

**Universidade Federal de Juiz de Fora**  
**Faculdade de Engenharia**  
**Programa de Pós-Graduação em Engenharia Elétrica**

**Antônio Sobrinho Campolina Martins**

**Fluxo de Potência Multifásico Probabilístico Baseado em *Clustering* e Recuperações Analíticas:** uma nova ferramenta para análise de redes de distribuição desequilibradas ativas com incertezas

**Juiz de Fora**  
**2025**

**Antônio Sobrinho Campolina Martins**

**Fluxo de Potência Multifásico Probabilístico Baseado em *Clustering* e Recuperações Analíticas:** uma nova ferramenta para análise de redes de distribuição desequilibradas ativas com incertezas

Tese apresentada ao corpo docente do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, na área de Sistemas de Energia Elétrica, como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Leandro Ramos de Araujo, D. Sc.

Coorientadora: Prof<sup>a</sup>. Débora Rosana Ribeiro Penido Araujo, D. Sc.

**Juiz de Fora  
2025**

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Martins, Antônio Sobrinho Campolina.

Fluxo de Potência Multifásico Probabilístico Baseado em Clustering e Recuperações Analíticas : uma nova ferramenta para análise de redes de distribuição desequilibradas ativas com incertezas / Antônio Sobrinho Campolina Martins. -- 2025. 188 f. : il.

Orientador: Leandro Ramos de Araujo

Coorientadora: Débora Rosana Ribeiro Penido Araujo  
Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2025.

1. Fluxo de Potência Probabilístico. 2. Sistemas Multifásicos. 3. Sistemas de Distribuição. 4. Clustering. 5. Análise de Redes. I. Araujo, Leandro Ramos de, orient. II. Araujo, Débora Rosana Ribeiro Penido, coorient. III. Título.

**Antônio Sobrinho Campolina Martins**

**Fluxo de Potência Multifásico Probabilístico Baseado em *Clustering* e Recuperações Analíticas:** uma nova ferramenta para análise de redes de distribuição desequilibradas ativas com incertezas

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Sistemas de Energia Elétrica

Aprovada em 03 de julho de 2025.

**BANCA EXAMINADORA**

**Prof. Dr. Leandro Ramos de Araujo** - Orientador

Universidade Federal de Juiz de Fora

**Profa. Dra. Débora Rosana Ribeiro Penido Araujo** - Coorientadora

Universidade Federal de Juiz de Fora

**Prof. Dr. Márcio Zamboti Fortes**

Universidade Federal Fluminense

**Prof. Dr. Kleber Melo e Silva**

Universidade de Brasília

**Prof. Dr. João Alberto Passos Filho**

Universidade Federal de Juiz de Fora

**Prof. Dr. Bruno Henriques Dias**

Universidade Federal de Juiz de Fora

Juiz de Fora, 25/06/2025.

---



Documento assinado eletronicamente por **Leandro Ramos de Araujo, Professor(a)**, em 03/07/2025, às 17:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Debora Rosana Ribeiro Penido Araujo, Professor(a)**, em 03/07/2025, às 17:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Joao Alberto Passos Filho, Professor(a)**, em 03/07/2025, às 17:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **MARCIO ZAMBOTI FORTES, Usuário Externo**, em 03/07/2025, às 17:30, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Bruno Henriques Dias, Professor(a)**, em 03/07/2025, às 18:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Kleber Melo e Silva, Usuário Externo**, em 03/07/2025, às 19:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no Portal do SEI-Uffj ([www2.uffj.br/SEI](http://www2.uffj.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **2467669** e o código CRC **445E2E36**.

---

*Dedico este trabalho  
aos meus sobrinhos  
Afonso José, Madalena e Celina.*

## AGRADECIMENTOS

Na Sua sabedoria eterna, Deus não fez a Criação de forma igualitária. Por algum motivo, Ele quis distribuir diferentes dons e carismas, tornando o mundo um lugar plural e diverso. “Há diversidade de dons, mas um só e o mesmo Espírito” (1 Cor 12,4). Uma consequência disso é a necessidade humana perene de depender uns dos outros para a sobrevivência. E mais do que isso: a plenitude só se alcança por meio da unidade de tudo e todos no mistério de Cristo, que cria, envolve e mantém o cosmos – e tudo o que nele há – a cada intervalo infinitesimal de tempo.

Nessa perspectiva, deve-se reconhecer que nenhum feito pode ser alcançado de maneira solitária. Dependemos imprescindivelmente das pessoas e de tudo o que nos envolve, de modo que qualquer coisa possa ser efetivamente realizada. Diante do enorme desafio de concluir um curso de Doutorado, repleto de dúvidas, incertezas, angústias e cruces, o único sentimento que pode prevalecer é o da gratidão. Sou eternamente grato às pessoas com as quais convivi nesse curto espaço-tempo que me foi confiado por Deus. Em especial, não posso deixar de agradecer nominalmente a algumas pessoas que foram importantíssimas nessa jornada.

Agradeço, em primeiro lugar, aos meus pais Antonio Henrique e Claudia, por terem proporcionado a vida, por abandonarem tudo para que eu pudesse existir, por me darem a fé, valor mais sagrado que possuo. Sou grato por todos os sacrifícios que fizeram por mim, por todo o esforço que fizeram para que a minha vida fosse a melhor possível.

Aos meus irmãos, Joana, André e Ciro, por serem sempre meus maiores exemplos e por estarem comigo nos momentos difíceis.

Aos meus orientadores, Leandro e Débora, pelos oito anos de convívio e amizade, pela instrução, pelos conselhos e por todos os momentos de aprendizado – tanto científico, quanto de vivência.

Aos professores do corpo docente do Programa de Pós-Graduação em Engenharia Elétrica, particularmente aos que aceitaram fazer parte da banca de avaliação do presente trabalho. Aos amigos do NUPESP, em especial: Glaucus, Robert, Bruna, Andrezza, Júnior, Pedro Henrique e Luiz Rogério, por todas as colaborações, aprendizados, mas também descontrações e risadas.

Agradeço à Universidade Federal de Juiz de Fora e à CAPES pelos auxílios prestados para a elaboração desse trabalho.

Encerro com o agradecimento mais especial: sou grato à minha namorada, Laura, por todo carinho, afeto e apoio. Obrigado por estar sempre comigo, nos momentos de alegria e nos de tristeza. Você é a minha maior inspiração e minha melhor companhia.

A todas as pessoas que, de alguma forma, me ajudaram nessa jornada, deixo meus agradecimentos. Que possamos, juntos, trabalhar para a maior glória de Deus.

*“Algum dia,  
quando tivermos dominado os ventos,  
as ondas, as marés e a gravidade,  
utilizaremos as energias do amor.  
Então, pela segunda vez na história do mundo,  
o homem descobrirá o fogo”.*  
Pierre Teilhard de Chardin (1881-1955)

## RESUMO

A presente tese apresenta como aspecto primordial a elaboração de uma nova ferramenta para simular o Fluxo de Potência Probabilístico, cuja formulação é direcionada a atender especificamente Sistemas de Distribuição de Energia Elétrica, de maneira a contemplar suas características inerentes. O nome do método é Fluxo de Potência Multifásico Probabilístico Baseado em *Clustering* e Recuperações Analíticas. Ele envolve a adoção de técnicas de *clustering* para reduzir as amostras de variáveis aleatórias pré-calculadas, a execução de problemas determinísticos que contemplam a modelagem multifásica das redes com o intuito de estimar os principais momentos estatísticos das variáveis aleatórias de saída e, finalmente, a execução de uma técnica chamada Expansão Gram-Charlier para recuperar analiticamente as funções densidade de probabilidade das variáveis de interesse. A ferramenta engloba um novo procedimento para definir o número ótimo de *clusters*, que tem por intuito buscar alta precisão dos resultados em consonância com um tempo computacional adequado para aplicações de planejamento e operação. Ele se baseia na construção de uma curva que envolve as distâncias *intra-cluster* e *inter-cluster* dos dados de entrada, encontrando uma quantidade ótima de *clusters* para cada aplicação em específico.

Para testar a sua aplicabilidade, foram simuladas situações diversas em que variáveis devem ser assumidas como aleatórias no Fluxo de Potência Multifásico. Os testes demonstram que o método apresenta: (i) capacidade de lidar com equacionamento multifásico das redes, incluindo sistemas multiterrados com modelagens explícitas dos condutores de neutro e aterramentos; (ii) alta precisão, para garantir resultados confiáveis, promovendo acurácia superior em relação a métodos consagrados na literatura; (iii) eficiente sob a perspectiva do tempo computacional, proporcionando soluções rápidas; (iv) flexível, possuindo abrangência quanto às condições das variáveis aleatórias de entrada, incluindo correlações espaciais e temporais, variáveis Gaussianas e não Gaussianas, e simulação de redes de distribuição desequilibradas com incertezas da geração fotovoltaica e do carregamento; (v) escalável, permitindo a simulação de problemas de grande porte e com um número alto de variáveis aleatórias, com diretrizes específicas para contornar a chamada ‘maldição da dimensionalidade’, que faz com que algoritmos de *clustering* percam precisão quando há o aumento do número de variáveis incertas.

**Palavras-chave:** Fluxo de Potência Probabilístico; Sistemas Multifásicos; Sistemas de Distribuição; *Clustering*; Análise de Redes.

## ABSTRACT

This thesis presents as its primary focus the formulation of a new tool for simulating the Probabilistic Power Flow, specifically designed for Electric Power Distribution Systems, in a manner that captures their inherent characteristics. The method is called the Clustering and Analytical Recovery-Based Multiphase Probabilistic Power Flow. It involves the use of clustering techniques to reduce precomputed random variable samples, the execution of deterministic analyses that consider the multiphase modeling of the network to estimate the primary statistical moments of the output random variables, and, finally, the application of a technique called the Gram–Charlier Expansion to analytically recover the probability density functions of the variables of interest. The tool incorporates a new procedure for determining the optimal number of clusters, aimed at achieving high-precision results while maintaining a computationally feasible time for planning and operational applications. This is based on the construction of a curve involving intra-cluster and inter-cluster distances of the input data, identifying the optimal number of clusters for each specific application.

To test its applicability, a range of scenarios were simulated in which variables must be treated as random within the Multiphase Power Flow context. The results demonstrate that the method: *(i)* is capable of addressing the multiphase formulation of the network, including multi-grounded systems with explicit modeling of neutral and ground conductors; *(ii)* delivers high precision, ensuring trustworthy results and providing superior accuracy compared with established methods in the literature; *(iii)* is efficient from a computational time perspective, yielding rapid solutions; *(iv)* is flexible, accommodating a broad range of conditions for the input random variables — including spatial and temporal correlations, Gaussian and non-Gaussian variables, and the simulation of unbalanced distribution networks with uncertainty in both photovoltaic generation and load demand; and *(v)* is scalable, allowing for the simulation of large-scale problems with a high number of random variables, including specific guidelines for addressing the so-called “curse of dimensionality” that causes clustering algorithms to lose precision as the number of variables increases.

**Keywords:** Probabilistic Power Flow; Multiphase Systems; Distribution Systems; Clustering; Electric Network Analysis.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Classificação dos Métodos de FP para SD de acordo com a técnica de solução. ....	32
Figura 2 – Classificação dos Métodos de FP para SD de acordo com a forma de operação....	33
Figura 3 – Classificação dos métodos de FP para SD quanto à inclusão de incertezas. ....	34
Figura 4 – Fluxograma geral da teoria de cópulas.....	37
Figura 5 – Fluxograma dos principais passos do FPPC. ....	45
Figura 6 – Fluxograma do FPMP-CRA.....	53
Figura 7 – Diagrama das características do FPMP-CRA. ....	54
Figura 8 – Esquema didático da representação de um nó no MICN. ....	58
Figura 9 – Aplicação do método do cotovelo para banco de dados tutorial.....	63
Figura 10 – Aplicação do método do cotovelo para dados normalmente distribuídos.....	64
Figura 11 – Fluxograma do método proposto para estimar número ótimo de clusters. ....	65
Figura 12 – Gráficos típicos para $G1(K)$ e $G2(K)$ .....	67
Figura 13 – Curva resultante $G(K) = G1(K) + G2(K)$ .....	68
Figura 14 – Sistema IEEE 33 Barras. ....	73
Figura 15 – IEEE 33: Resultados de tensão. (a) Barra 33, fase A (magnitude); (b) Barra 33, fase B (magnitude); (c) Barra 33, fase C (magnitude); (d) Barra 33, fase A (ângulo); (e) Barra 33, fase B (ângulo); (f) Barra 33, fase C (ângulo).....	75
Figura 16 – IEEE 33: Resultados de corrente. (a) Linha 1-2, fase A (magnitude); (b) Linha 1-2, fase B (magnitude); (c) Linha 1-2, fase C (magnitude); (d) Linha 1-2, fase A (ângulo); (e) Linha 1-2, fase B (ângulo); (f) Linha 1-2, fase C (ângulo). ....	77
Figura 17 – Topologia IEEE 4.....	82
Figura 18 – Amostras geradas pela cópula para cada caso.....	84
Figura 19 – Resultados das PDFs de tensões elétricas (IEEE 4).....	86
Figura 20 – Topologia IEEE 13.....	91
Figura 21 – (a) Número ótimo de clusters considerando diferentes dimensionalidades do banco de dados no IEEE 13; (b) Tempo computacional demandado, em segundos, de acordo com a dimensionalidade do banco de dados de entrada. ....	92
Figura 22 – Comparação FPMP-CRA x MCS: tensões elétricas na topologia IEEE 13.....	99
Figura 23 – Comparação FPMP-CRA x MCS: perdas ativas totais na topologia IEEE 13. ...	100
Figura 24 – Curvas de convergência para estimar número ótimo de clusters. ....	100
Figura 25 – Topologia IEEE 123.....	104

Figura 26 – Procedimento para obter amostras correlacionadas temporalmente. ....	107
Figura 27 – Curvas obtidas através dos dados: (a) Demandas; (b) Irradiação Solar.....	109
Figura 28 – Parâmetros das distribuições marginais. ....	110
Figura 29 – Amostras temporalmente correlacionadas de demanda. ....	113
Figura 30 – Amostras correlacionadas temporalmente de irradiação solar.....	114
Figura 31 – PDFs de magnitude de tensão, de acordo com a hora do dia: (a) 15-C; (b) 70-A; (c) 114-A. ....	117
Figura 32 – PDF de tensão para posição 114-A, às 12 horas. ....	120
Figura 33 – Probabilidades de violação por sobretensão. ....	121
Figura 34 – Tempo computacional demandado pelo FPMP-CRA correlacionado temporalmente. ....	122
Figura 35 – Momentos estatísticos das magnitudes de tensão: MCS x FPMP-CRA (00h). ..	124
Figura 36 – Momentos estatísticos das magnitudes de tensão: MCS x FPMP-CRA (12h). ..	125
Figura 37 - Momentos estatísticos das magnitudes de tensão: MCS x FPMP-CRA (18h)...	126
Figura 38 – Comparação das PDFs de magnitude de tensão obtidas via MCS e FPMP-CRA. ....	127
Figura 39 – Comparação do tempo computacional: MCS x FPMP-CRA.....	129
Figura 40 – Erros máximos: FPP correlacionado x independente.....	130
Figura 41 – Comparações PDFs: FPP correlacionado x independente. ....	131
Figura 42 – Topologia IEEE NEV. ....	135
Figura 43 – Comportamento probabilístico das magnitudes de tensões de neutro. ....	137
Figura 44 – Tensões de neutro: boxplots.....	138
Figura 45 – PDFs das tensões de neutro, recuperadas analiticamente usando EGC.....	140
Figura 46 – Topologia IEEE 8500.....	144
Figura 47 – Estagnação dos resultados estatísticos com a variação do número de clusters para IEEE 8500. ....	146
Figura 48 – Erros estatísticos para magnitudes de tensão, considerando-se a topologia IEEE 8500 e diferentes números de clusters no FPMP-CRA. ....	148
Figura 49 – Variação do tempo computacional do FPMP-CRA de acordo com o número de clusters. ....	149
Figura 50 – Conexões das cargas na topologia IEEE 8500.....	151
Figura 51 – Resultados estatísticos para tensões do IEEE 8500, considerando diferentes métodos probabilísticos. ....	152

Figura 52 – Comparação MEP x MCS para IEEE 8500. ....	153
Figura 53 – Comparação LHS x FPMP-CRA para topologia IEEE 8500 (magnitude de tensão).....	156
Figura 54 – Comparação LHS x FPMP-CRA para topologia IEEE 8500 (perdas ativas totais). .....	157
Figura 55 – Comportamento da distribuição uniforme. ....	180

## LISTA DE TABELAS

Tabela 1 – Variáveis aleatórias e suas PDFs mais adotadas.....	35
Tabela 2 – Comparações dos trabalhos recentes sobre FPPC em SD. ....	50
Tabela 3 – Breve resumo dos testes realizados e seus objetivos. ....	71
Tabela 4 – Resultados da análise de sensibilidade (modelo de rede versus natureza do FP), usando caso de 10% de desequilíbrio como referência. ....	79
Tabela 5 – Resultados da análise de sensibilidade (modelo de rede versus natureza do FP), usando caso de 50% de desequilíbrio como referência. ....	80
Tabela 6 – Resultados do FPMP-CRA para IEEE 4. ....	85
Tabela 7 – Probabilidades de subtensão de acordo com o caso (IEEE 4). ....	87
Tabela 8 – Comparação do FP determinístico versus FPMP-CRA. ....	88
Tabela 9 – Ganhos computacionais alcançados pelo FPMP-CRA de acordo com a dimensionalidade do banco de dados de entrada.....	94
Tabela 10 – Resultados para IEEE 13: valores esperados.....	96
Tabela 11 – Resultados para IEEE 13: desvios padrões.....	97
Tabela 12 – Dados dos geradores fotovoltaicos .....	104
Tabela 13 – Número ótimo de clusters por hora do dia.....	117
Tabela 14 – Erros percentuais médios em relação aos valores esperados e desvios padrões das magnitudes de tensão para topologia IEEE 8500. ....	154
Tabela 15 – Erros percentuais médios em relação aos valores esperados e desvios padrões dos ângulos de tensão para topologia IEEE 8500. ....	155
Tabela 16 – Tempo computacional demandado por método considerando IEEE 8500.....	158

## LISTA DE SIGLAS E ABREVIATURAS

CC	Coefficiente de Correlação
CDF	<i>Cumulative Distribution Function</i>
CLARA	<i>Clustering for Large Applications</i>
DM	Determinístico e Multifásico
D-STATCOM	<i>Distribution Static Compensator</i>
EGC	Expansão Gram-Charlier
FBS	<i>Forward/Backward Sweep</i>
FP	Fluxo de Potência
FPD	Fluxo de Potência Determinístico
FPMP	Fluxo de Potência Multifásico Probabilístico
FPMP-CRA	Fluxo de Potência Multifásico Probabilístico baseado em <i>Clustering</i> e Recuperações Analíticas
FPP	Fluxo de Potência Probabilístico
FPFC	Fluxo de Potência Probabilístico via <i>Clustering</i>
FPS	Fluxo de Potência Sequencial
GA	<i>Genetic Algorithm</i>
GC	Ganho Computacional
GD	Geração Distribuída
GS	Gauss-Seidel
IEEE	Instituto dos Engenheiros Eletricistas e Eletrônicos
LHS	<i>Latin Hypercube Simulation</i>
MC	Matriz Comprimida
MCS	<i>Monte Carlo Simulation</i>
MEP	Método da Estimativa dos Pontos
MICN	Método de Injeção de Correntes a N Condutores
MICT	Método de Injeção de Correntes Trifásico
MO	Matriz Original
NR	<i>Newton-Rhapson</i>
OpenDSS	<i>Open Distribution System Simulator</i>
PDF	<i>Probability Density Function</i>

PE	Probabilístico e Equilibrado
PSO	<i>Particle Swarm Optimization</i>
REF	Referência
SD	Sistemas de Distribuição
ST	Sistemas de Transmissão
UD	<i>Uniform Design</i>

## LISTA DE SÍMBOLOS

$Cov$	Covariância entre duas variáveis aleatórias
$\sigma$	Desvio padrão
$c$	Coefficiente de Gram-Charlier
$H$	Polinômio de Chebycheff-Hermite
$\phi$	Função densidade de probabilidade
$K$	Número de <i>clusters</i>
$I_{inj,k}$	Injeção de corrente líquida no nó $k$
$I_{X,k}$	Injeção de corrente devido a um elemento $X$ no nó $k$
$b$	Susceptância <i>shunt</i> de linha de distribuição
$y$	Admitância série de linha de distribuição
$V_k$	Tensão elétrica no nó $k$
$M_p$	Momento estatístico bruto de ordem $p$
$m_p$	Momento estatístico central de ordem $p$
$\mu_p$	Momento estatístico central normalizado de ordem $p$
$\alpha$	Distância total <i>intra-cluster</i> quando há apenas um agente
$\beta$	Distância total <i>inter-cluster</i> quando o número de <i>clusters</i> é igual à dimensão da base de dados original
$g$	Função atrelada a distâncias não normalizadas de <i>clusters</i>
$G$	Função atrelada a distâncias normalizadas de <i>clusters</i>
$pu$	Sistema por unidade
A	Ampère
W	Watts
s	Segundos
$\rho$	Coefficiente de correlação
Pr	Probabilidade
$N_T$	Número total de amostras
$\epsilon_1$	Erro de valor médio
$\epsilon_2$	Erro de desvio padrão

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>21</b>
1.1 CONTEXTUALIZAÇÃO .....	21
1.2 MOTIVAÇÕES .....	23
1.3 OBJETIVOS, METODOLOGIAS E CONTRIBUIÇÕES .....	25
1.4 PUBLICAÇÕES DECORRENTES.....	27
1.5 ESTRUTURA DA TESE.....	29
<b>2 REFERENCIAL TEÓRICO E REVISÃO BIBLIOGRÁFICA .....</b>	<b>30</b>
2.1 FLUXO DE POTÊNCIA PARA SISTEMAS DE DISTRIBUIÇÃO .....	30
2.2 FLUXO DE POTÊNCIA PROBABILÍSTICO .....	34
2.3 APLICAÇÕES RECENTES DO FPP EM SISTEMAS DE DISTRIBUIÇÃO .....	42
2.4 FPP VIA <i>CLUSTERING</i> .....	44
2.5 APLICAÇÕES DO FPPC EM SISTEMAS DE DISTRIBUIÇÃO .....	46
2.6 CONCLUSÕES PARCIAIS .....	48
<b>3 FLUXO DE POTÊNCIA MULTIFÁSICO PROBABILÍSTICO BASEADO EM <i>CLUSTERING</i> E RECUPERAÇÕES ANALÍTICAS (FPMP-CRA) .....</b>	<b>52</b>
3.1 ALGORITMO DO FPMP-CRA .....	52
3.2 MÉTODO PARA ESTIMAR NÚMERO ÓTIMO DE <i>CLUSTERS</i> .....	61
3.3 CONCLUSÕES PARCIAIS .....	69
<b>4 RESULTADOS OBTIDOS PELA APLICAÇÃO DO FPMP-CRA.....</b>	<b>70</b>
4.1 OBJETIVOS E CONDIÇÕES GERAIS DE SIMULAÇÃO .....	70
4.2 TESTE 1: ESTUDO DA MODELAGEM DE REDE NO ÂMBITO DO FPP .....	72
4.3 TESTE 2: ESTUDO DA MODELAGEM DA CARGA E CONSIDERAÇÃO DE DIFERENTES CORRELAÇÕES ESPACIAIS .....	81
4.4 TESTE 3: ESTUDO DA DIMENSIONALIDADE DO FPMP-CRA .....	89
4.5 TESTE 4: APLICAÇÕES COM GERAÇÃO DISTRIBUÍDA FOTOVOLTAICA....	102

4.6 TESTE 5: ESTUDOS PROBABILÍSTICOS DE SEGURANÇA DE TENSÃO DE NEUTRO .....	133
4.7 TESTE 6: APLICAÇÕES EM SISTEMAS DE GRANDE PORTE.....	143
<b>5 CONSIDERAÇÕES FINAIS.....</b>	<b>160</b>
5.1 CONCLUSÕES .....	160
5.2 TRABALHOS FUTUROS .....	164
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>166</b>
<b>APÊNDICE A – NOÇÕES DE ESTATÍSTICA .....</b>	<b>172</b>
<b>APÊNDICE B – MÉTODOS DE FPP .....</b>	<b>181</b>
<b>APÊNDICE C – ALGORITMOS DE <i>CLUSTERING</i> .....</b>	<b>186</b>

## 1 INTRODUÇÃO

A presente tese inicia-se com uma breve introdução, na qual é explicitado o tema pesquisado, destacando a relevância do assunto delimitado para o estado da arte. Para atingir tal fim, o capítulo é dividido em cinco seções: a *contextualização*, que situa o leitor em relação à problemática do objeto pesquisado; a *motivação*, que destaca a relevância do tema para o contexto citado; os *objetivos e metodologias*, que expõem o que se pretende alcançar e com quais meios o fazer; as *publicações decorrentes*, que consistem em uma lista dos trabalhos que foram publicados como consequência do desenvolvimento da tese apresentada; e, por fim, a *estrutura*, que enumera e explica sucintamente os capítulos do texto.

