

Aumentando o escrutínio da análise da zona avaliada na última seção, nota-se que o risco de suprimento da Barra 301 pode e deve ganhar atenção especial do agente de transmissão local. A forte presença de geração eólica na barra, apesar de oferecer elevada redundância de geradores, traz bastante incerteza à disponibilidade energética, o que acarreta inflexibilidade operativa e justifica uma análise exclusiva para a barra. Os resultados de tal análise podem ser observados na Tabela 3.5.

Tabela 3.5: Análise da Barra 301 (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio
LOLP [%]	$6,08 \times 10^{-3}$ ($\beta=1,99\%$)	$5,81 \times 10^{-3}$ ($\beta=2,21\%$)	4,44%
EENS [MWh/ano]	6,45 ($\beta=2,68\%$)	6,23 ($\beta=2,53\%$)	3,41%
LOLF [oc./ano]	0,218 ($\beta=5,00\%$)	0,222 ($\beta=5,00\%$)	1,83%
Performance	SMCC	SMCIS-GT	Ganho
Sorteios	41.335.790	5.208.363	7,94 pu
FPO	9.382.787	1.408.646	6,66 pu
Tempo	9,50 h	1,55 h	6,13 pu

Certamente, a falha da Barra 301 é mais rara que a da zona que a engloba, sendo adotado, então, um amortecimento mais arrojado de $\alpha = 0,50$, em concordância com a condição de $LOLP < 0,01\%$. Dessa forma, não há surpresa pela obtenção de um ganho considerável nesse teste. Deve-se ressaltar, porém, que o ganho está muito aquém do valor aproximado para o ganho máximo de 16 vezes. Esse fato, na verdade, promove uma discussão interessante sobre como a ferramenta SMCIS-GT é impactada pelo grau de dispersão da falha. Note que, assim como o resto do sistema, as barras da zona que envolve a Barra 301 estão fortemente conectadas entre si e com as demais regiões da Área 3. Isso leva a falha dessas barras a ser composta por muitos modos de falha, de forma que muitos equipamentos, inclusive de regiões distantes, são importantes para a confiabilidade da zona. Em consequência, a representação da falha na amostragem por importância se dissolve em muitos estados de falha de composição distinta, reduzindo a eficiência da ferramenta, mas não sua eficácia, como comprova o bom ganho computacional apresentado pela análise da Barra 301. Para a ferramenta SMCIS-G, por outro lado, destaca-se que mais uma vez não houve prejuízo nem ganho.

Repetindo o procedimento de análise de desvio que considera a faixa de incerteza da estimativa de índices, para EENS, tem-se que $6,23 \times 1,0253 = 6,39$ desvia $0,93\%$ da estimativa em SMCC, o que é menor que $\beta = 2,68\%$. Para o indicador LOLP, porém, tem-se que $5,81 \times 10^{-3} \times 1,0221 = 5,94 \times 10^{-3}$ desvia $2,30\%$ da estimativa em SMCC, o que é maior que $\beta = 1,99\%$. Todavia, deve-se reconhecer que o erro é pequeno e que para valores de probabilidade tão ínfimos exatidão não é prioridade, sendo satisfatório conhecer o patamar da LOLP.

Tendo apresentado análises de sistema, área/zona e barra, é possível, então, fazer uma importante observação sobre ganho computacional. Nas análises da zona A3SUL e Barra 301, em que a falha é pelo menos pouco rara (i.e., $LOLP < 1\%$), o ganho entre a SMCC e a SMCIS-GT para o número de sorteios é maior que o de execuções de FPO, que, por sua vez, é maior que o de tempo de processamento. Essa relação é absolutamente coerente, pois a SMCIS-GT amostra com mais frequência na região do espaço de estados em que reside a falha, onde os estados são, em geral, mais críticos. Assim, a frequência com que um FPO é executado é maior na SMCIS-GT, o que reduz o ganho em FPOs e, por consequência, em tempo de processamento. Adicionalmente, estados mais críticos entregam um problema de otimização mais complicado para o *solver* de programação linear, o que eleva o

custo computacional de FPO individualmente e restringe o ganho do tempo de processamento não apenas em relação ao de número de sorteios, mas também em relação ao de execuções de FPO. É esse entendimento que explica por que o amortecimento α é tão importante para a eficácia do método proposto. O correto dimensionamento de α previne que a simulação com IS atue adentro da região da falha, o que elevaria a carga de FPOs e o custo computacional de suas soluções. Alternativamente, a simulação é conduzida a amostrar com mais frequência na região de fronteira sucesso-falha, onde o evento de interesse tem maior probabilidade de ocorrência e, por isso, é melhor representado.

3.4.5. Estudo da Dispersão da Falha

Para averiguar a questão de dispersão da falha, três estudos de caso são apresentados com as barras: 301, 307 e 320. A ideia é avaliar a falha de cada barra primeiro sob o contexto do sistema RTS-96-REN-T, tal como definido na Subseção 3.4.1, e depois sob o contexto de uma nova configuração do sistema, para a qual a falha apresenta característica de dispersão oposta. O objetivo é demonstrar que a dispersão da falha é, depois do nível de raridade, o fator mais determinante para o ganho computacional. Quanto maior a dispersão, menor o ganho, e vice-versa.

3.4.5.1. Caso 1 – Barra 301

Na Subseção 3.4.4, foi sugerido que a falha da Barra 301 é dispersa e por isso o ganho computacional de sua análise foi contido. Para provar isso, a análise da Barra 301 para o sistema RTS-96-REN-T é comparada nesta subseção à sua análise em uma nova configuração de sistema, resultado de duas alterações. Primeiro, altera-se a topologia, retirando a linha de transmissão que liga a Barra 301 com a 303, de modo a isolar a Barra 301 do parque gerador ao norte da Área 3 e demais áreas, reduzindo o espectro de componentes do sistema que exercem influência sobre a falha do barramento em questão. A segunda alteração visa impedir que a falha da barra deixe de ser rara e, portanto, eleva a capacidade de transmissão do sistema trocando o fator de 0,45 (aplicado ao RTS-96 para gerar RTS-96-REN-T) para 0,50. Assim, tem-se o sistema RTS-96-REN-T#1, que faz referência ao estudo de caso

número 1 e cuja análise de confiabilidade para a Barra 301 pode ser observada na Tabela 3.6 (análise para o RTS-96-REN-T é apresentada novamente para facilitar a comparação). Novamente, o desvio da estimativa da EENS superou a incerteza alcançada na SMCC, e, mais uma vez, o exercício com a faixa de incerteza da estimativa em SMCIS-GT pode ser realizado. O valor $4,57 \times 1,0210 = 4,67$ desvia 1,27% da estimativa em SMCC, um desvio menor que $\beta = 2,55\%$.

Os resultados da Tabela 3.6 evidenciam o conceito de dispersão da falha. Observe que, mesmo com uma confiabilidade maior, a análise da Barra 301 em RTS-96-REN-T#1 tem a própria SMCC mais rápida do que em RTS-96-REN-T, onde há mais dispersão na falha, pois as execuções de FPO reduziram, apesar da amostra ter aumentado. Isso indica que embora seja necessário amostrar mais estados de sucesso para cada estado de falha, a proporção daqueles que são claramente de sucesso e não precisam de otimização de medidas corretivas é maior, enquanto os estados de falha são mais relevantes individualmente. Ou seja, a estimativa de índices converge mais rápido, pois há uma carga menor de estados de falha e de sucesso crítico. Assim, executa-se menos FPO, o que se traduz em custo computacional menor. Comparando a performance da SMCIS-GT, já se percebe que não só o montante de FPO é menor em RTS-96-REN-T#1, mas a amostra também, pois, claramente, a menor dispersão da falha leva a uma amostragem por importância mais concisa. Logo, o ganho computacional é bem maior e se aproxima bastante do fator de 16 vezes estimado para o ganho máximo.

Tabela 3.6: Análises da Barra 301 (RTS-96-REN-T e RTS-96-REN-T#1).

	RTS-96-REN-T			RTS-96-REN-T#1		
Índices	SMCC	SMCIS-GT	Desvio	SMCC	SMCIS-GT	Desvio
LOLP [%]	$6,08 \times 10^{-3}$ ($\beta=1,99\%$)	$5,81 \times 10^{-3}$ ($\beta=2,21\%$)	4,44%	$4,88 \times 10^{-3}$ ($\beta=1,90\%$)	$4,93 \times 10^{-3}$ ($\beta=1,85\%$)	1,02%
EENS [MWh/ano]	6,45 ($\beta=2,68\%$)	6,23 ($\beta=2,53\%$)	3,41%	4,73 ($\beta=2,55\%$)	4,57 ($\beta=2,10\%$)	3,38%
LOLF [oc./ano]	0,218 ($\beta=5,00\%$)	0,222 ($\beta=5,00\%$)	1,83%	0,161 ($\beta=5,00\%$)	0,165 ($\beta=5,00\%$)	2,48%
Performance	SMCC	SMCIS-GT	Ganho	SMCC	SMCIS-GT	Ganho
Sorteios	41.335.790	5.208.363	7,94 pu	56.454.044	3.734.657	15,12 pu
FPO	9.382.787	1.408.646	6,66 pu	6.491.142	480.681	13,50 pu
Tempo	9,50 h	1,55 h	6,13 pu	6,55 h	0,60 h	10,92 pu

Buscando materializar a característica de dispersão da falha, que até aqui foi apenas contextualizada pelo seu efeito na performance do método proposto, a Tabela 3.7 traz uma comparação entre a lista ranqueada gerada pela análise da Barra 301 em RTS-96-REN-T e RTS-96-REN-T#1. Tal lista ranqueia os equipamentos do sistema com base no novo parâmetro indisponibilidade (indisponibilidade-IS), obtido na Fase 1 da SMCIS-GT (vide Subseção 3.3.3). Em outras palavras, os equipamentos são qualificados quanto à relevância da sua condição operativa indisponível para com a ocorrência da falha. Importante destacar que apenas os dez equipamentos com maior indisponibilidade-IS são exibidos, pois são os mais importantes e necessários a entender a característica da falha, e que as listas apresentam as novas indisponibilidades sem amortecimento, de modo a fornecer a informação de importância para a falha na sua forma original.

Tabela 3.7: Comparação de lista ranqueada para a Barra 301.

nº	Comp.	Barra	Barras Terminais		Indisponibilidade		
			De	Para	IS	Original	Distorção
RTS-96-REN-T							
1	G-39	321	—	—	0,5041	0,1200	320,11%
2	G-38	318	—	—	0,4959	0,1200	313,22%
3	G-29	301	—	—	0,2810	0,1000	180,99%
4	G-42	323	—	—	0,2810	0,0800	251,24%
5	G-37	316	—	—	0,1983	0,0400	395,87%
6	C-96	—	310	312	0,1983	0,0018	11.231,82%
7	G-10	118	—	—	0,1818	0,1200	51,52%
8	G-14	123	—	—	0,1818	0,0800	127,27%
9	G-24	218	—	—	0,1736	0,1200	44,63%
10	G-11	121	—	—	0,1570	0,1200	30,85%
RTS-96-REN-T#1							
1	C-95	—	310	311	0,3476	0,0018	19.756,59%
2	C-96	—	310	312	0,2500	0,0018	14.182,81%
3	G-29	301	—	—	0,2470	0,1000	146,95%
4	C-80	—	301	302	0,1890	0,0004	43.040,09%
5	C-88	—	305	310	0,1768	0,0004	45.477,22%
6	G-25	221	—	—	0,1585	0,1200	32,11%
7	G-10	118	—	—	0,1463	0,1200	21,95%
8	G-39	321	—	—	0,1402	0,1200	16,87%
9	G-38	318	—	—	0,1341	0,1200	11,79%
10	G-11	121	—	—	0,1159	0,1200	-3,46%

Da lista para o sistema RTS-96-REN-T, nota-se que dos seis primeiros componentes, um é a estação geradora térmica da própria Barra 301, quatro são algumas

das estações térmicas da Área 3 com maior capacidade e pouca redundância, fechando com um dos transformadores que ligam as regiões norte e sul da área, na sexta posição. Já os demais componentes do top 10 são estações de grande porte de outras áreas. Isso corrobora a suposição de que elementos de regiões distantes impactam a falha da Barra 301, além do parque gerador ao norte da Área 3. A lista para o RTS-96-REN-T#1, por outro lado, tem a estação da Barra 301 e quatro circuitos nas primeiras 5 posições. Componentes da região de 230 kV da Área 3 e demais áreas já não apresentam tanta relevância nesse caso. Note que o sexto componente dessa lista seria apenas o último colocado na do sistema RTS-96-REN-T, e por pouco. Sobre os 4 circuitos, sua importância pode ser constatada pelo fato de que a saída individual dos mesmos pressiona bastante a Barra 301, a qual é responsável por atender parte do consumo da barra 305 além da sua própria.

Com base na análise das listas da Tabela 3.7, fica claro que para o sistema RTS-96-REN-T#1 a falha da Barra 301 apresenta causas mais óbvias e configurações mais simples, pois depende de menos componentes. Ou seja, a falha é de fato menos dispersa, o que se reflete no desempenho computacional melhor. Por fim, destaca-se que toda a análise feita até este estudo de caso foi publicada em [114].

3.4.5.2.

Caso 2 – Barra 307

A Barra 307 é muito interessante, pois é a única dentro da Área 3 cuja falha é verdadeiramente não dispersa no RTS-96-REN-T. Aliás, a responsabilidade por sua falha recai quase que totalmente sobre a sua própria estação geradora. A Barra 307 se conecta ao resto do sistema por apenas um circuito, via barra 308, e apresenta capacidade de geração elevada e redundante (3 unidades de 100 MW) para uma carga nominal de 112,5 MW. Como a capacidade do circuito que conecta as barras 307 e 308 é de 78,75 MW, qualquer condição operativa do sistema em que ao menos duas unidades da Barra 307 estão disponíveis nunca provocará corte de carga na barra, pois ainda que o sistema requirite a máxima contribuição de 78,75 MW, sobrarão 121,25 MW para atender a no máximo 112,5 MW de sua própria carga. Ou seja, é necessário que duas unidades saiam de operação concomitantemente para que a carga possa vir a ser cortada. Adicionalmente, observa-se que a Barra 307 tem primariamente a função de suprir energia, pois ela está conectada apenas à uma barra consumidora sem geração, e dificilmente terá sua carga atendida pelo sistema,

pois sua conexão com o sistema é fraca. Logo, a saída simultânea de dois geradores da Barra 307 é não só condicional para a ocorrência da falha na barra, como muito provavelmente levará ao corte de carga, sendo pouco relevante o estado em que os demais componentes do sistema se encontram.

Para criar uma situação em que a falha da Barra 307 seja dispersa, aplicam-se três modificações ao RTS-96-REN-T. Primeiro, a redundância da geração da Barra 307 é aumentada para que a importância de seus geradores seja diluída, enquanto a sua capacidade é reduzida para evitar que a falha da barra fique ainda mais rara (estação geradora passa de 3×100 para 4×50 MW). Por outro lado, para elevar a participação de outros componentes do sistema na falha da barra, a conexão com o barramento 308 é duplicada. Em terceiro lugar, visando manter o nível de raridade da falha, a capacidade de fluxo dos circuitos do sistema é reduzida, trocando o fator de 0,45 (aplicado ao RTS-96 para gerar RTS-96-REN-T) para 0,40. Assim, tem-se o sistema RTS-96-REN-T#2, que faz referência ao estado de caso número 2.

Tabela 3.8: Comparação de lista ranqueada para a Barra 307.

nº	Comp.	Barra	Barras Terminais		Indisponibilidade		
			De	Para	IS	Original	Distorção
RTS-96-REN-T							
1	G-33	307	—	—	0,9677	0,0400	2319,25%
2	G-24	218	—	—	0,1434	0,1200	19,51%
3	G-10	118	—	—	0,1318	0,1200	9,82%
4	G-25	221	—	—	0,1240	0,1200	3,36%
5	G-38	318	—	—	0,1202	0,1200	0,13%
6	G-39	321	—	—	0,1202	0,1200	0,13%
7	G-11	121	—	—	0,1163	0,1200	-3,10%
8	G-17	202	—	—	0,1143	0,1000	14,34%
9	G-29	301	—	—	0,1143	0,1000	14,34%
10	G-31	302	—	—	0,1124	0,1000	12,40%
RTS-96-REN-T#2							
1	G-33	307	—	—	0,4775	0,0400	1093,75%
2	G-38	318	—	—	0,2600	0,1200	116,67%
3	G-39	321	—	—	0,2400	0,1200	100,00%
4	G-11	121	—	—	0,2300	0,1200	91,67%
5	G-10	118	—	—	0,2100	0,1200	75,00%
6	G-24	218	—	—	0,2100	0,1200	75,00%
7	G-29	301	—	—	0,1750	0,1000	75,00%
8	G-25	221	—	—	0,1500	0,1200	25,00%
9	G-3	102	—	—	0,1300	0,1000	30,00%
10	G-15	201	—	—	0,1250	0,1000	25,00%

Da Tabela 3.8, todas as conjecturas feitas são confirmadas. No caso do sistema RTS-96-REN-T, nota-se que de fato a estação geradora da Barra 307 concentra quase toda a importância para a falha (indisponibilidade-IS é quase 1 pu). As distorções dos demais componentes listados são pouco relevantes, pois são residuais em comparação à estação G-33 e indicam apenas que nas falhas amostradas na Fase 1 da SMCIS-GT certos componentes estavam ocasionalmente fora de operação. Já para o RTS-96-REN-T#2, tem-se não apenas a G-33 com importância reduzida, mas inúmeros componentes com indisponibilidade-IS maior que 0,20 pu. Tais componentes, por sua vez, são justamente cinco das seis estações de maior capacidade (1×400 MW) e que mais falham (indisponibilidade original de 0,12 pu) de todo o sistema, o que é absolutamente coerente.

A Tabela 3.9 confirma que a falha da Barra 307 é mais dispersa no sistema RTS-96-REN-T#2, onde o ganho é bem menor. Mais uma vez, o exercício com a faixa de incerteza da estimativa em SMCIS-GT pode ser realizado para confirmar que todas as estimativas são satisfatórias.

Tabela 3.9: Análises da Barra 307 (RTS-96-REN-T e RTS-96-REN-T#2).

Índices	RTS-96-REN-T			RTS-96-REN-T#2		
	SMCC	SMCIS-GT	Desvio	SMCC	SMCIS-GT	Desvio
LOLP [%]	$2,23 \times 10^{-3}$ ($\beta=1,59\%$)	$2,26 \times 10^{-3}$ ($\beta=0,97\%$)	1,35%	$2,36 \times 10^{-3}$ ($\beta=2,86\%$)	$2,37 \times 10^{-3}$ ($\beta=2,87\%$)	0,42%
EENS [MWh/ano]	2,09 ($\beta=1,98\%$)	2,10 ($\beta=0,91\%$)	0,48%	1,97 ($\beta=4,81\%$)	1,96 ($\beta=4,14\%$)	0,51%
LOLF [oc./ano]	0,0513 ($\beta=5,00\%$)	0,0528 ($\beta=5,00\%$)	2,92%	0,176 ($\beta=5,00\%$)	0,179 ($\beta=5,00\%$)	1,70%
Performance	SMCC	SMCIS-GT	Ganho	SMCC	SMCIS-GT	Ganho
Sorteios	177.327.622	12.586.877	14,09 pu	51.657.746	6.400.485	8,07 pu
FPO	40.255.896	2.980.605	13,51 pu	18.155.049	2.466.419	7,42 pu
Tempo	40,40 h	3,30 h	12,24 pu	22,65 h	3,08 h	7,35 pu

3.4.5.3.

Caso 3 – Barra 320

A Barra 320, diferentemente da 301 e 307, está localizada na região de 230 kV da Área 3, onde o parque gerador é mais denso e a rede é ainda mais interligada. A sua falha é, então, ainda mais rara. No entanto, apesar de toda disponibilidade energética e flexibilidade operativa da sua região, a dispersão da falha não é tão

grande. Devido à presença de geradores de grande porte e elevada indisponibilidade, boa parte da falha é explicada pela saída de tais componentes.

Visando provar a consistência da relação entre dispersão da falha e desempenho computacional do método proposto (ou qualquer ferramenta baseada em amostragem por importância), um simples exemplo de falha pouco dispersa para a Barra 320 pode ser confeccionado, retirando o circuito entre as barras 316 e 319 do sistema RTS-96-REN-T. Assim, as barras 319 e 320 se tornam radiais e passam a depender bastante da conexão entre os barramentos 320 e 323. Para manter o nível de raridade da falha, a capacidade de fluxo dos circuitos do sistema é elevada, trocando o fator de 0,45 (aplicado ao RTS-96 para gerar RTS-96-REN-T) para 0,50. Assim, tem-se o sistema RTS-96-REN-T#3, referente ao estudo de caso número 3.

Tabela 3.10: Comparação de lista ranqueada para a Barra 320.

nº	Comp.	Barra	Barras Terminais		Indisponibilidade		
			De	Para	IS	Original	Distorção
RTS-96-REN-T							
1	G-39	321	—	—	0,9949	0,1200	729,12%
2	G-38	318	—	—	0,9848	0,1200	720,71%
3	G-37	316	—	—	0,5808	0,0400	1352,02%
4	G-25	221	—	—	0,1414	0,1200	17,85%
5	G-24	218	—	—	0,1364	0,1200	13,64%
6	G-31	302	—	—	0,1338	0,1000	33,84%
7	G-17	202	—	—	0,1035	0,1000	3,54%
8	G-15	201	—	—	0,1010	0,1000	1,01%
9	G-29	301	—	—	0,1010	0,1000	1,01%
10	G-1	101	—	—	0,0985	0,1000	-1,52%
RTS-96-REN-T#3							
1	C-115	—	320	323	0,5120	0,0004	>100.000%
2	C-114	—	320	323	0,4800	0,0004	>100.000%
3	G-10	118	—	—	0,1680	0,1200	40,00%
4	G-25	221	—	—	0,1440	0,1200	20,00%
5	G-17	202	—	—	0,1200	0,1000	20,00%
6	G-3	102	—	—	0,1120	0,1000	12,00%
7	G-24	218	—	—	0,1040	0,1200	-13,33%
8	G-31	302	—	—	0,1040	0,1000	4,00%
9	G-11	121	—	—	0,0960	0,1200	-20,00%
10	G-38	318	—	—	0,0960	0,1200	-20,00%

Da Tabela 3.10, todas as conjecturas feitas são confirmadas. Enquanto no sistema RTS-96-REN-T nota-se uma certa concentração de importância para as principais estações geradoras da Área 3, no RTS-96-REN-T#3 tem-se que os dois

circuitos que conectam a Barra 320 à 323 monopolizam totalmente a responsabilidade pela falha da Barra 320. Repare que a carga combinada das barras 319 e 320 é de 278,1 MW e que cada circuito entre 320 e 323 tem capacidade de fluxo de 250 MW, de modo que a saída individual dos circuitos C-114 e C-115 já provoca falha. Como a indisponibilidade simultânea de tais circuitos é um evento raríssimo, a falha da Barra 320 basicamente ocorre metade das vezes com apenas C-114 fora de operação e a outra metade com C-115 fora de operação. Isso justifica a indisponibilidade-IS desses circuitos estar em torno de 0,50 pu. Importante ressaltar que, no sistema RTS-96-REN-T, a indisponibilidade-IS estar próxima de 1 pu para as estações G-39 e G-38 não significa que a saída de seus geradores é fatal para o suprimento da Barra 320. Observe que, justamente por ser uma falha mais dispersa, suas causas não são tão óbvias e seus estados são compostos por múltiplas contingências, ainda que apenas um ou outro componente tenha participação muito mais frequente que os demais.

A Tabela 3.11 confirma que a falha da Barra 320 é menos dispersa no sistema RTS-96-REN-T#3, apresentando ganho bem maior. E, mais uma vez, o exercício com a faixa de incerteza da estimativa em SMCIS-GT pode ser realizado para confirmar que todas as estimativas são satisfatórias.

Tabela 3.11: Análises da Barra 320 (RTS-96-REN-T e RTS-96-REN-T#3).

	RTS-96-REN-T			RTS-96-REN-T#3		
Índices	SMCC	SMCIS-GT	Desvio	SMCC	SMCIS-GT	Desvio
LOLP [%]	$1,21 \times 10^{-3}$ ($\beta=1,92\%$)	$1,19 \times 10^{-3}$ ($\beta=1,77\%$)	1,65%	$1,31 \times 10^{-3}$ ($\beta=1,82\%$)	$1,30 \times 10^{-3}$ ($\beta=1,84\%$)	0,76%
EENS [MWh/ano]	2,49 ($\beta=2,43\%$)	2,42 ($\beta=1,98\%$)	2,81%	1,09 ($\beta=2,70\%$)	1,07 ($\beta=2,12\%$)	1,83%
LOLF [oc./ano]	0,0406 ($\beta=5,00\%$)	0,0417 ($\beta=5,00\%$)	2,71%	0,0401 ($\beta=5,00\%$)	0,0411 ($\beta=5,00\%$)	2,49%
Performance	SMCC	SMCIS-GT	Ganho	SMCC	SMCIS-GT	Ganho
Sorteios	224.114.025	23.703.157	9,46 pu	230.566.561	14.313.029	16,11 pu
FPO	50.876.036	5.539.983	9,18 pu	33.806.552	2.122.125	15,93 pu
Tempo	53,25 h	6,08 h	8,76 pu	27,65 h	2,40 h	11,52 pu

Do ganho observado no RTS-96-REN-T#3, tem-se um valor ligeiramente maior que o ganho máximo. Sobre isso, é válido destacar que é bem possível o ganho de algum caso superar o ganho máximo por uma margem ainda maior, da mesma forma que um ganho pode ficar aquém do valor máximo mesmo quando a

Fase 2 da SMCIS-GT é desprezível. Afinal, o ganho máximo é um valor aproximado (vide Subseção 3.3.3). Seu valor real depende dos valores de variância amostral e média amostral da função em estimação observados na conclusão da Fase 1 do método (vide eq. (3.30)). Em todas as análises apresentadas neste capítulo, o ganho entre a amostra da SMCC e a Fase 1 da SMCIS-GT está em torno de 16 vezes, desviando por cima em alguns casos. Especialmente no caso da Barra 320 para o sistema RTS-96-REN-T#3, o ganho da Fase 1 já era superior a 16 e permaneceu acima no desempenho total, porque a Fase 2 foi desprezível com apenas 559.486 sorteios, 3,9% do total gasto pela SMCIS-GT.

3.5. Conclusões

Este capítulo apresentou a primeira parte do método proposto em detalhe. Antes, porém, a metodologia genérica da ferramenta de entropia cruzada também foi detalhada, seguida de um resumo da aplicação existente mais bem sucedida em confiabilidade de geração e composta, de modo a apontar com clareza a deficiência da aplicação em confiabilidade composta.

A Subseção 3.3.3 confirmou as contribuições do método proposto listadas no início da Seção 3.3, exceto pelo item: flexibilidade no foco de análise, seja ele para sistema/área/barra. Sobre esse item, tem-se que a metodologia apresentada com foco na falha de suprimento global do sistema é facilmente estendida para quando o foco é um conjunto reduzido de barras, ou até mesmo uma barra em específico. Isso foi implementado, mas não foi descrito no algoritmo do método, pois é um aspecto de programação. Já no contexto dos trabalhos existentes, destaca-se que a análise localizada não é facilmente estendida, pois, ainda que a geração disponível fosse explicativa da falha de sistemas GT, não há forma técnica, genérica e automática de escolher as estações geradoras que devem ser consideradas na função performance de um certo grupo de barras consumidoras.

Na Seção 3.4, a metodologia proposta foi avaliada extensivamente. Uma versão modificada do sistema IEEE Reliability Test System – 1996 foi confeccionada de modo a obter um sistema tal como os sistemas interligados modernos, que apresentam participação renovável considerável e falha composta dominada por problemas de insuficiência de recursos de transmissão. A partir dessa versão, chamada de RTS-96-REN-T, três testes foram apresentados com sistema, zona e barra, para

demonstrar o comportamento da ferramenta proposta, a SMCIS-GT, nas típicas situações de falha não rara (sistema), pouco rara (área/zona) e bastante rara (barra). Depois, três estudos de caso, com as barras 301, 307 e 320, foram apresentados com novas configurações do sistema RTS-96-REN-T para avaliar e discutir o aspecto de dispersão da falha e seu impacto no desempenho da SMCIS-GT.

Foi demonstrado que a dispersão da falha é, depois do nível de raridade, o aspecto mais determinante na eficiência da ferramenta proposta. Isso não quer dizer, porém, que a dispersão pode anular o ganho computacional, tal como ocorre com o nível de raridade (falha não rara não possibilita ganho). Conquanto a falha for rara, haverá ganho. Para além da imprescindível discussão sobre dispersão da falha, o importante é destacar como a ferramenta proposta se prova eficaz em praticamente qualquer cenário, tornando a avaliação da confiabilidade composta exequível, seja a falha rara ou não e dispersa ou não, e seja o foco de análise localizado ou não. Ademais, essa eficácia da ferramenta é observada considerando a participação de renováveis, inclusive quando fontes que são intermitentes (como a eólica) se localizam próximas a consumidores com presença reduzida de geração com inércia, o que diminui a flexibilidade operativa e aumenta o risco de suprimento.

4. Confiabilidade Composta Otimizada por Aprendizado de Máquina

4.1. Introdução

Conforme discutido na Subseção 1.2.3, o uso de classificadores de estados pode ser acoplado à uma simulação otimizada por técnicas de redução de variância (VRTs – *variance reduction techniques*) para extrair mais do classificador. Todavia, apenas utilizar as ferramentas em conjunto não trará os benefícios desejados, sobretudo na avaliação de confiabilidade composta de sistemas modernos interligados para os quais o foco é a falha localizada e, portanto, um evento raro. Neste capítulo, é apresentada uma metodologia em que o próprio treinamento de um classificador é realizado a partir de uma simulação acelerada por VRT. Adicionalmente, a inteligência produzida pela VRT é empregada no desenvolvimento de variáveis explicativas de qualidade para o classificador. A VRT utilizada é a amostragem por importância (IS – *importance sampling*) via minimização da entropia cruzada (CE – *cross-entropy*), implementada tal como descrito no Capítulo 3, enquanto o classificador é uma ferramenta de aprendizado de máquina chamada regressão logística binária (BLR – *binary logistic regression*) [115].

Na próxima seção, a ferramenta BLR original é descrita, enquanto a seção seguinte traz a proposta de acoplamento da ferramenta ao método já estabelecido no Capítulo 3. Na quarta seção, os estudos de caso apresentados no capítulo anterior são revisitados, desta vez, traçando comparações com os resultados obtidos com a metodologia completa. A quinta seção se faz necessária para relatar tentativas realizadas ao longo da pesquisa desta tese de suplantando o BLR com uma rede neural de maior complexidade (e.g., redes multicamadas). Por fim, a sexta e última seção traz as conclusões do capítulo.

4.2. Regressão Logística Binária

Em estatística, um modelo logístico trata o *log-odds* de um evento como uma função linear de um grupo de variáveis independentes. A medida *log-odds* é o logaritmo da chance de ocorrência de um evento. A regressão logística, por sua vez, tem o papel de estimar os coeficientes da função linear do modelo logístico, o qual pode ser projetado para obter o *log-odds* de diferentes eventos (i.e., o modelo apresenta mais de uma variável dependente). No caso da regressão logística binária, tem-se interesse em apenas um evento e, portanto, há somente uma variável dependente no modelo. Tal variável corresponde à probabilidade de ocorrência do evento (diferente de chance de ocorrência) com saída variando de 0 pu (certeza de não ocorrência) a 1 pu (certeza de ocorrência). Já as variáveis independentes podem ser binárias ou contínuas.

Importante observar que o modelo logístico é baseado em *log-odds* de um evento, mas a regressão logística trabalha com probabilidade. Existe uma função, a função logística, que é responsável por converter *log-odds* em probabilidade. Adicionalmente, observa-se que a ferramenta ganhou bastante destaque em *machine learning* na forma de um classificador de aprendizado supervisionado. Em verdade, a ferramenta BLR pode ser vista como um *perceptron* (i.e., a rede neural artificial mais primitiva, de apenas um neurônio), mas de muita capacidade de classificação. A seguir, um desenvolvimento matemático da função logística [115], que é o objeto de regressão, é apresentado em detalhe.

4.2.1. Função Logística

Suponha que se quer estimar a probabilidade de ocorrência de um determinado evento de interesse $\{Y = A\}$ com base na realização de um grupo de m variáveis aleatórias independentes $\mathbf{X} = (X_1, \dots, X_m)$, tidas como variáveis explicativas do evento. Uma maneira simples de alcançar esse objetivo é por meio de regressão linear para os coeficientes $\mathbf{\Gamma} = (\Gamma_0, \Gamma_1, \dots, \Gamma_m)$ da função:

$$\mathbb{P}(Y = A) = \Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m \quad (4.1)$$

Contudo, sabe-se que a maioria dos problemas são não-lineares, e, portanto, uma curva não linear é mais indicada. A função exponencial é um exemplo de uma

não linearidade simples e de muita aplicabilidade, com uma curva capaz de promover bastante separação entre pontos com valor de probabilidade muito parecido. Assim, a probabilidade de $\{Y = A\}$ pode ser dada por:

$$\mathbb{P}(Y = A) = e^{(\Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m)} \quad (4.2)$$

Caso o expoente da eq. (4.2) tenda para $-\infty$ (i.e., as variáveis explicativas apresentam máxima influência para a não ocorrência de $\{Y = A\}$), a probabilidade tenderá corretamente para 0 pu, mas se o expoente tender para $+\infty$ (i.e., as variáveis explicativas apresentam máxima influência para a ocorrência de $\{Y = A\}$), a probabilidade também tenderá a $+\infty$, ao invés de 1 pu. Se no lugar de probabilidade for utilizada a grandeza chance de ocorrência de $\{Y = A\}$, a eq. (4.2) funciona tanto para o limite de $-\infty$ quanto de $+\infty$. Isso ocorre, pois diferentemente de probabilidade, que é dada pela razão entre o número de ocorrências do evento de interesse e o número total de possibilidades, a chance é dada pela razão entre o número de ocorrências e o número de não ocorrências possíveis do evento de interesse (que são números complementares e somam o total de possibilidades do espaço de estados). Logo, quando a probabilidade de um evento é igual a 0 pu, a chance também é igual zero, e quando a probabilidade é igual a 1 pu, a chance é igual a $+\infty$, que é o resultado do exponencial da eq. (4.2) quando o seu expoente tende para $+\infty$. Para esclarecer ainda mais essa relação, veja que a chance do evento $\{Y = A\}$ pode ser dada em função da sua probabilidade:

$$\mathbb{O}(Y = A) = \frac{\mathbb{P}(Y = A)}{1 - \mathbb{P}(Y = A)} \quad (4.3)$$

A eq. (4.3) é importante, pois se a chance $\mathbb{O}(Y = A)$ é substituída pelo exponencial da eq. (4.2), a probabilidade $\mathbb{P}(Y = A)$ pode ser obtida como uma função exponencial que é consistente para todos os valores possíveis das variáveis explicativas e tem imagem no intervalo $[0,1]$. Observe:

$$e^{(\Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m)} = \frac{\mathbb{P}(Y = A)}{1 - \mathbb{P}(Y = A)}$$

$$\mathbb{P}(Y = A) = \frac{e^{(\Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m)}}{1 + e^{(\Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m)}}$$

$$\mathbb{P}(Y = A) = \frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m)}} \quad (4.4)$$

A eq. (4.4) é chamada de função logística, responsável por converter *log-odds* em probabilidade. Para enxergar isso, basta observar que *log-odds*, dado pelo logaritmo neperiano de $\mathbb{O}(Y = A)$, é igual a $(\Gamma_0 + \Gamma_1 X_1 + \dots + \Gamma_m X_m)$, uma função linear das variáveis explicativas, tal como descrito no primeiro parágrafo desta seção. A função logística pode ser usada, então, como um bom classificador de estados para o evento $\{Y = A\}$. Para tanto, cria-se um limiar de classificação $\mathbb{P}^{thr}$ para a probabilidade $\mathbb{P}(Y = A)$, em que $\mathbb{P}(Y = A) \geq \mathbb{P}^{thr}$ indica a ocorrência de $\{Y = A\}$ e $\mathbb{P}(Y = A) < \mathbb{P}^{thr}$ indica a não ocorrência de $\{Y = A\}$. A eficiência deste classificador vai depender da qualidade das variáveis explicativas e da representatividade da amostra coletada para a regressão da função logística. A seguir, a metodologia de ajuste do classificador é apresentada.

4.2.2.

Ajuste do Classificador

A estimação de parâmetros de distribuições de probabilidade por meio de um conjunto de dados observado é comumente realizada via estimador de máxima verossimilhança (MLE – *maximum likelihood estimation*) [116]. A função logística, dada pela eq. (4.4), descreve uma distribuição exponencial, e, portanto, os coeficientes da BLR podem ser obtidos via MLE.

Considere as seguintes definições. A função distribuição de probabilidade (pdf – *probability density function*) mede, em meio a um espaço de estados definido, o quão provável é a realização de um conjunto de variáveis aleatórias $\mathbf{X}$. O vetor de parâmetros $\mathbf{v}$ da pdf é o que distingue o fenômeno descrito. Altere o vetor de parâmetros e outro fenômeno será descrito. Já a função de verossimilhança (LF – *likelihood function*) mede o quão provável é a realização de um conjunto de variáveis aleatórias dada a escolha de um vetor $\mathbf{v}$. Ou seja, faz-se uso da mesma função que a pdf, mas inverte-se o caráter de variável de $\mathbf{X}$ com o de parâmetro de $\mathbf{v}$. Seja a pdf dada por $f(\mathbf{X}; \mathbf{v})$, a LF é dada por $f(\mathbf{v}; \mathbf{X})$ ou $\mathcal{L}(\mathbf{v}; \mathbf{X})$.

No problema da regressão BLR quer-se determinar os parâmetros $(\Gamma_0, \Gamma_1, \dots, \Gamma_m)$ da eq. (4.4) de modo que a probabilidade de toda realização das variáveis explicativas $(X_1, \dots, X_m)$ que fazem par com $(Y = A)$ seja igual a um,

enquanto a probabilidade de toda realização das variáveis que fazem par com $(Y \neq A)$ seja igual a zero. Evidentemente, o fenômeno em questão não será descrito perfeitamente pela eq. (4.4), e, portanto, ao final da regressão, os pares $\mathbf{X}$ e $(Y = A)$ apresentarão apenas probabilidade mais próxima de 1 pu, enquanto os pares $\mathbf{X}$ e $(Y \neq A)$ apresentarão probabilidade mais próxima de 0 pu. Como a função LF dá a probabilidade da realização de $\mathbf{X}$ em função de $\mathbf{v}$, que na BLR é o vetor $\mathbf{\Gamma}$, pode-se obter $\mathbf{\Gamma}$ maximizando LF quando $(Y = A)$ e minimizando LF quando $(Y \neq A)$. Para alcançar esse objetivo, porém, é necessário antes desenvolver um pouco mais o modelo logístico apresentado nessa seção.

A função logística calcula $\mathbb{P}(Y = A)$ como uma distribuição exponencial, mas o modelo da regressão logística binária é essencialmente uma distribuição Bernoulli, sendo $p = \mathbb{P}(Y = A) = \mathbb{P}(Y = 1)$ e $1 - p = \mathbb{P}(Y \neq A) = \mathbb{P}(Y = 0)$. Assim, tem-se pdf e LF tal como as equações (4.5) e (4.6). Note que para uma distribuição Bernoulli convencional há apenas uma variável aleatória (i.e., Y) e um parâmetro (i.e., p). Já no modelo da regressão BLR, há um vetor de variáveis aleatórias dado pelo par $(\mathbf{X}, Y)$ e um vetor de parâmetros dado por $\mathbf{\Gamma}$. O parâmetro p da distribuição Bernoulli é, na verdade, uma função de $(\mathbf{X}, Y)$.

$$f((\mathbf{X}, Y); \mathbf{\Gamma}) = p^Y (1 - p)^{1-Y} \quad (4.5)$$

$$\mathcal{L}(\mathbf{\Gamma}; (\mathbf{X}, Y)) = p^Y (1 - p)^{1-Y} \quad (4.6)$$

A eq. (4.6) é importante, pois ao maximizá-la em $\mathbf{\Gamma}$ para todas as realizações do par $(\mathbf{X}, Y)$ é, então, possível alcançar o objetivo de maximizar LF quando $(Y = A)$ (i.e., $(Y = 1)$) e minimizar LF quando $(Y \neq A)$ (i.e., $(Y = 0)$). Isso ocorre, pois quando $(Y = 1)$, tem-se a maximização de p , e quando $(Y = 0)$, tem-se a maximização de $(1 - p)$. Todavia, a diferenciação da eq. (4.6) em $\mathbf{\Gamma}$ é muito complicada. Para resolver isso, é aplicado o logaritmo à função de verossimilhança, dado que a maximização do logaritmo de uma função tem efeito equivalente a maximizar a própria função. Assim, tem-se a função log-verossimilhança:

$$\ell(\mathbf{\Gamma}; (\mathbf{X}, Y)) = \log(\mathcal{L}(\mathbf{\Gamma}; (\mathbf{X}, Y)))$$

$$\ell(\mathbf{\Gamma}; (\mathbf{X}, Y)) = \log(p^Y (1 - p)^{1-Y})$$

$$\ell(\mathbf{\Gamma}; (\mathbf{X}, Y)) = \log(p^Y) + \log((1 - p)^{1-Y})$$

$$\ell(\Gamma; (\mathbf{X}, Y)) = Y \log(p) + (1 - Y) \log(1 - p) \quad (4.7)$$

A seguir, é apresentado o desenvolvimento matemático da maximização de $\ell(\Gamma; (\mathbf{X}, Y))$ para obter os valores do vetor Γ .