### 1.1 CONTEXTUALIZAÇÃO

O planejamento e a operação dos Sistemas Elétricos de Potência (SEP) são sustentados pelo problema de Fluxo de Potência (FP), que consiste em um conjunto de equações cuja solução informa as condições de regime permanente de determinada rede elétrica (Ramadhani *et al.*, 2020). Através do FP, é possível conhecer as tensões em magnitude e ângulo, além de quaisquer outras variáveis de interesse calculadas por intermédio das já citadas, como correntes elétricas e perdas ativas, por exemplo. Conhecendo-se o estado do sistema, podem ser feitas análises mais criteriosas que vão definir as ações específicas do planejamento e controle. Estudos que dependem da execução de problemas de FP incluem: análise de segurança de tensão, planejamento de reativos e alocação de capacitores, regulação de tensão, análise de perdas, entre outros (Abbasi; Mohammadi, 2023).

A formulação tradicional do FP considera apenas um cenário operativo, também chamado de Fluxo de Potência Determinístico (FPD). Isso significa que as variáveis conhecidas, ou variáveis de entrada, são reduzidas a um único conjunto de valores que modelam uma condição específica, nomeada comumente de *snapshot* ou fotografia do sistema (Ramadhani *et al.*, 2020). Tais parâmetros podem ser as gerações, cargas e a própria configuração da rede analisada (Abbasi; Mohammadi, 2023). Esta abordagem é particularmente útil quando as incertezas a que as variáveis do FP estão submetidas podem ser negligenciadas e, durante muito tempo, ela foi assumida como satisfatória para a maioria dos casos simulados. Muitas vezes, formulações mais conservadoras são até hoje adotadas, em que se consideram ‘piores casos’ segundo uma série histórica para modelar as variáveis de

entrada (Ramadhani *et al.*, 2020). Muito frequentemente, abordagens temporais também são empregadas, em que são simulados patamares discretizados que representam médias dos cenários operativos em sequência. Tal abordagem é denominada Fluxo de Potência Sequencial (FPS). As formas mais comuns englobam a representação diária com discretização horária (ou seja, através das modelagens das variáveis de entrada do FP de forma horária em um horizonte diário) ou a representação anual com discretização diária ou até horária.

Entretanto, mesmo considerando essa estratégia, as variáveis de entrada são definidas por valores determinísticos, ainda que uma simulação seja conduzida para cada patamar considerado. À medida em que as incertezas da topologia aumentam, o FPD passa a ser uma ferramenta inadequada para análise, uma vez que sua modelagem de apenas um único cenário pode ser restritiva, mesmo quando o FPS é considerado.

Recentemente, as incertezas do problema de FP têm aumentado significativamente, sobretudo em relação aos Sistemas de Distribuição (SD) de Energia Elétrica. Estes, que são os responsáveis por efetivamente alimentar as cargas e entregar a energia elétrica ao consumidor final, já possuem naturalmente uma imprevisibilidade no que tange à carga. Caracterizar o consumo de eletricidade desses sistemas é uma tarefa complexa que depende de fatores como comportamentos humanos e padrões de consumo incertos, variações climáticas e incertezas inerentes ao uso de certos equipamentos (Ramadhani *et al.*, 2020). Logo é impossível definir deterministicamente (ou seja, com 100% de certeza) como o consumo se comporta.

Todavia, os motivos que tornam o comportamento dos SD muito mais imprevisível se relacionam com as importantes quebras de premissas e mudanças tecnológicas que sucessivamente vêm ocorrendo e que transformam qualitativamente a operação dos SD (Chevalier; Schenato; Daniel, 2022; Ramadhani *et al.*, 2020). Particularmente, destacam-se a penetração cada vez maior de geração distribuída (GD) renovável (sobretudo solar fotovoltaica e eólica) nas redes e o advento de dispositivos armazenadores de energia (Abbasi; Mohammadi, 2023; Ramadhani *et al.*, 2020). Segundo Sun *et al.* (2022), os aumentos cada vez mais expressivos de recursos energéticos distribuídos (REDs) se justificam pelos seus benefícios quanto a questões climáticas, uma vez que representam gerações que ocasionam baixos impactos ambientais quanto a poluentes, além de proporcionarem ganhos em confiabilidade e flexibilidade dos sistemas. Já Constante-Flores e Illindala (2019) sugerem que os REDs também podem ser responsáveis pelo aumento da eficiência do sistema e de sua segurança. Além da tendência citada, outros avanços disruptivos em SD são citados em Chevalier, Schenato e Daniel (2022): advento de sensores automatizados e dispositivos

equipados com recursos de telecomunicação, tais como unidades de medição microfasorial ( $\mu$ -PMUs) e demais medidores inteligentes, que são capazes de responder reativamente aos sinais de preços do mercado de energia em tempo real. Tudo isso altera significativamente como a energia é produzida, direcionada e consumida. Gruosso *et al.* (2020) argumentam que os consumidores estão cada vez mais participando da geração ativa de energia elétrica, e a geração e o consumo tendem a acompanhar as tendências de mercado, o que agrava ainda mais a imprevisibilidade desses padrões.

A consequência direta das transformações citadas é o aumento expressivo das incertezas dos parâmetros do sistema (Singh; Moger; Jena, 2023), principalmente porque a geração e o consumo passam a depender fortemente das condições climáticas (como irradiação solar, temperatura, velocidade dos ventos) que têm seu comportamento altamente estocástico (Ramadhani *et al.*, 2020). Nesse sentido, a avaliação de um único cenário por meio do FPD passa a não ser indicada (Abbasi; Mohammadi, 2023; Singh; Moger; Jena, 2023), uma vez que a alta imprecisão existente ao modelar as condições de entrada de maneira singular pode comprometer a acurácia dos resultados encontrados, o que, por sua vez, pode conduzir engenheiros de planejamento e de operação a cometerem ações equivocadas. Por exemplo, a integração inadequada de gerações fotovoltaicas pode levar a sobretensões e prejuízos em qualidade de energia, ocasionando danos a equipamentos e perda de confiabilidade dos SD (Ramadhani *et al.*, 2020). Por esse motivo, estratégias devem ser implementadas com o intuito de introduzir no problema de FP as incertezas citadas e, conseqüentemente, ter resultados que também incluam incertezas, visando ações que contemplem um maior número de cenários operativos, sendo, por conseguinte, mais amplas e seguras.

## 1.2 MOTIVAÇÕES

Diante da problemática apresentada na seção anterior, o Fluxo de Potência Probabilístico (FPP) foi fundamentado (Borkowska, 1974), tendo se tornado uma das principais ferramentas para análise de SEP sob a presença de incertezas (Abbasi; Mohammadi, 2023). O FPP considera as variáveis de entrada como sendo aleatórias, e, por isso, as variáveis de saída também o são. Elas são modeladas por funções densidade de probabilidade (PDF, do inglês: *probability density functions*). Desta forma, é conferido ao problema de FP um contingente maior de informação, incluindo as incertezas de seus

parâmetros (Abbasi; Mohammadi, 2023). Os resultados, por serem funções de probabilidade, estabelecem uma maior segurança para os engenheiros de operação e planejamento, e ações mais relevantes que considerem muitos cenários podem ser obtidas. Logo, trata-se de uma alternativa ao FPD e é indicado para casos em que as incertezas no problema não podem ser negligenciadas (Sun *et al.*, 2022), como no caso de redes de distribuição ativas que possuem GD renovável em abundância ou quando deve-se incluir as imprevisibilidades da carga.

Ao longo das últimas décadas, muitos métodos foram desenvolvidos para solucionar o FPP. Tradicionalmente, eles podem ser divididos em três grupos (Sun *et al.*, 2022): os métodos *analíticos*, os métodos *numéricos* e os métodos *por aproximações*. O primeiro grupo é caracterizado pela solução por vias de transformações estocásticas de variáveis aleatórias de entrada para lidar com modelos físicos reais com equivalências estocásticas (Sandoval; Gong; Chung, 2023). No segundo grupo, simula-se o problema determinístico para diversas amostras das PDFs das variáveis de entrada, fazendo posteriormente as análises estatísticas em relação às respostas obtidas (Sun *et al.*, 2022). Finalmente, o terceiro grupo engloba uma mistura dos dois primeiros: as respostas são baseadas na simulação determinística para diversos pontos de operação; entretanto, eles são previamente calculados de forma analítica para promover maior eficiência computacional. Por isso, eles também são chamados de métodos *semi-analíticos* (Sandoval; Gong; Chung, 2023). As particularidades de implementação dos principais métodos serão discutidas no Capítulo 2.

Apesar dos diferentes desenvolvimentos apresentarem bons resultados, eles possuem notáveis desvantagens. Geralmente, os métodos analíticos têm sua dificuldade de implementação agravada devido à sua complexidade matemática (Sun *et al.*, 2022) e, por isso, as equações do FP são frequentemente linearizadas para facilitar o seu processo, o que pode levar a erros nos resultados. Apesar de formulações analíticas mais recentes conseguirem lidar com as não linearidades do FP, Sandoval, Gong e Chung (2023) revelam que, de forma geral, a precisão é sensível ao modelo equivalente da rede escolhido e depende também dos modelos de variáveis aleatórias, o que não os tornam indicados para muitas aplicações. Sun *et al.* (2022) destacam a dificuldade de alguns métodos analíticos em considerar correlações de variáveis aleatórias, por exemplo. Além disso, Li *et al.* (2023) citam que adequar métodos analíticos para apresentarem resultados precisos quando variáveis aleatórias não-Gaussianas são empregadas é desafiador. Os métodos numéricos, por sua vez, requerem grande tempo computacional para convergir porque precisam avaliar uma enorme quantidade de amostras para determinarem resultados confiáveis (Sandoval; Gong; Chung, 2023; Sun *et al.*, 2022). Já

os métodos por aproximações pecam na sua robustez, visto que, ao aproximarem os resultados, muita informação estatística é perdida. Sun *et al.* (2022) expõem que alguns métodos por aproximações falham em estimar momentos de ordem superior das variáveis aleatórias de saída, o que prejudica a representação adequada de suas PDFs. Em resumo, segundo Abbasi e Mohammadi. (2023), um método que seja ao mesmo tempo confiável, acessível e escalável ainda não foi desenvolvido. Essa é a motivação primordial da presente tese.

Outra razão motivadora é a escassez de métodos direcionados especificamente a SD. A maioria dos desenvolvimentos são aplicados a Sistemas de Transmissão (ST), o que significa que muitas particularidades de SD são negligenciadas. Logo, as reproduções e generalizações desses métodos não são garantidamente assegurados sem importantes modificações prévias em suas metodologias.

Por conseguinte, o principal tema delimitado para esse estudo é o problema de FPP aplicado especificamente a SD, além de suas muitas possibilidades de aplicações.

### 1.3 OBJETIVOS, METODOLOGIAS E CONTRIBUIÇÕES

O objetivo geral do presente trabalho é o desenvolvimento de um método novo de FPP para SD, que seja seguro, acessível e eficiente na perspectiva do tempo computacional. Deseja-se também que o mesmo seja robusto e escalável. Isto é, que sua aplicação se estenda para diferentes variáveis aleatórias e sistemas (incluindo sistemas pequenos, médios e grandes), apresentando sempre resultados precisos e confiáveis. Considera-se também como um objetivo importante o uso do método desenvolvido em diversas aplicações possíveis no que tange aos estudos de planejamento de SD.

Para atingir esse propósito, uma nova metodologia foi desenvolvida. A metodologia proposta segue os seguintes passos: (i) Desenvolvimento matemático de um novo método de FPP; (ii) Validação do método através de simulações computacionais e comparações com outros métodos; (iii) Exploração de aplicações práticas decorrentes do método proposto; (iv) Discussões críticas a respeito de suas vantagens e desvantagens.

Portanto, as principais contribuições da presente tese são:

- O desenvolvimento de um método novo para simular o Fluxo de Potência Multifásico Probabilístico (FPMP), que inclui o desequilíbrio das redes de distribuição e a modelagem de elementos particulares a esses tipos de sistemas. O método, chamado FPMP-CRA (Fluxo de

Potência Multifásico Probabilístico baseado em *Clustering* e Recuperações Analíticas), consiste em um método híbrido numérico-analítico. Nele, técnicas de *clustering* são usadas para tratar as amostras numéricas das variáveis aleatórias de entrada do FPP. Posteriormente, os principais momentos estatísticos das variáveis aleatórias de saída são estimados através da simulação numérica do FPD para cada *cluster*. A modelagem multifásica das redes de distribuição é empregada para aproximar os resultados de situações reais envolvendo SD. Por fim, as suas PDFs podem ser reconstruídas analiticamente usando a técnica Expansão Gram-Charlier (EGC). O uso de técnicas de *clustering*, aliado a uma mistura de abordagens numéricas e analíticas, justifica-se à medida que se deseja um método que apresente simplicidade de implementação, robustez, escalabilidade, eficiência computacional e enorme precisão.

- Um método novo para otimizar o número de *clusters* é desenvolvido especificamente para aplicações de FPP. Ele é baseado na avaliação das distâncias *intra-clusters* e *inter-clusters* e busca determinar a melhor configuração visando alta precisão dos resultados e tempo computacional satisfatório.
- São efetuadas investigações acerca do modelo de rede para averiguar a importância da inclusão do desequilíbrio em problemas de planejamento de SD. Para isso, são simuladas diferentes situações de desequilíbrio, incluindo uma abordagem equilibrada baseada na representação via sequência positiva, buscando-se avaliar criticamente o impacto dessa simplificação no âmbito probabilístico de análise de redes.
- A modelagem da carga é estudada considerando diferentes níveis de correlações espaciais entre cargas de fases diferentes, sendo o impacto dessas considerações avaliada.
- O FPMP-CRA é aplicado em diversos sistemas-teste para demonstrar sua abrangente aplicabilidade. Ressalta-se a simulação de um *framework* probabilístico em SD com condutores neutros e aterramentos explicitamente modelados. O método proposto é capaz de conduzir estudos probabilísticos acerca de segurança de tensão de neutro.
- São simuladas redes de distribuição ativas com geradores solares fotovoltaicos multifásicos. Técnicas de correlação entre variáveis aleatórias são utilizadas para modelar as suas respectivas correlações temporais. A simulação de cenários correlacionados e independentes, com variáveis aleatórias Gaussianas e não-Gaussianas, explicita a robustez do método.
- Diretrizes são propostas para adaptar o método para situações de alta escalabilidade, incluindo aplicações em sistemas de grande porte e em se tratando de um alto número de

variáveis aleatórias. O FPMP-CRA é testado no sistema-teste IEEE 8500, que conta com mais de 2000 variáveis aleatórias de entrada.

#### 1.4 PUBLICAÇÕES DECORRENTES

A presente tese originou publicações em anais de congressos e também em periódicos internacionais. Faz-se a divisão entre *artigos publicados que se relacionam diretamente com o assunto delimitado* e *artigos publicados que se relacionam indiretamente com o assunto delimitado*. No primeiro grupo, estão os que trabalham, de alguma forma, aspectos tangentes ao problema de FPP. Já o segundo grupo contém aqueles cujos desenvolvimentos não estão diretamente relacionados ao FPP, mas apresentam considerações que posteriormente foram utilizadas nesta tese sob uma perspectiva probabilística. Os trabalhos estão ordenados em ordem crescente cronológica, primeiro os artigos de periódicos e posteriormente os publicados em anais de congressos. Por fim, são listadas as publicações submetidas, mas que ainda se encontram em estágio de revisão.

##### 1.4.1 ARTIGOS PUBLICADOS QUE SE RELACIONAM DIRETAMENTE COM O ASSUNTO DELIMITADO

- MARTINS, Antônio Sobrinho Campolina; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. An active multiphase probabilistic power flow based on a clustering approach. **Electric Power Systems Research**, v. 230, p. 110264, 2024.
- MARTINS, Antônio Sobrinho Campolina; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow. **International Journal of Electrical Power & Energy Systems**, v. 157, p. 109861, 2024.
- MARTINS, Antônio Sobrinho Campolina; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. Quasi-Convex NoC Optimization in the Active Multiphase Probabilistic Power Flow. **IEEE Systems Journal**, v. 19, p. 294-304, 2025.

- MARTINS, Antônio Sobrinho Campolina; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. Probabilistic Optimal Power Flow Considering Load Uncertainties and Capacitor Allocation. **XIV Latin-American Congress on Electricity Generation and Transmission - CLAGTEE**, 2022.

#### 1.4.2 ARTIGOS PUBLICADOS QUE SE RELACIONAM INDIRETAMENTE COM O ASSUNTO DELIMITADO

- MARTINS, Antônio Sobrinho Campolina; ARAUJO, Leandro Ramos De; PENIDO, Débora Rosana Ribeiro. Sensibility Analysis with Genetic Algorithm to Allocate Distributed Generation and Capacitor Banks in Unbalanced Distribution Systems. **Electric Power Systems Research**, v. 209, p. 107962, 2022.

- CAVALARI, Carla Farage; MARTINS, Antônio Sobrinho Campolina; PENIDO, Débora Rosana Ribeiro; et al. An enhanced method for fault location in distribution systems based on optimal power flow. **International Journal of Electrical Power & Energy Systems**, v. 157, p. 109870, 2024.

- MARTINS, Antônio Sobrinho Campolina; ARAUJO, Leandro Ramos De; PENIDO, Débora Rosana Ribeiro. Alocação Ótima de Capacitores e Geradores Fotovoltaicos em Sistemas de Distribuição Considerando Variação Horária do Carregamento. **LIV Simpósio Brasileiro de Pesquisa Operacional - SBPO 2022**, 2022.

- MARTINS, Antônio Sobrinho Campolina; ARAUJO, Leandro Ramos De; PENIDO, Débora Rosana Ribeiro. Single-Phase Capacitor Allocation in Distribution Systems Considering Voltage Unbalance Constraints and Losses Minimization. **XV CBQEE - Conferência Brasileira sobre Qualidade de Energia Elétrica**, 2023.

#### 1.4.3 ARTIGOS SUBMETIDOS EM PROCESSO DE REVISÃO

- MARTINS, Antônio Sobrinho Campolina; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. Neutral Voltage Security Assessment by Using a Multiphase Probabilistic Power Flow Framework. **IEEE Transactions on Power Systems**, 2025.

- MARTINS, Antônio Sobrinho Campolina; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. Multiphase Probabilistic Power Flow with Temporal Correlations of Loads and Photovoltaic DGs in Active Distribution Networks. **IEEE Transactions on Smart Grid**, 2025.

## 1.5 ESTRUTURA DA TESE

A tese é dividida em cinco capítulos. O primeiro consiste na presente Introdução, cujo objetivo foi expor a caracterização geral do objeto trabalhado, ou seja, a problemática na qual está inserido, a delimitação do tema e do assunto diante dessa problemática, a relevância do assunto, as motivações, objetivos, contribuições e publicações decorrentes da pesquisa.

No Capítulo 2, há a exposição o referencial teórico, que caracteriza o desenvolvimento histórico dos tópicos pesquisados, além das tendências e avanços no atual estado da arte.

No Capítulo 3, o método proposto (FPMP-CRA) tem sua formulação matemática e algoritmo apresentados.

No Capítulo 4, os principais resultados obtidos por meio do FPMP-CRA são exibidos e discutidos criticamente.

Finalmente, no Capítulo 5, a tese é encerrada com as considerações finais, em que são expostas as conclusões e os trabalhos futuros.

O trabalho ainda contempla três apêndices. No Apêndice A, noções de estatística são apresentadas e conceitos dessa área que foram empregados ao longo da tese são definidos em detalhes. No Apêndice B, os procedimentos matemáticos e computacionais dos métodos de FPP utilizados para validação do FPMP-CRA são mostrados. Por fim, no Apêndice C, definições acerca dos algoritmos de *clustering* são explicitadas.

## 2 REFERENCIAL TEÓRICO E REVISÃO BIBLIOGRÁFICA

Este capítulo tem por finalidade apresentar as principais obras que se relacionam diretamente com a pesquisa, ou seja, com o FPP direcionado a SD. São destacados autores cujos trabalhos apresentam relevância para o estado da arte. Uma retrospectiva histórica acerca do assunto é conduzida, explicitando os motivos pelos quais as metodologias foram criadas. Posteriormente, desenvolvimentos mais recentes são referenciados, visando mostrar as tendências e avanços na área. Os principais métodos de FPP são apresentados e seus algoritmos são exibidos. Procura-se relacioná-los de maneira crítica, comparando-os com o intuito de extrair suas vantagens e desvantagens. As lacunas ainda não exploradas são enfatizadas. Por fim, tabelas que comparam os principais aspectos dos artigos contidos na literatura são mostradas, através das quais pode-se ter uma compreensão qualitativa e visual das contribuições da presente tese.

### 2.1 FLUXO DE POTÊNCIA PARA SISTEMAS DE DISTRIBUIÇÃO

O problema de FP consiste em um conjunto de equações cuja solução informa as condições de regime permanente do sistema analisado. Tais equações são construídas tendo-se como referência uma topologia específica traduzida em modelos matemáticos de seus elementos elétricos, como linhas, cargas, gerações, transformadores e muitos outros (Ramadhani *et al.*, 2020). Existem inúmeras abordagens de FP, no que diz respeito à modelagem de rede e de seus elementos, forma de equacionamento e técnicas de solução. Deve-se avaliar criteriosamente quais são os fenômenos que se deseja representar e adotar as premissas que vão proporcionar os resultados mais adequados aos objetivos do usuário (Balamurugan; Srinivasan, 2011).

No caso de SD, existem uma série de aspectos que os diferenciam em relação a ST, como as redes primordialmente radiais, alta razão R/X das linhas de distribuição, a capacidade de mudança da topologia tendo em vista reconfigurações, ramais monofásicos e bifásicos, linhas não-transpostas e, principalmente, o alto desequilíbrio da carga (Heidarifar; Andrianesis; Caramanis, 2019). Formulações tradicionais de FP apresentam problemas de convergência quando tais aspectos são levados em consideração (Abbasi; Mohammadi, 2023). Portanto, para buscar resultados confiáveis e realistas, métodos específicos foram desenvolvidos para tratar do FP em redes de distribuição. Abbasi e Mohammadi (2023)

formularam um artigo de *review* no qual é proposta uma classificação dos métodos de FP direcionados a SD tendo-se como base três aspectos.

O primeiro aspecto relaciona-se com a técnica de solução. São consideradas três categorias: métodos iterativos, métodos não-iterativos e métodos evolucionários.

Os métodos iterativos englobam a técnica de varredura (também conhecida como *Forward/Backward Sweep Method* ou FBS), mas também métodos baseados em Newton-Rhapson (NR), Gauss-Seidel (GS) e métodos diretos. O método FBS promove bons resultados para sistemas radiais, além de ser relativamente simples de se implementar. Foi primeiramente proposto por Berg, Hawkins e Pleines (1967) e, desde então, muitos avanços foram feitos, incluindo aplicações que envolvem sistemas malhados (Shirmohammadi *et al.*, 1988), simulações de sistemas com cargas que dependem da tensão (Haque, 1996), aplicações com geração distribuída (Zhu; Tomsovic, 2002) e, finalmente, sua abordagem multifásica (Ciric; Feltrin; Ochoa, 2003). Segundo Abbasi e Mohammadi (2023), as principais desvantagens dos métodos iterativos são: sensibilidade quanto a estruturas de topologias, dependência da condição inicial, cálculo da matriz Jacobiana para cada iteração. Apesar do método FBS não requerer cálculo de Jacobiana, ele apresenta tempo de simulação muito alto nos casos de aplicações de grande porte, tornando-se pouco eficiente. O método GS é também insatisfatório para aplicações de larga escala, pois requer muitas iterações. Já o método direto só é aplicável quando as cargas são lineares, o que frequentemente não é o caso na prática (Abbasi; Mohammadi, 2023). Buscando maior flexibilidade e robustez, o Método de Injeção de Correntes Trifásico (MICT) foi proposto em Garcia *et al.* (2000). Nele, cada contribuição de corrente devido a cada elemento elétrico é calculada, o que permite um tratamento individualizado dos nós, facilitando a representação trifásica e desequilibrada. Em seguida, o método foi aprimorado para modelar sistemas de quatro condutores (Penido *et al.*, 2008) e, posteriormente, generalizado para sistemas de N condutores (Penido *et al.*, 2013). O Método de Injeção de Correntes a N Condutores (MICN) é muito flexível no que se refere à simulação de topologias, permitindo sistemas radiais ou malhados, com condutores de neutro e aterramentos, sendo recomendado para aplicações em que se deseja uma proximidade maior com a realidade (Penido *et al.*, 2013). Além disso, possui vantagens quanto à convergência computacional e tempo de simulação em relação ao FBS (Araujo *et al.*, 2010).

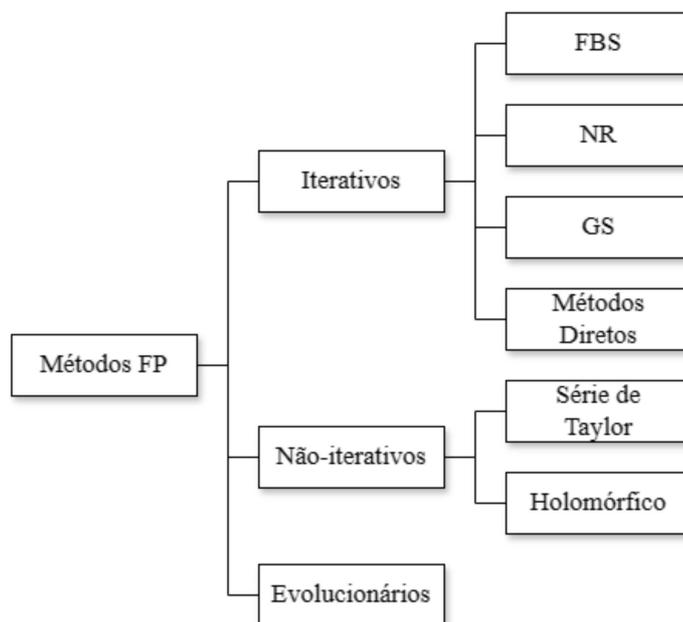
Os métodos não-iterativos incluem expansões em Série de Taylor e métodos holomórficos. Foram desenvolvidos para superar algumas das dificuldades citadas dos métodos iterativos (Galiana; Banakar, 1981). Particularmente, o método holomórfico elimina

as dependências de Jacobiana e de condição inicial. Em Sun *et al.* (2018), um procedimento holomórfico foi criado para resolver um FP trifásico que contém cargas ZIP e GDs. Em Asl *et al.* (2019), foi incluída a modelagem específica de geradores de indução no equacionamento holomórfico para melhor representar as GDs. Em Shamseldein (2022), simulações considerando SD malhados foram conduzidas através do método holomórfico.

Os métodos baseados em técnicas evolutivas como PSO (do inglês: *Particle Swarm Optimization*) e GA (do inglês: *Genetic Algorithm*) foram utilizadas em certo contexto, principalmente para superar a questão da condição inicial a que os métodos iterativos estão submetidos (Ting; Wong; Chung, 2008; Wong; Li; Law, 1997). Entretanto, essas técnicas são altamente dependentes de sua população inicial no processo evolutivo e a convergência local do FP não é garantida (Abbasi; Mohammadi, 2023). Por isso, os métodos tradicionais são vistos como mais seguros do que estes.

Na Figura 1, é apresentado um esquema didático da classificação dos métodos de FP para SD, de acordo com a técnica de solução.

Figura 1 – Classificação dos Métodos de FP para SD de acordo com a técnica de solução.

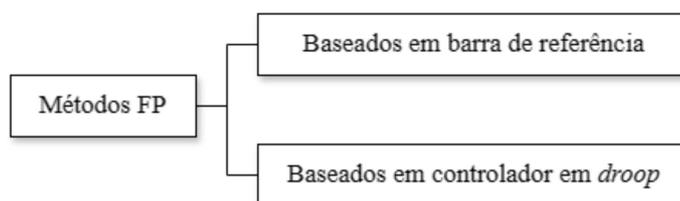


Fonte: elaborada pelo autor, baseada em Abbasi e Mohammadi (2023).

O segundo aspecto de classificação apontado por Abbasi e Mohammadi (2023) é segundo a forma de operação. Os SD podem operar de maneira conectada à rede ou de modo

isolado, o que pode ser alcançado com sistemas inteligentes ou microrredes. Segundo os autores, métodos convencionais não podem ser aplicados nesse último caso, porque a frequência e a tensão da barra de referência são assumidas como constantes. Dessa forma, é preciso adicionar equações que modelam o controle em *droop*, isto é, que relacionam a potência ativa dos geradores com a frequência da microrrede. Esses métodos ainda estão sob estudo, principalmente devido à escassez de algoritmos que considerem o desequilíbrio das redes. Na Figura 2, a classificação dos métodos segundo as formas de operação é exposta.

Figura 2 – Classificação dos Métodos de FP para SD de acordo com a forma de operação.



Fonte: elaborada pelo autor, baseada em Abbasi e Mohammadi (2023).

Por fim, o último aspecto considerado é em relação à consideração de incertezas no FP. Os métodos, então, podem ser determinísticos, estocásticos ou de inteligência artificial (Abbasi; Mohammadi, 2023). Os métodos determinísticos são os mais empregados em estudos de operação e planejamento (Singh; Moger; Jena, 2022). Neles, as variáveis de entrada do FP são assumidas como um conjunto fixo e, para cada mudança nessas condições, um novo problema de FP deve ser simulado. No FPD, a acurácia dos resultados depende da precisão dos dados informados (Abbasi; Mohammadi, 2023). Incertezas não são levadas em consideração (Singh; Moger; Jena, 2022). Todavia, no SD há muitas fontes de incerteza, como a variação da potência demandada, a configuração da rede, mudanças na geração, flutuações de potências geradas devido a fontes renováveis, como solar fotovoltaica e eólica, e, por fim, erros de modelagem e erros de medições (Abbasi; Mohammadi, 2023). A principal desvantagem do FPD reside no fato de que não é possível investigar todos os cenários possíveis de carga e geração em determinados contextos. Por causa disso, pode-se dizer que os resultados fornecidos pelo FPD são pouco confiáveis (Abbasi; Mohammadi, 2023).

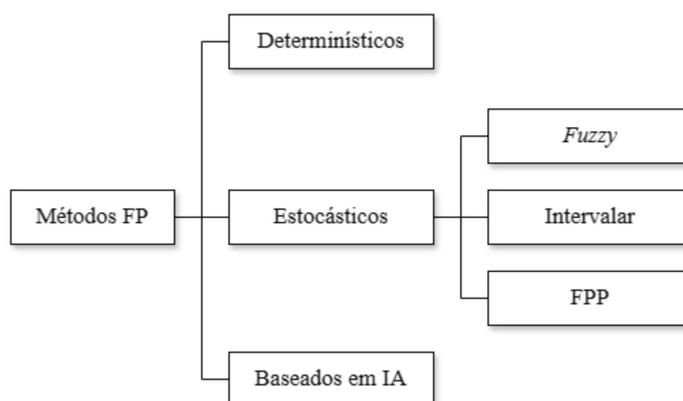
Os métodos estocásticos são aqueles que levam em consideração as incertezas dos parâmetros de entrada do FP. Sua formulação inicial foi em Borkowska (1974) e, desde então, muitos procedimentos foram implementados, incluindo lógica *fuzzy*, aritmética intervalar e, finalmente, o Fluxo de Potência Probabilístico (FPP). Este último é o tema delimitado da

presente tese, de maneira que a Seção 2.2 terá como enfoque específico a sua caracterização, seu histórico de desenvolvimento, principais métodos, suas vantagens e desvantagens, suas tendências e avanços.

Por fim, autores se dedicam às aplicações de métodos baseados em inteligência artificial para resolver o problema de FP em SD. Redes Neurais são treinadas tendo como referência uma série de condições operativas, incluindo contingências (Müller; Rider; Castro, 2010), de maneira que as variáveis de saída podem ser estimadas com base nos resultados fornecidos no treinamento.

Na Figura 3, as principais classificações quanto à inclusão de incertezas no FP são apresentadas de forma didática.

Figura 3 – Classificação dos métodos de FP para SD quanto à inclusão de incertezas.



Fonte: elaborada pelo autor, baseado em Abbasi e Mohammadi (2023).

Conforme mencionado, a presente tese se dedica ao estudo do FPP para SD. Dessa forma, na próxima seção estão contidos os referenciais teóricos e os pressupostos metodológicos desta técnica.

## 2.2 FLUXO DE POTÊNCIA PROBABILÍSTICO

O propósito do FPP é obter as condições de regime permanente de um sistema analisado tal que as incertezas dos seus parâmetros estão contidas em sua formulação. Dessa forma, as variáveis de entrada são modeladas por PDFs. Portanto, o processo do FPP engloba três etapas primordiais: (i) modelagem das incertezas das variáveis de entrada; (ii) solução do

modelo; (iii) determinação das variáveis de saída (Ramadhani *et al.*, 2020). Cada uma das etapas será explicada de maneira detalhada em seções particulares.

### 2.2.1 MODELAGEM DAS VARIÁVEIS COM INCERTEZAS

A primeira etapa, isto é, a modelagem das incertezas das variáveis de entrada, envolve a decisão sobre quais variáveis do FP serão consideradas aleatórias e quais serão suas modelagens. Essas escolhas devem ser feitas de maneira criteriosa segundo os objetivos almejados pelo usuário, pois uma escolha equivocada pode alterar significativamente os resultados (Abbasi; Mohammadi, 2023). Tradicionalmente, são consideradas como variáveis aleatórias de entrada as gerações e cargas. No caso de GDs renováveis, como geradores fotovoltaicos e eólicos, a consideração das incertezas torna-se crucial para a precisão dos resultados, devido às suas características inerentemente estocásticas associadas às condições climáticas (Ramadhani *et al.*, 2020). Além da definição das variáveis aleatórias, deve-se escolher também a função densidade de probabilidade que melhor modela os fenômenos que se deseja representar no FPP. Ramadhani *et al.* (2020) e Abbasi e Mohammadi. (2023) relacionam as variáveis mais vistas na literatura com suas funções de probabilidade adotadas, algumas dentre as quais estão presentes na Tabela 1<sup>1</sup>.

Tabela 1 – Variáveis aleatórias e suas PDFs mais adotadas.

Elemento elétrico	Variável aleatória	PDF escolhida
Carga	Demanda	Normal, lognormal binomial, beta, exponencial
GD fotovoltaica	Irradiação solar	Normal, Beta, Uniforme
	Temperatura	Normal
GD eólica	Velocidade do vento	Weibull
Bateria	Capacidade da bateria	Normal
	Demanda de carga/descarga	Normal, Weibull, Rayleigh

Fonte: (Ramadhani *et al.*, 2020) e (Abbasi; Mohammadi, 2023).

<sup>1</sup> Definições estatísticas e as descrições matemáticas completas das variáveis aleatórias empregadas na presente tese podem ser vistas com detalhes no Apêndice A.

Estudos atuais dedicam-se ainda a estabelecer correlações entre as variáveis adotadas. Ou seja, não basta definir quais são as variáveis aleatórias do problema, mas deve-se também estabelecer como essas variáveis se relacionam. Isso deve ser definido por estudos estatísticos, uma vez que a distribuição de probabilidade de uma variável pode afetar a distribuição de probabilidade de outra variável. Diz-se, então, que as variáveis são correlacionadas ou dependentes. Segundo Ramadhani *et al.* (2020), tais correlações afetam significativamente o resultado do FPP. Cargas podem estar correlacionadas devido às rotinas dos seres humanos, que definem um padrão comum de consumo. As gerações renováveis podem estar correlacionadas devido à localização geográfica das mesmas, uma vez que unidades mais próximas vão estar sujeitas às condições climáticas semelhantes. No entanto, cargas e gerações também podem estar correlacionadas. Tudo isso impacta nas incertezas do FPP e deve-se avaliar cuidadosamente quais serão as condições adotadas das variáveis. Os métodos de correlação mais adotados são correlação linear, decomposição *Cholesky* e cópulas estatísticas.

Geralmente, a correlação linear entre duas variáveis aleatórias é definida por sua covariância. De modo qualitativo, a covariância positiva indica que as variáveis correlacionadas aumentam conjuntamente e uma covariância negativa indica que as variáveis diminuem conjuntamente. Para duas variáveis aleatórias  $X$  e  $Y$ , a covariância, denotada por  $Cov(X, Y)$ , é dada por (2.1), onde  $\bar{x}$  e  $\bar{y}$  são as suas respectivas médias.

$$Cov(X, Y) = \sum_{i=1}^N \frac{(x_i - \bar{x})(y_i - \bar{y})}{N} \quad (2.1)$$

A covariância normalizada é chamada coeficiente de correlação (CC), dada por (2.2), onde  $\sigma_x$  e  $\sigma_y$  são os desvios padrões de  $X$  e  $Y$ , respectivamente.

$$CC(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y} \quad (2.2)$$

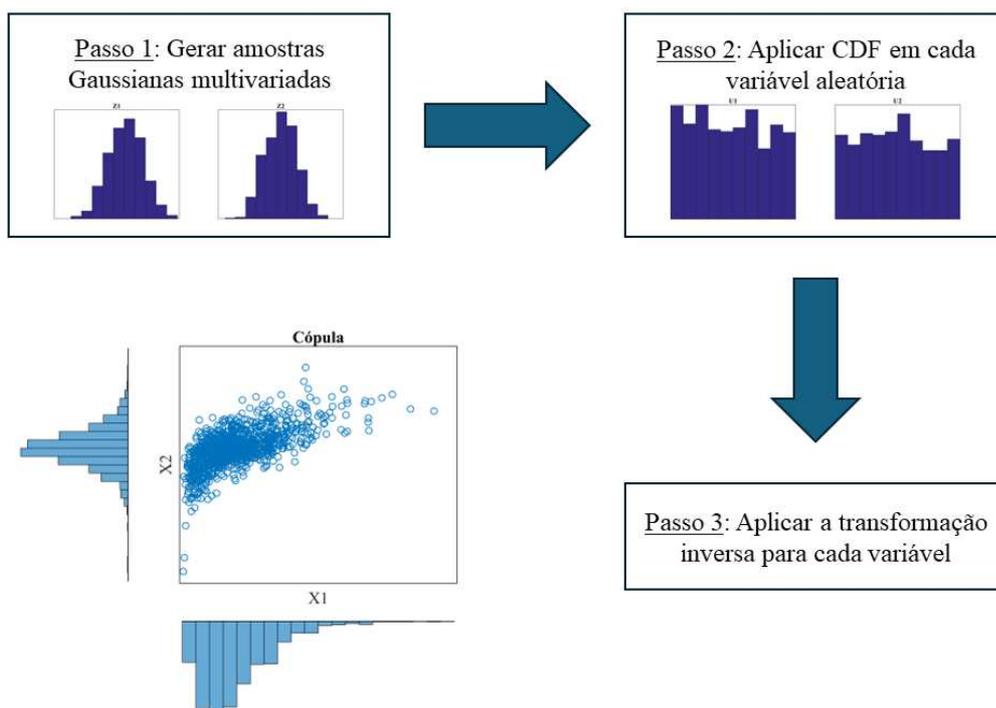
O CC indica que uma variação em uma variável aleatória tende a estar associada a uma oscilação, em mesma ou em oposta direção, em uma outra variável aleatória. Em alguns casos, as variáveis têm dependência não linear, de maneira que o CC sofre alterações chamadas *rankings*. Alguns exemplos de *ranking* são os CC de Spearman e de Kendall. A

dependência linear foi primeiramente implementada no FPP em Hoese e Garcés (1999) e em seguida foi aprimorada em Bhat, Prusty e Jena (2017).

Apesar dessa formulação ser muito popular devido a sua simplicidade de implementação, ela não consegue modelar cenários mais complexos de dependência entre variáveis aleatórias. A geração de correlações entre variáveis não-Gaussianas é particularmente desafiadora. Autores usam a teoria de cópulas estatísticas para alcançar tal fim.

Cópulas podem ser definidas como um procedimento estatístico que transforma distribuições marginais em distribuições conjuntas e correlacionadas. Trata-se de uma ferramenta capaz de descrever dependências entre variáveis e providenciam uma maneira de criar distribuições que modelam dados multivariados dependentes entre si.

Figura 4 – Fluxograma geral da teoria de cópulas.



Fonte: elaborada pelo autor (2025).

A geração de amostras correlacionadas, por meio do uso prático de cópulas estatísticas, para simulação de variáveis aleatórias dependentes segue três passos principais (Palahalli; Maffezzoni; Grusso, 2021), mostrados didaticamente na Figura 4 (com duas variáveis aleatórias, buscando melhor compreensão):

*Passo 1:* considere  $\mathbf{Z} = [Z_1 Z_2 \dots Z_n]$ , um conjunto de variáveis aleatórias  $Z_i$  que seguem funções Gaussianas multivariadas<sup>2</sup> tais que os CC são dados e conhecidos.

*Passo 2:* gere  $\mathbf{U} = [U_1 U_2 \dots U_n]$ , um novo conjunto de variáveis aleatórias  $U_i$  tal que  $\mathbf{U} = \Phi(\mathbf{Z})$ , onde  $\Phi$  é o operador função de probabilidade acumulada (ou CDF, do inglês: *Cumulative Density Function*). Note que a aplicação desse operador em  $\mathbf{Z}$  faz com que  $\mathbf{U}$  seja composto por distribuições uniformes.

*Passo 3:* obtenha  $\mathbf{X}$ , um novo conjunto de variáveis aleatórias, tal que  $\mathbf{X} = G^{-1}(\mathbf{U}) = [G_1^{-1}(U_1) G_2^{-1}(U_2) \dots G_n^{-1}(U_n)]$ , onde  $G_i$  é um operador relacionado a alguma distribuição desejada e  $G_i^{-1}$  sua transformação inversa. Note que  $G_i^{-1}$  aplicado em  $U_i$  gera, por definição, densidades de probabilidade que seguem a mesma  $G$ .

Em outras palavras,  $\mathbf{X}$  é uma distribuição multivariada de probabilidades tal que as correlações são mantidas como sendo iguais às informadas inicialmente. Além disso, as suas distribuições marginais são iguais às desejadas, e seguem os operadores  $G_i$  informados. Dessa forma, é possível gerar quaisquer distribuições correlacionadas<sup>3</sup>.

No FPP, cópulas foram usadas para estabelecer correlações entre cargas, geradores fotovoltaicos e baterias em Hernandez *et al.* (2017) e Ruiz-Rodriguez *et al.* (2018). Foi utilizada para correlacionar propriedades de geradores eólicos e fotovoltaicos em Yang *et al.* (2013). Em Palahalli, Maffezzoni e Gruosso (2021), a técnica foi usada para estabelecer correlações espaciais entre geradores fotovoltaicos em sistemas desequilibrados. Esses trabalhos mostraram a importância de se considerar distribuições multivariadas correlacionadas e o impacto que essa consideração traz aos resultados do FPP. Dessa maneira, é possível obter resultados mais precisos e mais próximos da realidade. Segundo Ramadhani *et al.* (2020), essa teoria é bastante flexível e pode ser aplicada em diversos métodos de FPP, os quais serão descritos na próxima seção.

### 2.2.2 SOLUÇÃO DO MODELO

O objetivo do FPP é obter as variáveis de saída com a contabilização das incertezas informadas nos parâmetros de entrada. Para isso, deve-se adotar um *método de solução*. O método diz respeito a como essa solução será obtida.

<sup>2</sup> Existem outros métodos de cópulas que não são provenientes de funções Gaussianas multivariadas, porém a abordagem mais empregada é a relatada na presente tese.

<sup>3</sup> A demonstração matemática relacionada aos passos da teoria de cópulas está presente no Apêndice A.

De maneira geral, pode-se dividir os métodos de FPP em três categorias principais: (i) métodos analíticos; (ii) métodos numéricos; e (iii) métodos de aproximação.

Os métodos analíticos resolvem o problema realizando operações aritméticas nas PDFs de entrada. Geralmente, é necessária uma simplificação matemática das equações de FP, como a linearização. O primeiro método formulado para solucionar o FPP foi visto em Borkowska (1974), no qual um procedimento analítico de convolução foi introduzido para tratar as incertezas das injeções ativas de potência. No entanto, esse tipo de abordagem apresenta uma grande desvantagem associada à enorme quantidade de cálculos, mesmo utilizando Transformada de Fourier, o que aumenta substancialmente o tempo computacional. Para solucionar esse problema, os métodos baseados em cumulantes foram implementados. A ideia é determinar os cumulantes de cada variável de entrada e, a partir disso, formar uma matriz que relaciona a sensibilidade das entradas com as saídas. Dessa maneira, pode-se obter diretamente os cumulantes das variáveis de saída e, finalmente, suas PDFs usando alguma técnica de aproximação ou séries polinomiais (Hasan; Preece; Milanovic, 2017). Inicialmente, os métodos analíticos não eram capazes de lidar com variáveis correlacionadas. Hoese e Garcés (1999) alteraram algumas premissas para conseguir simular problemas de variáveis interdependentes. Porém, até em estudos recentes, tais métodos não apresentam resultados precisos quando um número muito grande de variáveis aleatórias se faz presente, principalmente nos casos correlacionados. Além disso, a necessidade de simplificação das equações do FP que esses métodos frequentemente requerem pode prejudicar a acurácia das PDFs das variáveis de saída (Ramadhani *et al.*, 2020).

Os métodos numéricos são aqueles que se baseiam na simulação do FPD, dadas amostras das variáveis aleatórias de entrada. Um dos métodos mais comuns dessa categoria é o Método de Monte Carlo (MCS, do inglês: *Monte Carlo Simulation*). No MCS clássico, amostras aleatórias são selecionadas e problemas determinísticos são simulados para cada uma. Posteriormente, análises estatísticas são conduzidas de sorte a determinar as variáveis aleatórias de saída, tendo como referência as respostas individuais de cada amostra. A convergência do MCS pode se dar pelo número máximo de amostras ou pela estagnação de algum parâmetro de saída pré-selecionado. Sua maior desvantagem é o seu enorme esforço computacional, uma vez que são necessárias muitas simulações para alcançar resultados confiáveis (Ramadhani *et al.*, 2020). Por isso, alterações desse método foram propostas para aprimorar seu tempo de execução computacional. Essas modificações residem principalmente na forma pela qual a amostragem é feita. Em Yu *et al.* (2009) a técnica LHS (do inglês: *Latin*

*Hypercube Simulation*) é sugerida para realizar a amostragem. Ela é baseada em uma amostragem por intervalos, de maneira que não há amostras que cobrem um mesmo intervalo cuja discretização é pré-definida. Alcança-se, dessa maneira, uma melhor cobertura do domínio e uma redução do número de amostras necessárias para a convergência do FPP (Yu *et al.*, 2009). Outra possibilidade é a técnica UD (do inglês: *Uniform Design*), proposta em Cai *et al.* (2014), em que as amostras são uniformemente espalhadas pelo domínio. Outras abordagens foram propostas: o método *Quasi-Monte Carlo* (Wong *et al.*, 2013) e o MCS baseado em cadeias de Markov (Mori; Wenjun Jiang, 2009). As vantagens dos métodos numéricos são sua facilidade de implementação, sua flexibilidade quanto a variáveis aleatórias (são aceitas quaisquer PDFs e correlações) e sua grande confiabilidade. Por isso, é frequentemente utilizado para validar outros métodos. No entanto, sua maior desvantagem é seu tempo computacional demasiadamente alto, mesmo em suas variações de amostragem, o que o descredencia para diversas aplicações (Ramadhani *et al.*, 2020).

Por fim, a última categoria corresponde aos métodos por aproximação. Eles também são baseados na simulação determinística do FP para pontos de operação específicos. No entanto, ao invés de determinar esses pontos por amostragem, tal como nos métodos numéricos, são adotados procedimentos analíticos para tal fim. Logo, trata-se de um “meio termo” entre os métodos analíticos e numéricos. O método de FPP por aproximações mais comum é o Método da Estimativa dos Pontos (MEP), em que os principais pontos de concentração das variáveis aleatórias são calculados tendo como base os seus momentos estatísticos. Na literatura, existem vários esquemas associados ao MEP, como os de Hong (1998), chamados métodos  $2m$ ,  $2m+1$ ,  $3m$ ,  $4m+1$  e outros. Nesses esquemas, o parâmetro  $m$  é relacionado ao número de variáveis aleatórias do problema. Já o ‘nome’ do esquema revela o número de FPD necessários para se alcançar a convergência. Logo, nota-se de imediato a principal desvantagem do MEP, e dos métodos por aproximação em geral: o número de FPD necessários são diretamente ligados ao número de variáveis aleatórias do problema. Logo, aplicações de grande porte apresentam tempo computacional excessivamente alto e, em algumas situações, não há ganho em relação aos métodos numéricos usuais. Outras desvantagens são: os momentos de alta ordem são estimados com pouca precisão, a necessidade de alterar o esquema analítico quando se deseja alterar o número de pontos de concentração e a ausência de possibilidade de simular variáveis correlacionadas sem alteração prévia do método e de suas premissas.