4.2.2.1. Máxima log-verossimilhança

A eq. (4.7) fornece a função log-verossimilhança, que é dada para um par $(\mathbf{X}, Y)$. Na regressão BLR, tem-se um conjunto de pares observados, de modo que os coeficientes da função logística precisam ser dimensionados para maximizar $\ell(\Gamma; (\mathbf{X}, Y))$ para todos os pares amostrados. Logo, o problema de otimização a ser resolvido é dado por:

$$\max_{\Gamma} L = \max_{\Gamma} \sum_{k \in [1, NS]} \ell(\Gamma; (\mathbf{X}_k, Y_k)) \quad (4.8)$$

Substituindo $\ell(\Gamma; (\mathbf{X}, Y))$ pela eq. (4.7):

$$\max_{\Gamma} L = \max_{\Gamma} \sum_{k \in [1, NS]} [Y_k \log(p_k) + (1 - Y_k) \log(1 - p_k)] \quad (4.9)$$

E como $p = \mathbb{P}(Y = A)$, da eq. (4.4) tem-se que:

$$\begin{aligned} \max_{\Gamma} L = \max_{\Gamma} \sum_{k \in [1, NS]} & \left[Y_k \log \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right. \\ & \left. + (1 - Y_k) \log \left(\frac{1}{1 + e^{+(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right] \end{aligned} \quad (4.10)$$

Cada coeficiente Γ_i , $i \in [0, m]$ pode, então, ser obtido igualando a sua derivada parcial de L a zero:

$$\begin{aligned} \frac{dL}{d\Gamma_i} &= \frac{d \left(\sum_{k=1}^{NS} \left[Y_k \log \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) + (1 - Y_k) \log \left(\frac{1}{1 + e^{+(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right] \right)}{d\Gamma_i} \\ \frac{dL}{d\Gamma_i} &= \sum_{k=1}^{NS} \left[(Y_k) \frac{d \left(\log \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right)}{d\Gamma_i} + (1 - Y_k) \frac{d \left(\log \left(\frac{1}{1 + e^{+(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right)}{d\Gamma_i} \right] \end{aligned}$$

Chamando $a = \log \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right)$ e $b = \log \left(\frac{1}{1 + e^{+(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right)$, tem-se que:

se que:

$$\frac{dL}{d\Gamma_i} = \sum_{k=1}^{NS} \left[(Y_k) \frac{da}{d\Gamma_i} + (1 - Y_k) \frac{db}{d\Gamma_i} \right]$$

De modo que se pode resolver $\frac{da}{d\Gamma_i}$ e $\frac{db}{d\Gamma_i}$ separadamente, tal como:

$$\begin{aligned} \frac{da}{d\Gamma_i} &= \frac{d \left(\log \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right)}{d\Gamma_i} \\ \frac{da}{d\Gamma_i} &= \frac{d \left(\log \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right) \right)}{d \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right)} \times \frac{d \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right)}{d\Gamma_i} \\ \frac{da}{d\Gamma_i} &= (1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}) \times \frac{d \left(\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right)}{d(1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})})} \\ &\quad \times \frac{d(1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})})}{d\Gamma_i} \\ \frac{da}{d\Gamma_i} &= (1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}) \times \frac{(-1)}{(1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})})^2} \\ &\quad \times \frac{d(1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})})}{d(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})} \times \frac{d(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}{d\Gamma_i} \\ \frac{da}{d\Gamma_i} &= \frac{-1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \times (-1) \times e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})} \times \begin{cases} 1, & \text{se } i = 0 \\ X_{k,i}, & \text{se } i > 0 \end{cases} \\ \frac{da}{d\Gamma_i} &= \begin{cases} \frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}}, & \text{se } i = 0 \\ \frac{X_{k,i}}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}}, & \text{se } i > 0 \end{cases} \end{aligned}$$

E, analogamente:

$$\frac{db}{d\Gamma_i} = \begin{cases} -\frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}}, & \text{se } i = 0 \\ -\frac{X_{k,i}}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}}, & \text{se } i > 0 \end{cases}$$

Logo:

$$\frac{dL}{d\Gamma_i} = \begin{cases} \sum_{k=1}^{NS} \left[(Y_k) \frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} - (1 - Y_k) \frac{1}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right], & \text{se } i = 0 \\ \sum_{k=1}^{NS} \left[(Y_k) \frac{X_{k,i}}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} - (1 - Y_k) \frac{X_{k,i}}{1 + e^{-(\Gamma_0 + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right], & \text{se } i > 0 \end{cases}$$

Mas, se a variável $X_{k,0} = 1$ for introduzida, tem-se que:

$$\frac{dL}{d\Gamma_i} = \sum_{k=1}^{NS} X_{k,i} \times \left[\frac{Y_k}{1 + e^{+(\Gamma_0 X_{k,0} + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} - \frac{1 - Y_k}{1 + e^{-(\Gamma_0 X_{k,0} + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right] \quad (4.11)$$

E com a eq. (4.11), é, enfim, possível obter o coeficiente Γ_i ótimo fazendo:

$$\frac{dL}{d\Gamma_i} = \sum_{k=1}^{NS} X_{k,i} \times \left[\frac{Y_k}{1 + e^{+(\Gamma_0 X_{k,0} + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} - \frac{1 - Y_k}{1 + e^{-(\Gamma_0 X_{k,0} + \Gamma_1 X_{k,1} + \dots + \Gamma_m X_{k,m})}} \right] = 0 \quad (4.12)$$

A expressão da eq. (4.12), porém, não é resolvida analiticamente, e, assim, recorre-se a métodos numéricos. A seguir o algoritmo de ajuste do classificador BLR é disponibilizado.

4.2.3. Algoritmo

Adotando o método numérico de otimização de gradiente crescente (pois se está maximizando L), os coeficientes ótimos da eq. (4.10) podem ser obtidos por meio de um mecanismo iterativo simples:

Passo 1) Estipular valores iniciais para os coeficientes, e.g.: $\Gamma_i(j = 0) = 1$, onde j é o contador de iterações;

Passo 2) Fazer $j = j + 1$;

Passo 3) Atualizar os coeficientes: $\Gamma_i(j) = \Gamma_i(j - 1) + \alpha^{LR} \times \frac{dL}{d\Gamma_i}$, onde α^{LR} é o parâmetro taxa de aprendizagem (do inglês: *learning rate*);

Passo 4) Recalcular L ;

Passo 5) Se $|L| \leq tol$ pare; senão, retorne ao Passo 2).

Importante destacar, que na maioria dos casos a tolerância do Passo 5) nunca será alcançada, por limitações da própria função logística em modelar adequadamente o fenômeno em estudo e, principalmente, por limitações impostas pela qualidade das variáveis explicativas utilizadas. Esse problema de convergência do método é resolvido com a busca pela mínima diferença entre a função L calculada na iteração atual e a anterior. Assim, ter-se-á o melhor ajuste possível para as variáveis explicativas escolhidas, ainda que o classificador resultante seja de qualidade insatisfatória. Com essa consideração, tem-se um novo algoritmo:

Passo 1) Estipular valores iniciais para os coeficientes, e.g.: $\Gamma_i(j = 0) = 1$, e calcular $L(j = 0)$, onde j é o contador de iterações;

Passo 2) Fazer $j = j + 1$;

Passo 3) Atualizar os coeficientes: $\Gamma_i(j) = \Gamma_i(j - 1) + \alpha^{LR} \times \frac{dL}{d\Gamma_i}$, onde α^{LR} é o parâmetro taxa de aprendizagem;

Passo 4) Calcular $L(j)$;

Passo 5) Se $|L(j) - L(j - 1)| \leq tol$ pare; senão, retorne ao Passo 2).

4.3. Proposta

A proposta de uma nova ferramenta de simulação Monte Carlo (SMC) otimizada por redução de variância e aprendizado de máquina que acopla um classificador a uma VRT, ao invés de apenas utilizá-los em conjunto, é, enfim, formalizada nessa seção. Para tanto, estabelece-se a seguinte descrição do método completo:

Fase 1: Simulação Monte Carlo convencional para estimação da distribuição amostragem por importância, ou pdf-IS (pdf - *importance sampling*), com parada em β^{F1} (sendo um valor para a estimativa do índice de probabilidade e outro para a do índice de frequência).

Fase 2: Simulação Monte Carlo com IS para estimação de índices, com parada em β^{F2} . Esta fase é subdividida em duas partes:

- *Fase 2.1:* A simulação executa até β^{F1} , para que o classificador BLR possa ser ajustado;
- *Fase 2.2:* A simulação retoma sua execução com o auxílio do classificador (estados classificados como de sucesso não têm sua adequação avaliada) até que a incerteza da estimativa de todos os índices alcance β^{F2} .

Note que, o que foi estabelecido como Fase 1 do método proposto na Subseção 3.3.3 assim permanece no método completo, enquanto a Fase 2 é reestruturada em duas partes. O critério de parada para a Fase 2.1 é o mesmo da Fase 1, de modo que a amostra utilizada para o treinamento do classificador apresentará boa representatividade. Relembrando, a estimativa do índice LOLP (*loss of load probability*)

deve alcançar a incerteza de $\beta_{LOLP}^{F1} = 10\%$, enquanto a do índice LOLF (*loss of load frequency*), que converge por último, deve alcançar $\beta_{LOLF}^{F1} = 20\%$.

No caso de um classificador perfeito, a Fase 2.2 apresentará custo computacional desprezível e a Fase 2 será por volta de 16 vezes mais rápida do que a Fase 2 do método apresentado na Subseção 3.3.3. Observa-se que o número 16 é o valor de ganho máximo estimado entre uma amostra obtida a uma incerteza de $\beta^{F2} = 5\%$ e outra com $\beta_{LOLF}^{F1} = 20\%$. Todavia, caso a otimização CE seja perfeita, o conjunto total será por volta de 16 vezes mais rápido que uma simulação convencional, independente do que ocorre com o classificador. Contudo, nem a otimização CE nem o classificador será perfeito, de modo que o classificador terá um potencial de impacto condicionado tanto à sua própria qualidade, quanto ao impacto já garantido pelo método CE. A seguir, a metodologia de seleção de variáveis explicativas é apresentada.

4.3.1. Variáveis Explicativas

É na definição de variáveis explicativas inteligentes que reside a principal contribuição do uso de classificadores proposto nesta tese. As variáveis são divididas em dois grupos: um grupo convencional, composto pelos valores de carga momentânea e de energia disponível para fontes renováveis; e outro grupo com variáveis produzidas a partir do método CE.

Em princípio, a disponibilidade de cada equipamento do sistema poderia configurar uma variável explicativa. No entanto, isso elevaria a complexidade do classificador, de modo que as informações que de fato contribuem para a identificação de um desejado padrão serão poluídas por informação inútil. Ademais, para sistemas reais, os quais dispõem de centenas a milhares de geradores e circuitos, o uso de todos os equipamentos elevaria muito o custo de memória computacional. Para resolver, então, o desafio de selecionar um conjunto reduzido de variáveis explicativas de qualidade, é proposto empregar a inteligência codificada na estimativa da pdf-IS. Afinal, quanto maior a indisponibilidade-IS v_j (vide equações (3.33) e (3.34)) de um equipamento, maior é sua relevância para com a ocorrência da falha. Logo, é proposto agrupar equipamentos por nível de indisponibilidade-IS. São 100 níveis possíveis, percentis do valor de v_j do equipamento com maior

indisponibilidade-IS, definido como v_j^{ref} (ou indisponibilidade-IS de referência). Cada grupo deve, então, conter os equipamentos com v_j menor ou igual ao respectivo percentil de v_j^{ref} . Adicionalmente, é importante que geradores e circuitos sejam diferenciados para o classificador e, portanto, seus grupos são separados. Assim, tem-se que os grupos, em adição às variáveis convencionais de carga e energia renovável, podem formar mais de 200 variáveis explicativas. Isso não deve ocorrer, porém, ao passo que os valores de indisponibilidade-IS dos equipamentos não são uniformemente distribuídos entre si, o que faz com que a maioria dos níveis fiquem vazios. Portanto, o total de variáveis explicativas deverá ficar abaixo de 100, independentemente do tamanho do sistema.

4.3.2.

Amostra de Treinamento e Limiar de Classificação

Antes de apresentar o algoritmo final do método completo proposto nesta tese de doutorado, cabe discutir como são tratados a amostra de treinamento e um parâmetro fundamental para a eficácia do classificador, o limiar de classificação.

O classificador é empregado com o cunho de diferenciar estados de falha de estados de sucesso que são considerados críticos. Numa definição mais ampla, estados de sucesso críticos são aqueles que residem próximos à região de fronteira entre sucesso-falha (i.e., que estão próximos de falhar). No contexto deste trabalho, tais estados são definidos estritamente como aqueles que após os dois primeiros estágios da análise de adequação (vide Seção 2.7) apresentam violação de capacidade de fluxo de circuitos e, portanto, acionam a execução de fluxo de potência ótimo (FPO). Logo, a amostra de treinamento deve incluir somente os estados que demandam a execução de FPO.

Estados de sucesso ditos saudáveis não precisam ser classificados, pois já não há a execução de FPO na sua análise de adequação, a qual se encerra com apenas um fluxo de potência não otimizado, cujo custo computacional é negligenciável em comparação ao FPO. Pode ser questionado, porém, que se a amostra de treinamento incluísse todos os tipos de estado de sucesso, o classificador poderia ser empregado para todos os estados com maior impacto no desempenho. Isso não procede, pois os estados de sucesso saudáveis são muito distintos daqueles que residem na fronteira sucesso-falha, de modo que um classificador binário como o BLR não seria

capaz de detectar tantos padrões. Nesse caso, um classificador de aprendizado não-supervisionado é mais indicado. Contudo, isso não coincide com o interesse deste trabalho, que é evitar execuções de FPO.

Outra questão sobre a amostra de treinamento é de qual proporção de estados de falha e sucesso deve ser considerada. Nos estudos de caso publicados em [117], foi empregada a proporção real, de modo que todos os estados para os quais um FPO foi executado na Fase 2.1 do método foram selecionados. Isso não restringe a eficácia do classificador BLR, tendo em vista que o evento de interesse falha de suprimento está bem representado, pois a amostra coletada tem estimativa para a LOLP com incerteza de 10%. No entanto, para sistemas interligados modernos, de falha rara, a razão sucesso/falha é demasiadamente elevada, mesmo na região de fronteira. Consequentemente, foi observado que a eficiência do classificador pode ser aumentada se uma proporção menor de estados de sucesso para de falha for utilizada. Afinal, se a falha que é rara já está bem representada com tão poucos estados, o sucesso não precisa de muitos estados para também ser bem representado. Logo, propõe-se selecionar todos os estados de falha, mais o dobro em estados de sucesso (i.e., proporção 1:2). O cuidado que se deve ter é em selecionar estados de sucesso que representam todas as partes da região de fronteira. Caso contrário, o classificador não saberá diferenciar estados de falha dos de sucesso cujas características ele não foi exposto no treinamento.

O dimensionamento do limiar de classificação é bem direto. Os parâmetros estimados na avaliação de confiabilidade composta dizem respeito à falha de suprimento de alguma área de interesse. Portanto, é desejado que o classificador erre o mínimo para estados de falha. Ou seja, falsos negativos têm prioridade sobre os falsos positivos. Assim, foi determinado que o limiar de classificação $\mathbb{P}^{thr}$ deve ser igual à probabilidade estimada pela função logística (vide eq. (4.4)) do estado de falha da amostra de treinamento que apresentou a menor probabilidade. Isso é bastante conservador, pois indica tolerância zero para falso negativo, mas é adequado, dado que é muito difícil criar uma heurística para definir um limiar que possa extrair o máximo de desempenho do classificador sem prejudicar a estimativa de índices de confiabilidade.

4.3.3. Algoritmo

Para facilitar a visualização do método completo, o algoritmo final é representado na forma de um fluxograma, vide Figura 4.1. Os passos de cada fase do método (Fase 1, Fase 2.1 e Fase 2.2) são destacados.

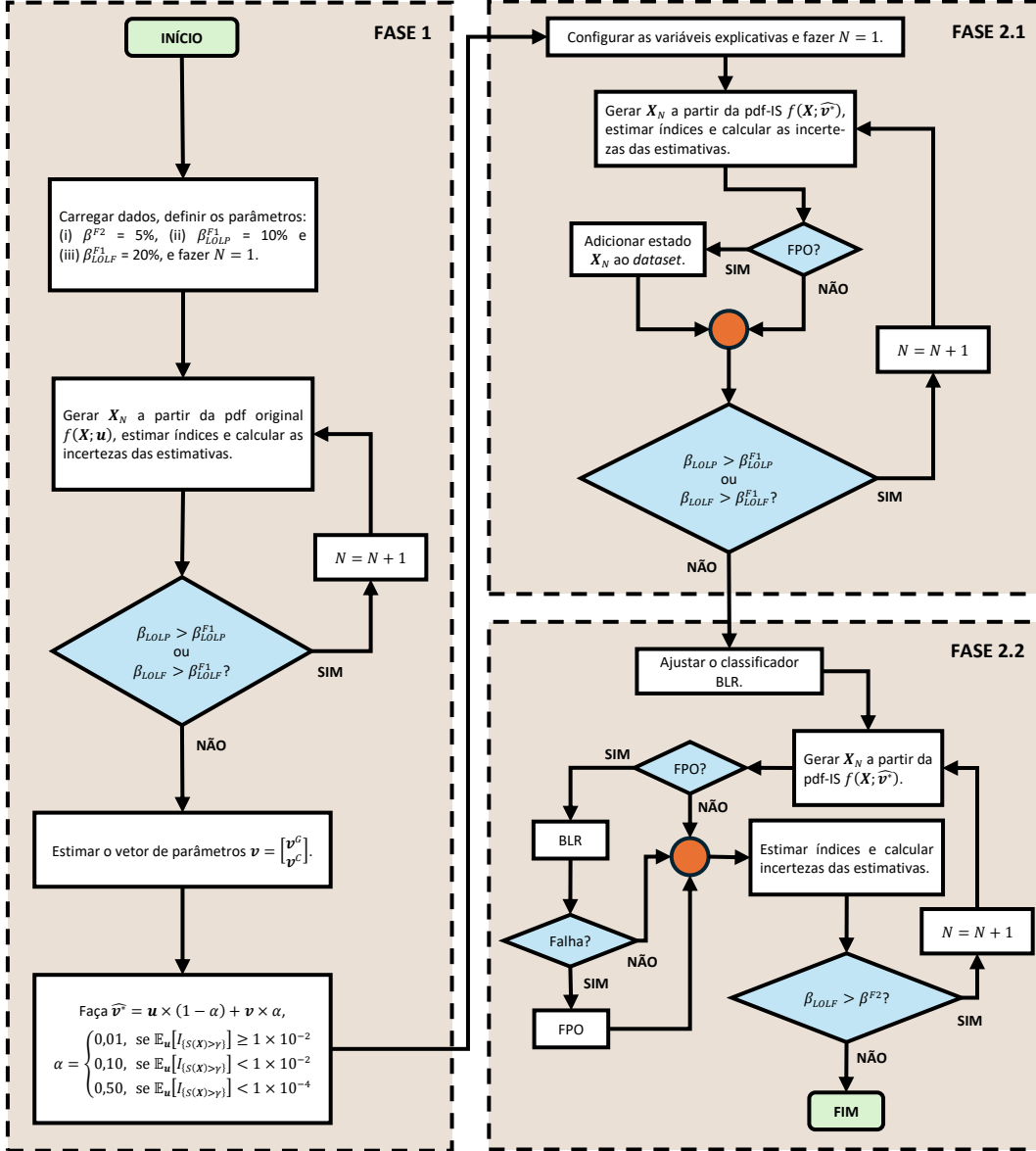


Figura 4.1: Fluxograma do método completo.

Pelo fluxograma, observam-se as duas contribuições para o uso proposto de classificadores de estados na simulação de falhas raras. Primeiro, tem-se que a coleta da amostra de treinamento (ou *dataset*) é acelerada por uma simulação com amostragem por importância. A segunda contribuição, também associada ao treinamento do classificador, está atrelada à produção de variáveis explicativas de

qualidade, as quais carregam inteligência produzida pela VRT método CE, na Fase 1. Por fim, destaca-se que dentro do passo “Ajustar o classificador BLR” está inclusa uma heurística para seleção dos estados de sucesso que integram o *dataset* de treinamento. Tal heurística corresponde a qualquer processo que produz uma seleção heterogênea dos estados de sucesso crítico, fornecendo uma boa representação das suas diferentes características.

4.3.4. Complementariedade entre VRT e Classificador

Definida a metodologia completa, até o seu algoritmo, agora é possível fazer uma importante discussão sobre a existência de complementariedade entre a ferramenta VRT (IS via método CE) e o classificador BLR.

A qualidade do classificador tem relação linear positiva com a qualidade da estimativa da pdf-IS, pois parte das variáveis explicativas utilizadas dependem da pdf-IS. Contudo, o potencial do classificador que é realizado no desempenho total da ferramenta proposta tem relação complementar com o impacto da simulação-IS. O classificador atua apenas na Fase 2.2 do método, ao passo que o seu ganho real deve ser medido em referência a todas as fases (Fase 1, 2.1 e 2.2). Logo, quanto mais rápida for a simulação-IS, que engloba toda a Fase 2, menor será a parcela do ganho produzido pelo classificador que é aproveitada, não obstante a sua qualidade. O inverso também é verdadeiro e de grande interesse, pois indica que quando a amostragem por importância tiver dificuldade, o classificador BLR será mais bem aproveitado e complementarizará o desempenho da simulação-IS.

Seja G^{total} , G^{IS} e G^{BLR} os ganhos total, da simulação-IS e do classificador BLR, e FPO^{SMCC} , FPO^{SMCIS} , FPO^{F1} , $FPO^{F2.1}$ e $FPO^{F2.2}$ os números de FPO executados em uma SMC convencional, uma SMC com IS e nas fases 1, 2.1 e 2.2 de uma SMC com a metodologia proposta. Então, os ganhos são dados por:

$$G^{total} = \frac{FPO^{SMCC}}{FPO^{F1} + FPO^{F2.1} + FPO^{F2.2}} \quad (4.13)$$

$$G^{IS} = \frac{FPO^{SMCC}}{FPO^{SMCIS}} \quad (4.14)$$

$$G^{BLR} = \frac{FPO^{SMCIS} - FPO^{F1} - FPO^{F2.1}}{FPO^{F2.2}} \quad (4.15)$$

O ganho G^{total} é dado pela razão de execuções de FPO entre a simulação de referência (i.e., sem auxílio de VRT e classificador) e a simulação completa do método proposto. O ganho G^{IS} é dado pela razão entre a simulação de referência e uma simulação com a ajuda da amostragem por importância, mas sem classificador. Já G^{BLR} é dado pela razão entre o que seria executado na Fase 2.2 do método proposto caso o classificador não fosse utilizado e o que de fato foi executado.