Apesar de métodos modernos e atualizados terem sido desenvolvidos nos últimos anos, a maioria deles podem ser encaixados nessas três categorias. Outras técnicas podem ser destacadas, como expansões polinomiais (Ding *et al.*, 2017), Redes Neurais (Baghaee *et al.*, 2017) e *clustering* (Sadeghian *et al.*, 2020). O FPP via *clustering* é o tema central da presente tese, e seu referencial teórico será abordado de forma específica na Seção 2.4<sup>4</sup>.

### 2.2.3 DETERMINAÇÃO DAS VARIÁVEIS DE SAÍDA

Após a solução do modelo, muitas técnicas podem ser utilizadas para traçar as PDFs das variáveis de saída do FPP. A forma mais simples é definir *a priori* uma distribuição de probabilidade e calcular apenas os parâmetros da mesma, usando os dados de solução do modelo. Por exemplo: define-se que uma dada magnitude de tensão é *a priori* modelada através de uma distribuição normal, de maneira que basta calcular sua média e desvio padrão, que são extraídos da solução do modelo.

No entanto, existem métodos mais sofisticados para isso, visto que não necessariamente a variável que se deseja estudar segue o comportamento estabelecido *a priori*. No exemplo anterior, temos que não necessariamente tal tensão será normalmente distribuída, pois existem diversos fatores intrínsecos ao FP que podem alterar seu comportamento.

Para determinar de modo mais realista as PDFs das variáveis de saída, pode-se adotar séries que dependem dos momentos estatísticos, de maneira que a PDF a ser recuperada pode incorporar distorções causadas por esses momentos de ordem mais alta, o que permite resultados mais precisos do que a consideração anterior. Destacam-se as expansões Gram-Charlier, Corinsh-Fisher, Edgeworth e Legendre.

A expansão Gram-Charlier (EGC) consiste em uma série infinita que reconstrói uma PDF desconhecida com base em uma conhecida. De uma forma geral, ela pode ser expressa segundo (2.3), tendo como base o trabalho de Usaola (2009)<sup>5</sup>.

$$f(x) = \sum_{j=1}^{\infty} (c_j(\boldsymbol{\mu})H_j(x))\phi(x) \quad (2.3)$$

<sup>4</sup> Os desenvolvimentos matemáticos e os algoritmos detalhados dos métodos utilizados na presente tese como critério comparativo para o método proposto podem ser vistos com detalhes no Apêndice B. A saber: MCS, MEP 2m, LHS e FPP via K-Means.

<sup>5</sup> A demonstração matemática da EGC, assim como as fórmulas dos coeficientes e dos polinômios podem ser vistas com detalhes no Apêndice A.

Onde:  $f(x)$  é a PDF desconhecida que se deseja estimar;  $c_j$  é o  $j$ -ésimo coeficiente de Gram-Charlier, associado ao  $j$ -ésimo momento estatístico, que dependem dos momentos estatísticos denotados por  $\mu$ ;  $H_j$  é o  $j$ -ésimo polinômio de Chebyshev-Hermite, associado ao  $j$ -ésimo momento estatístico;  $\phi(x)$  é a distribuição normal,  $j \in \{1, \dots, \infty\}$ .

Geralmente essa série é truncada na terceira (Usaola, 2009) ou na quarta parcela (Barik; Das, 2019).

Alguns trabalhos abordaram a união de métodos com a recuperação analítica via EGC. Em Xu *et al.* (2017), o método dos cumulantes é aplicado para calcular os principais momentos das variáveis aleatórias de saída, e, posteriormente, as suas PDFs são estimadas via EGC. Em Barik *et al.* (2019), os momentos são estimados via MEP e as PDFs são recuperadas usando a mesma expansão. Logo, é uma técnica que é muito flexível e pode ser aplicada conjuntamente com outros métodos, aprimorando-os.

### 2.3 APLICAÇÕES RECENTES DO FPP EM SISTEMAS DE DISTRIBUIÇÃO

Apesar de muitos métodos e aplicações de FPP terem sido propostos na literatura, a sua grande maioria destina-se a simular ST. Muitos desses trabalhos estão compilados no *review* proposto por Singh *et al.* (2022). Não se pode garantir, no entanto, que tais procedimentos possam ser generalizados e replicados automaticamente para SD, uma vez que muitas considerações de modelagem específicas para esses sistemas não são incluídas, como a modelagem multifásica, o alto desequilíbrio das cargas, topologias predominantemente radiais, linhas assimétricas, ramais monofásicos e bifásicos, conexões não usuais de transformadores, capacitores monofásicos e desequilibrados, reguladores, entre outras (Martins; De Araujo; Penido, 2024; Penido *et al.*, 2013).

No entanto, alguns autores direcionaram seus estudos e desenvolveram formulações do FPP exclusivamente para SD. Alguns trabalhos relevantes são destacados e discutidos. Em Constante-Flores e Illindala (2019), o impacto da alocação de geradores fotovoltaicos em redes de distribuição é estudado, considerando-se a irradiação solar como uma variável aleatória que impacta indiretamente na potência produzida por esses geradores. Adota-se uma distribuição de probabilidade não Gaussiana para modelar esse parâmetro, baseada no estimador de *kernel*. Além disso, a carga e a velocidade dos ventos em geradores eólicos também são consideradas variáveis aleatórias. O método adotado de solução do FPP é o MCS, adotando-se, para tal, 10000 simulações determinísticas. Não há consideração de correlações

entre as variáveis aleatórias e o sistema adotado para validar o modelo é um sistema de pequeno porte de 33 barras, representado por sua sequência positiva.

Em Gruosso *et al.* (2020), uma técnica analítica baseada em uma expansão polinomial chamada *Polynomial Chaos Method* é proposta para solucionar o FPP, onde a variável aleatória considerada é a carga do sistema. Os autores consideram a modelagem da carga via funções Gaussianas. Além disso, adotam a representação multifásica da rede, através da utilização do *software* OpenDSS para executar os problemas de FP. Resultados probabilísticos de desequilíbrio de tensão são avaliados. Para validação do modelo, os autores simulam testes no sistema *IEEE European Low Voltage System*, que possui 906 nós e 55 cargas. No entanto, há uma consideração de que as cargas estão agregadas em três regiões e, dentro dessas regiões, as variáveis aleatórias são perfeitamente correlacionadas. Em regiões diferentes, elas são independentes. Logo, apesar do sistema simulado apresentar certa complexidade, há uma redução substancial do número de variáveis aleatórias consideradas por conta da simplificação adotada. Dessa maneira, a escalabilidade do método proposto não foi validada. Correlações não foram tratadas, tampouco variáveis aleatórias associadas a geradores renováveis.

Um método baseado em expansão polinomial também foi proposto em Chevalier *et al.* (2022). Chama-se *Neumann Series Expansion* e foi adotado para acelerar a convergência dos problemas determinísticos em um ambiente de FPP numérico. Simulações foram conduzidas em uma rede de distribuição de grande porte, IEEE 8500, onde a carga é considerada aleatória. Há a consideração da modelagem multifásica, onde a matriz de admitância nodal é construída no OpenDSS e exportada para alimentar o modelo de FP construído pelos autores. Assim como no trabalho anterior, os autores também levaram em conta uma série de simplificações: apesar do sistema possuir em sua topologia mais de 2000 cargas, apenas 25 foram assumidas como aleatórias e as demais permaneceram fixas. Não foram feitos testes com as demais cargas sendo aleatórias, o que prejudica a noção da precisão do método. Além disso, o FPP com um número grande de variáveis aleatórias não foi testado, comprometendo a escalabilidade do método proposto. Por fim, simulações considerando correlações não foram conduzidas e geradores renováveis não foram incluídos na modelagem.

O trabalho de Sun *et al.* (2022) propõe uma formulação analítica baseada em cumulantes para solucionar o FPP no qual o interesse é a avaliação do impacto de conversores que conectam geradores fotovoltaicos à rede. A irradiação solar é modelada por uma distribuição de probabilidade Beta, enquanto as cargas são modeladas por distribuições

normais. Correlações entre diferentes geradores fotovoltaicos são levados em consideração, adotando-se decomposição *Cholesky*. As PDFs das variáveis de saída são recuperadas usando-se a EGC. A validação é feita através de simulações tendo como referência a topologia IEEE 34 que, nesse trabalho, é modelada por sua sequência positiva. Logo, estudos de desequilíbrio e escalabilidade não foram feitos.

Em Sandoval *et al.* (2023), é proposta uma técnica analítica para solucionar o FPP em sistemas multifásicos onde as cargas são modeladas pela abordagem ZIP. As parcelas correspondentes à impedância, corrente e potência são modeladas cada uma por uma variável aleatória, o que aumenta a segurança dos resultados e os aproximam de situações práticas. As variáveis são consideradas correlacionadas temporalmente e estudos com GDs fotovoltaicas são abordados. A validação é feita utilizando o sistema-teste IEEE 37 barras. Logo, aplicações de grande porte e a escalabilidade do número de variáveis aleatórias não foram tratadas.

#### 2.4 FPP VIA *CLUSTERING*

Ao longo dos anos, pesquisadores formularam métodos para solucionar o FPP, buscando solucionar os problemas apontados dos procedimentos tradicionais. Primordialmente, é desejável que um método seja:

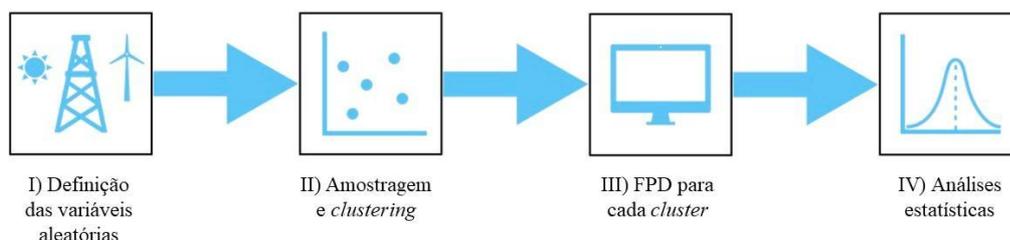
1. Acessível, isto é, que apresente uma implementação simples.
2. Flexível, ou seja, que seja apto a simular diferentes sistemas, com diferentes considerações de variáveis aleatórias no que diz respeito às suas modelagens e correlações, sem necessidade de alteração de seus pressupostos metodológicos ou adotar técnicas híbridas para tal (como a união de dois ou mais métodos).
3. Confiável, ou seja, que visa resultados precisos e abrangentes, que traduza situações mais realistas.
4. Escalável, isto é, que consiga performar de forma satisfatória considerando-se aplicações de grande porte e, principalmente, com um grande número de variáveis aleatórias.
5. Rápido, porque deseja-se que o método seja computacionalmente eficiente.

Segundo Abbasi e Mohammadi (2023), um método que reúne todos esses predicados ainda não foi desenvolvido.

Nos últimos anos, alguns autores adotaram formulações baseadas em *clustering*, pois eles são de implementação simples (Abbasi; Mohammadi, 2023), alcançando resultados precisos em um tempo computacional favorável (Sadeghian *et al.*, 2020). O processo do FPP

via *clustering* (FPPC) pode ser visto como um método por aproximações. As técnicas mais usadas no FPPC são as particionais, particularmente, K-Means e K-Medoids<sup>6</sup>.

Figura 5 – Fluxograma dos principais passos do FPPC.



Fonte: elaborado pelo autor (2025).

O FPPC é caracterizado por, de uma forma geral, seguir os principais passos mostrados na Figura 5 e descritos a seguir: (i) modelagem do sistema e das variáveis de entrada; (ii) amostragem das variáveis de entrada ou aquisição de dados reais e *clustering*; (iii) execução do FPD para cada *cluster*; (iv) análises estatísticas dos resultados.

Algumas vantagens a respeito do FPPC são a possibilidade de se trabalhar com dados reais, o que traduz situações muito práticas, o tempo computacional favorável e a enorme flexibilidade em termos de sistemas e variáveis de entrada.

As principais desvantagens são a necessidade de se informar o número de *clusters* antes da execução do FPPC; a definição da partição inicial pode levar a convergências ruins, e, principalmente, a *maldição da dimensionalidade*. Esta última é a principal desvantagem e se manifesta principalmente em algoritmos de *clustering* baseados em distâncias, como K-Means e K-Medoids. Ela ocorre, pois, ao aumentar muito a dimensionalidade dos dados a serem comprimidos, as distâncias calculadas no processo de *clustering* começam a ficar cada vez mais ineficientes, de maneira que o agrupamento dos dados fica comprometido. Isso significa que o FPPC vai se tornando cada vez mais impreciso à medida que o número de variáveis aleatórias de entrada aumenta.

Além disso, Wan (2022) aborda mais três motivos que prejudicam uma partição de algoritmos de *clustering* baseados em distâncias, principalmente K-Means: (i) essas técnicas

<sup>6</sup> K-Means e K-Medoids são duas técnicas de *clustering* particionais. Dado um conjunto de dados, ambas particionam os dados em K grupos, onde K é informado pelo usuário e o critério de afinidade são distâncias (que podem ser euclidianas ou não). A diferença reside nos centros de *clusters* obtidos em cada um dos métodos. Enquanto K-Means obtém *centroides* (pontos médios de cada grupo), K-Medoids apresenta *medoides*, que são pontos que realmente existem no banco de dados (Arora; Deepali; Varshney, 2016). Mais detalhes sobre a formulação matemática de ambos podem ser vistos no Apêndice C.

são muito sensíveis a *outliers* que podem existir nos dados de entrada; (ii) a convergência do método é local e depende substancialmente da partição inicial informada; (iii) o número de *clusters* informado pode não ser apropriado para comprimir eficientemente a base de dados informada.

Logo, dos cinco itens citados que caracterizam um bom método de FPP, a escalabilidade passa a ser um aspecto crítico do FPPC. Por isso, procedimentos e modificações que melhorem sua performance sob essas condições são necessários para o atual estado da arte.

## 2.5 APLICAÇÕES DO FPPC EM SISTEMAS DE DISTRIBUIÇÃO

Alguns trabalhos exploraram o FPPC e o usaram para simular especificamente SD. Em Galvani *et al.* (2019), o FPP é solucionado para averiguar o estado probabilístico de uma rede de distribuição onde estão presentes cargas não lineares aleatórias que produzem distorções harmônicas. A técnica K-Means é empregada para agrupar os dados de entrada e então o FPD é conduzido para cada centroide. Os resultados obtidos foram comparados com uma simulação tradicional de MCS, em que pode ser observada a precisão das funções densidade de probabilidade obtidas e também a redução do esforço computacional. Não foram conduzidos testes utilizando variáveis correlacionadas. Além disso, simulações com gerações renováveis intermitentes também não foram feitas. O sistema-teste empregado foi a topologia desequilibrada IEEE 37 barras. Apesar da modelagem ser multifásica, trata-se de um sistema de pequeno porte. A escalabilidade do método não foi testada, uma vez que um número grande de variáveis aleatórias não foi considerado. Portanto, a questão da dimensionalidade de problemas de *clustering* não foi tratada nem discutida.

Sadeghian *et al.* (2020) aprimoram a formulação de FPPC por meio de um refinamento dos valores encontrados de centroides. Para isso, várias execuções da técnica K-Means são realizadas para estimar a melhor partição possível. Esse ajuste resulta em melhorias práticas da precisão das variáveis de saída em um problema em que as potências de saída de geradores eólicos são assumidas como variáveis aleatórias. Testes são conduzidos em topologias equilibradas modeladas por sequência positiva. Estudos de correlação e escalabilidade também não foram abordados, o que não garante que o método pode ser eficiente em situações em que se deseja modelar um número muito grande de variáveis aleatórias.

Em Rouhani *et al.* (2022), uma técnica de *soft clustering* é empregada para solucionar o problema de FPP em redes de distribuição. As variáveis aleatórias incluem cargas, geradores fotovoltaicos e eólicos, além de veículos elétricos conectados à rede. Simulações com correlações espaciais e temporais são feitas e os testes são validados nas topologias desequilibradas IEEE 34 e IEEE 123. As correlações são estabelecidas via decomposição de *Cholesky*. No entanto, as condições e ajustes da técnica de *clustering* empregada são definidas através de simulações usando a primeira topologia citada e também comparando os resultados obtidos com um MCS pré-calculado. Posteriormente, tais definições se mantêm para a simulação de IEEE 123. Isso significa que o método é dependente de uma outra técnica que deve ser aplicada em antecedência, o que prejudica sua flexibilidade e, principalmente, sua generalidade. Além disso, sistemas maiores do que o IEEE 123 não foram simulados e as definições adotadas não necessariamente vão ser precisas para aplicações de grande porte. Por fim, o tempo computacional do método não apresenta ganhos tão expressivos em relação a outras técnicas analíticas, como a EGC. Comparações com outras técnicas de *clustering* também não foram feitas.

No trabalho de Salehi e Rezaei (2023), um algoritmo híbrido baseado em *clustering* e no MEP é proposto para estimar de modo mais preciso os momentos estatísticos e as PDFs das variáveis aleatórias de saída. As potências demandadas pelas cargas são assumidas como variáveis aleatórias. No entanto, apenas sistemas equilibrados de pequeno porte foram considerados nas simulações. Além disso, correlações entre variáveis não foram estudadas. Simulações considerando GDs não foram levadas em consideração. Esses aspectos enfraquecem a comprovação da robustez e da escalabilidade do método proposto. Por fim, o número selecionado de *clusters* foi conduzido de uma forma exaustiva, ou seja, simulando o mesmo problema com diferentes quantidades até que ocorra uma estagnação de um parâmetro de saída, o que aumenta substancialmente o tempo computacional demandado e compromete a generalidade do método.

No trabalho de Putra, Sarjiya e Setyonegoro (2024), a incerteza inerente aos geradores fotovoltaicos foi modelada no problema de FPP por meio de uma abordagem que mistura MCS e K-Means. No entanto, o sistema selecionado para a validação do método foi um alimentador equilibrado de pequeno porte, que não representa de forma fidedigna as características inerentes de SD, o que compromete a análise. Além disso, a escalabilidade de variáveis aleatórias não foi considerada, fator esse crucial para se testar adequadamente métodos baseados em *clustering*. Por fim, variáveis correlacionadas não foram tratadas.

Quanto às características implementadas do algoritmo de *clustering*, o número escolhido de *clusters* não foi justificado analiticamente, o que levanta dúvidas acerca da reprodutibilidade e precisão do método.

Alguns trabalhos exploram aplicações de otimização tais que as restrições envolvem as equações do FPPC. Em Wang *et al.* (2019), são buscadas as localizações e as capacidades de estações de carregamento de veículos elétricos, geradores renováveis e baterias. Os objetivos são minimizar o custo de investimento dessas ações de planejamento e os custos de confiabilidade, e maximizar a capacidade de carregamento dos veículos elétricos. Para solucionar esse modelo multi-objetivo, uma técnica de otimização metaheurística chamada *Multi-objective Neutral Aggregation Algorithm* é adotada. Os cenários probabilísticos dos geradores renováveis são criados utilizando-se a técnica de *clustering* K-Medoids. São consideradas correlações temporais dos dados dos geradores eólicos e fotovoltaicos. A formulação é testada em dois sistemas de pequeno porte equilibrados, um de 25 barras e outro de 54 barras.

Em Baructu *et al.* (2020), uma otimização para determinar a maior penetração de potência oriunda de geradores fotovoltaicos em SD é proposta. Os cenários de carga e geração são sorteados pelo MCS e depois comprimidos usando a técnica K-Means. Os cenários obtidos pelo *clustering* são os cenários a serem otimizados, empregando-se, para tal, o Método dos Pontos Interiores. O maior sistema simulado possui apenas 25 nós, de maneira que aplicações direcionadas a sistemas de grande porte não foram estabelecidas. Além disso, estudos que levam em consideração as funções de probabilidade das variáveis de saída não foram realizados.

Em Razaiean-Marjani *et al.* (2020), a alocação de unidades D-STATCOM (compensadores estáticos de potência reativa para SD) é proposta. A otimização considera variáveis aleatórias correlacionadas associadas às cargas e às gerações eólica e fotovoltaica. A técnica K-Means é usada para selecionar cenários pré-calculados pelo método LHS. A otimização é conduzida pela técnica metaheurística PSO. A validação do modelo ocorre por meio de simulações utilizando o sistema IEEE 69 barras, que é modelado por sua sequência positiva.

## 2.6 CONCLUSÕES PARCIAIS

Os SD possuem particularidades que inspiraram autores a formularem métodos específicos de FP. A modelagem multifásica é fundamental para representar adequadamente o alto desequilíbrio presente nesses sistemas. Além disso, a consideração de variáveis aleatórias é cada vez mais requisitada para alcançar resultados mais próximos da realidade, uma vez que as redes estão dotadas de equipamentos que aumentam a incerteza das condições de operação da rede, como geradores renováveis distribuídos e baterias.

O FPP é uma das principais ferramentas para lidar com incertezas no FPP. Em sua formulação, os seus parâmetros de entrada podem ser modelados por PDFs. Para solucionar tal modelo, empregam-se métodos que podem ser divididos em três grupos principais: métodos analíticos (baseados em operações aritméticas diretamente nas PDFs de entrada); métodos numéricos (baseados no sorteio de amostras das PDFs e na execução do FPD para cada amostra); métodos por aproximações (similar ao anterior, porém as amostras são calculadas previamente por métodos analíticos). Porém, um método que seja, ao mesmo tempo, acessível, seguro, escalável e rápido não foi ainda desenvolvido.

Buscando uma solução para isso, autores formularam métodos baseados em *clustering*. Usando bancos de dados reais ou amostras oriundas de PDFs, a ideia é comprimir as entradas e melhorar o tempo computacional em relação aos métodos numéricos tradicionais, mantendo sua precisão. Nos últimos anos, alguns trabalhos propuseram diferentes aplicações para o FPPC direcionados a SD e conseguiram alcançar alta precisão mediante um excelente tempo computacional. No entanto, existem ainda lacunas a serem tratadas, tanto no âmbito da simulação quanto no âmbito da modelagem do problema.

No âmbito de simulação, nota-se que a maioria dos trabalhos não consideram tratamento das principais causas de imprecisão do método de *clustering*: metodologias para refinar a partição inicial e a otimização do número de *clusters*. Sem ambos, resultados podem ser díspares (isto é, podem produzir partições muito diferentes de uma execução para outra) e de baixa qualidade (ou seja, a partição não representa estatisticamente a base de dados original), o que pode levar a erros consideráveis nos resultados das variáveis de saída. Dos trabalhos citados, apenas em Sadeghian *et al.* (2020) há o tratamento da partição inicial. Nenhum aborda a otimização do número de *clusters* para o FPP. Além disso, nota-se também que a maioria dos trabalhos que empregam o FPPC não têm a preocupação de recuperar as PDFs das variáveis de saída através de métodos analíticos, o que pode ocasionar equívocos em suas funções. Um maior rigor estatístico é requisitado para alcançar resultados mais fidedignos. Por fim, ressalta-se a importância de se formular métodos que sejam escaláveis,

principalmente no contexto de técnicas de *clustering*, e, sobretudo, no que tange aos métodos baseados em distâncias. Sabe-se que estes são afetados pela chamada ‘maldição da dimensionalidade’, em que eles perdem muita precisão quando do aumento do número de variáveis de entrada. No entanto, os trabalhos citados na literatura não abordam simulações que contemplem sistemas de grande porte munidos de um grande número de variáveis aleatórias. Portanto, não se sabe se tais algoritmos são escaláveis sem necessidade de alteração prévia de seus pressupostos metodológicos.

Tabela 2 – Comparações dos trabalhos recentes sobre FPPC em SD.

Trabalho	Condições de Simulação					Clustering	
	I	II	III	IV	V	VI	VII
Galvani <i>et al.</i> (2019)	✓						
Wang <i>et al.</i> (2019)		✓			✓		
Barutcu <i>et al.</i> (2020)	✓	✓			✓		
Razaeian-Marjani <i>et al.</i> (2020)		✓			✓		
Sadeghian <i>et al.</i> (2020)		✓				✓	
Rouhani <i>et al.</i> (2022)		✓			✓		
Salehi e Rezaei (2023)							
Putra <i>et al.</i> (2024)		✓					
Proposto	✓	✓	✓	✓	✓	✓	✓

Fonte: elaborada pelo autor (2025).

Legenda: o caractere ‘✓’ é usado quando o trabalho possui a característica em questão.

I – Modelagem multifásica.

II – Geração distribuída.

III – Sistemas de grande porte (sistemas com mais de 1000 nós).

IV – Escalabilidade das variáveis aleatórias (mais de 1000 variáveis aleatórias).

V – Correlações.

VI – Tratamento da partição inicial.

VII – Otimização do número de *clusters*.

No âmbito de modelagem, percebe-se que há uma escassez de trabalhos que adotam o FPPC para sistemas de representação multifásica. A maioria deles ainda apresenta sistemas modelados sem desequilíbrio, o que pode ocasionar severos erros. Posteriormente, aplicações que exploram variáveis aleatórias correlacionadas também são restritas. No âmbito de sistemas desequilibrados, são completamente ausentes. Essas considerações são importantes para que se possa obter resultados mais próximos da realidade e, com isso, tomadas de decisão mais seguras.

Nesse sentido, as lacunas observadas são tratadas na presente tese de maneira a aprimorar o FPPC no que se refere à sua validação, escalabilidade e modelagem.

De forma a resumir os principais trabalhos citados neste capítulo, na Tabela 2 estão contidos os seus principais aspectos de simulação e de modelagem. Ao final, são apresentados também estes aspectos em relação à presente tese, para estimular visualmente a comparação do estado da arte com os desenvolvimentos aqui propostos, assim como suas contribuições para a literatura.

### 3 FLUXO DE POTÊNCIA MULTIFÁSICO PROBABILÍSTICO BASEADO EM CLUSTERING E RECUPERAÇÕES ANALÍTICAS (FPMP-CRA)

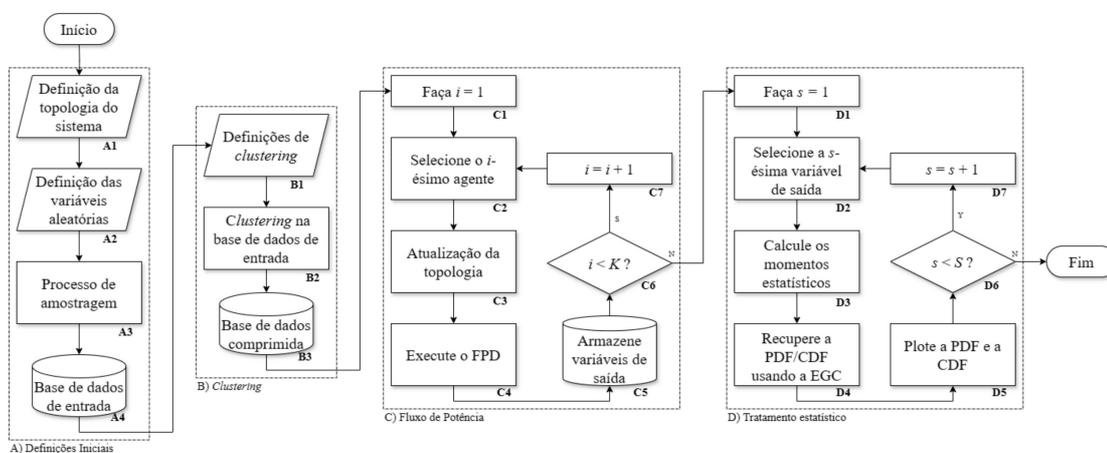
Este capítulo tem por objetivo demonstrar os principais passos e desenvolvimentos matemáticos do método proposto para avaliar o FPP em redes de distribuição ativas. Ele é denominado Fluxo de Potência Multifásico Probabilístico baseado em *Clustering* e Recuperações Analíticas (FPMP-CRA). Nele, técnicas de *clustering* são usadas para gerar cenários probabilísticos que serão avaliados numericamente pelo FPD multifásico. Finalmente, os momentos estatísticos das variáveis aleatórias de saída podem ser estimados e suas PDFs podem ser recuperadas usando-se a EGC. É desejado um procedimento que proporcione acessibilidade (ou seja, simples implementação), segurança e confiabilidade (isto é, que seja suficientemente preciso em relação a outras técnicas presentes na literatura), rapidez (que seja eficiente do ponto de vista do tempo computacional requerido para convergência) e escalável (que possa ser aplicado em sistemas de grande porte ou em sistemas que contenham um número elevado de variáveis aleatórias de entrada).

#### 3.1 ALGORITMO DO FPMP-CRA

O fluxograma com os principais passos compreendidos pelo FPMP-CRA pode ser visto por meio da Figura 6. Pode-se constatar que ele é formado por quatro grupos principais de passos a serem executados. Tais grupos são chamados de *módulos*. Cada um desses módulos tem uma função específica no âmbito do processo. De forma geral, a importância deles pode ser expressa como listado abaixo.

- A. Definições iniciais: onde se define qual sistema será simulado e quais serão as variáveis aleatórias adotadas e como serão suas modelagens. O processo de amostragem dessas variáveis também é conduzido nesse módulo.
- B. *Clustering*: o banco de dados original, formado pelas amostras obtidas no módulo A, é comprimido através de um processo de *clustering*.
- C. Fluxo de Potência: um problema de FPD é conduzido para cada *cluster* obtido no módulo B. Os resultados das variáveis aleatórias de saída para cada FPD são armazenados.
- D. Tratamento estatístico: com base nos resultados encontrados por meio dos problemas de FPD conduzidos no módulo C, são calculados os momentos estatísticos das variáveis de saída de interesse. Além disso, suas PDFs são recuperadas analiticamente usando a EGC.

Figura 6 – Fluxograma do FPMP-CRA.



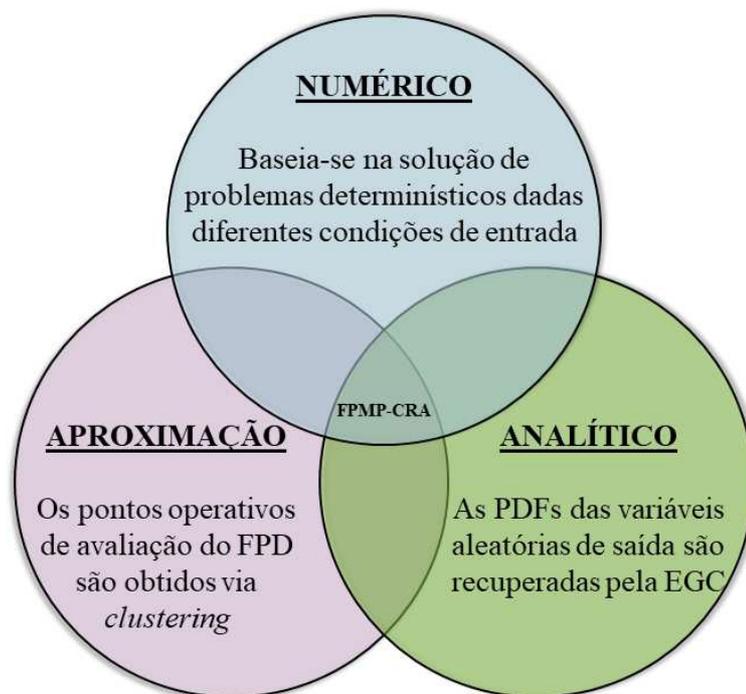
Fonte: elaborada pelo autor (2025).

Conforme mencionado nos Capítulos 1 e 2, a aplicação de técnicas de *clustering* aprimoram o desempenho computacional quanto ao tempo de convergência do método, diminuindo-se o número de cenários a serem avaliados. A precisão do método, no entanto, é garantida pela recuperação analítica das funções de probabilidade que, por sua vez, dependem dos momentos estatísticos das variáveis de saída. Dadas as características citadas do método, pode-se dizer que se trata de um método híbrido, possuindo tanto aspectos numéricos, de aproximação e analíticos. A natureza híbrida do método é mostrada de forma didática na Figura 7.

Pode-se dizer que o FPMP-CRA tem caráter numérico, pois envolve a solução de problemas determinísticos dadas diferentes condições das variáveis aleatórias de entrada. No entanto, não são determinadas amostras aleatórias para estimar o estado probabilístico do sistema, mas sim pontos aproximados, obtidos por meio de técnicas de *clustering*. Portanto, lhe é conferido caráter de um método por aproximações. Todavia, o cálculo das PDFs das variáveis de saída não é obtido diretamente do resultado das simulações determinísticas, mas sim os seus resultados são utilizados para estimar os principais momentos estatísticos (a saber: média aritmética, desvio padrão, coeficiente de assimetria e coeficiente de curtose<sup>7</sup>), e então as suas densidades de probabilidade podem ser recuperadas utilizando-se a técnica analítica EGC, conferindo ao método caráter analítico.

<sup>7</sup> O coeficiente de assimetria é aquele que informa o quão simétrica ou assimétrica é uma PDF. PDFs simétricas como a Gaussiana possuem coeficiente de assimetria nulo. Já o coeficiente de curtose informa a distorção da PDF em relação ao comportamento de uma Gaussiana. Mais detalhes sobre momentos estatísticos estão presentes no Apêndice A.

Figura 7 – Diagrama das características do FPMP-CRA.



Fonte: elaborada pelo autor (2025).

Logo, na sua confecção, o FPMP-CRA explora as principais vantagens de cada uma dessas categorias primordiais do FPP, de maneira a garantir um método flexível e eficiente.

Visando uma melhor compreensão do método, os passos de cada módulo mostrado na Figura 6 serão individualmente detalhados a seguir.

### 3.1.1 MÓDULO A: DEFINIÇÕES INICIAIS

O FPMP-CRA inicia-se em A1, onde é definido o sistema que se deseja simular. Com isso, todas as características de sua topologia também são definidas. Deve-se adotar criteriosamente a modelagem de rede de acordo com os desígnios da simulação. Neste trabalho, são enfocadas as análises em SD, de maneira que é priorizada a modelagem multifásica da rede. Dessa forma, fenômenos inerentes a esses tipos de sistema devem estar contemplados no equacionamento. Particularmente, destaca-se a inclusão do desequilíbrio do sistema. No entanto, podem ser considerados também ST e equacionamentos que considerem a rede como equilibrada e representadas por sua sequência positiva. Na Seção 4.2, os

resultados probabilísticos serão analisados sob a ótica da modelagem de rede e consequências da adoção de um modelo simplificado serão discutidas criticamente.

Em A2, são definidas quais são as variáveis do problema que tem seu comportamento aleatório. Na formulação desse método, quaisquer variáveis de entrada de um FP tradicional podem ser aleatórias, dependendo dos objetivos de simulação. Também deve-se fazer essa escolha com cuidado, pois os resultados dependerão fortemente dos pressupostos adotados. Neste trabalho, são escolhidas como variáveis aleatórias as gerações e cargas do sistema. Nesse passo, também são definidas as formas através das quais as variáveis selecionadas se relacionam. Elas podem ser independentes, isto é, quando a ocorrência de um evento relacionado a uma variável não interfere na ocorrência de um evento relacionado a uma outra variável, ou podem ser dependentes ou correlacionadas, caso contrário. Ressalta-se que ambas as possibilidades podem ser adotadas no presente método, aumentando sua flexibilidade em relação a outros, como o MEP, por exemplo, que não permite variáveis correlacionadas. Em A2, também são definidas as modelagens de cada variável aleatória. Elas podem ter seu comportamento definido diretamente por um banco de dados de medições reais ou podem ser modeladas por amostras oriundas de distribuições de probabilidade. Uma terceira possibilidade é a modelagem via distribuições de probabilidade cujos parâmetros são calculados por meio de valores reais medidos. No caso de variáveis correlacionadas, pode-se adotar distribuições multivariadas ou técnicas mais sofisticadas, como cópulas estatísticas.

Sendo as modelagens das variáveis definidas, em A3 é feito um processo de amostragem para obter um número  $n$  de cenários. Tal processo pode ser feito de forma aleatória, como no MCS, em que tais cenários são sorteados indiscriminadamente, ou outras técnicas podem ser usadas, como o LHS, por exemplo. Se forem consideradas variáveis correlacionadas, deve-se adotar uma técnica de amostragem condizente com esse pressuposto, como a decomposição *Cholesky*, por exemplo. No caso das cópulas, os vetores originados já podem ser vistos como os cenários amostrados.

Em A4, as amostras são armazenadas em uma única matriz, chamada *banco de dados original* ou *matriz original* (MO). Trata-se de uma matriz de dimensão  $n \times r$ , onde  $n$  é o número de cenários amostrados e  $r$  o número de variáveis aleatórias de entrada. Sua representação matemática é expressa por meio de (3.1), onde  $x_{ij}$  é a observação referente à  $j$ -ésima variável do  $i$ -ésimo cenário amostrado.

$$MO = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1r} \\ x_{21} & x_{22} & \dots & x_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nr} \end{bmatrix} \quad (3.1)$$

### 3.1.2 MÓDULO B: *CLUSTERING*

O segundo módulo visa empregar técnicas de *clustering* para comprimir a MO de amostragens obtida no módulo anterior. Neste trabalho, foram adotados dois algoritmos particionais baseados em distâncias: K-Means e K-Medoids<sup>8</sup>. Em ambos, divide-se iterativamente uma base de dados em grupos menores, chamados *clusters*, em que cada um deles é representado por um centro de *cluster*. No caso do K-Means, esse centro é chamado *centroide*, e trata-se de um ponto médio que minimiza as distâncias entre ele mesmo e os outros elementos dentro de um *cluster*. No caso do K-Medoids, é chamado de *medoide*, e trata-se de um ponto real contido na MO que também minimiza as distâncias entre os medoides e os demais pontos do *cluster*. Esses centros de *clusters* são chamados genericamente de agentes. Logo, agentes são os centroides, caso K-Means seja adotado, ou medoides, caso K-Medoids seja o método empregado. Nesse trabalho, o K-Means é utilizado para problemas com um pequeno número de variáveis aleatórias e um K-Medoids modificado é adotado para aplicações de grande porte. Isso é justificado pela já citada “maldição da dimensionalidade”, que faz com que algoritmos de *clustering* percam sua precisão à medida que o número de variáveis aumenta. Logo, modificações são imprescindíveis para garantir a escalabilidade do método de FPP proposto. Nesse sentido, é adotado o algoritmo K-Medoids CLARA (do inglês: *Clustering for Large Applications*)<sup>9</sup> para alcançar esse propósito nesses casos mais complexos.

Em B1 são escolhidas as definições do algoritmo de *clustering*. A saber: o método empregado; a métrica escolhida, ou seja, qual distância será definida para agrupar os dados (euclidiana, *manhattan*, *minkovsky*, entre outras); o procedimento adotado para definir a partição inicial e o número de *clusters*. Segundo Wan (2022), esses três quesitos são fundamentais para se obter uma boa partição. Por isso, foi desenvolvido um método particular para estimar o número ótimo de *clusters* para o FPMP-CRA. Esse método foi construído para

<sup>8</sup> Os algoritmos e desenvolvimentos matemáticos dos dois algoritmos de *clustering* utilizados nesse trabalho (K-Means e K-Medoids) podem ser vistos com detalhes no Apêndice C.

<sup>9</sup> O algoritmo K-Medoids CLARA é uma alteração do tradicional K-Medoids para ser aplicado em problemas de grande porte. Ele é baseado na procura exaustiva dos medoides considerando subgrupos aleatórios do banco de dados original a ser comprimido, o que melhora sua eficiência. Mais detalhes podem ser vistos no Apêndice C.

tratar especificamente do problema de FPPC, onde a definição do número ótimo de *clusters* é um fator crítico para se obter a precisão desejada, e é uma contribuição da presente tese. O desenvolvimento matemático deste método é apresentado em uma seção particular dada sua importância. Trata-se da Seção 3.2.

Após suas respectivas definições, o algoritmo de *clustering* escolhido é aplicado em B2 para comprimir a MO obtida no Módulo A. Dessa maneira, diminui-se o número de linhas de  $n$  para  $K$ , onde  $K \ll n$  e  $K$  é o número de *clusters* adotado. Obtém-se, portanto, uma *matriz comprimida* (MC), formada pelos  $K$  agentes calculados por meio do algoritmo de *clustering*. Sua representação matemática é expressa por (3.2), onde  $\alpha_{ij}$  é o elemento referente à  $j$ -ésima variável do  $i$ -ésimo agente de *clustering*.

$$MC = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{K1} & \alpha_{K2} & \dots & \alpha_{Kr} \end{bmatrix} \text{ tal que } K \ll n \quad (3.2)$$

Essa matriz é armazenada em B3, onde encerra-se o módulo B.

### 3.1.3 MÓDULO C: FLUXO DE POTÊNCIA

O Módulo C é responsável pela execução de problemas de FPD para cada agente que está contido na MC, obtida, por sua vez, no Módulo B. Em C1, inicia-se o contador  $i$  dos agentes em  $i = 1$ . Em C2, toma-se a  $i$ -ésima linha da MC, que corresponde ao  $i$ -ésimo agente. Ele é formado por observações para cada variável aleatória de entrada. Logo, são  $r$  valores selecionados. Em C3, a topologia do sistema é atualizada de acordo com o  $i$ -ésimo agente.

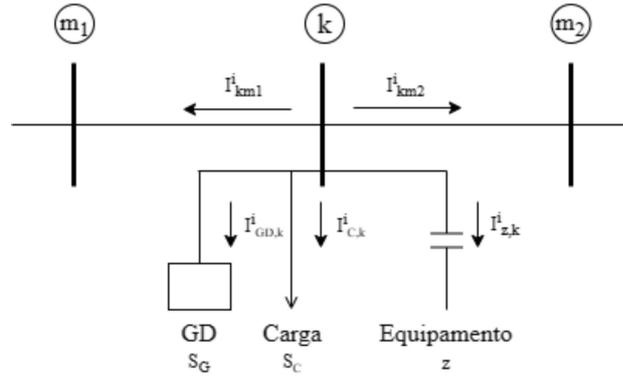
Por exemplo, se um problema de FPP possui duas variáveis aleatórias, sendo uma a potência ativa de uma carga e a outra a sua potência reativa, o valor de  $r$  é igual a 2. Por isso, cada agente possui dois valores, um correspondente à potência ativa e o outro relacionado à potência reativa. No  $i$ -ésimo cenário (ou *cluster*), a topologia será atualizada de acordo com o  $i$ -ésimo agente, que é a  $i$ -ésima linha da MC. Isso quer dizer que, em tal cenário, a potência ativa da carga será o valor contido na  $i$ -ésima linha e primeira coluna da MC. Similarmente, a potência reativa, nesse caso, será o valor contido na  $i$ -ésima linha e segunda coluna.

Em C4, é executado um problema de FPD considerando o  $i$ -ésimo cenário. Novamente, é ressaltado que se deve escolher criteriosamente um método de FP que esteja

consoante com os objetivos do usuário e com as modelagens de rede adotadas no módulo A. No presente trabalho, foi escolhido o MICN (Penido *et al.*, 2013) devido à sua capacidade de trabalhar de forma eficiente com sistemas multifásicos, tal como mencionado no Capítulo 2.

No MICN, todas as injeções de corrente correspondentes aos elementos elétricos são contabilizadas para cada nó, de maneira que se cria um sistema de equações não lineares de acordo com a Lei de Kirchhoff para os nós. As contribuições de corrente são escritas em coordenadas retangulares e as variáveis de estado são as componentes reais e imaginárias de todas as tensões nodais. Para melhor visualização, considere o desenho esquemático de um nó genérico  $s$ , localizado em uma barra genérica  $k$ , mostrado por meio da Figura 8.

Figura 8 – Esquema didático da representação de um nó no MICN.



Fonte: elaborada pelo autor (2025).

De acordo com a Figura 8, no  $i$ -ésimo cenário (ou *cluster*), a injeção total de corrente no nó representado é igual a  $I_{inj_k}^i = I_{km_1}^i + I_{km_2}^i + I_{GD,k}^i + I_{C,k}^i + I_{z,k}^i$ . Mais detalhadamente, a expressão mostrada na equação (3.3).

$$I_{inj_{k,s}}^i = \sum_{m \in \Omega_k} \sum_{t \in \alpha_P} (j b_{km}^{st} V_{k,t}^i + (V_{k,t}^i - V_{m,t}^i) y_{km}^{st}) + \left( \frac{S_{G_{k,s}}^i - S_{C_{k,s}}^i}{V_k^{s,i}} \right)^* + \frac{V_{k,s}^i}{Z_{k,s}^i} \quad (3.3)$$

Onde:

$V_{k,t}^i$  é a tensão da barra  $k$ , entre a fase  $t$  e terra, no cenário  $i$ ;

$V_{m,t}^i$  é a tensão da barra  $m$ , entre a fase  $t$  e terra, no cenário  $i$ ;

$b_{km}^{st}$  é a susceptância *shunt* da linha  $k$ - $m$ ;

$y_{km}^{st}$  é a admitância série da linha  $k-m$ ;

$S_{G_{k,s}}^i$  é a potência complexa gerada por uma GD conectada à barra  $k$ , fase  $s$ , no cenário  $i$ ;

$S_{C_{k,s}}^i$  é a potência complexa demandada por uma carga conectada à barra  $k$ , fase  $s$ , no cenário  $i$ ;

$z_k^s$  é a impedância de um equipamento conectado à barra  $k$ , fase  $s$ , como capacitores e indutores, por exemplo;

$\Omega_k$  é o conjunto de todas as barras conectadas à barra  $k$ ;

$\alpha_p$  é o conjunto de todas as fases do sistema, ou seja  $\{a,b,c,n\}$ .

Finalmente, a solução do FPD para o  $i$ -ésimo cenário (ou *cluster*) é obtida através do sistema de equações construído por meio da contabilização de todas as injeções de corrente para todos os nós, igualando-as a zero. Isso é expresso matematicamente pelas equações (3.4) e (3.5).

$$\text{Re}[I_{inj}^i(\mathbf{V}_{Re}^i, \mathbf{V}_{Im}^i, \mathbf{P}^i, \mathbf{Q}^i)] = \mathbf{0} \quad (3.4)$$

$$\text{Imag}[I_{inj}^i(\mathbf{V}_{Re}^i, \mathbf{V}_{Im}^i, \mathbf{P}^i, \mathbf{Q}^i)] = \mathbf{0} \quad (3.5)$$

Após a solução do FPD, todas as variáveis de saída de interesse são armazenadas em C5. Tais variáveis consistem nas tradicionalmente calculadas em um FP, como tensões em magnitudes e ângulos, correntes em magnitudes e ângulos, potências ativas e reativas que transitam nas linhas, perdas ativas ou quaisquer outras que se possa calcular por meio do equacionamento exposto.

Em seguida, averigua-se se todos os  $K$  cenários foram percorridos. Caso não seja verdade, incrementa-se o contador em C6 e o processo retorna ao passo C2, onde o próximo agente será tomado. Caso todos os agentes tenham sido simulados, o módulo C se encerra.

### 3.1.4 MÓDULO D: TRATAMENTO ESTATÍSTICO

O módulo D é o responsável pelo tratamento estatístico dos resultados obtidos no módulo anterior. Em D1, o contador de variáveis aleatórias de saída é inicializado em  $s = 1$ .

Em D2, são selecionados os resultados da  $s$ -ésima variável de saída, calculadas por meio dos FPD do módulo anterior.

Em D3, são calculados os momentos estatísticos dessa variável. Seja  $Y$  uma variável aleatória de saída e sejam  $Y_i$  as respostas dessa variável para cada agente  $i$  simulado, para todo  $i \in \{1, \dots, K\}$ . O momento bruto de primeira ordem de  $Y$  é dado por (3.6) e também é conhecido como média aritmética, e denota a tendência central, ou valor esperado, dessa variável aleatória.

$$M_1 = \sum_{i=1}^K \xi_i Y_i \quad (3.6)$$

Onde  $\xi_i$  é o fator de peso de  $Y_i$ , dado pela razão do número de elementos contidos no *cluster*  $i$ , denotado por  $n_i$ , pelo número total de elementos, denotado por  $n$ . Isto é  $\xi_i = n_i/n$ .

Os momentos centrais de ordem  $p$  são dados por (3.7).

$$m_p = \sum_{i=1}^K \xi_i (Y_i - M_1)^p \quad (3.7)$$

O momento central de ordem 2 tem importância particular e é chamado de variância da variável aleatória. Ela denota a dispersão da mesma em torno de sua tendência central. A raiz quadrada da variância é chamada de desvio padrão, indicada por  $\sigma$ .

Os momentos centrais normalizados de ordem  $p$  são calculados por meio de (3.8).

$$\mu_p = \frac{m_p}{\sigma^p} \quad (3.8)$$

Os momentos centrais normalizados de ordem 3 e 4 têm importância particular, e são chamados de coeficiente de assimetria e curtose, respectivamente. O primeiro mensura a simetria da PDF da variável aleatória e o segundo está atrelado à distorção da forma da mesma em relação a uma Gaussiana<sup>10</sup>.