Colocando FPO^{SMCIS} em evidência na eq. (4.15) e o substituindo na eq. (4.14), tem-se que:

$$G^{IS} = \frac{FPO^{SMCC}}{G^{BLR} \times FPO^{F2.2} + FPO^{F1} + FPO^{F2.1}} \quad (4.16)$$

E se FPO^{SMCC} é colocado em evidência na eq. (4.16) e substituído na eq. (4.13), tem-se que:

$$G^{total} = \frac{G^{IS} \times (G^{BLR} \times FPO^{F2.2} + FPO^{F1} + FPO^{F2.1})}{FPO^{F1} + FPO^{F2.1} + FPO^{F2.2}} \quad (4.17)$$

Note que o denominador da eq. (4.17) corresponde ao total execuções de FPO da ferramenta proposta, de modo que a parcela de cada fase dividida pelo total fornece a participação p que cada fase tem. Observe:

$$G^{total} = G^{IS} \times (G^{BLR} \times p^{F2.2} + p^{F1} + p^{F2.1}) \quad (4.18)$$

Como a soma das parcelas é igual 1 pu, tem-se que $p^{F1} + p^{F2.1} = 1 - p^{F2.2}$ e, portanto:

$$G^{total} = G^{IS} \times (p^{F2.2} \times (G^{BLR} - 1) + 1) \quad (4.19)$$

Da eq. (4.19) conclui-se que G^{total} é diretamente proporcional à G^{IS} , $p^{F2.2}$ e G^{BLR} . Uma pdf-IS de maior qualidade tende a produzir um classificador de maior qualidade. Isso significa que se G^{IS} aumenta, G^{BLR} também aumenta e, com isso, G^{total} aumenta em duas frentes. O valor de $p^{F2.2}$, por sua vez, é inversamente proporcional ao êxito da VRT, pois quanto mais rápida é a simulação-IS, menor é o número de execuções de FPO na Fase 2 e, por conseguinte, menor é o valor de $p^{F2.2}$. Deve-se observar, porém, que apenas a contribuição de G^{BLR} é regulada por $p^{F2.2}$. Caso o classificador não apresente nenhum ganho (ou nem seja empregado), G^{BLR} será igual a 1 pu e G^{total} será igual a G^{IS} apenas. Portanto, é na parcela $p^{F2.2} \times (G^{BLR} - 1)$ que se percebe a complementariedade entre a VRT e o

classificador. Afinal, ainda que um valor menor de $p^{F2.2}$ também signifique que G^{BLR} é maior, porque a amostragem por importância é bem-sucedida, o ganho do classificador tem uma parcela menor sendo aproveitada.

Para o fim de uma análise mais elaborada do impacto do classificador no método proposto, define-se mais uma estatística:

$$G^{BLR*} = \frac{G^{total}}{G^{IS}} = \frac{FPO^{SMCIS}}{FPO^{F1} + FPO^{F2.1} + FPO^{F2.2}} \quad (4.20)$$

O ganho G^{BLR*} é o ganho real que o classificador BLR produz, dado pela razão entre o ganho total e o associado à simulação-IS, ou pela razão do número de FPOs executados por uma simulação-IS sem classificador e o observado na ferramenta proposta. Assim, é possível obter o aproveitamento percentual do ganho do classificador, tal como eq. (4.21). Essa estatística é fundamental na demonstração do aspecto de complementariedade, na medida que conforme o êxito da VRT aumenta, o aproveitamento ϕ deve reduzir, e vice-versa.

$$\phi = \frac{(G^{BLR*} - 1)}{(G^{BLR} - 1)} \times 100 \quad (4.21)$$

A conclusão mais importante a ser observada nesta subseção é a de que o ganho produzido por um classificador é maior, mas é menos aproveitado, quando o método menos precisa (i.e., quando o ganho produzido pela simulação-IS é maior), enquanto é menor, mas é mais aproveitado, quando o método mais precisa (i.e., quando o ganho produzido pela simulação-IS é menor). Isso denota a existência de complementariedade entre a VRT e o classificador para o método proposto.

4.4. Avaliação do Método Proposto

Nesta seção, o método completo proposto nesta tese de doutorado é avaliado. Para tanto, todos os estudos de caso apresentados no Capítulo 3 são reavaliados, de modo que a ferramenta final possa ser comparada não apenas à simulação convencional, mas, também, à ferramenta proposta no Capítulo 3. Mais uma vez, destaca-se que os índices exibidos nos resultados das análises são: LOLP (*loss of load probability*), EENS (*expected energy not supplied*) e LOLF (*loss of load frequency*), enquanto os indicadores de performance são: número de estados amostrados, número de execuções de fluxo de potência ótimo e tempo de processamento.

A ferramenta proposta no Capítulo 3 é referida como SMCIS-GT (SMC quase-sequencial com IS via método CE para sistemas de geração e transmissão), e a simulação convencional é SMCC (SMC convencional). Já o método proposto neste capítulo é chamado de SMCIS-BLR (SMC quase-sequencial com IS via método CE e classificador BLR para sistemas de geração e transmissão). Sobre a implementação computacional, todas as considerações feitas no capítulo anterior seguem para os estudos apresentados nesta seção.

Por fim, tem-se que o sistema teste utilizado, tal como definido na Subseção 3.4.1, é uma versão modificada do IEEE Reliability Test System – 1996 (RTS-96) [111], chamado de RTS-96-REN-T. O RTS-96-REN-T é bem representativo de sistemas interligados modernos, pois apresenta participação renovável elevada, falha composta basicamente igual à falha T (i.e., falha por insuficiência de recursos de transmissão) e uma adequada disposição de níveis de raridade para sistema, área/zona e barramento (i.e., a falha do sistema é não rara, de área/zona é pouco rara e de barra é bastante rara).

4.4.1. Análise de Sistema

Na análise da falha do sistema da Subseção 3.4.2 foi denotado que o método SMCIS-GT não apresenta ganho real, mas também não causa prejuízo à performance computacional e estimativa de índices. Isso é tratado de forma natural, pois na situação de falha não rara ($LOLP \geq 1\%$) não é esperado ganho computacional decorrente da atuação de VRTs. Por outro lado, na Tabela 4.1, já se nota um *speedup* com algum impacto para o SMCIS-BLR. Observe que não há ganho adicional para sorteios, tendo em vista que o classificador atua apenas na redução de execuções de FPO. O ganho em FPOs é maior que o de tempo de processamento, pois no tempo da SMCIS-BLR está incluído o tempo gasto para ajustar o classificador (em torno de 5 segundos), bem como o tempo gasto para classificar estados, que não é nulo, ainda que muito pequeno. Ademais, destaca-se que o custo de solução de FPO para estados de falha é, em geral, maior que o de estados de sucesso.

Nas fases 1, 2.1 e 2.2 do SMCIS-BLR foram executados 2826, 3121 e 42.138 FPOs, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 66.004 e 63.717 execuções. Logo, os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 1,37 pu, 1,04 pu

e 1,37 pu. O fato de G^{total} e G^{BLR} apresentarem valores iguais é mera coincidência. O ganho real do classificador ficou em $G^{BLR*} = 1,32$ pu, de modo que o aproveitamento ficou em $\phi = 86,49\%$. Observe que, por a falha ser não rara, o amortecimento aplicado na formação do vetor de parâmetros da pdf-IS é elevado (vide Subseção 3.3.3), o que reduz a carga de informação acerca da falha em estudo que é codificada na pdf-IS. Assim, as variáveis explicativas produzidas a partir da pdf-IS são menos explicativas e a qualidade do classificador é reduzida. Em contrapartida, o ganho do classificador é muito bem aproveitado, justo quando a simulação-IS tem menos força.

Os resultados da análise de sistema sinalizam que o método SMCIS-BLR pode, para além de ser utilizado indiscriminadamente (sem temer eventual prejuízo computacional), apresentar ganho real mesmo em situação de falha não rara, ainda que obter desempenho nessa circunstância não seja fundamental. Com o intuito de exemplificar o processo de construção das variáveis explicativas que são fornecidas ao classificador BLR, o Apêndice B traz uma análise das variáveis geradas e seus coeficientes para esse caso de falha do sistema RTS-96-REN-T.

Tabela 4.1: Análise do sistema (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	1,23 ($\beta=1,67\%$)	1,23 ($\beta=1,68\%$)	0,00%	1,22 (1,68%)	0,81%
EENS [MWh/ano]	1580 ($\beta=2,97\%$)	1610 ($\beta=3,01\%$)	1,90%	1600 ($\beta=2,99\%$)	1,27%
LOLF [oc./ano]	31,1 ($\beta=5,00\%$)	32,6 ($\beta=5,00\%$)	4,82%	32,1 ($\beta=5,00\%$)	3,22%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	289.893	277.333	1,05 pu	282.152	1,03 pu
FPO	66.004	63.717	1,04 pu	48.085	1,37 pu
Tempo	4,27 min	4,12 min	1,04 pu	3,23 min	1,32 pu

4.4.2. Análise de Zona

A Tabela 4.2 mostra os resultados para a avaliação da zona A3SUL, que engloba as barras 301, 302, 304, 305, 307 e 308 (vide Figura 3.2). Note que nesse caso a falha é mais rara, ainda que pouco rara ($0,01\% < \text{LOLP} < 1\%$), o que leva a simulação-IS a ser mais rápida, com ganho acima de 2 pu. Consequentemente, a

qualidade do classificador aumenta, ao passo que a parcela que é aproveitada do seu ganho diminui. De qualquer forma, o impacto do aumento da qualidade do classificador é maior e, com isso, o ganho real é mais expressivo do que o obtido na análise do sistema.

Nas fases 1, 2.1 e 2.2 do SMCIS-BLR foram executados 267.944, 88.083 e 417.848 FPOs, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 3.456.715 e 1.508.261 execuções. Logo, os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 4,47 pu, 2,29 pu e 2,76 pu. O ganho real do classificador ficou em $G^{BLR*} = 1,95$ pu, de modo que o aproveitamento ficou em $\phi = 53,98\%$. Apesar do aproveitamento bem menor neste caso, o ganho real é claramente maior que os 1,32 pu da análise do sistema. Isso indica que a maior qualidade da pdf-IS produz mais aumento de G^{BLR*} , devido à maior qualidade do classificador, do que redução, devido à simulação-IS mais rápida. Esse comportamento é desejável, pois se quer tão mais desempenho quanto maior for a raridade da falha em estudo.

Tabela 4.2: Análise da zona A3SUL (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$2,42 \times 10^{-2}$ ($\beta=1,65\%$)	$2,40 \times 10^{-2}$ ($\beta=1,45\%$)	0,83%	$2,37 \times 10^{-2}$ ($\beta=1,52\%$)	2,07%
EENS [MWh/ano]	39,6 ($\beta=2,46\%$)	38,1 ($\beta=2,06\%$)	3,79%	37,5 ($\beta=2,16\%$)	5,30%
LOLF [oc./ano]	0,595 ($\beta=5,00\%$)	0,603 ($\beta=5,00\%$)	1,34%	0,588 ($\beta=5,00\%$)	1,18%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	15.213.052	6.448.294	2,36 pu	5.958.086	2,55 pu
FPO	3.456.715	1.508.261	2,29 pu	773.875	4,47 pu
Tempo	3,84 h	1,84 h	2,10 pu	1,11 h	3,46 pu

4.4.3. Análise de Barra

Revisitando a Barra 301, têm-se os resultados da Tabela 4.3. Conforme discutido na Subseção 3.4.4, a falha desta barra é um evento bastante raro (LOLP < 0,01%), de modo que o ganho da ferramenta SMCIS-GT deveria ter se aproximado da estimativa do ganho máximo de 16 vezes (vide Subseção 3.3.3). Isso não ocorre, pois a falha da Barra 301 é uma falha dispersa, com muita diversidade de causas, o que leva à Fase 2 do método a ser menos eficiente do que se esperava. É aqui,

portanto, que a complementariedade entre a simulação-IS e o classificador BLR encontra a maior oportunidade de contribuição. Quando a falha é rara, situação em que um elevado ganho computacional é mais desejado, mas a simulação-IS tem dificuldades devido à dispersão na falha, o classificador fornece o “impulso” que faltava para que o método proposto possa se aproximar do seu potencial.

Nas fases 1, 2.1 e 2.2 do SMCIS-BLR foram executados 402.293, 62.853 e 211.626 FPOs, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 9.382.787 e 1.408.646 execuções. Logo, os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 13,86 pu, 6,66 pu e 4,46 pu. O ganho real do classificador ficou em $G^{BLR*} = 2,08$ pu, de modo que o aproveitamento ficou em $\phi = 31,21\%$. Note que, de fato, quanto mais rara é a falha em análise, maior é a qualidade da pdf-IS e, com isso, do classificador. Em contrapartida, o aproveitamento do ganho produzido pelo classificador diminui, mas com intensidade menor. Na ordem sistema, zona A3SUL e Barra 301 (falha não rara, pouco rara e muito rara) tem-se G^{BLR} de 1,37 pu, 2,76 pu e 4,46 pu, enquanto os aproveitamentos do ganho do classificador ficam em 86,49%, 53,98% e 31,21%. Logo, o ganho real do classificador tende a sempre aumentar (1,32 pu, 1,95 pu e 2,08 pu) conforme o nível de raridade aumenta. Ou seja, o classificador contribui tão mais quanto mais desempenho é desejado em termos gerais (i.e., quanto maior é o nível de raridade da falha). No entanto, deve-se constatar que o classificador é tão mais aproveitado quanto mais a simulação-IS precisa (i.e., quanto mais alguém do seu potencial ela performa). Na próxima subseção, os estudos de caso que foram projetados para explorar o aspecto de dispersão da falha são reapresentados, de modo a reforçar o conceito de complementariedade presente na ferramenta SMCIS-BLR.

Tabela 4.3: Análise da Barra 301 (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$6,08 \times 10^{-3}$ ($\beta=1,99\%$)	$5,81 \times 10^{-3}$ ($\beta=2,21\%$)	4,44%	$5,83 \times 10^{-3}$ ($\beta=2,20\%$)	4,11%
EENS [MWh/ano]	6,45 ($\beta=2,68\%$)	6,23 ($\beta=2,53\%$)	3,41%	6,23 ($\beta=2,52\%$)	3,41%
LOLF [oc./ano]	0,218 ($\beta=5,00\%$)	0,222 ($\beta=5,00\%$)	1,83%	0,217 ($\beta=5,00\%$)	0,46%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	41.335.790	5.208.363	7,94 pu	5.233.268	7,90 pu
FPO	9.382.787	1.408.646	6,66 pu	676.772	13,86 pu
Tempo	9,50 h	1,55 h	6,13 pu	0,89 h	10,67 pu

4.4.4. Estudo da Dispersão da Falha

Na Subseção 3.4.5, três estudos de caso foram apresentados com as barras 301, 307 e 320, para averiguar a questão de dispersão da falha. Já nesta subseção, os casos são reavaliados pela SMCIS-BLR com o objetivo de demonstrar que o classificador BLR terá seu ganho tão mais bem aproveitado quanto maior for a dispersão da falha em análise.

4.4.4.1. Caso 1 – Barra 301

Na Subseção 3.4.5.1, foi demonstrado que a falha da Barra 301 é dispersa no RTS-96-REN-T e pouco dispersa no RTS-96-REN-T#1. Na Tabela 4.4, o impacto de reduzir a dispersão da falha na análise da SMCIS-BLR pode ser observado. É possível confirmar, então, que o ganho real do classificador é menor quando a dispersão da falha é menor, observe a análise dos números no próximo parágrafo.

Tabela 4.4: Análise da Barra 301 (RTS-96-REN-T#1).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$4,88 \times 10^{-3}$ ($\beta=1,90\%$)	$4,93 \times 10^{-3}$ ($\beta=1,85\%$)	1,02%	$4,86 \times 10^{-3}$ ($\beta=1,84\%$)	0,41%
EENS [MWh/ano]	4,73 ($\beta=2,55\%$)	4,57 ($\beta=2,10\%$)	3,38%	4,55 ($\beta=2,10\%$)	3,81%
LOLF [oc./ano]	0,161 ($\beta=5,00\%$)	0,165 ($\beta=5,00\%$)	2,48%	0,171 ($\beta=5,00\%$)	6,21%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	56.454.044	3.734.657	15,12 pu	3.742.074	15,09 pu
FPO	6.491.142	480.681	13,50 pu	373.603	17,37 pu
Tempo	6,55 h	0,60 h	10,92 pu	0,46 h	14,24 pu

Nas fases 1, 2.1 e 2.2 do SMCIS-BLR foram executados 331.386, 7119 e 35.098 FPOs, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 6.491.142 e 480.681 execuções. Logo, os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 17,37 pu, 13,50 pu e 4,05 pu. O ganho real do classificador ficou em $G^{BLR*} = 1,29$ pu, de modo que o aproveitamento ficou em $\phi = 9,51\%$. Note que a falha da Barra 301 permaneceu no mesmo nível de raridade, de modo que a qualidade do classificador não altera muito. Por outro lado, como a falha é bem menos dispersa no

sistema RTS-96-REN-T#1, a simulação-IS é tão mais rápida que o ganho real do classificador ficou menor até que o obtido na análise do sistema RTS-96-REN-T. A redução de aproveitamento em mais de três vezes (de 31,21 para 9,51%) confirma isso. Deve-se observar, porém, que o ganho da SMCIS-GT já é bem alto e se aproxima do ganho máximo estimado de 16 pu. Ou seja, tem-se um exemplo muito claro da complementariedade entre VRT e classificador. Os classificadores associados a duas falhas distintas (falha da Barra 301 no sistema RTS-96-REN-T e RTS-96-REN-T#1) apresentam qualidade similar, mas um produz ganho real maior quando é mais necessário (simulação-IS menos eficiente), enquanto o outro produz ganho real menor quando é menos necessário (simulação-IS mais eficiente).

4.4.4.2.

Caso 2 – Barra 307

Conforme destacado na Subseção 3.4.5.2, a Barra 307 é a única dentro da Área 3 cuja falha é verdadeiramente não dispersa no RTS-96-REN-T, sendo, inclusive, uma falha com causas quase totalmente associadas à estação geradora da própria barra. Já no sistema RTS-96-REN-T#2, a falha ganha caráter de dispersão. A Tabela 4.5 e a Tabela 4.6 trazem os resultados da análise da Barra 307 nos sistemas RTS-96-REN-T e RTS-96-REN-T#2.

Tabela 4.5: Análise da Barra 307 (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$2,23 \times 10^{-3}$ ($\beta=1,59\%$)	$2,26 \times 10^{-3}$ ($\beta=0,97\%$)	1,35%	$2,25 \times 10^{-3}$ ($\beta=0,97\%$)	0,90%
EENS [MWh/ano]	2,09 ($\beta=1,98\%$)	2,10 ($\beta=0,91\%$)	0,48%	2,11 ($\beta=0,91\%$)	0,96%
LOLF [oc./ano]	0,0513 ($\beta=5,00\%$)	0,0528 ($\beta=5,00\%$)	2,92%	0,0510 ($\beta=5,00\%$)	0,58%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	177.327.622	12.586.877	14,09 pu	12.570.006	14,11 pu
FPO	40.255.896	2.980.605	13,51 pu	2.838.347	14,18 pu
Tempo	40,40 h	3,30 h	12,24 pu	3,13 h	12,91 pu

Para a análise da barra no sistema RTS-96-REN-T, foram executados 2.636.822, 18.132 e 183.393 FPOs nas fases 1, 2.1 e 2.2 do SMCIS-BLR, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 40.255.896 e 2.980.605 execuções. Os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 14,18 pu, 13,51 pu e 1,78 pu,

enquanto o ganho real do classificador ficou em $G^{BLR^*} = 1,05$ pu, o que confere um aproveitamento de $\phi = 6,41\%$. Já para a análise da barra no sistema RTS-96-REN-T#2, foram executados 1.573.454, 60.972 e 231.038 FPOs nas fases 1, 2.1 e 2.2 do SMCIS-BLR, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 18.155.049 e 2.466.419 execuções. Os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 9,73 pu, 7,42 pu e 3,60 pu, enquanto o ganho real do classificador ficou em $G^{BLR^*} = 1,32$ pu, o que confere um aproveitamento de $\phi = 12,31\%$. Desse modo, constata-se mais uma vez que o classificador é mais bem aproveitado quando a simulação-IS é mais fraca.

Tabela 4.6: Análise da Barra 307 (RTS-96-REN-T#2).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$2,36 \times 10^{-3}$ ($\beta=2,86\%$)	$2,37 \times 10^{-3}$ ($\beta=2,87\%$)	0,42%	$2,27 \times 10^{-3}$ ($\beta=2,87\%$)	3,81%
EENS [MWh/ano]	1,97 ($\beta=4,81\%$)	1,96 ($\beta=4,14\%$)	0,51%	1,92 ($\beta=4,13\%$)	2,54%
LOLF [oc./ano]	0,176 ($\beta=5,00\%$)	0,179 ($\beta=5,00\%$)	1,70%	0,170 ($\beta=5,00\%$)	3,41%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	51.657.746	6.400.485	8,07 pu	6.457.739	8,00 pu
FPO	18.155.049	2.466.419	7,42 pu	1.865.464	9,73 pu
Tempo	22,65 h	3,08 h	7,35 pu	2,68 h	8,45 pu

4.4.4.3.

Caso 3 – Barra 320

O último estudo de caso traz uma falha, da Barra 320, de grau de dispersão intermediário, com características distintas das demais barras apresentadas. Sendo considerada um exemplo de falha dispersa no sistema RTS-96-REN-T, modificações são feitas de modo que a falha da barra apresente grau bem reduzido de dispersão no sistema RTS-96-REN-T#3. A Tabela 4.5 e a Tabela 4.6 trazem os resultados da análise da Barra 320 nos sistemas RTS-96-REN-T e RTS-96-REN-T#3.