Por meio desses momentos, é possível estimar a PDF da variável aleatória de interesse usando a técnica analítica EGC, em D4. Conforme já mencionado no Capítulo 2, trata-se de uma expansão polinomial infinita em que cada parcela está associada a um momento

---

<sup>10</sup> Mais detalhes sobre os momentos estatísticos podem ser acessados no Apêndice A.

estatístico. Logo, ela é capaz de recuperar a PDF considerando-se esses momentos de ordem maior que podem, porventura, distorcer seus gráficos e impactar nos resultados. Por isso, essa etapa é de suma importância, pois garante a precisão das funções estimadas. A expressão analítica da EGC até a quarta ordem é mostrada em (3.9)<sup>11</sup>.

$$f(x) \cong \left[ 1 + \frac{1}{3!} \mu_3 (\bar{x} - 3\bar{x}) + \frac{1}{4!} (\mu_4 - 3)(\bar{x}^4 - 6\bar{x}^2 + 3) \right] \phi(x, m_1, \sigma) \quad (3.9)$$

Onde  $\bar{x} = (x - m_1)/\sigma$  e  $\phi(x, m_1, \sigma)$  é uma PDF Gaussiana de média  $m_1$  e desvio padrão  $\sigma$ .

A CDF da variável aleatória pode ser obtida por integração numérica da expressão (3.9). Em D5, a PDF e a CDF da variável de saída de interesse são plotadas. Em D6, é averiguado se todas as variáveis de saída de interesse tiveram suas PDFs e CDFs estimadas e plotadas, respectivamente. Caso não seja verdade, incrementa o contador em D7 e passe para a próxima variável em D2. Caso contrário, o processo do FPMP-CRA é encerrado.

### 3.2 MÉTODO PARA ESTIMAR NÚMERO ÓTIMO DE *CLUSTERS*

Segundo Wan (2022), existem três fatores que prejudicam o desempenho de um algoritmo de *clustering* de partições baseado em distâncias: a sua sensibilidade quanto a *outliers*; a definição aleatória de sua partição inicial e a definição inadequada do número de *clusters*. Logo, segundo o mesmo autor, a definição correta do número de *clusters* é fundamental para se obter resultados precisos das partições. Quando uma quantidade de agentes é muito inferior ao requisitado, geralmente a MC não representa estatisticamente a MO. No caso concreto do FPP, isso leva a resultados imprecisos das PDFs das variáveis de saída. Portanto, isso pode conduzir usuários a tomarem decisões equivocadas de planejamento ou operação.

#### 3.2.1 ESTADO DA ARTE

Existem três formas básicas de estimar um número adequado de *clusters*. A primeira refere-se à simulação do problema várias vezes com quantidades aleatórias de *clusters*. Aqui,

---

<sup>11</sup> Mais detalhes acerca da sua demonstração matemática podem ser conferidos no Apêndice A.

essa abordagem será chamada de ‘método exaustivo’. No caso do FPP, é a alternativa mais vista na literatura, como em Sadeghian *et al.* (2020). Ou seja, simula-se o FPP  $N$  vezes, aumentando-se progressivamente o número de *clusters*, até que o resultado probabilístico de uma variável de saída analisada tenha sua estagnação. No entanto, essa abordagem apresenta uma série de problemas. Primeiramente, ela requer um tempo computacional exageradamente alto devido às inúmeras simulações que devem ser feitas. Em segundo lugar, para estimar um valor muito preciso de *clusters*, deve-se utilizar um intervalo muito discretizado, aumentando-se demasiadamente o número  $N$ . Como isso geralmente não é feito, é muito comum que os resultados sejam sobredimensionados. Em Sadeghian *et al.* (2020), por exemplo, são testados 8 grupos de *clusters*, variando de 25 em 25, até um limite de 200 *clusters*. Apesar de a simulação final apresentar grande diferença de tempo computacional em relação ao MCS (no artigo é reportado que a abordagem de *clustering* levou aproximadamente 1,60 segundo contra 17,24 segundos do método numérico, considerando a topologia IEEE 118 barras), quando são somados todos os tempos computacionais demandados até se chegar no número ideal de 100 *clusters*, é visto que a diferença não é tão grande assim (14,40 *versus* 17,24 segundos<sup>12</sup>). Logo, tal abordagem exaustiva se apresenta como pouco efetiva, considerando também que o número de *clusters* varia de acordo com a aplicação desejada.

A segunda forma básica de se estimar o número de *clusters* é através de métodos que levam em consideração as distâncias *intra-clusters*. Ou seja, distâncias dos elementos de um *cluster* até o seu agente (centroide/medoide) mais próximo. O método mais comum é o conhecido como “método do cotovelo”, e alguns trabalhos utilizam essa técnica, como em Singh *et al.* (2023), onde o FPP é simulado para avaliar o estado de ST. Neste método, calcula-se a distância total *intra-cluster*, isto é, a distância de todos os pontos até seus centros de *cluster* mais próximos, à medida que o número de *clusters* aumenta, traçando-se uma curva que tem um formato semelhante ao de um cotovelo. O número adequado de *clusters* será aquele no qual a partir dele as distâncias *intra-clusters* não variam significativamente. Apesar de ser o método mais comum para avaliar o número requerido de *clusters* em muitas aplicações, Schubert (2023) argumenta que ele apresenta desvantagens substanciais: primeiramente, ele é muito eficiente para bancos de dados cujos pontos se acumulam em regiões bem separadas umas das outras. Mas, em casos de dados sobrepostos, o algoritmo não

---

<sup>12</sup> Esse tempo computacional não foi apresentado no trabalho de Sadeghian *et al.* (2020). Porém pode ser estimado através de uma aproximação do tempo computacional médio para a convergência do FPD, dados os tempos demandados totais informados e o número de *clusters*.

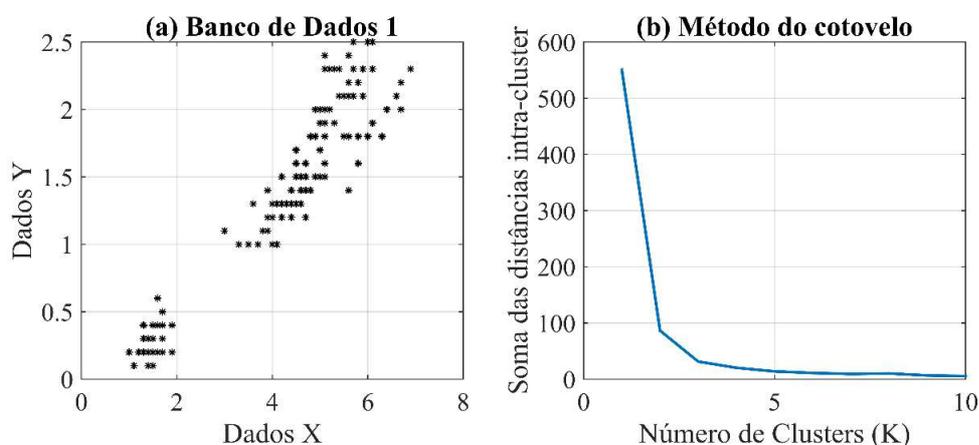
é tão eficiente, pois em sua própria curva do cotovelo a escolha do número de *clusters* se apresenta como ambígua. Isso pode ser visto nas comparações mostradas nas Figuras 9 e 10.

A Figura 9 (a) mostra um banco cujos dados não são tão sobrepostos e são relativamente bem espaçados no plano. Isso faz com que o método do cotovelo aplicado a ele, tal como mostrado na Figura 9 (b) aponte sem maiores dúvidas que o número ideal de clusters é igual a 5, uma vez que, após essa quantidade, quase não há variação nas distâncias *intra-cluster*. Segundo Schubert (2023), este é o comportamento esperado de uma “curva do cotovelo”.

Por outro lado, a Figura 10 (a) apresenta um banco de dados normalmente distribuído. Aqui, os dados tendem a se sobrepor mais em relação à situação anterior. Por isso, o método do cotovelo, cujos resultados são vistos na Figura 10 (b) são inconclusivos. Não se pode apontar com clareza qual é o número de *clusters* indicado, pois não há um valor que cause uma diminuição tamanha nas distâncias *intra-cluster*.

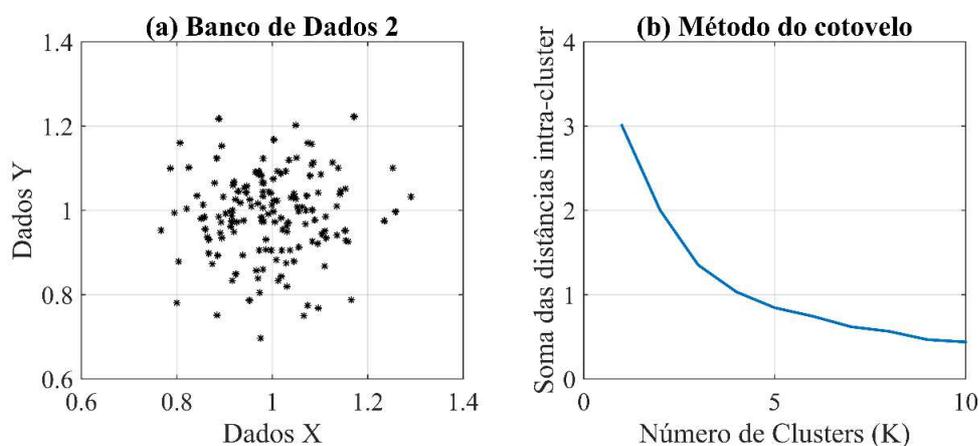
Schubert (2023) argumenta que isso faz com que tal método não seja eficiente para lidar com dados distribuídos de maneira mais abrangente, como dados uniformemente distribuídos ou normalmente distribuídos. Isto posto, esse método não é visto como adequado para estimar número de *clusters* para o problema de FPP, pois a maioria dos dados têm suas distribuições de probabilidade definidas de maneira que eles se tornam muitas vezes sobrepostos, o que faz com que o método do cotovelo proporcione um número de *clusters* ambíguo.

Figura 9 – Aplicação do método do cotovelo para banco de dados tutorial.



Fonte: elaborada pelo autor (2025).

Figura 10 – Aplicação do método do cotovelo para dados normalmente distribuídos.



Fonte: elaborada pelo autor (2025).

A terceira forma mais comum para estimar o número de *clusters* é por via de métodos que considerem tanto a distância *intra-cluster* quanto a distância *inter-cluster*. Ou seja, são levadas em conta não só as distâncias dos pontos até seus centros, mas também as distâncias entre os próprios *clusters*. Apesar de essa abordagem não ser comum na aplicação do FPP, na literatura diz-se que ela é mais segura e abrangente em relação ao método do cotovelo. Particularmente, destaca-se o “método da silhueta” ou “método da sombra” (conhecido em inglês como *silhouette method*). Nele, busca-se minimizar as distâncias *intra-clusters* e maximizar as distâncias *inter-clusters* por meio da avaliação de “coeficientes de silhueta” para cada dado, garantindo, dessa forma, semelhança para elementos dentro de um mesmo *cluster* e maior diferença para elementos de *clusters* diferentes.

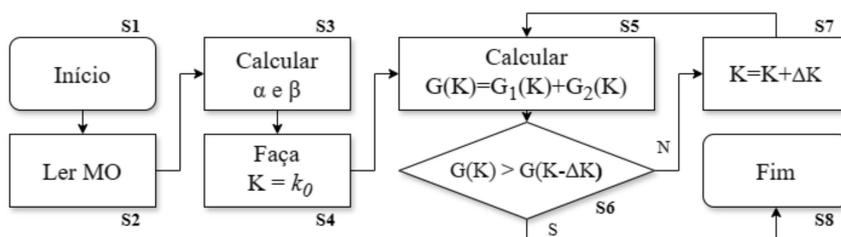
No caso do FPP, essa abordagem também incorre em grandes desvantagens: deve-se adotar um intervalo de *clusters* pré-definido em que esses coeficientes de silhueta serão avaliados e isso pode ser muito custoso computacionalmente, especialmente para bancos de dados de alta dimensionalidade (Martins; De Araujo; Penido, 2024). A principal desvantagem é em relação à curva de silhueta, que não é unimodal, o que força o usuário a simular todos os pontos informados *a priori* para se obter o melhor coeficiente de silhueta.

### 3.2.2 MÉTODO PROPOSTO

Tendo em vista as vantagens e desvantagens de todas as principais abordagens para estimar o número ótimo de *clusters*, busca-se, no presente trabalho, uma formulação que seja

benéfica para o problema do FPP em específico. Em outras palavras, busca-se um procedimento que seja eficiente computacionalmente como o “método do cotovelo”, mas que seja abrangente como o “método da silhueta”. O procedimento desenvolvido é uma contribuição da presente tese e é eficiente para estimar o número de *clusters* para o FPFC. Seus principais passos podem ser vistos na Figura 11.

Figura 11 – Fluxograma do método proposto para estimar número ótimo de *clusters*.



Fonte: elaborada pelo autor (2025).

O método se baseia no cálculo das distâncias *intra-cluster* e *inter-cluster*, assim como no método da silhueta. Porém, busca-se uma formulação que estabeleça uma curva de avaliação convexa (ou, no mínimo, *quasi-convexa*), para melhorar o desempenho computacional em relação à técnica citada.

Em S1, o procedimento é iniciado. Em S2, a MO que contém os dados a serem comprimidos é lida. Recorda-se que essa matriz é obtida no módulo A do algoritmo do FPMP-CRA. Em seguida, são calculados dois parâmetros importantes  $\alpha$  e  $\beta$ , em S3.

O parâmetro  $\alpha$  consiste na distância total de todos os pontos em relação a um ponto médio de MO. Em outras palavras, é a distância *intra-cluster* total no caso em que há apenas um único *cluster*. Seu cálculo pode ser visto na expressão (3.10), onde  $\mu$  é o vetor médio de MO.

$$\alpha = \sum_{x \in MO} \|x - \mu\|^2 \quad (3.10)$$

Já o parâmetro  $\beta$  consiste na distância total de todos os pontos de MO, tomados dois a dois. Em outras palavras, é a distância total *inter-cluster* no caso em que o número de *clusters* é igual à dimensão da própria MO. Seu cálculo é expresso em (3.11).

$$\beta = \sum_{\substack{i,j \in \{1, \dots, n\} \\ i \neq j}} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (3.11)$$

Em S4 é iniciado o contador de *clusters*  $K = k_0$ . É recomendado que se comece em  $k_0 = 2$ . Em S5, são calculadas duas funções  $g_1$  e  $g_2$ , que, por sua vez, dependem da partição dos dados da MO em  $K$  *clusters*. Logo, após a aplicação da técnica de *clustering* adotada, são colhidas duas informações:

- $g_1(K)$ : é a distância total *intra-cluster* de todos os pontos em relação aos seus centros de *cluster* mais próximos. Trata-se do mesmo parâmetro visto no “método do cotovelo” tradicional. Isto é, a distância total *intra-cluster* considerando-se  $K$  *clusters*.
- $g_2(K)$ : é a distância total entre todos os  $K$  centros de *clusters*, tomados dois a dois. Em outras palavras, a distância total *inter-cluster*, considerando-se  $K$  *clusters*.

As expressões de ambas as funções estão representadas em (3.12) e (3.13), respectivamente.

$$g_1(K) = \sum_{i=1}^K \sum_{j \in \Omega_i} \|\mathbf{x}_{ij} - \boldsymbol{\mu}_i\|^2 \quad (3.12)$$

$$g_2(K) = \sum_{\substack{i,j \in \{1, \dots, K\} \\ i \neq j}} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2 \quad (3.13)$$

Onde:  $\mathbf{x}_{ij}$  é o vetor que denota o  $j$ -ésimo elemento contido no  $i$ -ésimo *cluster*;  $\boldsymbol{\mu}_i$  é o vetor centro do  $i$ -ésimo *cluster*;  $\Omega_i$  é o conjunto de elementos contidos no  $i$ -ésimo *cluster*.

Dessa forma, são observadas algumas considerações importantes:

- $g_1(K = 1) \cong \alpha$ , pois trata-se justamente da distância *intra-cluster* entre todos os elementos da MO em relação a um único centro de *cluster*, justamente a definição de  $\alpha$ .
- $g_1(K = n) \cong 0$ , pois cada ponto da MO será aproximadamente igual a um centro de *cluster*, de maneira que a distância *intra-cluster* total torna-se aproximadamente zero.
- $g_2(K = 1) \cong 0$ , pois trata-se da distância entre dois elementos iguais, uma vez que há apenas um centro de *cluster*.

- $g_2(K = n) \cong \beta$ , pois trata-se da distância *inter-cluster* total entre todos os elementos da MO, que é justamente a definição de  $\beta$ .

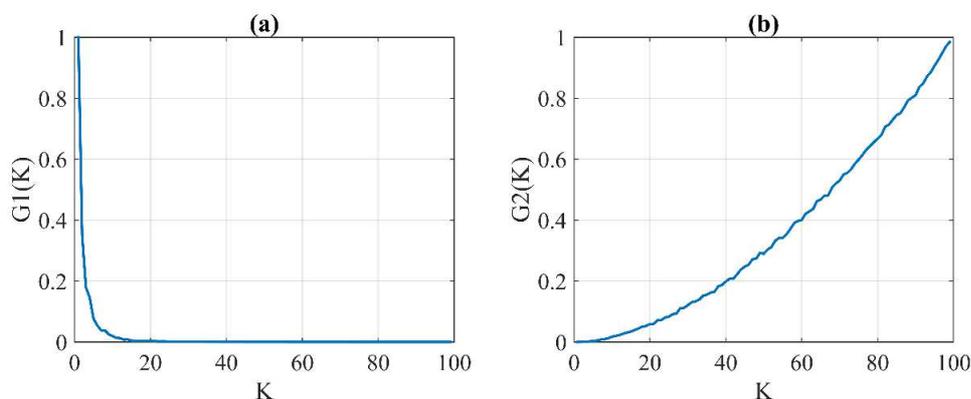
Logo, pode-se dividir  $g_1(K)$  por  $\alpha$  e  $g_2(K)$  por  $\beta$  a fim de normalizar tais curvas.

Denota-se  $\frac{g_1(K)}{\alpha} = G_1(K)$  e  $\frac{g_2(K)}{\beta} = G_2(K)$ .

Agora, nota-se que  $G_1(K)$  é uma curva que contém as distâncias normalizadas *intra-clusters* e trata-se de uma curva que varia de 1 até sua convergência em 0, tal como mostrado na Figura 12 (a). Já  $G_2(K)$  é uma curva que contém as distâncias normalizadas *inter-clusters* e trata-se de uma curva que varia de 0 até sua convergência em 1, tal como revelado por meio da Figura 12 (b).

Na Figura 12, ambas as curvas são obtidas considerando-se uma MO tutorial formada por 100 amostras de uma distribuição normal. O intervalo analisado de *clusters*, portanto, foi de 1 a 100.

Figura 12 – Gráficos típicos para  $G_1(K)$  e  $G_2(K)$ .



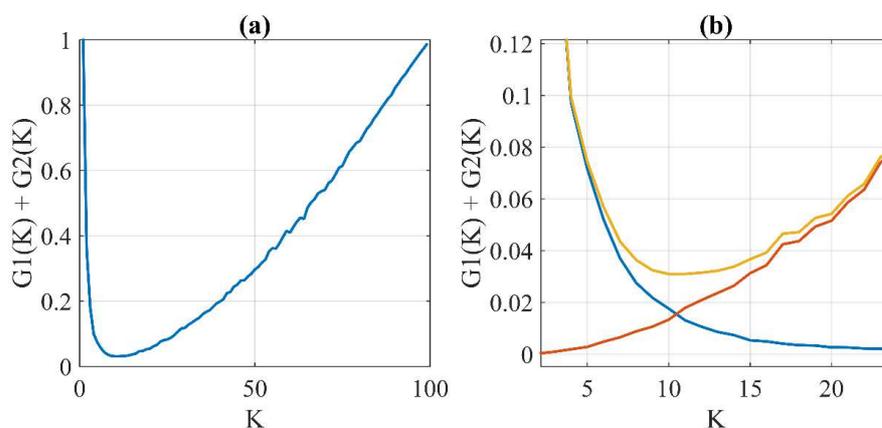
Fonte: elaborada pelo autor (2025).

Ora, já que as curvas estão normalizadas, é possível somá-las, calculando-se uma única curva  $G(K) = G_1(K) + G_2(K)$ , em S5. O comportamento dessa função resultante pode ser visto na Figura 13 (a). Já na Figura 13 (b) é apresentado um *zoom* da função perto do seu ponto de mínimo global.

Algumas considerações podem ser observadas. Em um primeiro momento, a função decresce, o que significa que  $G_1(K) > G_2(K)$ . Em outras palavras, as distâncias *intra-clusters* são maiores do que as distâncias *inter-clusters*, o que significa que as partições não são precisas o suficiente. Em um segundo momento, a função atinge um ponto de mínimo global

e começa a crescer. Nesse momento,  $G_2(K) > G_1(K)$ , o que significa que as distâncias *intra-clusters* já são reduzidas o suficiente e as distâncias entre os *clusters* já são superiores às distâncias dos elementos dentro dos *clusters*. Isso, de acordo com a teoria do método da silhueta, indica que a partição é bem-sucedida. Em outros termos, a MC representa estatisticamente a MO.

Figura 13 – Curva resultante  $G(K) = G_1(K) + G_2(K)$ .



Fonte: elaborada pelo autor (2025).

Note que, à medida que o número de *clusters* cresce, maior será o tempo computacional requerido pelo FPMP-CRA convergir, pois tal quantidade indica o número de FPD necessários para estimar os momentos das variáveis aleatórias de saída. Então, é preciso determinar o menor número de *clusters* possível tal que há a garantia de que a partição será bem-sucedida.

Ora, sob o ponto de vista do método desenvolvido, o número indicado de *clusters* será justamente o ponto de mínimo global da curva  $G(K)$ , pois, nesse ponto, as distâncias *intra-clusters* são pequenas o suficiente e, além disso, é o ponto no qual as distâncias *inter-clusters* são superiores às *intra-clusters*, garantindo uma partição bem-sucedida. Ressalta-se que valores maiores do que ponto de mínimo, denotado a partir de agora como  $K^*$ , podem resultar em partições mais precisas, porém o tempo computacional do FPMP-CRA também sofrerá acréscimos. Logo, a configuração em  $K^*$  é aquela que garante a melhor qualidade de solução mediante o tempo computacional favorável.

Para melhorar ainda mais o tempo computacional de execução do método, pode-se avaliar o número de *clusters* até que tal ponto de mínimo global seja atingido. Essa condição é

verificada em S6. Se o ponto de mínimo for atingido, o processo se encerra em S8. Caso contrário, uma quantidade  $\Delta K$  de *clusters* é adicionada em S7, e a função  $G(K + \Delta K)$  é avaliada para essa nova configuração.

### 3.3 CONCLUSÕES PARCIAIS

Foi desenvolvido um novo método de FPP para ser aplicado em SD desequilibrados. Ele é chamado Fluxo de Potência Multifásico Probabilístico Baseado em *Clustering* e Recuperações Analíticas (FPMP-CRA). Nele, são consideradas formulações numéricas e analíticas para garantir maior robustez, eficiência computacional e escalabilidade do método. Primeiramente, as variáveis aleatórias de entrada são definidas e amostradas através de técnicas convencionais numéricas. Em segundo lugar, algoritmos de *clustering* são executados para comprimir o número de amostras. Posteriormente, um FPD é executado para cada centro de *cluster* obtido, armazenando-se as respectivas variáveis de saída de interesse. Em seguida, os principais momentos estatísticos dessas últimas são estimados tendo como base os resultados determinísticos para cada *cluster*. Finalmente, as PDFs das variáveis são recuperadas utilizando-se a técnica analítica EGC.

Para que o método alcance sucesso, é necessário informar com antecedência o número correto de *clusters*. Métodos convencionais podem ser divididos em procedimentos exaustivos, procedimentos *intra-cluster* e procedimentos *inter-cluster*. Neste trabalho, é sugerida uma formulação que se baseia no cálculo dos dois últimos citados, garantindo que os centros de *clusters* obtidos representam estatisticamente o banco de dados original mediante um bom tempo computacional.

## 4 RESULTADOS OBTIDOS PELA APLICAÇÃO DO FPMP-CRA

Neste capítulo, os principais resultados alcançados mediante a aplicação do método proposto, isto é, o FPMP-CRA, são expostos, além de discussões críticas acerca dos mesmos. Primeiramente, são informadas as condições gerais de simulação. Em seguida, os resultados e análises referentes a cada um dos testes são apresentados em seções individuais.

### 4.1 OBJETIVOS E CONDIÇÕES GERAIS DE SIMULAÇÃO

Os testes são realizados mediante dois grupos de objetivos gerais:

1. **Objetivos técnicos:** avaliar criticamente os resultados do FPMP-CRA considerando-se situações próprias a sistemas desequilibrados, variando-se as condições das topologias como modelagens de redes, desequilíbrios, cargas, presença ou ausência de GD renovável, correlações entre variáveis aleatórias, neutros e aterramentos, entre outras. Muitas dessas situações ainda não foram consideradas na literatura e a simulação desses casos é imprescindível para enriquecer o estado da arte de métodos probabilísticos para sistemas multifásicos, além de justificar a robustez do método.
2. **Objetivos computacionais:** demonstrar os benefícios do método proposto sob a perspectiva de sua implementação e do tempo computacional demandado, destacando sua eficiência e comparando com outros métodos presentes na literatura, como o MCS e o MEP. A análise da escalabilidade do método torna-se fundamental para verificar a robustez do método desenvolvido, para justificar aplicações mais próximas de sistemas reais e para averiguar questões relacionadas à “maldição da dimensionalidade”, citada nos Capítulos 2 e 3. Um método robusto e eficiente deve apresentar boa precisão para situações nas quais um número alto de variáveis aleatórias é introduzido, o que é altamente requisitado para o atual contexto de métodos probabilísticos.