Para a análise da barra no sistema RTS-96-REN-T, foram executados 3.720.641, 82.101 e 1.177.000 FPOs nas fases 1, 2.1 e 2.2 do SMCIS-BLR, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 50.876.036 e 5.539.983 execuções. Os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 10,22 pu, 9,18 pu e 1,48 pu,

enquanto o ganho real do classificador ficou em $G^{BLR^*} = 1,11$ pu, o que confere um aproveitamento de $\phi = 22,92\%$. Já para a análise da barra no sistema RTS-96-REN-T#2, foram executados 2.016.137, 4592 e 35.578 FPOs nas fases 1, 2.1 e 2.2 do SMCIS-BLR, respectivamente, enquanto a SMCC e SMCIS-GT realizaram 33.806.552 e 2.122.125 execuções. Os ganhos G^{total} , G^{IS} e G^{BLR} ficaram em 16,44 pu, 15,93 pu e 2,85 pu, enquanto o ganho real do classificador ficou em $G^{BLR^*} = 1,03$ pu, o que confere um aproveitamento de $\phi = 1,62\%$.

Tabela 4.7: Análise da Barra 320 (RTS-96-REN-T).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$1,21 \times 10^{-3}$ ($\beta=1,92\%$)	$1,19 \times 10^{-3}$ ($\beta=1,77\%$)	1,65%	$1,20 \times 10^{-3}$ ($\beta=1,93\%$)	0,83%
EENS [MWh/ano]	2,49 ($\beta=2,43\%$)	2,42 ($\beta=1,98\%$)	2,81%	2,45 ($\beta=2,18\%$)	1,61%
LOLF [oc./ano]	0,0406 ($\beta=5,00\%$)	0,0417 ($\beta=5,00\%$)	2,71%	0,0402 ($\beta=5,00\%$)	0,99%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	224.114.025	23.703.157	9,46 pu	22.519.658	9,95 pu
FPO	50.876.036	5.539.983	9,18 pu	4.979.742	10,22 pu
Tempo	53,25 h	6,08 h	8,76 pu	5,77 h	9,23 pu

Tabela 4.8: Análise da Barra 320 (RTS-96-REN-T#3).

Índices	SMCC	SMCIS-GT	Desvio	SMCIS-BLR	Desvio
LOLP [%]	$1,31 \times 10^{-3}$ ($\beta=1,82\%$)	$1,30 \times 10^{-3}$ ($\beta=1,84\%$)	0,76%	$1,29 \times 10^{-3}$ ($\beta=1,86\%$)	1,53%
EENS [MWh/ano]	1,09 ($\beta=2,70\%$)	1,07 ($\beta=2,12\%$)	1,83%	1,06 ($\beta=2,17\%$)	2,75%
LOLF [oc./ano]	0,0401 ($\beta=5,00\%$)	0,0411 ($\beta=5,00\%$)	2,49%	0,0420 ($\beta=5,00\%$)	4,74%
Performance	SMCC	SMCIS-GT	Ganho	SMCIS-BLR	Ganho
Sorteios	230.566.561	14.313.029	16,11 pu	14.314.733	16,11 pu
FPO	33.806.552	2.122.125	15,93 pu	2.056.307	16,44 pu
Tempo	27,65 h	2,40 h	11,52 pu	2,36 h	11,72 pu

Claramente, se confirma mais uma vez que o classificador é mais bem aproveitado quando a simulação-IS é mais fraca. A diferença, nesse caso, é que a qualidade do classificador no cenário de falha mais dispersa (RTS-96-REN-T) ficou ruim, com apenas 1,48 pu de ganho. Isso se explica pela ocorrência de um *outlier* na amostra de treinamento que levou o limiar de classificação $\mathbb{P}^{thr}$ para um nível mais baixo. Deve-se lembrar que $\mathbb{P}^{thr}$ é modelado para receber o valor de

classificação do estado da amostra de treinamento que é de falha e apresentou o menor valor de classificação. Refazendo a análise da falha da Barra 320 no sistema RTS-96-REN-T desconsiderando o referido estado *outlier* da amostra de treinamento, obteve-se um ganho para o classificador de 7,14 pu, de modo que $G^{BLR^*} = 1,37$ pu e $\phi = 6,03\%$. Esse era o desempenho esperado para o classificador, mas a ferramenta proposta não é projetada para lidar com *outliers*, algo que pode e deve ser investigado no futuro. Cabe ainda destacar que em todas as análises existem estados *outliers* em alguma capacidade, mas o caso da Barra 320 foi o mais significativo, sendo o escolhido para exemplificar esse aspecto.

4.5. Redes Neurais Multicamadas

Com o objetivo de elevar a capacidade de classificação em meio à metodologia proposta, buscou-se implementar redes neurais de maior complexidade. Apesar de não ter sido criado com essa finalidade, o classificador BLR é efetivamente uma rede neural artificial. Contudo, toda a sua excelente simplicidade é também o que limita, em princípio, sua capacidade de reconhecer padrões de complexidade maior. Afinal, a rede BLR possui apenas um neurônio, ao passo que redes de múltiplos neurônios e múltiplas camadas permitem mais combinações entre as variáveis explicativas designadas ao problema. Ou seja, redes mais profundas podem reconhecer um volume maior de modos de falha que compõem uma dada falha de interesse. Desse modo, recorreu-se a uma rede já provada na literatura, a *group method of data handling* (GMDH) [31-32], a qual se apresenta como uma das redes multicamadas mais simples e de outras qualidades interessantes, como ser uma rede do tipo *self-organizing*; i.e., a topologia da rede é construída de forma automática. A seguir, tem-se um breve resumo da metodologia da rede GMDH.

4.5.1. Rede GMDH

A rede GMDH surge como uma rede polinomial auto construtiva, desenvolvida pelo matemático ucraniano Alexey Ivakhnenko. Ela foi um dos primeiros métodos a lidar com as chamadas redes profundas [118], que compõem a área de *deep learning*. A principal característica do GMDH é o uso de raciocínio indutivo. O

raciocínio dedutivo parte do geral para inferir conclusões específicas, enquanto o raciocínio indutivo parte do particular para formar conclusões mais abrangentes. Em geral, as redes neurais têm sua topologia e variáveis explicativas dimensionadas a partir da experiência do usuário acerca do problema, o que configura um método dedutivo. Já a rede GMDH se propõe a organizar a sua topologia automaticamente, do número de camadas e de neurônios por camada à escolha das variáveis explicativas fornecidas, o que configura um método indutivo.

A formação da topologia da rede GMDH é dada pela seguinte lógica. Primeiro, define-se que todo neurônio, de qualquer camada, deve receber a saída de dois neurônios da camada anterior, além de ter como função de ativação o polinômio de Ivaknenko, tal como eq. (4.22). A partir disso, conclui-se que cada camada terá um número de neurônios nnc dado pela combinação dois a dois do número de neurônios da camada anterior. Isso criaria um problema combinatório explosivo, e, portanto, nnc é limitado pelo hiperparâmetro nnc^{max} , que é menor ou igual ao número de variáveis explicativas inicialmente assinaladas pelo usuário. Desse modo, a cada nova camada, nnc^{max} neurônios devem ser selecionados dentre aqueles que são gerados pela combinação dos neurônios da camada anterior. Essa seleção é feita com base na aptidão do neurônio em classificar corretamente os estados da amostra de treinamento. Isso, por sua vez, indica que para toda nova camada os novos neurônios são treinados individualmente, ao passo que os demais neurônios da rede permanecem com o ajuste que lhes foi conferido na geração da sua respectiva camada.

$$Y(U, V) = A + BU + CV + DU^2 + EV^2 + FUV \quad (4.22)$$

A convergência da construção da rede GMDH ocorre quando a estatística de aptidão do melhor neurônio da nova camada não é melhor que a do melhor neurônio da camada anterior, a qual é designada como a camada de saída. Como cada neurônio possui apenas duas entradas, toda camada deverá ter neurônios descartados, inclusive a de entrada. As mais próximas à saída da rede terão mais neurônios descartados, enquanto as mais próximas à entrada da rede terão menos. Na Figura 4.2 tem-se um exemplo ilustrativo de rede GMDH, no qual se verifica o seu caráter auto construtivo. Os neurônios em preto representam aqueles que de fato compõem a rede final dentre todos os que foram gerados durante o processo de treinamento.

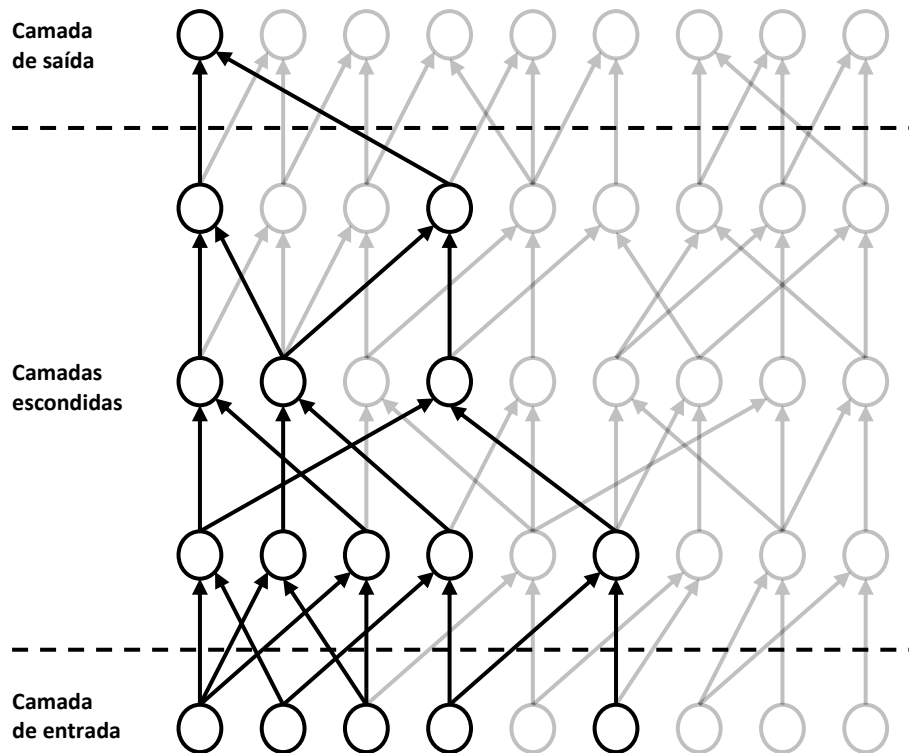


Figura 4.2: Exemplo ilustrativo de topologia de uma rede GMDH.

4.5.2. Implementação

Implementando uma rede GMDH tal como descrito na Subseção 4.5.1 não resultou em bom desempenho. O classificador GMDH em si apresentou qualidade satisfatória, mas o custo computacional para treinar a rede se mostrou proibitivo. Além disso, foi observado que as melhores variáveis explicativas tendem a dominar a formação dos neurônios, de modo que a geração de novas camadas rapidamente se torna redundante. Com essas questões em mente, uma versão adaptada foi confeccionada buscando extrair mais do GMDH. Nesta versão, foi determinado que na seleção dos neurônios para toda nova camada não se pode repetir nenhum neurônio da camada anterior. Isto é, ainda que a combinação de um novo neurônio ainda não tenha sido selecionada, se um dos dois componentes já tiver aparecido em outra combinação, tal neurônio deve ser descartado. Assim, tem-se duas consequências:

- i) toda nova camada apresentará um número de neurônios igual à metade do montante presente na camada anterior;
- ii) todas as variáveis explicativas que compõem uma entrada de um dado neurônio são diferentes de todas as variáveis que compõem a outra entrada.

A primeira consequência resolve a questão de custo computacional proibitivo, enquanto a segunda resolve a questão de redundância de variáveis explicativas dominantes. Por fim, entendendo que a função logística, por calcular uma probabilidade, tem maior afinidade com a tarefa de classificar estados dicotômicos (tal como são os estados de falha/sucesso do problema de confiabilidade), o polinômio de Ivakhnenko foi preterido pela função logística como a função de ativação dos neurônios da rede GMDH.

4.5.3. Resultados

Na Tabela 4.9 e Tabela 4.10, é apresentado um resumo comparativo de desempenho entre a ferramenta SMCIS-BLR e aquela que emprega GMDH no lugar de BLR, a qual é denominada SMCIS-GMDH (SMC quasi-sequencial com IS via método CE e classificador GMDH para sistemas de geração e transmissão). As análises apresentadas são do sistema, da zona A3SUL e das barras 301, 307 e 320, considerando o sistema RTS-96-REN-T. As estatísticas de desempenho apresentadas são o ganho do classificador G^{class} , o ganho real do classificador G^{class*} e o aproveitamento percentual do ganho do classificador ϕ . Tais estatísticas dizem respeito à execução de FPO da ferramenta em questão para com a simulação de referência SMCC e seguem, respectivamente, as equações (4.15), (4.20) e (4.21).

Tabela 4.9: Análises do sistema e da zona A3SUL (RTS-96-REN-T).

Sistema	SMCIS-BLR	SMCIS-GMDH
G^{class}	1,37 pu	1,13 pu
G^{class*}	1,32 pu	1,11 pu
ϕ	86,49%	84,62%
Zona A3SUL	SMCIS-BLR	SMCIS-GMDH
G^{class}	2,76 pu	2,70 pu
G^{class*}	1,95 pu	1,93 pu
ϕ	53,98%	54,71%

Nota-se que o desempenho da SMCIS-GMDH ficou realmente inferior apenas nas análises de sistema e Barra 301, enquanto as demais análises apresentaram desempenho basicamente igual. A análise da Barra 320, por sua vez, foi a única em que a SMCIS-GMDH desempenhou melhor que a SMCIS-BLR. Portanto, apesar

de bons resultados, a ferramenta proposta auxiliada por um classificador GMDH não foi capaz de superar aquela auxiliada por um classificador BLR. Isso diz muito sobre a qualidade da simples rede BLR, a qual atingiu um desempenho semelhante, senão superior, ao de uma rede multicamadas, cuja expectativa de retorno é maior.

Tabela 4.10: Análises das barras 301, 307 e 320 (RTS-96-REN-T).

Barra 301	SMCIS-BLR	SMCIS-GMDH
G^{class}	4,46 pu	2,21 pu
G^{class*}	2,08 pu	1,58 pu
ϕ	31,21%	47,93%
Barra 307	SMCIS-BLR	SMCIS-GMDH
G^{class}	1,78 pu	1,65 pu
G^{class*}	1,05 pu	1,04 pu
ϕ	6,41%	6,15%
Barra 320	SMCIS-BLR	SMCIS-GMDH
G^{class}	1,48 pu	1,71 pu
G^{class*}	1,11 pu	1,15 pu
ϕ	22,92%	21,13%

Os resultados apresentados sinalizam também que a rede GMDH, tal como foi implementada, pode não estar produzindo uma arquitetura suficientemente profunda, de modo a identificar os inúmeros modos de falha que compõem um dado evento de interesse. Adicionalmente, tem-se a questão do ajuste individual de neurônios que pode levar à perda de informação entre uma camada e a próxima. Evidentemente, esses são aspectos que dão à rede GMDH o caráter de método indutivo e auto construtivo. No entanto, há ainda muitas redes multicamadas profundas de caráter dedutivo que podem ser exploradas e podem produzir bons resultados, algo que será indicado nas propostas de trabalhos futuros desta tese.

4.6. Conclusões

Este capítulo apresentou a segunda e última parte do método proposto em detalhe. A metodologia genérica da ferramenta de regressão logística binária foi apresentada, seguida da proposta de acoplar tal ferramenta na forma de um classificador de estados à simulação com amostragem por importância estabelecida no Capítulo 3. Esse acoplamento, por sua vez, sugere não apenas usar uma técnica de

redução de variância e outra de aprendizado de máquina em conjunto, mas de empregar conhecimento adquirido em uma fase no desenvolvimento da outra.

Duas novas contribuições são geradas pela parcela do método proposta neste capítulo. A primeira está atrelada a ter o processo de treinamento de um classificador de estados acelerado por uma simulação com amostragem por importância. Já a segunda, diz respeito a elevar a qualidade do classificador com a produção de novas variáveis explicativas a partir de informação associada ao evento de interesse que é codificada na estimativa de parâmetros de distribuição para a técnica de amostragem por importância.

Na Seção 4.4, a metodologia proposta foi avaliada extensivamente, revisitando todos os estudos de caso realizados no capítulo anterior, a fim de comparar o desempenho da ferramenta proposta sem e com o classificador. As análises comprovaram não somente a efetividade do uso do classificador, mas também uma relação de complementariedade entre o impacto da simulação-IS e do classificador no desempenho final. Em resumo, o aproveitamento do ganho do classificador é inversamente proporcional à eficiência da amostragem por importância, ainda que a qualidade das ferramentas seja diretamente proporcional. Essa complementariedade é importante, pois indica que o classificador tem sua contribuição tão mais bem aproveitada quanto mais a simulação-IS precisa.

A Seção 4.5 traz a experiência realizada com um classificador, em tese, mais potente. Para tanto, recorreu-se ao método GMDH que é uma rede neural artificial multicamadas e, portanto, de maior capacidade em capturar não-linearidades de um problema. Os resultados foram, em geral, tão bons quanto os obtidos com regressão logística, mas não foram superiores como se esperava, o que reforça o quão exitosa foi a adaptação do classificador logístico binário, o qual corresponde a uma rede básica, de apenas um neurônio.

5. Conclusões e Propostas de Trabalhos Futuros

5.1. Conclusões

Esta tese de doutorado apresenta uma nova abordagem para o emprego do método da entropia cruzada (CE – *cross-entropy*) no problema de confiabilidade de sistemas compostos de geração e transmissão (GT). Também é proposto acoplar um classificador de estados de uma maneira tal que o seu treinamento possa ser aprimorado pelo uso da inteligência obtida na otimização do método CE. A seguir, faz-se um pequeno resumo dos capítulos da tese, para então se estabelecer de forma mais explícita as principais contribuições alcançadas pelo trabalho.

O Capítulo 1 aponta como as inúmeras adaptações do método CE já existentes ainda encontram dificuldade em modelar corretamente as restrições físicas da rede de transmissão, sendo esse um dos principais pontos endereçados pelo método proposto nesta tese. O principal ponto se dá pela ausência na literatura de trabalhos que propõem empregar inteligência produzida por uma técnica de redução de variância no treinamento de um classificador de estados. Ao final do capítulo, é introduzida a proposta de discutir esses pontos em um contexto de sistemas interligados modernos e de falha rara. No Capítulo 2, a modelagem do problema objetivo é estabelecida, sendo a ferramenta simulação Monte Carlo quasi-sequencial a escolhida para lidar com sistemas GT com elevada participação de recursos renováveis. Adicionalmente, destaca-se que a falha de suprimento é apropriadamente desacoplada em: insuficiência de geração (falha G) e insuficiência de transmissão (falha T). Essa definição é fundamental para a correta interpretação do problema de confiabilidade composta e a subsequente adaptação do método CE e do classificador de estados. No Capítulo 3, o método CE é descrito em detalhe, seguido da principal aplicação em confiabilidade de geração e composta com o intuito de evidenciar a limitação ainda existente na abordagem para a confiabilidade de sistemas GT. Embasado nessa observação, a nova metodologia é apresentada e seu algoritmo descrito. A ferramenta proposta é apreciada por meio de um sistema teste com características

semelhantes às de sistemas modernos interligados. Já no Capítulo 4, a regressão logística binária (BLR – *binary logistic regression*) é descrita em detalhe, de modo a introduzir o classificador de estados escolhido para integrar o método proposto nesta tese. Os mesmos testes realizados no capítulo anterior são reavaliados com o método completo e uma experiência com a rede neural artificial *group method of data handling* (GMDH) no lugar da rede BLR é apresentada.

A análise de adequação de sistemas GT demanda a solução de fluxo de potência e, caso o suprimento total não possa ser verificado, a otimização de medidas corretivas. Tanto o fluxo de potência convencional quanto o ótimo (FPO) não fornecem uma informação qualitativa para estados de sucesso. Como o procedimento de níveis múltiplos do método CE necessita de tal informação, optou-se, nos trabalhos existentes, por utilizar a capacidade de geração disponível como variável explicativa. Isso é problemático, pois sistemas interligados modernos de grande porte geralmente apresentam falha composta basicamente igual à falha T (a falha G é desprezível), e a falha T não é explicada pela capacidade de geração disponível. Alternativamente, a primeira parte do método proposto nesta tese, cuja ferramenta computacional é chamada de SMCIS-GT, estima a distribuição amostragem por importância (pdf-IS) através de uma única simulação e em função de estados de falha apenas. Tal simulação é controlada pelas incertezas das estimativas de probabilidade e de frequência da falha, as quais devem ser maiores que as exigidas para a estimativa definitiva dos índices.

Com base na discussão apresentada no parágrafo anterior, é possível apontar quatro contribuições da primeira parte do método proposto, descrita no Capítulo 3. São elas: (i) emprego efetivo do método CE na avaliação de sistemas GT; (ii) potencial de ganho computacional previamente dimensionável; (iii) procedimento de estimação da pdf-IS com critério de parada automático; e (iv) menos parâmetros para definição do usuário. Outra contribuição, que, diferentemente das demais, não é explicitada no algoritmo do método, é a flexibilidade no foco de análise, seja ele para sistema, área/zona ou barra. Essa contribuição é justificada pelo fato de que, apesar da falha localizada ser justamente a mais rara e, portanto, a mais apropriada para aplicação do método CE, os trabalhos encontrados na literatura avaliam a falha do sistema. Isso ocorre pela dificuldade de se adaptar a função performance dada pela capacidade de geração disponível para um determinado grupo de barras.

Afinal, seria necessário definir quais estações geradoras explicam a falha de interesse e devem ser consideradas, uma tarefa muito pouco clara.

Um aspecto importante a destacar e que impacta a eficiência da ferramenta, mas não está sob controle do usuário, é a dispersão da falha. A conectividade da malha que envolve um dado contingente de consumidores tem relação direta com a quantidade e a “distância elétrica” dos componentes do sistema que impactam o suprimento da região em análise. Quanto maior a conectividade, mais confiável deve ser o suprimento, mas isso vem com maior dispersão no espectro de estados de falha que são relevantes para a composição dos índices de confiabilidade. Uma falha mais dispersa induz uma amostragem por importância mais diluída e, por consequência, uma SMCIS-GT menos eficiente. Apesar disso, deve-se observar que a eficácia da ferramenta é mantida. Em verdade, a principal conclusão que pode ser feita é que, independentemente das circunstâncias, a ferramenta SMCIS-GT nunca apresenta prejuízo quando a falha é não rara e sempre apresenta ganho quando a falha é rara.

No âmbito do uso de classificadores de estados para reduzir execuções de FPO, têm-se dificuldades em duas frentes. A coleta da amostra de treinamento é muitas vezes submetida a um contexto de evento raro, o que torna esse processo muito oneroso computacionalmente. A outra questão é que as variáveis explicativas do evento de interesse, bem como ocorre com as técnicas de redução de variância, são muitas vezes submetidas a uma incorreta interpretação da falha de sistemas GT e focam nas informações do sistema de geração, pouco explicativas dos problemas ocasionados por inflexibilidade operativa da transmissão, ou falha T, que dominam a composição dos índices de confiabilidade. Foi proposto, então, que um classificador tenha sua amostra de treinamento coletada em uma simulação com amostragem por importância (simulação-IS) e que novas variáveis explicativas da falha de sistemas GT sejam produzidas a partir da pdf-IS. Logo, a segunda parte do método proposto traz duas contribuições: (i) acelerar a coleta da amostra de treinamento; e (ii) produzir variáveis explicativas que aumentam a qualidade do classificador.

Considerando o método completo, destaca-se o aspecto de complementariedade entre o aproveitamento do classificador e o êxito da simulação-IS. Um melhor desempenho do método CE produz uma pdf-IS de maior qualidade, que, por sua vez, produz variáveis explicativas melhores e um classificador de mais qualidade. Por outro lado, o classificador atua em uma parte da simulação já acelerada pela

amostragem por importância, o que reduz seu impacto global. Observe que o classificador ser muito pouco aproveitado em uma análise para a qual o custo da simulação-IS é desprezível, onde ganhos maiores que 10 pu já são observados sem o classificador, não representa nenhum ônus relevante. Contrariamente, o classificador ser mais aproveitado quando a simulação-IS fica aquém do seu potencial, onde ganhos chegam a ficar abaixo da metade do potencial estimado de 16 pu, é muito bem-vindo e representa um bônus relevante.

Por fim, chama-se a atenção ao estudo feito com uma rede neural artificial multicamadas, a GMDH, no lugar da rede BLR. A rede GMDH apresentou desempenho, em geral, pouco inferior ao da rede BLR. Isso é interessante, pois aponta que um classificador simples e barato computacionalmente como o BLR é suficiente para incrementar a velocidade da avaliação de confiabilidade composta de sistemas GT via simulação Monte Carlo.

5.2. Propostas de Trabalhos Futuros

Esta tese de doutorado se insere no contexto de otimização de algoritmos de avaliação de confiabilidade composta. Visando o futuro da pesquisa nessa área nota-se, primeiro, a necessidade de apreciar a metodologia proposta em sistemas modernos reais de grande porte, para que o verdadeiro escopo do impacto que a ferramenta pode apresentar possa ser materializado. Depois, pensa-se em como a ferramenta ainda pode ser aprimorada. Por mais que a tentativa com redes neurais multicamadas não tenha se provado superior à rede BLR, entende-se que outras redes *deep learning* podem trazer o impulso esperado. Redes mais convencionais como a própria MLP (*multilayer perceptron*), ou outras do estado da arte como as redes convolucionais, podem ser implementadas.

Visando objetivos diferentes, e até mais ambiciosos, tem-se que o objeto de classificação pode ser repensado como o próprio corte de carga de um estado e não apenas se ele é de falha ou não. Outra proposta é o uso de aprendizado não-supervisionado para estratificar a região de fronteira falha-sucesso e, com isso, obter mais precisão de classificação, bem como informações interessantes sobre a saúde de um sistema. Análise de bem-estar de sistemas já existe na literatura, mas com apenas três níveis: saudável, marginal e falha, ao passo que um classificador não-supervisionado pode ser adaptado para oferecer ainda mais níveis.

6. Referências Bibliográficas

1. SCHILLING, M. T.; LEITE DA SILVA, A. M. Conceptual Investigation on Probabilistic Adequacy Protocols: Brazilian Experience. **IEEE Transactions on Power Systems**, v. 29(3), p. 1270-1278, May 2014.
2. BILLINTON, R. **Bibliography on application of probability methods in the evaluation of generating capacity requirements**. Proc. IEEE Winter Power Meeting. New York, USA. 1966. p. 62-66.
3. BILLINTON, R. Bibliography on the Application of Probability Methods In Power System Reliability Evaluation. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-91(2), p. 649-660, March 1972.
4. POWER SYSTEM ENGINEERING COMMITTEE. Bibliography on the Application of Probability Methods in Power System Reliability Evaluation 1971-1977. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-97(6), p. 2235-2242, November 1978.
5. ALLAN, R. N.; BILLINTON, R.; LEE, S. H. Bibliography on the Application of Probability Methods in Power System Reliability Evaluation 1977-1982. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-103(2), p. 275-282, February 1984.
6. ALLAN, R. N.; BILLINTON, R.; SHAHIDEHPOUR, S. M.; SINGH, C. Bibliography on the application of probability methods in power system reliability evaluation: 1982-1987. **IEEE Transactions on Power Systems**, v. 3(4), p. 1555-1564, November 1988.
7. SCHILLING, M. T.; BILLINTON, R.; LEITE DA SILVA, A. M.; EL-KADY, M. A. Bibliography on composite system reliability. **IEEE Transactions on Power Systems**, v. 4(3), p. 1122-1132, August 1989.
8. ALLAN, R. N.; BILLINTON, R.; BREIPOHL, A. M.; GRIGG, C. H. Bibliography on the application of probability methods in power system reliability evaluation: 1987-1991. **IEEE Transactions on Power Systems**, v. 9(1), p. 41-49, February 1994.
9. ALLAN, R. N.; BILLINTON, R.; BREIPOHL, A. M.; GRIGG, C. H. Bibliography on the application of probability methods in power system reliability evaluation. **IEEE Transactions on Power Systems**, v. 14(1), p. 51-57, February 1999.

10. BILLINTON, R.; FOTUHI-FIRUZABAD, M.; BERTLING, L. Bibliography on the application of probability methods in power system reliability evaluation 1996-1999. **IEEE Transactions on Power Systems**, v. 16(4), p. 595-602, November 2001.
11. PAPIC, M.; LOGAN, D. **Bibliography on Composite System Reliability Assessment 2000-2020**. 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Manchester, UK. 2022. p. 1-8.
12. BILLINTON, R.; ALLAN, R. N. **Reliability Evaluation of Power Systems**. 2nd. ed. New York, NY, USA: Plenum Press, 1996.
13. HALL, J. D.; RINGLEE, R. J.; WOOD, A. J. Frequency and Duration Methods for Power System Reliability Calculations: I - Generation System Model. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-87(9), p. 1787-1796, September 1968.
14. LEITE DA SILVA, A. M.; MELO, A. C. G.; CUNHA, S. H. F. Frequency and duration method for reliability evaluation of large-scale hydrothermal generating systems. **IEE Proceedings on Generation, Transmission and Distribution - Part C**, v. 138(1), p. 94-102, January 1991.
15. BILLINTON, R.; BOLLINGER, K. E. Transmission System Reliability Evaluation Using Markov Processes. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-87(2), p. 538-547, February 1968.
16. GAVER, D. P.; MONTMEAT, F. E.; PATTON, A. D. Power System Reliability I-Measures of Reliability and Methods of Calculation. **IEEE Transactions on Power Apparatus and Systems**, v. 83(7), p. 727-737, July 1964.
17. PEREIRA, M. V. F.; BALU, N. J. Composite generation/transmission reliability evaluation. **Proceedings of the IEEE**, v. 80(4), p. 470-491, April 1992.
18. MELLO, J. C. O.; LEITE DA SILVA, A. M.; PEREIRA, M. V. F. Efficient loss-of-load cost evaluation by combined pseudo-sequential and state transition simulation. **IEE Proceedings on Generation, Transmission and Distribution**, v. 144(2), p. 147-154, March 1997.
19. 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(1), p. 73-80, February 2000.
20. LEITE DA SILVA, A. M.; GONZÁLEZ-FERNÁNDEZ, R. A.; SALES, W. S.; MANSO, L. A. F. **Reliability assessment of time-dependent systems via quasi-sequential Monte Carlo simulation**. 2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems. Singapore. 2010. p. 697-702.

21. MELO, A. C. G.; PEREIRA, M. V. F.; LEITE DA SILVA, A. M. Frequency and duration calculations in composite generation and transmission reliability evaluation. **IEEE Transactions on Power Systems**, v. 7(2), p. 469-476, May 1992.
22. MELO, A. C. G.; PEREIRA, M. V. F.; LEITE DA SILVA, A. M. A conditional probability approach to the calculation of frequency and duration indices in composite reliability evaluation. **IEEE Transactions on Power Systems**, v. 8(3), p. 1118-1125, August 1993.
23. MANSO, L. A. F.; LEITE DA SILVA, A. M. Modelagem de cargas variantes no tempo na avaliação de confiabilidade composta via simulação Monte Carlo não-seqüencial. **Revista SBA: CONTROLE & AUTOMAÇÃO**, v. 15(1), p. 93-100, March 2004.
24. LEITE DA SILVA, A. M.; RESENDE, L. C.; MANSO, L. A. F.; BILLINTON, R. Well-being analysis for composite generation and transmission systems. **IEEE Transactions on Power Systems**, v. 19(4), p. 1763-1770, November 2004.
25. BORGES, C. L. T.; FALCÃO, D. M.; MELLO, J. C. O.; MELO, A. C. G. Composite reliability evaluation by sequential Monte Carlo simulation on parallel and distributed processing environments. **IEEE Transactions on Power Systems**, v. 16(2), p. 203-209, May 2001.
26. AKHAVEIN, A.; FOTUHI-FIRUZABAD, M.; BILLINTON, R.; FAROKHZAD, D. Adequacy equivalent development of composite generation and transmission systems using network screening. **IET Generation, Transmission & Distribution**, v. 5(11), p. 1141-1148, November 2011.
27. YONG, P.; ZHANG, N.; KANG, C.; XIA, Q.; LU, D. MPLP-Based Fast Power System Reliability Evaluation Using Transmission Line Status Dictionary. **IEEE Transactions on Power Systems**, v. 34(2), p. 1630-1640, March 2019.
28. SAMAAAN, N.; SINGH, C. **A new method for composite system annualized reliability indices based on genetic algorithms**. IEEE Power Engineering Society Summer Meeting. Chicago, USA. 2002. p. 850-855.
29. MIRANDA, V.; DE MAGALHAES CARVALHO, L.; DA ROSA, M. A.; LEITE DA SILVA, A. M.; SINGH, C. Improving Power System Reliability Calculation Efficiency With EPSO Variants. **IEEE Transactions on Power Systems**, v. 24(4), p. 1772-1779, November 2009.
30. AMARASINGHE, P. A. G. M.; ABEYGUNAWARDANE, S. K.; SINGH, C. Adequacy Evaluation of Composite Power Systems Using an Evolutionary Swarm Algorithm. **IEEE Access**, v. 10, p. 19732-19741, February 2022.

31. LEITE DA SILVA, A. M.; RESENDE, L. C.; MANSO, L. A. F.; MIRANDA, V. Composite Reliability Assessment Based on Monte Carlo Simulation and Artificial Neural Networks. **IEEE Transactions on Power Systems**, v. 22(3), p. 1202-1209, August 2007.
32. LEITE DA SILVA, A. M.; RESENDE, L. C.; MANSO, L. A. F.; MIRANDA, V. Well-being Analysis for Composite Generation and Transmission Systems Based on Pattern Recognition Techniques. **IET Generation, Transmission & Distribution**, v. 2(2), p. 202-208, March 2008.
33. URGUN, D.; SINGH, C. **Composite System Reliability Analysis using Deep Learning enhanced by Transfer Learning**. 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Liege, Belgium. 2020. p. 1-6.
34. KAMRUZZAMAN, M.; BHUSAL, N.; BENIDRIS, M. A convolutional neural network-based approach to composite power system reliability evaluation. **International Journal of Electrical Power & Energy Systems**, v. 135(1), p. 1-10, February 2022.
35. ZHU, Y.; SINGH, C. Assessing Bulk Power System Reliability by End-to-End Line Maintenance-Aware Learning. **IEEE Access**, v. 11(1), p. 49639-49649, May 2023.
36. ZHANG, Y.; KARVE, P. M.; MAHADEVAN, S. **Power grid operational risk assessment using graph neural network surrogates**. 2024 IEEE Power & Energy Society General Meeting (PESGM). Seattle, USA. 2024. p. 1-5.
37. LUO, X.; SINGH, C.; PATTON, A. D. Power system reliability evaluation using learning vector quantization and Monte Carlo simulation. **Electric Power Systems Research**, v. 66(2), p. 163-169, August 2003.
38. URGUN, D.; SINGH, C. **Multi Label RBF Classification Method for Composite System Reliability Evaluation**. 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Boise, USA. 2018. p. 1-5.
39. GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA, USA: Press, MIT, 2016.
40. ROCCO, C. M.; MORENO, J. A. Fast Monte Carlo reliability evaluation using support vector machine. **Reliability Engineering & System Safety**, v. 76(3), p. 237-243, June 2002.
41. PINDORIYA, N. M.; JIRUTITIJAROEN, P.; SRINIVASAN, D.; SINGH, C. Composite Reliability Evaluation Using Monte Carlo Simulation and Least Squares Support Vector Classifier. **IEEE Transactions on Power Systems**, v. 26(4), p. 2483-2490, November 2011.

42. RESENDE, L. C.; MANSO, L. A. F.; DUTRA, W. D.; LEITE DA SILVA, A. M. **Support Vector Machine application in composite reliability assessment**. 2015 18th International Conference on Intelligent System Application to Power Systems (ISAP). Porto, Portugal. 2015. p. 1-6.
43. ZHU, Y.; SINGH, C. **End-to-End Topology-Aware Machine Learning for Power System Reliability Assessment**. 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Manchester, UK. 2022. p. 1-6.
44. URGUN, D.; SINGH, C. A Hybrid Monte Carlo Simulation and Multi Label Classification Method for Composite System Reliability Evaluation. **IEEE Transactions on Power Systems**, v. 34(2), p. 908-917, March 2019.
45. CAMPOS, F. S. V.; ASSIS, F. A.; LEITE DA SILVA, A. M.; COELHO, A. J. C.; MOURA, R. A. R.; SCHROEDER, M. A. O. Reliability evaluation of composite generation and transmission systems via binary logistic regression and parallel processing. **International Journal of Electrical Power & Energy Systems**, v. 142-B, p. 1-11, November 2022.
46. LUO, X.; SINGH, C.; PATTON, A. D. **Power system reliability evaluation using self organizing map**. 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.00CH37077). Singapore. 2000. p. 1103-1108.
47. SONG, Y.; BU, G.; ZHANG, R. **A fast method probabilistic reliability assessment of bulk power systems using FSOM neural network as system states filters**. 2005 IEEE/PES Transmission & Distribution Conference & Exposition: Asia and Pacific. Dalian, China. 2005. p. 1-6.
48. SINGH, C.; LUO, X.; KIM, H. **Power System Adequacy and Security Calculations Using Monte Carlo Simulation incorporating Intelligent System Methodology**. 2006 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Stockholm, Sweden. 2006. p. 1-9.
49. FREITAG, S. C.; COSTA, P. F. S.; SPERANDIO, M. Composite reliability analysis of electric power systems considering climatic and spatial conditions using self-organizing map. **Sustainable Energy, Grids and Networks**, v. 39(1), p. 1-11, September 2024.
50. ASSIS, F. A.; COELHO, A. J. C.; REZENDE, L. D.; LEITE DA SILVA, A. M.; RESENDE, L. C. Unsupervised machine learning techniques applied to composite reliability assessment of power systems. **International Transactions on Electrical Energy Systems**, v. 31(11), p. 1-18, November 2021.
51. BILLINTON, R.; JONNAVITHULA, A. **Variance reduction techniques for use with sequential Monte Carlo simulation in bulk power system**

- reliability evaluation.** Proceedings of 1996 Canadian Conference on Electrical and Computer Engineering. Calgary, Canada. 1996. p. 416-419.
52. BILLINTON, R.; JONNAVITHULA, A. Composite system adequacy assessment using sequential Monte Carlo simulation with variance reduction techniques. **IEE Proceedings on Generation, Transmission and Distribution**, v. 144(1), p. 1-6, January 1997.
53. PINTO, L. M. V. G.; PEREIRA, M. V. F. **A variance reduction technique to the reliability analysis of a generation/transmission system.** 1991 IEEE International Symposium on Circuits and Systems (ISCAS). Singapore. 1991. p. 892-895.
54. PEREIRA, M. V. F.; PINTO, L. M. V. G. A new computational tool for composite reliability evaluation. **IEEE Transactions on Power Systems**, v. 7(1), p. 258-264, February 1992.
55. TANG, Y.; LI, W.; ZHAO, Y. **A cross-entropy-based control variate method for power system reliability assessment.** 2017 IEEE Power & Energy Society General Meeting. Chicago, USA. 2017. p. 1-5.
56. DUCHESNE, L.; KARANGELOS, E.; WEHENKEL, L. **Using Machine Learning to Enable Probabilistic Reliability Assessment in Operation Planning.** 2018 Power Systems Computation Conference (PSCC). Dublin, Ireland. 2018. p. 1-8.
57. SINGH, C.; MITRA, J. Composite system reliability evaluation using state space pruning. **IEEE Transactions on Power Systems**, v. 12(1), p. 471–479, February 1997.
58. GREEN, R. C.; WANG, L.; WANG, Z.; ALAM, M.; SINGH, C. **Power system reliability assessment using intelligent state space pruning techniques: A comparative study.** 2010 International Conference on Power System Technology. Zhejiang, China. 2010. p. 1-8.
59. JIRUTITIJAROEN, P.; SINGH, C. Comparison of Simulation Methods for Power System Reliability Indexes and Their Distributions. **IEEE Transactions on Power Systems**, v. 23(2), p. 486-493, May 2008.
60. SHU, Z.; JIRUTITIJAROEN, P. Latin Hypercube Sampling Techniques for Power Systems Reliability Analysis With Renewable Energy Sources. **IEEE Transactions on Power Systems**, v. 26(4), p. 2066-2073, November 2011.
61. SHU, Z.; JIRUTITIJAROEN, P.; LEITE DA SILVA, A. M.; SINGH, C. Accelerated state evaluation and latin hypercube sequential sampling for composite system reliability assessment. **IEEE Transactions on Power Systems**, v. 29(4), p. 1692–1700, July 2014.

62. GUO, J.; FENG, T.; CAI, Z.; YU, Z.; GU, Y.; QIAN, T.; HUANG, K.; TANG, W. **Security Risk Assessment of Power System Based on Latin Hypercube Sampling and Daily Peak Load Forecasting**. IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). Wuhan, China. 2020. p. 2787-2792.
63. CAI, J.; HAO, L.; XU, Q.; ZHANG, K. Reliability assessment of renewable energy integrated power systems with an extendable Latin hypercube importance sampling method. **Sustainable Energy Technologies and Assessments**, v. 50(1), p. 1-13, March 2022.
64. CERVAN, D.; CORONADO, A. M.; LUYO, J. E. Cluster-based stratified sampling for fast reliability evaluation of composite power systems based on sequential Monte Carlo simulation. **International Journal of Electrical Power & Energy Systems**, v. 147(1), p. 1-14, May 2023.
65. HUA, B.; BIE, Z.; AU, S.; LI, W.; WANG, X. Extracting Rare Failure Events in Composite System Reliability Evaluation Via Subset Simulation. **IEEE Transactions on Power Systems**, v. 30(2), p. 753-762, March 2015.
66. ZHAO, Y.; HAN, Y.; LIU, Y.; XIE, K.; LI, W.; YU, J. Cross-Entropy-Based Composite System Reliability Evaluation Using Subset Simulation and Minimum Computational Burden Criterion. **IEEE Transactions on Power Systems**, v. 36(6), p. 5198-5209, November 2021.
67. LIN, Z.; TAO, L.; WANG, S.; YONG, N.; XIA, D.; WANG, J.; GE, D. A subset simulation analysis framework for rapid reliability evaluation of series-parallel cold standby systems. **Reliability Engineering & System Safety**, v. 241(1), p. 1-16, January 2024.
68. RUBINSTEIN, R. Y.; KROESE, D. P. **The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning**. 1st. ed. New York, NY, USA: Springer Science+Business Media, 2004.
69. LIEBER, D.; NEMIROVSKII, A.; RUBINSTEIN, R. Y. A fast Monte Carlo method for evaluating reliability indexes. **IEEE Transactions on Reliability**, v. 48(3), p. 256-261, September 1999.
70. YAN, C.; DING, T.; BIE, Z.; WANG, X. A Geometric Programming to Importance Sampling for Power System Reliability Evaluation. **IEEE Transactions on Power Systems**, v. 32(2), p. 1568-1569, March 2017.
71. LUKASHEVICH, A.; MAXIMOV, Y. Power Grid Reliability Estimation via Adaptive Importance Sampling. **IEEE Control Systems Letters**, v. 6(1), p. 1010-1015, June 2022.
72. 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(1), p. 129-137, February 2010.
73. 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(4), p. 2381-2389, November 2011.
74. CARVALHO, L. D. M.; GONZÁLEZ-FERNÁNDEZ, R. A.; LEITE DA SILVA, A. M.; DA ROSA, M. A.; MIRANDA, V. Simplified Cross-Entropy Based Approach for Generating Capacity Reliability Assessment. **IEEE Transactions on Power Systems**, v. 28(2), p. 1609-1616, May 2013.
75. TÓMASSON, E.; SÖDER, L. Generation Adequacy Analysis of Multi-Area Power Systems With a High Share of Wind Power. **IEEE Transactions on Power Systems**, v. 33(4), p. 3854-3862, July 2018.
76. GONZÁLEZ-FERNÁNDEZ, R. A.; LEITE DA SILVA, A. M.; RESENDE, L. C.; SCHILLING, M. T. Composite Systems Reliability Evaluation Based on Monte Carlo Simulation and Cross-Entropy Methods. **IEEE Transactions on Power Systems**, v. 28(4), p. 4598-4606, November 2013.
77. WANG, Y.; GUO, C.; WU, Q. H. A Cross-Entropy-Based Three-Stage Sequential Importance Sampling for Composite Power System Short-Term Reliability Evaluation. **IEEE Transactions on Power Systems**, v. 28(4), p. 4254-4263, November 2013.
78. LEITE DA SILVA, A. M.; GONZÁLEZ-FERNÁNDEZ, R. A.; FLÁVIO, S. A.; MANSO, L. A. F. **Composite reliability evaluation with renewable sources based on quasi-sequential Monte Carlo and cross entropy methods**. 2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Durham, UK. 2014. p. 1-6.
79. YAN, C.; LUCA, L. G.; BIE, Z.; DING, T.; LI, G. **A three-stage CE-IS Monte Carlo algorithm for highly reliable composite system reliability evaluation based on screening method**. 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Beijing, China. 2016. p. 1-6.
80. TÓMASSON, E.; SÖDER, L. Improved Importance Sampling for Reliability Evaluation of Composite Power Systems. **IEEE Transactions on Power Systems**, v. 32(3), p. 2426-2434, May 2017.
81. LEITE DA SILVA, A. M.; DE CASTRO, A. M. Risk Assessment in Probabilistic Load Flow via Monte Carlo Simulation and Cross-Entropy Method. **IEEE Transactions on Power Systems**, v. 34(2), p. 1193-1202, March 2019.