Logo, cada teste tem seus objetivos específicos que se condensam nos dois objetivos gerais citados. Para melhor organização do texto, cada um é apresentado em uma seção particular, em que são introduzidas as características da topologia adotada, as condições específicas de simulação, além dos objetivos próprios de cada uma delas. De modo resumido, os testes com seus respectivos intuitos estão listados na Tabela 3.

Tabela 3 – Breve resumo dos testes realizados e seus objetivos.

Topologia	Condições de Simulação	Objetivos	Variáveis aleatórias
IEEE 33	Simular o FPMP-CRA considerando-se dois modelos de rede: um baseado em sequência positiva e outro usando a abordagem multifásica, com diferentes níveis de desequilíbrio.	Averiguar criticamente o impacto da modelagem multifásica no âmbito probabilístico de análise de redes de distribuição.	Cargas.
IEEE 4	Simular o FPMP-CRA considerando-se cargas com diferentes correlações espaciais entre variáveis aleatórias de entrada.	Averiguar criticamente o impacto nas PDFs das variáveis de saída das correlações entre cargas.	Cargas.
IEEE 13	Simular o FPMP-CRA com diferentes números de variáveis aleatórias de entrada. Propor diretrizes para redução de dimensionalidade.	Averiguar criticamente o impacto do número de variáveis aleatórias na quantidade requerida de <i>clusters</i> , tempo computacional do método e precisão de resultados. Testar diferentes técnicas de amostragem e <i>clustering</i> .	Cargas.
IEEE 123	Simular o FPMP-CRA com inserção de GDs fotovoltaicas e considerando correlações temporais.	Averiguar criticamente o comportamento do FPMP-CRA quando variáveis aleatórias relacionadas a GDs modeladas por funções	Cargas e irradiação solar.

---

		não Gaussianas são inseridas; análises dos resultados probabilísticos quanto a violações de tensão em redes de distribuição ativas.	
IEEE NEV	Simular o FPMP-CRA em sistemas cujos condutores de neutro e aterramentos são modelados explicitamente.	Averiguar criticamente como se comportam as PDFs das tensões de neutro e utilizar o FPMP-CRA como ferramenta para acessar segurança de tensão de neutro.	Resistências de aterramento.
IEEE 8500	Simular o FPMP-CRA considerando um sistema de altíssimo porte.	Averiguar criticamente a escalabilidade do método proposto. Comparar com outros métodos.	Cargas.

---

Fonte: elaborada pelo autor (2025).

Em todos os testes, as cargas são representadas por seu modelo “potência constante”.

Todos os casos foram simulados em um computador de processador Intel®Core™ i7-7200 U, 2.50 GHz CPU com dois núcleos, 8GB RAM. Os códigos foram implementados em Matlab®.

#### 4.2 TESTE 1: ESTUDO DA MODELAGEM DE REDE NO ÂMBITO DO FPP

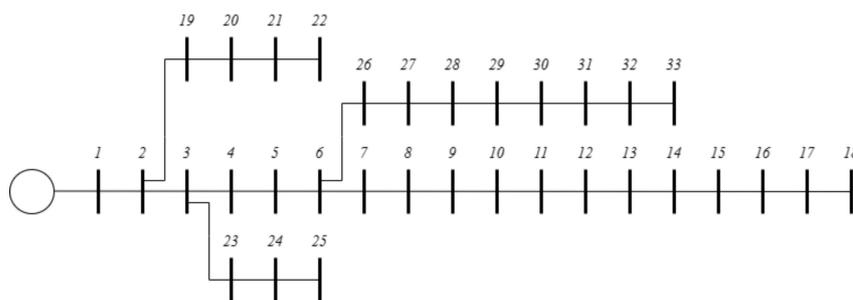
Neste primeiro teste, o objetivo é averiguar quais os impactos que a modelagem de rede ocasiona no cálculo probabilístico do estado de SD. Em outras palavras, busca-se verificar a diferença de resultados das funções de probabilidade das variáveis de saída quando modelagens simplificadas para representar as redes de distribuição são adotadas, em contraste

com uma modelagem multifásica. Para efetuar os testes, é utilizada a topologia IEEE 33 barras, que é um sistema muito empregado na literatura para simulação de métodos probabilísticos na distribuição.

#### 4.2.1 APRESENTAÇÃO DA TOPOLOGIA IEEE 33 E CONDIÇÕES DE SIMULAÇÃO

A topologia unifilar do sistema-teste IEEE 33 Barras pode ser vista na Figura 14. Trata-se de um sistema radial que contém 33 barras e 32 cargas trifásicas, sendo seu carregamento total originalmente igual a  $3715 + j2300$  kVA. Os dados completos do sistema podem ser acessados em Baran e Wu (1989).

Figura 14 – Sistema IEEE 33 Barras.



Fonte: elaborada pelo autor (2025).

Trata-se de um sistema originalmente modelado pela sua sequência positiva, o que significa que não há consideração do desequilíbrio inerente aos SD. Apesar de ser uma modelagem simplificada, muitos autores a empregam para efetuar simulações. No âmbito probabilístico, ainda é muito comum que sistemas como esse sejam usados para validação dos métodos, como em Constante-Flores e Illindala (2019), Sadeghian *et al.* (2020), e Uniyal e Sarangi (2021). Conforme visto no Capítulo 2, a desconsideração do desequilíbrio pode ocasionar sérios erros nos resultados das variáveis de saída do FP, o que leva a decisões equivocadas de planejamento e operação das redes. Partindo dessa premissa, busca-se, por meio da presente simulação, atestar qual o impacto da modelagem de rede também no âmbito probabilístico, visto que não são poucos os autores que ainda utilizam estas simplificações nos cálculos de métodos dessa natureza.

Para averiguar criticamente os resultados utilizando-se diferentes modelagens de rede, foram definidos três casos a serem examinados sob a perspectiva probabilística:

- Caso A: o sistema é considerado equilibrado, tal como a topologia originalmente se apresenta e tal como os autores citados a empregam nas suas validações dos métodos probabilísticos. Em outras palavras, as cargas são distribuídas de forma igualitária entre as fases.
- Caso B: o sistema é concebido de uma maneira multifásica, de maneira que as cargas são distribuídas de maneira desigual entre as fases. Neste caso, as cargas conectadas na fase A são aumentadas em 10%, enquanto as cargas conectadas na fase C são reduzidas em 10%. Note que o valor total do carregamento continua o mesmo do Caso 1, alterando-se apenas a distribuição das cargas entre as fases e, conseqüentemente, o desequilíbrio do sistema.
- Caso C: as condições são as mesmas das vistas no Caso B, porém é sugerido um desequilíbrio extremo. As cargas conectadas na fase A são aumentadas em 50% e as cargas conectadas na fase C são reduzidas em 50%. Embora seja uma situação crítica, que inclusive deve-se evitar em planejamentos de redes, esse cenário de altíssimo desequilíbrio é incluído para analisar criticamente o desempenho da modelagem de rede no âmbito probabilístico.

Em relação às condições das variáveis aleatórias de entrada, as cargas são consideradas incertas em todos os casos. Não há a inclusão de geração distribuída porque o objetivo desta primeira simulação é verificar a precisão dos resultados probabilísticos tendo em vista diferentes modelagens de rede. As cargas são consideradas perfeitamente correlacionadas, o que significa que todas elas têm seu comportamento atrelado a uma única distribuição de probabilidade. Em outras palavras, a variável aleatória empregada são fatores multiplicativos que se aplicam igualmente a todas as cargas dentro de um mesmo cenário, mantendo-se a proporção original entre elas. Isso garante que o desequilíbrio original dos casos é mantido, o que permite a avaliação do modelo da rede, que é o objetivo primordial da presente simulação. Alternativas serão estudadas nas seções seguintes, como variáveis independentes e correlacionadas, dependendo dos objetivos de cada simulação.

Para modelar a carga, é utilizada uma distribuição normal cuja média é 1,00 pu (ou seja, já assumindo valores normalizados) e seu intervalo de confiança adotado varia de aproximadamente 0,90 a 1,10, isto é, admite-se no máximo 10% de variação da carga. Isso resulta em um desvio padrão de 0,032 pu. Frisa-se que as cargas são representadas por seu modelo “potência constante”.

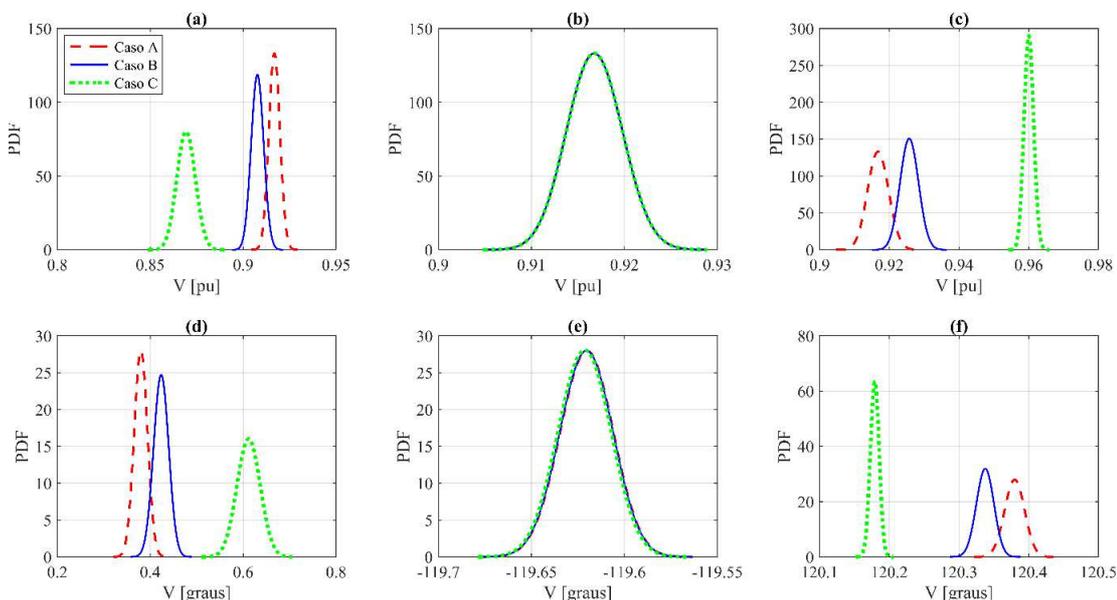
São coletadas 1000 amostras aleatórias dessa PDF para construir o banco de dados original que alimenta o FPMP-CRA. A compressão dos dados é feita através do método K-Means e o número ótimo de *clusters* é determinado usando o procedimento mostrado na

Seção 3.2.2. Como a base de dados de entrada é a mesma para todos os casos, o número ótimo de *clusters* é mantido o mesmo para as três simulações. Ressalta-se que apenas o carregamento base é alterado nos casos, sendo que a aleatoriedade da carga se manifesta de maneira igual, porém o desequilíbrio será diferente devido à diferente disposição das cargas originais entre as fases de acordo com o caso.

#### 4.2.2 IEEE 33: RESULTADOS DO FPMP-CRA

O método desenvolvido FPMP-CRA foi aplicado ao sistema IEEE 33 sob as condições expressas na Seção 4.2.1. Foi obtido um número ótimo de 33 *clusters*. O procedimento de otimização do número de *clusters* levou 0,84 segundo para convergir. O processo completo do FPMP-CRA levou 9,83 segundos para convergir. Na Seção 4.7, as comparações a respeito do tempo computacional em relação a outros métodos serão exploradas minuciosamente.

Figura 15 – IEEE 33: Resultados de tensão. (a) Barra 33, fase A (magnitude); (b) Barra 33, fase B (magnitude); (c) Barra 33, fase C (magnitude); (d) Barra 33, fase A (ângulo); (e) Barra 33, fase B (ângulo); (f) Barra 33, fase C (ângulo).



Fonte: elaborada pelo autor (2025).

As PDFs de tensão, em magnitude e ângulo, obtidas após a recuperação analítica via EGC são mostradas na Figura 15. Por questões de conveniência de aproveitamento do espaço

da presente tese, optou-se por mostrar os resultados relacionados à barra 33, mas as conclusões podem ser generalizadas para quaisquer posições. Na Figura 15 (a), (b) e (c), os resultados das PDFs de magnitude de tensão estão demonstrados, enquanto nas figuras (d), (e) e (f) são revelados os ângulos. Os resultados dos Casos A, B e C estão representados nas cores vermelho, azul e verde, respectivamente.

É possível perceber que as funções de probabilidade obtidas são muito distintas, dependendo do caso.

Quanto à magnitude, os resultados para o Caso A são exatamente iguais para as três fases, o que é esperado, pois neste caso o sistema é tratado como equilibrado. No entanto, à medida que o desequilíbrio aumenta, as funções de probabilidade passam a se comportar de maneira muito diferente em relação ao caso simplificado. Considerando 10% de desequilíbrio, por exemplo, nota-se que a maior parte dos cenários de fases A e C não são levados em consideração quando a modelagem equilibrada é adotada (pois a curva em azul está quase totalmente à esquerda da curva vermelha, na fase A, ou quase totalmente à direita, na fase C). Em outros termos, quando há um desequilíbrio considerável no sistema, a maior parte da densidade de probabilidade das tensões não é constatada quando uma abordagem equilibrada é utilizada. Como os SD são altamente desequilibrados, a consideração de uma modelagem multifásica é imprescindível para contabilizar corretamente os resultados probabilísticos das tensões. Analisando-se as respostas obtidas no Caso C, essa discrepância se apresenta como crítica: enquanto na modelagem equilibrada quase não são detectadas tensões abaixo de 0,90 pu para a fase A, a modelagem desequilibrada aponta que há 100% de certeza acerca de subtensão para esse nó. Isso reforça ainda mais a preocupação de engenheiros de planejamento em se adotar modelagens mais completas para representar a rede no FPP, de maneira que esses erros na avaliação sejam minimizados. Os resultados para a fase B são aproximadamente idênticos, uma vez que o carregamento dessa fase foi mantido igual nos três casos.

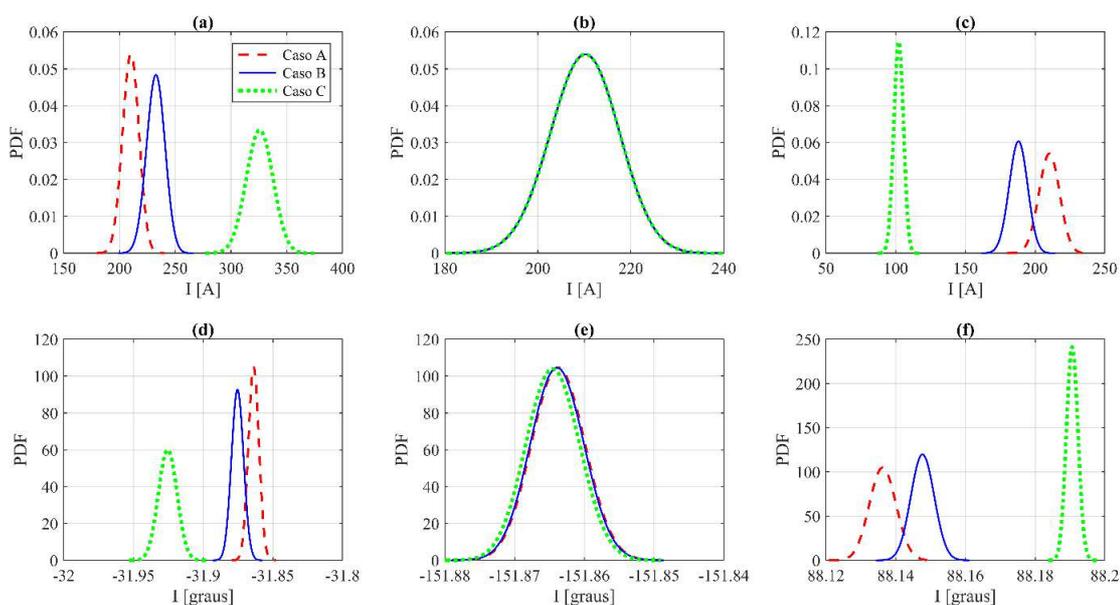
Conclusões similares podem ser feitas quanto aos ângulos: diferentes modelagens de rede levam a resultados completamente diferentes. É importante ressaltar que erros grosseiros no tratamento dos ângulos podem levar a imprecisões graves nos cálculos das correntes elétricas que fluem nos ramos do sistema.

Comparações acerca das correntes elétricas são mostradas na Figura 16. Novamente, por questão de aproveitamento de espaço, optou-se por demonstrar os resultados referentes à

linha que conecta as barras 1 e 2. Porém, as conclusões podem ser generalizadas para as demais.

Análises análogas às anteriores podem ser conduzidas para as correntes elétricas. Quando o desequilíbrio é muito alto, como no Caso C, a corrente na fase de maior carregamento será muito mais expressiva em relação às demais, o que pode ser constatado na Figura 16 (a). O engenheiro de planejamento deve estar ciente desse efeito, principalmente para que os condutores possam ser corretamente dimensionados. Além disso, a probabilidade da violação do valor máximo de corrente pode ser corretamente estimada caso uma modelagem mais ampla das redes, tal como a multifásica, for considerada no planejamento probabilístico, o que subsidia também formulações de Fluxo de Potência Ótimo.

Figura 16 – IEEE 33: Resultados de corrente. (a) Linha 1-2, fase A (magnitude); (b) Linha 1-2, fase B (magnitude); (c) Linha 1-2, fase C (magnitude); (d) Linha 1-2, fase A (ângulo); (e) Linha 1-2, fase B (ângulo); (f) Linha 1-2, fase C (ângulo).



Fonte: elaborada pelo autor (2025).

#### 4.2.3 COMPARAÇÕES ENTRE MODELOS DE REDE À LUZ DA NATUREZA DO FLUXO DE POTÊNCIA

Conforme visto no Capítulo 2 e na introdução do presente capítulo, ainda é largamente empregada a modelagem equilibrada das redes de distribuição em trabalhos que

tratam de planejamento de SD por meio do FPP. Na seção anterior, foi constatado um enorme erro nos resultados probabilísticos de tensão e corrente quando uma modelagem equilibrada é adotada, sendo esse erro cada vez maior à medida que cresce o desequilíbrio do sistema, o que pode prejudicar as ações de planejamento.

Portanto, o objetivo da presente seção é propor uma análise de sensibilidade para averiguar se o usuário incorre em um erro maior se empregar um modelo determinístico, porém multifásico, ou se adotar um modelo probabilístico, porém monofásico.

Para efetivar a análise, são propostos dois casos:

- Caso Determinístico e Multifásico (denotado por DM): é simulado um FP determinístico, considerando carregamento médio para o IEEE 33 Barras, dotado de modelagem multifásica, em que há 10% de desequilíbrio (carregamento base do Caso 2 da seção anterior).
- Caso Probabilístico e Equilibrado (denotado por PE): é equivalente ao Caso 1 da seção anterior. Ou seja, a rede é considerada equilibrada, porém com tratamento probabilístico. O FPMP-CRA é usado para simular probabilisticamente.

Para comparar ambos os casos, é preciso adotar uma métrica de erro que seja aplicável tanto no caso determinístico quanto probabilístico. A métrica escolhida é a seguinte: o erro será a maior diferença percentual possível no valor encontrado no caso simulado em relação a um caso de referência, considerando-se uma variável de saída do FP. O caso de referência adotado é justamente o Caso 2 da seção anterior, denotado por REF, que contempla tanto a modelagem probabilística quanto a multifásica. A variável considerada é a tensão em magnitude na barra 33, fase A.

No caso determinístico, só há um valor de saída. Logo, o erro será a maior diferença entre tal valor e o 5<sup>o</sup> ou 95<sup>o</sup> percentil<sup>13</sup> da CDF calculada no FMPF-CRA. No caso probabilístico, o erro será a maior diferença entre os 5<sup>os</sup> ou 95<sup>os</sup> percentis das CDFs calculadas em ambas as situações. Essas expressões são matematicamente definidas em (4.1) e (4.2). O intervalo entre o 5<sup>o</sup> e 95<sup>o</sup> percentil constitui um intervalo de confiança no qual 90% dos cenários estão incluídos. Ou seja, são excluídos *outliers* ou cenários altamente improváveis para promover uma comparação mais conservadora.

---

<sup>13</sup> Os valores de 5<sup>o</sup> e 95<sup>o</sup> percentis são adotados para se estimar valores mínimos e máximos para a variável aleatória considerada. 5<sup>o</sup> percentil: trata-se do valor cuja probabilidade de a variável aleatória ser menor ou igual vale 5%. 95<sup>o</sup> percentil: trata-se do valor cuja probabilidade de a variável aleatória ser maior vale também 5%.

$$\epsilon_{DM} = \max \left\{ \frac{|V_{DM} - V_{REF}^5|}{V_{REF}^5}, \frac{|V_{DM} - V_{REF}^{95}|}{V_{REF}^{95}} \right\} \quad (4.1)$$

$$\epsilon_{PE} = \max \left\{ \frac{|V_{PE}^5 - V_{REF}^5|}{V_{REF}^5}, \frac{|V_{PE}^{95} - V_{REF}^{95}|}{V_{REF}^{95}} \right\} \quad (4.2)$$

Onde:  $\epsilon_{DM}$  e  $\epsilon_{PE}$  são os erros cometidos usando as abordagens DM e PE, respectivamente.  $V_{DM}$  é a magnitude de tensão na barra 33, fase A, usando a abordagem DM.  $V_{PE}^5$  e  $V_{PE}^{95}$  são, respectivamente, o 5<sup>o</sup> e 95<sup>o</sup> percentil da CDF da magnitude de tensão da barra 33, fase A, calculados usando a abordagem PE.  $V_{REF}^5$  e  $V_{REF}^{95}$  são, respectivamente, o 5<sup>o</sup> e 95<sup>o</sup> percentil da CDF da magnitude de tensão da barra 33, fase A, calculados usando a abordagem de referência. Os resultados obtidos estão apresentados na Tabela 4.

Tabela 4 – Resultados da análise de sensibilidade (modelo de rede *versus* natureza do FP), usando caso de 10% de desequilíbrio como referência.

Caso	Valor/intervalo encontrado [pu]	Intervalo de confiança (Caso 2) [pu]	Erro máximo [%]
DM	0,9073	[0,9023 – 0,9132]	0,646%
PE	[0,9120 – 0,9217]	[0,9023 – 0,9132]	1,075%

Fonte: elaborada pelo autor (2025).

É possível perceber que o erro é maior quando uma formulação equilibrada é proposta, ainda que haja tratamento probabilístico do FP. Isso significa que, nessa situação, a utilização da modelagem multifásica das redes é mais imprescindível para alcançar resultados precisos do que o próprio equacionamento probabilístico. Isso demonstra a importância da consideração de um equacionamento que esteja apto a representar o desequilíbrio, inerente a esse tipo de sistema.

Tais imprecisões aumentam à medida que o desequilíbrio do sistema também se torna mais significativo. Pode-se fazer a mesma análise quando o Caso 3 é o referencial, ou seja, considerando agora 50% de desequilíbrio. Neste caso, a abordagem DM adota o carregamento médio do Caso 3. Os resultados estão disponibilizados na Tabela 5.

Mais uma vez, o erro é maior quando a modelagem equilibrada é adotada. Novamente, conclui-se que, nesse caso, a formulação multifásica do FP é mais importante do que o tratamento estatístico. Com um desequilíbrio crítico, existe um agravante: enquanto a tensão

assume valores de 0,8612 a 0,8775 pu em 90% dos cenários, o que é um panorama extremamente crítico, uma formulação equivocada que desconsidera o desequilíbrio chega à conclusão que a tensão varia de 0,9120 a 0,9217 pu. Logo, há um claro erro de avaliação que pode conduzir a erros drásticos de planejamento e operação, que foi gerado por uma decisão equivocada da modelagem da rede.

Tabela 5 – Resultados da análise de sensibilidade (modelo de rede *versus* natureza do FP), usando caso de 50% de desequilíbrio como referência.

Caso	Valor/intervalo encontrado [pu]	Intervalo de confiança (Caso 3) [pu]	Erro máximo [%]
DM	0,8687	[0,8612 – 0,8775]	1,003%
PE	[0,9120 – 0,9217]	[0,8612 – 0,8775]	5,900%

Fonte: elaborada pelo autor (2025).