82. GENG, L.; ZHAO, Y.; LI, W. Enhanced Cross Entropy Method for Composite Power System Reliability Evaluation. **IEEE Transactions on Power Systems**, v. 34(4), p. 3129-3139, July 2019.
83. CAI, J.; XU, Q.; CAO, M.; YANG, B. A novel importance sampling method of power system reliability assessment considering multi-state units and correlation between wind speed and load. **International Journal of Electrical Power & Energy Systems**, v. 109(1), p. 217-226, July 2019.
84. ZHAO, Y.; TANG, Y.; LI, W.; YU, J. Composite Power System Reliability Evaluation Based on Enhanced Sequential Cross-Entropy Monte Carlo Simulation. **IEEE Transactions on Power Systems**, v. 34(5), p. 3891-3901, September 2019.
85. CAI, J.; ZHANG, Z.; WANG, M.; GE, X. An improved sequential importance sampling method for reliability assessment of renewable power systems with energy storage. **Energy Reports**, v. 9-7, p. 1637-1646, September 2023.
86. ZHAO, Y.; HU, J.; XIE, K.; TANG, J. Improved cross-entropy-based power system reliability evaluation using GMM with highly-efficient closed-form parameter updating. **International Journal of Electrical Power & Energy Systems**, v. 166(1), p. 1-10, May 2025.
87. YU, X.; SINGH, C. **Expected power loss calculation including protection failures using importance sampling and SOM**. IEEE Power Engineering Society General Meeting. Denver, USA. 2004. p. 206-211.
88. BAKKIYARAJ, R. A.; KUMARAPPAN, N. **Reliability design of composite generation and transmission system based on Latin Hypercube Sampling with GRNN state adequacy evaluation**. IET Chennai Fourth International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2013). Chennai, India. 2013. p. 8-14.
89. URGUN, D.; SINGH, C. **Composite Power System Reliability Evaluation Using Importance Sampling and Convolutional Neural Networks**. 2019 20th International Conference on Intelligent System Application to Power Systems (ISAP). New Delhi, India. 2019. p. 1-6.
90. URGUN, D.; SINGH, C.; VITTAL, V. Importance Sampling Using Multilabel Radial Basis Classification for Composite Power System Reliability Evaluation. **IEEE Systems Journal**, v. 14(2), p. 2791-2800, June 2020.
91. ANSARI, O. A.; GONG, Y.; LIU, W.; CHUNG, C. Y. Data-driven Operation Risk Assessment of Wind-integrated Power Systems via Mixture Models and Importance Sampling. **Journal of Modern Power Systems and Clean Energy**, v. 8(3), p. 437-445, May 2020.

92. LEITE DA SILVA, A. M.; PEREIRA, M. V. F.; SCHILLING, M. T. **Power System Analysis under Uncertainties - Concepts and Techniques**. II Symposium of Specialists in Electric Operational and Expansion Planning (SEPOPE), Invited Paper, IP-21. São Paulo, Brasil. 1989. p. 1-12.
93. MANSO, B. A. S. **Avaliação da Confiabilidade de Sistemas de Geração e Transmissão Baseada em Técnicas de Enumeração de Estados e Amostragem por Importância**. Dissertação (Mestrado em Engenharia Elétrica) - Puc-Rio. Rio de Janeiro-RJ, Brasil, p. 134. 2021.
94. MANSO, B. A. S.; MANSO, L. A. F.; LEITE DA SILVA, A. M.; CASTRO, J. F. C. **Lista Ranqueada de Equipamentos via Amostragem por Importância para Estudos de Confiabilidade de Sistemas de Geração e Transmissão**. XV Simpósio Brasileiro de Automação Inteligente - SBAI. Rio Grande, RS, Brasil (Evento Virtual): Sociedade Brasileira de Automática (SBA). 2021. p. 1472-1477.
95. MANSO, B. A. S.; MANSO, L. A. F.; LEITE DA SILVA, A. M.; CASTRO, J. F. C. **Avaliação da Confiabilidade Composta via Enumeração de Estados Auxiliada por Amostragem por Importância**. XV Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica (SEPOPE). Foz do Iguaçu, PR, Brasil. 2022. p. 1-9.
96. GONZÁLEZ-FERNÁNDEZ, R. A.; LEITE DA SILVA, A. M. **Comparison between Different Cross-Entropy Based Methods Applied to Generating Capacity Reliability**. 2012 12th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Istanbul, Turkey. 2012. p. 10-14.
97. LEITE DA SILVA, A. M.; CASTRO, J. F. C.; BILLINTON, R. Probabilistic assessment of spinning reserve via cross-entropy method considering renewable sources and transmission restrictions. **IEEE Transactions on Power Systems**, v. 33(4), p. 4574-4582, July 2018.
98. SCHNEIDER, A. W.; RAKSANY, J.; GUNDERSON, R. O.; FONG, C. C.; BILLINGTON, R.; O'NEILL, P. M.; SILVERSTEIN, B. Bulk system reliability-measurement and indices. **IEEE Transactions on Power Systems**, v. 4(3), p. 829-835, August 1989.
99. BILLINTON, R.; KHAN, E. A security based approach to composite power system reliability evaluation. **IEEE Transactions on Power Systems**, v. 7(1), p. 65-72, February 1992.
100. BILLINTON, R.; LIAN, G. Composite power system health analysis using a security constrained adequacy evaluation procedure. **IEEE Transactions on Power Systems**, v. 9(2), p. 936-941, May 1994.

101. BILLINTON, R.; FOTUHI-FIRUZABAD, M. A basic framework for generating system operating health analysis. **IEEE Transactions on Power Systems**, v. 9(3), p. 1610-1617, August 1994.
102. MANSO, L. A. F.; LEITE DA SILVA, A. M.; MELLO, J. C. O. Comparison of alternative methods for evaluating loss of load costs in generation and transmission systems. **Electric Power Systems Research**, v. 50(2), p. 107-114, May 1999.
103. MELLO, J. C. O. **Determinação do Valor da Confiabilidade em Sistemas Compostos de Geração e Transmissão**. Tese (Doutorado em Engenharia Elétrica) - PUC-Rio. Rio de Janeiro-RJ, Brasil. 1994.
104. MANSO, L. A. F. **Um Ambiente para Planejamento de Sistemas de Geração e Transmissão Utilizando o Valor da Confiabilidade**. Tese (Doutorado em Engenharia Elétrica) - UNIFEI. Itajubá-MG, Brasil. 1999.
105. TINNEY, W. F.; HART, C. E. Power Flow Solution by Newton's Method. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-86(11), p. 1449-1460, November 1967.
106. MONTICELLI, A. J. **Fluxo de Carga em Redes de Energia Elétrica**. São Paulo, SP, Brasil: Edgard Blücher Ltda, 1983.
107. ASSIS, F. A.; LEITE DA SILVA, A. M.; MANSO, L. A. F.; CASTRO, J. F. C. Transmission expansion planning of large power networks via constructive metaheuristics with security constraints and load uncertainty analysis. **International Transactions on Electrical Energy Systems**, v. 31(1), p. 1-19, January 2021.
108. MANSO, B. A. S.; ASSIS, F. A.; LEITE DA SILVA, A. M.; MILHORANCE, A. **Otimização do Processo de Avaliação da Confiabilidade de Sistemas Compostos por Geração e Transmissão**. XXV Congresso Brasileiro de Automática (CBA). Rio de Janeiro, RJ, Brasil: Sociedade Brasileira de Automática (SBA). 2024. p. 1-7.
109. MANSO, B. A. S.; LEITE DA SILVA, A. M.; MILHORANCE, A.; ASSIS, F. A. **Confiabilidade de Sistemas Compostos via Simulação Monte Carlo Não Sequencial e Entropia Cruzada: Avaliação por Área e Barramento**. X Simpósio Brasileiro de Sistemas de Elétricos (SBSE). Manaus, AM, Brasil: Sociedade Brasileira de Automática (SBA). 2023. p. 816-822.
110. GNU. **GLPK (GNU Linear Programming Kit)**. Disponível em: <<https://www.gnu.org/software/glpk/>>. Acesso em: 25 Setembro 2023.
111. GRIGG, C.; WONG, P.; ALBRECHT, P.; ALLAN, R.; BHAVARAJU, M.; BILLINTON, R.; CHEN, Q.; FONG, C.; HADDAD, S.; KURUGANTY, S.; LI, W.; MUKERJI, R.; PATTON, D.; RAU, N.; REPPEN, D.; SCHNEIDER, A.; SHAHIDEHPOUR, M.; SINGH, C. The IEEE Reliability Test System-

1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee. **IEEE Transactions on Power Systems**, v. 14(3), p. 1010-1020, August 1999.
112. PROBABILITY METHODS SUBCOMMITTEE. IEEE Reliability Test System. **IEEE Transactions on Power Apparatus and Systems**, v. PAS-98(6), p. 2047-2054, November 1979.
113. RTS-96-REN-T_SYSTEM_DATA. **GitHub repository**, 2023. Disponível em: <https://github.com/BASM95/RTS-96-REN-T_system_data>. Acesso em: 25 Abril 2023.
114. MANSO, B. A. S.; LEITE DA SILVA, A. M.; MILHORANCE, A.; ASSIS, F. A. Composite reliability assessment of systems with grid-edge renewable resources via quasi-sequential Monte Carlo and cross-entropy techniques. **IET Generation, Transmission & Distribution**, v. 18(2), p. 326-336, January 2024.
115. HOSMER JR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. 3rd. ed. Hoboken, NJ, USA: John Wiley & Sons, 2013.
116. CHAMBERS, R. L.; STEEL, D. G.; WANG, S.; WELSH, A. **Maximum Likelihood Estimation for Sample Surveys**. 1st. ed. Boca Raton, FL, USA: CRC Press, 2012.
117. MANSO, B. A. S.; LEITE DA SILVA, A. M.; ASSIS, F. A.; MILHORANCE, A. **Binary Logistic Regression and Cross-Entropy Techniques Applied to Composite Reliability Evaluation of Power Systems**. 2024 22nd International Conference on Intelligent Systems Applications to Power Systems (ISAP). Budapest, Hungary. 2024. p. 1-6.
118. IVAKHNENKO, A. G. Polynomial Theory of Complex Systems. **IEEE Transactions on Systems, Man, and Cybernetics**, v. SMC-1(4), p. 364-378, October 1971.

Apêndice A. Dados do Sistema RTS-96-REN-T

A.1. Dados de Barra

Tabela A.1: Dados de Barra.

Barra	Carga [MW]	Cap. [MW]	Barra	Carga [MW]	Cap. [MW]	Barra	Carga [MW]	Cap. [MW]
101	97,20	837,50	201	97,20	492,50	301	97,20	752,50
102	87,30	208,00	202	87,30	200,00	302	87,30	232,00
103	162,00	—	203	162,00	—	303	162,00	—
104	66,60	—	204	66,60	—	304	66,60	—
105	63,90	—	205	63,90	—	305	63,90	—
106	122,40	—	206	122,40	—	306	122,40	—
107	112,50	300,00	207	112,50	300,00	307	112,50	300,00
108	153,90	—	208	153,90	—	308	153,90	—
109	157,50	—	209	157,50	—	309	157,50	—
110	175,50	—	210	175,50	—	310	175,50	—
111	—	—	211	—	—	311	—	—
112	—	—	212	—	—	312	—	—
113	238,50	636,00	213	238,50	759,00	313	238,50	1056,00
114	174,60	—	214	174,60	—	314	174,60	—
115	285,30	872,50	215	285,30	520,00	315	285,30	785,00
116	90,00	155,00	216	90,00	155,00	316	90,00	155,00
117	—	—	217	—	—	317	—	—
118	299,70	400,00	218	299,70	400,00	318	299,70	400,00
119	162,90	—	219	162,90	—	319	162,90	—
120	115,20	—	220	115,20	—	320	115,20	—
121	—	400,00	221	—	400,00	321	—	400,00
122	—	300,00	222	—	300,00	322	—	300,00
123	—	660,00	223	—	660,00	323	—	660,00
124	—	—	224	—	—	324	—	—
						325	—	—
						Total	7695	13.996

*Carga: valores de carga pico.
 **Cap.: capacidade de geração.

A.2.

Dados de Estações Geradoras

Tabela A.2: Dados de Estações Geradoras.

Estação	Barra	Tipo	Unidades		TF [oc./ano]	MTTR [h]
			nº	Cap. [MW]		
1	101	Térmica	2	20,00	19,4667	50,0000
2	101	Eólica	319	2,50	4,0000	90,0000
3	102	Térmica	2	20,00	19,4667	50,0000
4	102	Hidro	2	84,00	6,0444	12,6839
5	107	Térmica	3	100,00	7,3000	50,0000
6	113	Hidro	3	212,00	7,6212	24,3902
7	115	Térmica	5	12,00	2,9796	60,0000
8	115	Eólica	325	2,50	4,0000	90,0000
9	116	Térmica	1	155,00	9,1250	40,0000
10	118	Térmica	1	400,00	7,9636	150,0000
11	121	Térmica	1	400,00	7,9636	150,0000
12	122	Hidro	6	50,00	4,4242	20,0000
13	123	Térmica	2	155,00	9,1250	40,0000
14	123	Térmica	1	350,00	7,6174	100,0000
15	201	Térmica	2	20,00	19,4667	50,0000
16	201	Eólica	181	2,50	4,0000	90,0000
17	202	Térmica	2	20,00	19,4667	50,0000
18	202	Hidro	2	80,00	6,0444	12,6839
19	207	Térmica	3	100,00	7,3000	50,0000
20	213	Hidro	3	253,00	7,6212	24,3902
21	215	Térmica	5	12,00	2,9796	60,0000
22	215	Eólica	184	2,50	4,0000	90,0000
23	216	Térmica	1	155,00	9,1250	40,0000
24	218	Térmica	1	400,00	7,9636	150,0000
25	221	Térmica	1	400,00	7,9636	150,0000
26	222	Hidro	6	50,00	4,4242	20,0000
27	223	Térmica	2	155,00	9,1250	40,0000
28	223	Térmica	1	350,00	7,6174	100,0000
29	301	Térmica	2	20,00	19,4667	50,0000
30	301	Eólica	285	2,50	4,0000	90,0000
31	302	Térmica	2	20,00	19,4667	50,0000
32	302	Hidro	2	96,00	6,0444	12,6839
33	307	Térmica	3	100,00	7,3000	50,0000
34	313	Hidro	3	352,00	7,6212	24,3902
35	315	Térmica	5	12,00	2,9796	60,0000
36	315	Eólica	290	2,50	4,0000	90,0000
37	316	Térmica	1	155,00	9,1250	40,0000
38	318	Térmica	1	400,00	7,9636	150,0000
39	321	Térmica	1	400,00	7,9636	150,0000
40	322	Hidro	6	50,00	4,4242	20,0000
41	323	Térmica	2	155,00	9,1250	40,0000
42	323	Térmica	1	350,00	7,6174	100,0000

*TF: taxa de falha; MTTR: *mean time to failure*.

A.3. Dados de Circuito

Tabela A.3: Dados de Circuito.

Circuito	Barras Terminais		Reatância [%]	Cap. [MW]	TF [oc./ano]	MTTR [h]
	De	Para				
1	101	102	1,39	78,75	0,24	16
2	101	103	21,12	78,75	0,51	10
3	101	105	8,45	78,75	0,33	10
4	102	104	12,67	78,75	0,39	10
5	102	106	19,20	78,75	0,48	10
6	103	109	11,90	78,75	0,38	10
7	103	124	8,39	180,00	0,02	768
8	104	109	10,37	78,75	0,36	10
9	105	110	8,83	78,75	0,34	10
10	106	110	6,05	78,75	0,33	35
11	107	108	6,14	78,75	0,30	10
12	107	203	16,13	78,75	0,44	10
13	108	109	16,51	78,75	0,44	10
14	108	110	16,51	78,75	0,44	10
15	109	111	8,39	180,00	0,02	768
16	109	112	8,39	180,00	0,02	768
17	110	111	8,39	180,00	0,02	768
18	110	112	8,39	180,00	0,02	768
19	111	113	4,76	225,00	0,40	11
20	111	114	4,18	225,00	0,39	11
21	112	113	4,76	225,00	0,40	11
22	112	123	9,66	225,00	0,52	11
23	113	123	8,65	225,00	0,49	11
24	113	215	7,49	225,00	0,47	11
25	114	116	5,89	225,00	0,38	11
26	115	116	1,73	225,00	0,33	11
27	115	121	4,90	225,00	0,41	11
28	115	121	4,90	225,00	0,41	11
29	115	124	5,19	225,00	0,41	11
30	116	117	2,59	225,00	0,35	11
31	116	119	2,31	225,00	0,34	11
32	117	118	1,44	225,00	0,32	11
33	117	122	10,53	225,00	0,54	11
34	118	121	2,59	225,00	0,35	11
35	118	121	2,59	225,00	0,35	11
36	119	120	3,96	225,00	0,38	11
37	119	120	3,96	225,00	0,38	11
38	120	123	2,16	225,00	0,34	11
39	120	123	2,16	225,00	0,34	11
40	121	122	6,78	225,00	0,45	11
41	123	217	7,35	225,00	0,46	11
42	201	202	1,39	78,75	0,24	16

Circuito	Barras Terminais		Reatância [%]	Cap. [MW]	TF [oc./ano]	MTTR [h]
	De	Para				
43	201	203	21,12	78,75	0,51	10
44	201	205	8,45	78,75	0,33	10
45	202	204	12,67	78,75	0,39	10
46	202	206	19,20	78,75	0,48	10
47	203	209	11,90	78,75	0,38	10
48	203	224	8,39	180,00	0,02	768
49	204	209	10,37	78,75	0,36	10
50	205	210	8,83	78,75	0,34	10
51	206	210	6,05	78,75	0,33	35
52	207	208	6,14	78,75	0,30	10
53	208	209	16,51	78,75	0,44	10
54	208	210	16,51	78,75	0,44	10
55	209	211	8,39	180,00	0,02	768
56	209	212	8,39	180,00	0,02	768
57	210	211	8,39	180,00	0,02	768
58	210	212	8,39	180,00	0,02	768
59	211	213	4,76	225,00	0,40	11
60	211	214	4,18	225,00	0,39	11
61	212	213	4,76	225,00	0,40	11
62	212	223	9,66	225,00	0,52	11
63	213	223	8,65	225,00	0,49	11
64	214	216	5,89	225,00	0,38	11
65	215	216	1,73	225,00	0,33	11
66	215	221	4,90	225,00	0,41	11
67	215	221	4,90	225,00	0,41	11
68	215	224	5,19	225,00	0,41	11
69	216	217	2,59	225,00	0,35	11
70	216	219	2,31	225,00	0,34	11
71	217	218	1,44	225,00	0,32	11
72	217	222	10,53	225,00	0,54	11
73	218	221	2,59	225,00	0,35	11
74	218	221	2,59	225,00	0,35	11
75	219	220	3,96	225,00	0,38	11
76	219	220	3,96	225,00	0,38	11
77	220	223	2,16	225,00	0,34	11
78	220	223	2,16	225,00	0,34	11
79	221	222	6,78	225,00	0,45	11
80	301	302	1,39	78,75	0,24	16
81	301	303	21,12	78,75	0,51	10
82	301	305	8,45	78,75	0,33	10
83	302	304	12,67	78,75	0,39	10
84	302	306	19,20	78,75	0,48	10
85	303	309	11,90	78,75	0,38	10
86	303	324	8,39	180,00	0,02	768
87	304	309	10,37	78,75	0,36	10
88	305	310	8,83	78,75	0,34	10

Circuito	Barras Terminais		Reatância [%]	Cap. [MW]	TF [oc./ano]	MTTR [h]
	De	Para				
89	306	310	6,05	78,75	0,33	35
90	307	308	6,14	78,75	0,30	10
91	308	309	16,51	78,75	0,44	10
92	308	310	16,51	78,75	0,44	10
93	309	311	8,39	180,00	0,02	768
94	309	312	8,39	180,00	0,02	768
95	310	311	8,39	180,00	0,02	768
96	310	312	8,39	180,00	0,02	768
97	311	313	4,76	225,00	0,40	11
98	311	314	4,18	225,00	0,39	11
99	312	313	4,76	225,00	0,40	11
100	312	323	9,66	225,00	0,52	11
101	313	323	8,65	225,00	0,49	11
102	314	316	5,89	225,00	0,38	11
103	315	316	1,73	225,00	0,33	11
104	315	321	4,90	225,00	0,41	11
105	315	321	4,90	225,00	0,41	11
106	315	324	5,19	225,00	0,41	11
107	316	317	2,59	225,00	0,35	11
108	316	319	2,31	225,00	0,34	11
109	317	318	1,44	225,00	0,32	11
110	317	322	10,53	225,00	0,54	11
111	318	321	2,59	225,00	0,35	11
112	318	321	2,59	225,00	0,35	11
113	319	320	3,96	225,00	0,38	11
114	319	320	3,96	225,00	0,38	11
115	320	323	2,16	225,00	0,34	11
116	320	323	2,16	225,00	0,34	11
117	321	322	6,78	225,00	0,45	11
118	325	121	9,65	225,00	0,52	11
119	318	223	10,37	225,00	0,53	11
120	323	325	0,92	324,90	0,02	768

*TF: taxa de falha; MTTR: *mean time to failure*.

A.4. Séries Hídricas

Tabela A.4: Localização dos Reservatórios.

Reservatório	Barra
Hidro_1	102
Hidro_2	113
Hidro_3	122
Hidro_4	202
Hidro_5	213
Hidro_6	222
Hidro_7	302
Hidro_8	313
Hidro_9	322

Tabela A.5: Séries Hídricas (aproveitamento mensal em pu).

Ano 1									
Mês	Hidro_1	Hidro_2	Hidro_3	Hidro_4	Hidro_5	Hidro_6	Hidro_7	Hidro_8	Hidro_9
1	0,795803	0,939726	0,982956	0,961801	0,976606	0,552488	0,967561	0,792700	0,967769
2	0,774087	0,969270	0,968493	0,961106	0,980652	0,605069	0,972460	0,747095	0,967987
3	0,817404	0,973547	0,956949	0,960727	0,974308	0,588769	0,967063	0,773752	0,966417
4	0,855691	0,970077	0,968721	0,959844	0,967179	0,460736	0,964805	0,807616	0,963564
5	0,854634	0,965816	0,972692	0,950562	0,960710	0,467308	0,968572	0,790587	0,961850
6	0,853274	0,955441	0,975535	0,944627	0,950096	0,505166	0,970434	0,779090	0,963261
7	0,845782	0,941389	0,975535	0,947468	0,930923	0,472566	0,967455	0,789527	0,960937
8	0,823483	0,932228	0,969511	0,947910	0,911032	0,483608	0,964980	0,794342	0,955935
9	0,795889	0,921170	0,967112	0,944879	0,895647	0,545653	0,960645	0,798569	0,944729
10	0,770654	0,903684	0,968607	0,942480	0,880092	0,624786	0,952554	0,795936	0,936333
11	0,759598	0,892505	0,966762	0,947405	0,880816	0,592450	0,945058	0,786545	0,948945
12	0,765516	0,899593	0,957508	0,957065	0,904237	0,532508	0,948722	0,778089	0,965779
Ano 2									
Mês	Hidro_1	Hidro_2	Hidro_3	Hidro_4	Hidro_5	Hidro_6	Hidro_7	Hidro_8	Hidro_9
1	0,775145	0,912941	0,944641	0,961864	0,921061	0,557747	0,960421	0,766550	0,965830
2	0,783143	0,915751	0,944004	0,954477	0,925978	0,493861	0,964699	0,777522	0,963339
3	0,801634	0,909930	0,954037	0,945511	0,922598	0,412362	0,961931	0,794342	0,962592
4	0,815916	0,905368	0,967806	0,943427	0,916713	0,497016	0,960636	0,801229	0,959867
5	0,830597	0,905016	0,978381	0,945890	0,922543	0,483871	0,961889	0,805070	0,967844
6	0,851715	0,905874	0,976767	0,948289	0,934313	0,443384	0,962707	0,795517	0,976268
7	0,850597	0,898068	0,976461	0,944437	0,930211	0,529616	0,961679	0,785518	0,974144
8	0,830865	0,883284	0,978181	0,943112	0,911186	0,494387	0,958236	0,799776	0,958744
9	0,802521	0,869202	0,975535	0,940712	0,892557	0,428399	0,952676	0,805235	0,933521
10	0,779139	0,856632	0,960862	0,945069	0,888728	0,422878	0,944254	0,778306	0,923803
11	0,771856	0,873721	0,939571	0,957760	0,919407	0,495439	0,947747	0,747857	0,945944
12	0,773367	0,910021	0,937278	0,961738	0,954588	0,579305	0,965391	0,776079	0,970288

Ano 3									
Mês	Hidro_1	Hidro_2	Hidro_3	Hidro_4	Hidro_5	Hidro_6	Hidro_7	Hidro_8	Hidro_9
1	0,766476	0,935502	0,938659	0,961864	0,968388	0,546442	0,972946	0,792531	0,970866
2	0,765516	0,942185	0,944431	0,952519	0,961499	0,510161	0,968236	0,778437	0,958753
3	0,789258	0,932815	0,950554	0,943112	0,944559	0,470989	0,962667	0,785732	0,951438
4	0,828402	0,934813	0,962772	0,946900	0,952928	0,412362	0,964457	0,786331	0,965027
5	0,854566	0,939822	0,971719	0,949425	0,963476	0,532245	0,969991	0,782941	0,981955
6	0,853665	0,929000	0,971828	0,943617	0,954708	0,570366	0,970659	0,769723	0,978191
7	0,843556	0,914831	0,970844	0,936924	0,945958	0,500960	0,967412	0,774412	0,971663
8	0,821019	0,898749	0,968493	0,933830	0,929443	0,485974	0,963681	0,786929	0,967676
9	0,776948	0,878720	0,964010	0,932441	0,909785	0,463365	0,958139	0,781214	0,961016
10	0,738214	0,859131	0,954785	0,933136	0,893382	0,456792	0,950502	0,785775	0,953784
11	0,737604	0,840216	0,947946	0,935977	0,873234	0,508584	0,942570	0,786246	0,939671
12	0,749945	0,822684	0,950219	0,937555	0,866641	0,555906	0,934022	0,785518	0,928698
Ano 4									
Mês	Hidro_1	Hidro_2	Hidro_3	Hidro_4	Hidro_5	Hidro_6	Hidro_7	Hidro_8	Hidro_9
1	0,778223	0,898184	0,951377	0,925243	0,935471	0,642401	0,958409	0,761851	0,956971
2	0,765059	0,893988	0,947579	0,929474	0,916324	0,643190	0,962035	0,771145	0,932540
3	0,771465	0,886818	0,956240	0,931810	0,903512	0,672898	0,963957	0,755837	0,926849
4	0,795218	0,900746	0,969286	0,943490	0,933738	0,670269	0,967217	0,738904	0,953077
5	0,823979	0,916800	0,971719	0,942859	0,959739	0,633462	0,965447	0,752249	0,967272
6	0,835978	0,913864	0,975845	0,936419	0,954432	0,616374	0,960248	0,764208	0,958503
7	0,825257	0,904574	0,977679	0,944437	0,943984	0,627415	0,956571	0,769367	0,951013
8	0,792997	0,891942	0,971392	0,945637	0,937996	0,607435	0,952697	0,776342	0,943173
9	0,751571	0,879601	0,966292	0,943680	0,930872	0,551437	0,949106	0,775071	0,935161
10	0,719810	0,868709	0,952653	0,942101	0,908943	0,589821	0,944204	0,758375	0,930395
11	0,757185	0,883242	0,934880	0,949931	0,916587	0,667902	0,942828	0,778611	0,960124
12	0,786317	0,931492	0,964376	0,959844	0,960195	0,568525	0,960505	0,815870	0,981092
Ano 5									
Mês	Hidro_1	Hidro_2	Hidro_3	Hidro_4	Hidro_5	Hidro_6	Hidro_7	Hidro_8	Hidro_9
1	0,734428	0,871288	0,942661	0,941786	0,884971	0,658438	0,936372	0,783156	0,933130
2	0,742351	0,862861	0,952653	0,941217	0,880476	0,676315	0,931823	0,760574	0,923483
3	0,761019	0,860525	0,972045	0,940333	0,894878	0,655283	0,931047	0,771898	0,928472
4	0,804658	0,867770	0,979473	0,937934	0,915562	0,573784	0,933972	0,788336	0,941861
5	0,835167	0,875611	0,978181	0,934651	0,928695	0,555643	0,936593	0,772561	0,950814
6	0,828030	0,877023	0,976256	0,932946	0,931214	0,567211	0,937137	0,772428	0,950681
7	0,801160	0,869829	0,971174	0,930294	0,927345	0,600600	0,936601	0,767402	0,945796
8	0,771149	0,855414	0,964982	0,933641	0,923651	0,639509	0,935970	0,768073	0,941190
9	0,746985	0,845578	0,958878	0,936166	0,918284	0,642138	0,935469	0,735880	0,934322
10	0,734917	0,843898	0,958607	0,938250	0,927283	0,609275	0,935183	0,688452	0,942498
11	0,755721	0,843868	0,953886	0,942228	0,939892	0,598496	0,935807	0,691824	0,956736
12	0,775943	0,844293	0,944431	0,942543	0,941509	0,642664	0,936464	0,696468	0,960142

Apêndice B.

Exemplo de Rede BLR – Sistema RTS-96-REN-T

A análise da falha do sistema teste RTS-96-REN-T (Subseção 4.4.1) com a ferramenta SMCIS-BLR (Seção 4.3) foi escolhida para exemplificar o processo de construção das variáveis explicativas do classificador BLR, que são derivadas dos valores de indisponibilidade-IS dos equipamentos do sistema. A Tabela B.1 apresenta a lista ranqueada de equipamentos com base na indisponibilidade-IS (apenas equipamentos com valores não nulos são disponibilizados).

Tabela B.1: Lista ranqueada para o sistema RTS-96-REN-T.

nº	Comp.	Barra	Barras Terminais		Indisponibilidade		
			De	Para	IS	Original	Distorção
1	G-38	318	—	—	0,1676	0,1200	39,64%
2	G-3	102	—	—	0,1486	0,1000	48,65%
3	G-10	118	—	—	0,1459	0,1200	21,62%
4	G-1	101	—	—	0,1459	0,1000	45,95%
5	G-39	321	—	—	0,1459	0,1200	21,62%
6	G-11	121	—	—	0,1297	0,1200	8,11%
7	G-24	218	—	—	0,1243	0,1200	3,60%
8	G-25	221	—	—	0,1189	0,1200	-0,90%
9	G-29	301	—	—	0,1081	0,1000	8,11%
10	G-42	1323	—	—	0,1081	0,0800	35,14%
11	G-28	1223	—	—	0,1027	0,0800	28,38%
12	G-17	202	—	—	0,0946	0,1000	-5,41%
13	G-15	201	—	—	0,0919	0,1000	-8,11%
14	G-31	302	—	—	0,0919	0,1000	-8,11%
15	C-89	—	306	310	0,0865	0,0013	6468,14%
16	G-14	1123	—	—	0,0703	0,0800	-12,16%
17	C-51	—	206	210	0,0703	0,0013	5236,62%
18	G-13	123	—	—	0,0649	0,0400	62,16%
19	G-19	207	—	—	0,0631	0,0400	57,66%
20	G-27	223	—	—	0,0595	0,0400	48,65%
21	G-23	216	—	—	0,0541	0,0400	35,14%
22	C-10	—	106	110	0,0486	0,0013	3594,58%
23	G-9	116	—	—	0,0432	0,0400	8,11%
24	G-37	316	—	—	0,0432	0,0400	8,11%
25	G-41	323	—	—	0,0405	0,0400	1,35%
26	G-22	1215	—	—	0,0404	0,0395	2,26%
27	G-8	1115	—	—	0,0403	0,0395	2,09%

nº	Comp.	Barra	Barras Terminais		Indisponibilidade		
			De	Para	IS	Original	Distorção
28	G-16	1201	—	—	0,0402	0,0395	1,83%
29	G-30	1301	—	—	0,0397	0,0395	0,47%
30	G-5	107	—	—	0,0396	0,0400	-0,90%
31	G-2	1101	—	—	0,0394	0,0395	-0,28%
32	G-36	1315	—	—	0,0392	0,0395	-0,79%
33	G-33	307	—	—	0,0360	0,0400	-9,91%
34	G-4	1102	—	—	0,0351	0,0087	304,97%
35	C-7	—	103	124	0,0324	0,0018	1752,91%
36	G-20	213	—	—	0,0306	0,0208	47,41%
37	G-32	1302	—	—	0,0270	0,0087	211,52%
38	C-5	—	102	106	0,0270	0,0005	4835,13%
39	C-46	—	202	206	0,0270	0,0005	4835,13%
40	G-21	215	—	—	0,0238	0,0200	18,92%
41	G-7	115	—	—	0,0216	0,0200	8,11%
42	G-35	315	—	—	0,0216	0,0200	8,11%
43	C-17	—	110	111	0,0216	0,0018	1135,27%
44	C-86	—	303	324	0,0216	0,0018	1135,27%
45	G-34	313	—	—	0,0198	0,0208	-4,61%
46	G-18	1202	—	—	0,0162	0,0087	86,91%
47	C-96	—	310	312	0,0162	0,0018	826,45%
48	C-99	—	312	313	0,0162	0,0005	3130,12%
49	G-6	113	—	—	0,0162	0,0208	-21,96%
50	C-48	—	203	224	0,0108	0,0018	517,64%
51	C-57	—	210	211	0,0108	0,0018	517,64%
52	C-106	—	315	324	0,0108	0,0005	2000,92%
53	G-12	122	—	—	0,0099	0,0100	-0,90%
54	G-26	222	—	—	0,0099	0,0100	-0,90%
55	G-40	322	—	—	0,0090	0,0100	-9,91%
56	C-15	—	109	111	0,0054	0,0018	208,82%
57	C-19	—	111	113	0,0054	0,0005	976,71%
58	C-25	—	114	116	0,0054	0,0005	1033,35%
59	C-39	—	120	123	0,0054	0,0004	1166,62%
60	C-42	—	201	202	0,0054	0,0004	1133,65%
61	C-50	—	205	210	0,0054	0,0004	1293,23%
62	C-55	—	209	211	0,0054	0,0018	208,82%
63	C-65	—	215	216	0,0054	0,0004	1204,99%
64	C-74	—	218	221	0,0054	0,0004	1130,45%
65	C-82	—	301	305	0,0054	0,0004	1335,43%
66	C-84	—	302	306	0,0054	0,0005	887,03%
67	C-2	—	101	103	0,0054	0,0006	829,00%
68	C-95	—	310	311	0,0054	0,0018	208,82%
69	C-3	—	101	105	0,0054	0,0004	1335,43%
70	C-102	—	314	316	0,0054	0,0005	1033,35%
71	C-116	—	320	323	0,0054	0,0004	1166,62%
72	C-118	—	325	121	0,0054	0,0007	728,36%

Conforme descrito na Subseção 4.3.1, 100 clusters são gerados como percentis da maior indisponibilidade-IS (0,1676 pu, da estação geradora 38, barra 318). Cada grupo de equipamentos, cuja indisponibilidade-IS se enquadra dentro de algum percentil, forma uma variável explicativa. Como estações geradoras e circuitos são considerados separadamente, podem-se formar um máximo de 200 variáveis. Contudo, muitos percentis ficam vazios, de modo que o número de variáveis explicativas geradas é muito menor que 200 (em geral, esse número é menor do que 50). Adicionalmente, define-se uma variável para representar a carga e uma para cada zona de cada tipo de fonte renovável (são 9 zonas para fontes hídricas e 3 para eólicas). Na Tabela B.2, pode-se observar a relação de variáveis explicativas geradas e os valores ajustados para os respectivos coeficientes. Ao se amostrar um estado para o sistema, cada variável recebe um valor de disponibilidade em pu. Para as variáveis com estações geradoras, é dada a razão entre número de geradores disponíveis e número de geradores existentes, enquanto, para circuitos, tem-se apenas 0 (circuito indisponível) ou 1 (circuito disponível). Já para as variáveis referentes a fontes renováveis, é fornecido o valor de disponibilidade energética amostrado, enquanto, para a carga, tem-se um fator, dado em relação ao valor de pico.

Tabela B.2: Variáveis explicativas geradas para o sistema RTS-96-REN-T.

nº	Composição	Coeficiente
1	G-38	-0,2252
2	G-3	-1,3542
3	G-10; G-1; G-39	-2,5733
4	G-11	0,2445
5	G-24	0,9377
6	G-25	0,5832
7	G-29; G-42	-0,2489
8	G-28	-0,1128
9	G-17	1,0795
10	G-15; G-31	0,9549
11	G-14	0,6139
12	G-13	-0,5777
13	G-19	-1,3209
14	G-27	0,1106
15	G-23	-0,6881
16	G-9; G-37	-0,3188
17	G-41; G-22; G-8	3,664
18	G-16; G-30; G-5; G-2; G-36	3,9886
19	G-33	-0,7802
20	G-4	-5,035
21	G-20	0,3241

nº	Composição	Coefficiente
22	G-32	-3,0338
23	G-21	2,1876
24	G-7; G-35	1,7548
25	G-34	1,4949
26	G-18; G-6	1,8257
27	G-12; G-26; G-40	4,3257
28	C-89	-8,6214
29	C-51	-8,3814
30	C-10	-7,7677
31	C-7	-4,5197
32	C-5; C-46	-9,6748
33	C-17; C-86	-7,2915
34	C-96; C-99	-4,2534
35	C-48; C-57; C-106	-3,5071
36	C-15; C-19; C-25; C-39; C-42; C-50; C-55; C-65; C-74; C-82; C-84; C-2; C-95; C-3; C-102; C-116; C-118	0,5289
37	zona hidro 1	0,3854
38	zona hidro 2	-0,0098
39	zona hidro 3	2,3597
40	zona hidro 4	4,2781
41	zona hidro 5	3,8496
42	zona hidro 6	5,6951
43	zona hidro 7	2,2922
44	zona hidro 8	2,1456
45	zona hidro 9	4,6368
46	zona eólica 1	-2,5819
47	zona eólica 2	-0,5727
48	zona eólica 3	-0,7289
49	carga	23,824

Pela equação (4.4), que descreve a função logística (Subseção 4.2.1), tem-se que quanto mais positivo for o coeficiente, maior é a probabilidade do estado ser de falha conforme a respectiva variável aumenta, enquanto quanto mais negativo for o coeficiente, menor é a probabilidade do estado ser de falha conforme a respectiva variável aumenta. Para os coeficientes de valor mais próximo de zero, a relação entre a variável e a probabilidade é atenuada, tanto para um lado, quanto para o outro; ou seja, o classificador “enxerga” tais variáveis com certa indiferença.

A variável que demonstrou maior influência sobre a probabilidade de falha é a da carga, o que era esperado. As variáveis de fontes hídricas apresentaram coeficientes positivos (exceto as zonas 1 e 2, com coeficientes virtualmente nulos), o que indicaria que sua disponibilidade tem relação direta com a falha do sistema. Isso deve ser relevado, porém, tendo em vista que os coeficientes se equilibram entre si.

Por exemplo, é possível que as fontes hídricas apresentem mais energia disponível quando as fontes eólicas, que possuem mais capacidade de geração, apresentam menos. Deve-se observar que é muito difícil tirar conclusões sobre certas variáveis num cenário tão complexo como o de confiabilidade composta.

No caso de equipamentos, que é o foco da metodologia proposta, são os circuitos que se destacam, enquanto as estações geradoras apontam coeficientes mais próximos de zero. Muitas variáveis apresentam coeficiente negativo, pois o classificador recebe informação de disponibilidade dos equipamentos. Quanto maior a disponibilidade, menor a probabilidade de falha, e vice-versa. Um aspecto importante a ser discutido, porém, é por qual motivo os geradores são mais bem colocados na lista ranqueada da Tabela B.1, mas menos importantes para o classificador. Os geradores falham bem mais que circuitos e, por isso, eles aparecem indisponíveis em um volume maior de estados de falha, o que eleva sua indisponibilidade-IS em relação aos circuitos, os colocando na frente na lista ranqueada. Por outro lado, apesar de em efeito combinado a indisponibilidade dos geradores ser determinante para a falha, individualmente, em geral, apenas alguns circuitos são fatais ou quase fatais para a ocorrência do corte de carga. Ainda que esses circuitos saiam de operação com muito menos frequência, as suas eventuais contingências invariavelmente levarão à ocorrência da falha de interesse. Isso faz com que esses circuitos se destaquem dos demais equipamentos na perspectiva do classificador.

A análise de variáveis explicativas apresentada exemplifica a necessidade de usar inteligência artificial para a tarefa de identificar padrões de comportamento entre componentes de um sistema elétrico de potência e um dado evento de interesse. Afinal, o problema da falha de suprimento por inflexibilidade de operação da rede de transmissão é multivariado e apresenta tendências diversas e não-óbvias.