Em resumo, o engenheiro de planejamento deve estar ciente das características do sistema a ser modelado e adotar as melhores condições possíveis para representá-lo, de maneira a minimizar possíveis erros em relação ao estado real desses sistemas. No caso de redes de distribuição, a formulação multifásica é imprescindível para que o desequilíbrio possa ser devidamente incluído no equacionamento e nos resultados.

Ressalta-se, também, que os erros da formulação determinística em relação à probabilística foram reduzidos devido a alguns motivos, a saber: a carga foi considerada perfeitamente correlacionada, o desvio padrão adotado para a carga foi baixo e não foram incluídas GDs de alta incerteza, como as baseadas em geração renovável. Tudo isso colabora para que os desvios padrões das tensões elétricas sejam limitadas, o que aproxima os resultados probabilísticos dos obtidos por vias determinísticas. Porém, nem sempre isso ocorre, como ficará claro nas próximas seções, em que diferentes condições de entrada serão estudadas.

#### 4.2.4 CONCLUSÕES PARCIAIS DO TESTE 1

Dos resultados do teste 1, pode-se concluir que a modelagem de rede é fator muito importante de se considerar com rigor para que ações precisas possam ser tomadas. Como os SD são inerentemente desequilibrados, é imprescindível que uma formulação multifásica seja

adotada. Apesar de muitos autores ainda trabalharem com modelagens equilibradas nos seus trabalhos relacionados ao FPP, foi visto que o erro cometido ao adotar simplificações na rede pode ser maior do que o próprio efeito estocástico que se pretende modelar. Logo, os resultados obtidos carecem de qualquer relevância, pois não representam adequadamente a realidade prática.

Dessa maneira, deve-se estar ciente sobre quais fenômenos são necessários modelar e representá-los da maneira mais fiel possível. Diante disso, foi visto que, na presença de desequilíbrios extremos, as funções densidade de probabilidade de tensões e correntes podem ser muito distintas em relação a uma modelagem equilibrada, o que pode gerar equívocos notáveis nas ações de planejamento e/ou operação.

Vista a importância de tal questão, a partir de agora, todos os testes serão executados mediante aplicação da formulação multifásica das redes. Vale a pena ressaltar também que o FPMP-CRA apresenta grande flexibilidade, podendo simular tanto sistemas equilibrados quanto desequilibrados sem necessidade de alteração de seus pressupostos metodológicos.

#### 4.3 TESTE 2: ESTUDO DA MODELAGEM DA CARGA E CONSIDERAÇÃO DE DIFERENTES CORRELAÇÕES ESPACIAIS

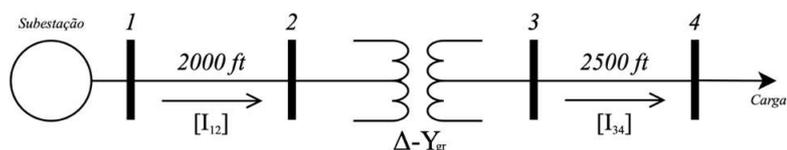
O objetivo do teste 2 é verificar o impacto da representação da carga no FPP, considerando diferentes modelagens probabilísticas e níveis de correlações em cargas localizadas em diferentes fases. Conforme apontado na revisão bibliográfica conduzida no Capítulo 2, muitos trabalhos presentes na literatura consideram as cargas ou totalmente independentes (isto é, a distribuição de probabilidade de uma carga não impacta na distribuição de probabilidade das demais), ou totalmente correlacionadas (ou seja, todas as cargas variam mediante uma única função de probabilidade). Diante disso, busca-se compreender o impacto de se adotar um ou outro, ou até correlações intermediárias, nos resultados referentes às PDFs das variáveis de saída.

O método FPMP-CRA consegue lidar com entradas correlacionadas ou independentes, o que o torna extremamente versátil. A demonstração de sua aplicabilidade sob essas condições também é objetivo da presente seção. As simulações são realizadas considerando-se a topologia IEEE 4 Barras.

#### 4.3.1 APRESENTAÇÃO DA TOPOLOGIA IEEE 4 E CONDIÇÕES DE SIMULAÇÃO

A topologia IEEE 4 é um alimentador teste utilizado para simulações de SD. Trata-se de uma topologia simples, porém desequilibrada, o que permite uma representação mais adequada das redes de distribuição. Possui 4 barras trifásicas, duas linhas, um transformador e uma carga desequilibrada conectada na barra 4. A potência total vale, originalmente,  $1275 + j791$  kVA,  $1800 + j873$  kVA e  $2375 + j782$  kVA, nas fases A, B e C, respectivamente. A topologia em sua forma unifilar pode ser vista na Figura 17. Informações detalhadas acerca da topologia podem ser vistas em *IEEE PES Test Feeders* (2024).

Figura 17 – Topologia IEEE 4.



Fonte: elaborada pelo autor (2025).

Assim como no teste 1, as cargas aqui também são consideradas variáveis aleatórias. No entanto, o objetivo será a simulação de diferentes condições de entrada para o FPMP-CRA no que tange à modelagem probabilística desses elementos. Em todos os casos, as cargas são modeladas por meio de distribuições Gaussianas multivariadas<sup>14</sup>. Optou-se por considerar, como carregamento médio, a metade do valor original, para evitar o sorteio de cenários que não possuem convergência (pois podem exceder, justamente, o ponto de maior carregamento). Adotando-se um carregamento médio menor, essas ocorrências são mitigadas. O desvio padrão adotado para as cargas, por sua vez, é igual a 10% do valor médio. A análise proposta consiste na simulação do FPMP-CRA considerando-se diferentes correlações entre cargas modeladas por funções Gaussianas.

Para gerar as amostras correlacionadas, é adotado o método de cópulas estatísticas, cujos passos são mostrados no Capítulo 2. As cargas, nesse caso, são modeladas como PDFs Gaussianas. Logo, basta informar os coeficientes de correlação (CC) entre as três cargas, tomadas duas a duas. Para facilitar e padronizar a notação, o símbolo  $\rho_{XY}$  denota o CC entre

<sup>14</sup> A distribuição de probabilidade Gaussiana multivariada consiste em uma função densidade de probabilidade normalmente distribuída para mais de uma variável aleatória, sendo, portanto, uma função de probabilidade conjunta. Nela, as distribuições marginais são Gaussianas, e o coeficiente de correlação determina seu comportamento. Mais detalhes no Apêndice A.

as cargas localizadas nas fases X e Y. Para efetuar uma análise de sensibilidade, são analisados quatro casos. A saber:

- Caso A: as cargas são todas fracamente correlacionadas. Isto é, a distribuição de probabilidade de uma carga não afeta intensamente a distribuição das demais. Todos os CC são iguais a  $\rho = 0,1$ .
- Caso B: as cargas são moderadamente correlacionadas. Nesse caso, todos os CC são iguais a  $\rho = 0,5$ .
- Caso C: as cargas são fortemente correlacionadas. Nesse caso, todos os CC são iguais a  $\rho = 0,9$ .
- Caso D: o comportamento das cargas apresenta correlações diferentes de acordo com as fases nas quais estão conectadas. Foram adotados os seguintes CC:  $\rho_{AB} = 0,4$ ,  $\rho_{CA} = 0,2$  e  $\rho_{BC} = -0,8$ . Note que o último é um valor negativo, o que significa que a correlação é inversa em relação aos outros pares. Apesar de não ser uma tendência do comportamento humano no que tange à demanda de potência, essa consideração foi incluída para testar o FPMP-CRA sob diferentes condições de entrada.

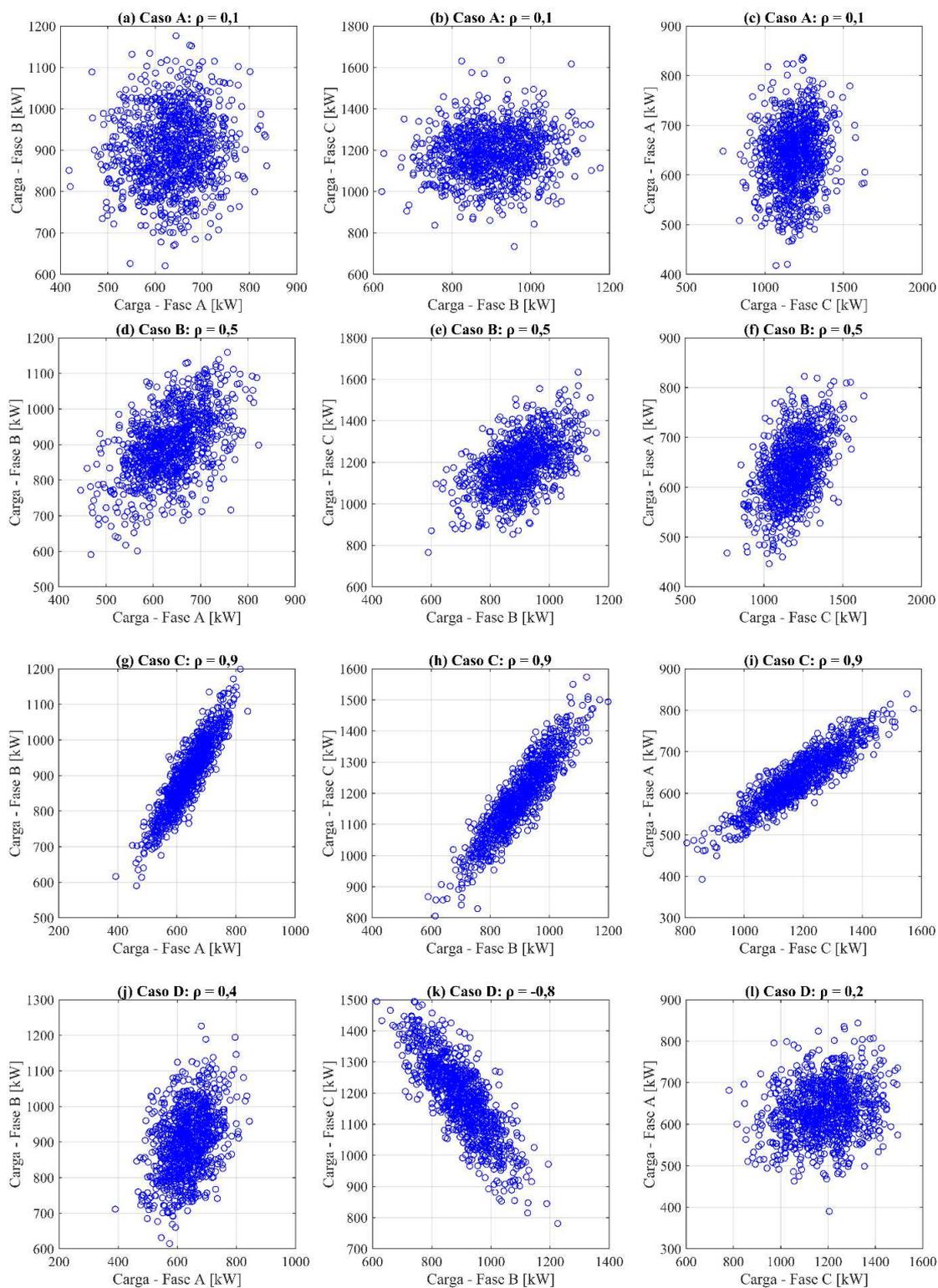
Em todos os casos, foram geradas 1000 amostras das cópulas resultantes, que podem ser vistas na Figura 18. Nela, estão expostas as amostras com referência à potência ativa de cada caso, mas recorda-se que o fator de potência sempre é mantido constante em relação aos valores originais da topologia.

Em outras palavras, a Figura 18 apresenta os espaços a serem comprimidos pela técnica de *clustering* que, para as simulações da presente seção, é o algoritmo K-Means.

#### 4.3.2 IEEE 4: RESULTADOS DO FPMP-CRA

O número ótimo de *clusters*, o tempo de execução da técnica de otimização do número de *clusters* e o tempo computacional total demandado pelo FPMP-CRA estão apresentados na Tabela 6, considerando cada um dos quatro casos. Em todos os casos, o número ótimo de *clusters* é muito maior em relação ao teste 1, em que foi constatado um valor de 33 *clusters*. Isso é devido à inclusão de uma PDF por carga, o que triplica o número de variáveis aleatórias do problema em relação ao caso perfeitamente correlacionado.

Figura 18 – Amostras geradas pela cópula para cada caso.



Fonte: elaborada pelo autor (2025).

Tabela 6 – Resultados do FPMP-CRA para IEEE 4.

Caso	Número ótimo de <i>clusters</i>	Tempo requerido para otimizar número de <i>clusters</i> [s]	Tempo demandado total do FPMP-CRA [s]
A	143	2,58	8,65
B	123	1,85	7,16
C	82	0,89	4,79
D	92	1,01	5,39

Fonte: elaborada pelo autor (2025).

Conforme já mencionado nos Capítulos 1 e 2, o aumento do número de variáveis aleatórias acarreta uma maior dificuldade para o algoritmo de *clustering* obter êxito, o que causa o aumento da quantidade necessária de grupos para representar adequadamente o espaço a ser comprimido. No entanto, nota-se também que, mesmo no pior caso, alcança-se uma redução considerável de amostras por meio do método proposto (reduz-se de 1000 amostras para 143, ou seja, 85,7% dos pontos a serem avaliados foram eliminados).

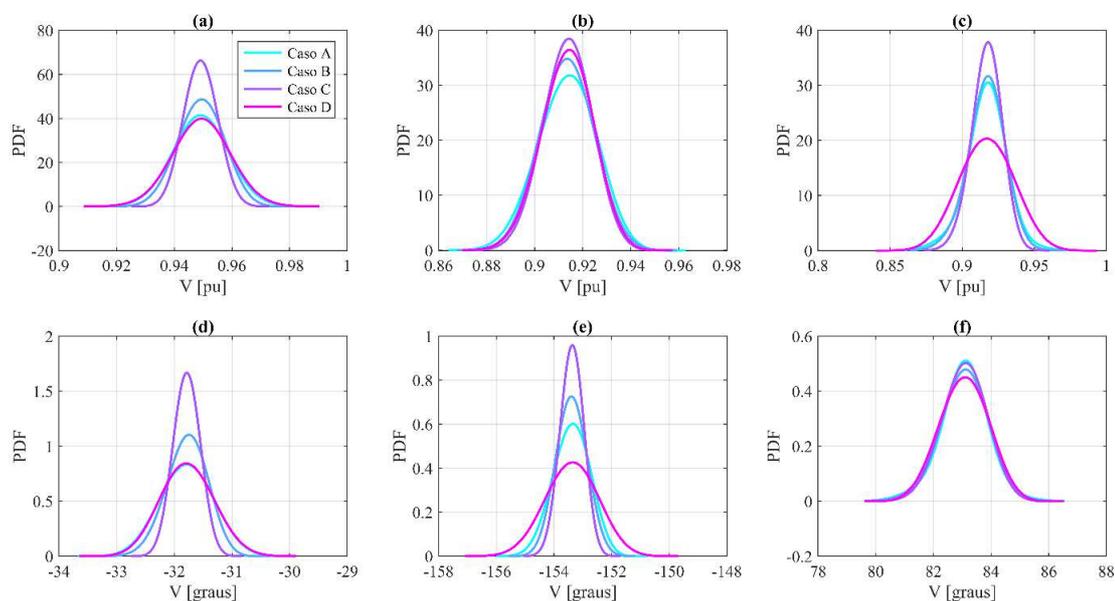
Outro aspecto que se pode intuir ao analisar os resultados é que os casos com menor coeficiente de correlação apresentam um maior número ótimo de *clusters*. Ora, isso é perfeitamente coerente com a teoria de *clustering*, visto que quanto mais as amostras são independentes, mais elas tendem a se espalhar espacialmente. Logo, as distâncias entre os pontos aumentam, o que leva à necessidade de se ter um número maior de *clusters* para representar adequadamente esse espaço de amostras. De forma análoga, quanto maior o coeficiente de correlação entre as variáveis aleatórias, maior a tendência das amostras se localizarem em uma mesma região espacial. Isso explica a diferença considerável de *clusters* em cada um dos casos. No último caso, apenas um par de cargas é fortemente correlacionado, o que faz o número ótimo de *clusters* ser um valor intermediário entre os alcançados no Caso A e no Caso C. Quanto ao tempo computacional, todas as simulações convergiram em menos de 10 segundos, o que indica um excelente desempenho. Nas próximas seções, serão apresentadas diretrizes para diminuir ainda mais o tempo demandado e técnicas para acelerar o método, principalmente para aplicações de grande porte.

As PDFs relacionadas às magnitudes e ângulos das tensões elétricas estão representadas na Figura 19, para cada caso simulado. Por questão de conveniência de

aproveitamento do espaço, optou-se por mostrar os resultados referentes à barra 4, que é justamente a posição na qual a carga, que é a variável aleatória adotada, está conectada.

De maneira análoga aos resultados vistos na Seção 4.2, em que foi estudada a modelagem da rede, pode-se verificar que também na modelagem da carga diferentes considerações podem levar a PDFs de tensão bem diferentes, sobretudo no que tange ao desvio padrão. De uma maneira geral, infere-se que quanto maior o coeficiente de correlação das cargas, menor será o desvio padrão das tensões – o que pode ser constatado visualmente por meio das curvas em cor roxa na Figura 19, que possuem geralmente um maior ‘pico’ de densidade de probabilidade e, conseqüentemente, menor desvio padrão. Isso significa que cargas correlacionadas tendem a produzir um intervalo menor de operação das tensões, devido ao seu baixo desvio padrão. Em suma, os engenheiros devem estar atentos às caracterizações das cargas para alcançarem resultados precisos que vão subsidiar os estudos de planejamento e de operação. Nesse sentido, é imprescindível que os mesmos adotem métodos capazes de incluir correlações de variáveis aleatórias como o FPMP-CRA.

Figura 19 – Resultados das PDFs de tensões elétricas (IEEE 4).



Fonte: elaborada pelo autor (2025).

Para ilustrar essa questão, é calculada, para cada caso simulado, a probabilidade de violação de tensão (subtensão) por meio das PDFs obtidas usando o FPMP-CRA. Em outras

palavras, calcula-se a probabilidade de a tensão ser menor do que 0,90 pu para cada caso. Os resultados estão dispostos na Tabela 7.

Apesar de todos os níveis de correlação apontarem que não há qualquer chance de subtensão para a barra 4, fase A, nota-se que nas demais posições a probabilidade de violação é muito distinta. Isso reforça a ideia de que resultados díspares podem ser obtidos mediante diferentes níveis de correlação entre as cargas.

Outro fator importante de ser ressaltado é em relação às probabilidades de violação do Caso D, que são muito diferentes dos demais. Em Chevalier *et al.* (2022), é citado que basta calcular o estado probabilístico sob as duas correlações extremas (ou seja, perfeitamente correlacionado e perfeitamente independente) para se averiguar a precisão de um método de FPP, pois todas as outras correlações estariam em posições intermediárias a ambas. No entanto, é possível perceber que a probabilidade de violação no nó 4C em se tratando do Caso D é superior em relação ao Caso A. Em outras palavras, foi visto que a consideração de cargas totalmente independentes não se configura como o mais conservador possível, conforme alguns trabalhos na literatura apontam. Apesar de não ser o mais desafiador sob a perspectiva do *clustering*, o Caso D aponta maior chance de violação. Isso sublinha ainda mais a importância de uma modelagem de rede multifásica e, sobretudo, de uma modelagem apropriada da carga no que tange às suas correlações. A primeira porque permite a representação de cargas conectadas em diferentes fases e, portanto, a inclusão do desequilíbrio. A segunda porque introduz a forma como essas cargas variam sob uma perspectiva probabilística mais realista.

Tabela 7 – Probabilidades de subtensão de acordo com o caso (IEEE 4).

Caso	Pr ( $V < 0,90$ ) [%]		
	4A	4B	4C
A	0,00	14,08	10,96
B	0,00	12,05	10,22
C	0,00	9,45	6,20
D	0,00	11,06	19,21

Fonte: elaborada pelo autor (2025).

Enfatiza-se que um cálculo preciso das PDFs das variáveis de saída é imprescindível para o condicionamento correto de restrições em problemas de otimização estocásticos, em

que as probabilidades de violação de parâmetros de redes associados à qualidade de energia são introduzidas para que se possa obter soluções de planejamento boas e confiáveis. Nesse sentido, é de suma importância a consideração correta da modelagem de rede, vista no teste 1 e da carga, vista no presente teste.

Por fim, é feita uma análise comparativa acerca do FP determinístico em oposição ao FPMP-CRA. Em oposição aos resultados encontrados no teste 1, é possível perceber, agora, que o erro cometido ao utilizar uma abordagem que considera apenas um cenário pode ser muito grave, devido ao maior desvio padrão e às correlações das cargas. Os resultados estão dispostos na Tabela 8, em que a grandeza analisada é a magnitude de tensão na barra 4.

Tabela 8 – Comparação do FP determinístico *versus* FPMP-CRA.

Caso	Valor determinístico			Valor máximo			Valor mínimo		
	[pu]			utilizando FPMP-			utilizando FPMP-		
	4A	4B	4C	CRA [pu]			CRA [pu]		
	4A	4B	4C	4A	4B	4C	4A	4B	4C
A	0,95	0,91	0,92	0,98	0,95	0,97	0,92	0,88	0,85
B	0,95	0,91	0,91	0,97	0,95	0,95	0,93	0,89	0,87
C	0,95	0,91	0,92	0,97	0,94	0,95	0,93	0,88	0,88
D	0,95	0,91	0,92	0,99	0,94	0,97	0,92	0,87	0,86

Fonte: elaborada pelo autor (2025).

Logo, as tensões podem assumir valores muito mais altos ou muito mais baixos em relação aos constatados pela abordagem determinística. Inclusive, podem assumir valores extremamente críticos, como os mínimos verificados para o nó 4C, que chegam a 0,85 pu, ao passo que o FP convencional aponta valores próximos do mínimo estabelecido no PRODIST<sup>15</sup>, não detectando, portanto, violações tão contundentes. Isso ocorre devido aos patamares de carga não avaliados por meio da abordagem determinística, que se limita a um único cenário. Portanto, torna-se imprescindível para os estudos de planejamento modernos considerarem uma ferramenta de análise mais ampla, tal como o FPMP-CRA.

<sup>15</sup> Os “Procedimentos de Distribuição” (ou PRODIST) referem-se às normas estabelecidas pela Agência Nacional de Energia Elétrica (ANEEL) no que tange às condições de operação dos SDs. Mais especificamente, o Módulo 8 trata dos requisitos mínimos de qualidade de energia elétrica, em que há considerações acerca de subtensão, sobretensão, entre outros distúrbios elétricos de tensão (ANEEL, 2022).

### 4.3.3 CONCLUSÕES PARCIAIS DO TESTE 2

Após a constatação de que a modelagem correta da rede é fundamental para alcançar resultados precisos no âmbito do FPP, pode-se concluir também que a representação da carga possui igual importância. Além de assumir a carga como variável aleatória, é importante a correta modelagem de como tais elementos se comportam entre si, isto é, como a distribuição de probabilidade de uma carga afeta as demais, e como seus diferentes comportamentos afetam as distribuições de probabilidade das variáveis aleatórias de saída do FPP.

No caso do FPMP-CRA, a consideração de diferentes graus de correlação entre as cargas acarreta diferentes números ótimos de *clusters*, o que impacta no tempo computacional demandado pelo método. No entanto, percebe-se que, ao assumir correlações muito fortes entre as cargas, as variáveis aleatórias de saída tendem a apresentar PDFs mais limitadas. Em outras palavras, seus desvios padrões tendem a ser menores, reduzindo o intervalo de confiança dessas funções. Logo, cenários importantes de operação podem equivocadamente ser negligenciados. Em suma, a consideração de cargas fortemente correlacionadas, embora faça com que o algoritmo tenha uma melhor performance computacional, pode ocasionar em perda de informação e, conseqüentemente, ações insatisfatórias de planejamento.

Foi visto que correlações diferentes em cargas localizadas em fases distintas podem levar a resultados muito diferentes em relação às PDFs de casos fortemente correlacionados ou fortemente independentes – que são as abordagens mais comuns na literatura e muitas vezes assumidas como casos extremos e conservadores. Isso ressalta ainda mais a grande importância de adotar uma modelagem multifásica que proporcione a representação de cargas conectadas em fases diferentes do sistema.

Por fim, foi constatada a enorme diferença entre a abordagem convencional determinística e o FPMP-CRA, enfatizando a relevância de técnicas probabilísticas quando a carga possui uma modelagem mais completa, que inclui as correlações estatísticas e, sobretudo, quando o desvio padrão desses elementos é grande o suficiente. Nesses casos, é imprescindível que métodos probabilísticos que estejam aptos a contemplar tais condições de entrada sejam adotados.

### 4.4 TESTE 3: ESTUDO DA DIMENSIONALIDADE DO FPMP-CRA

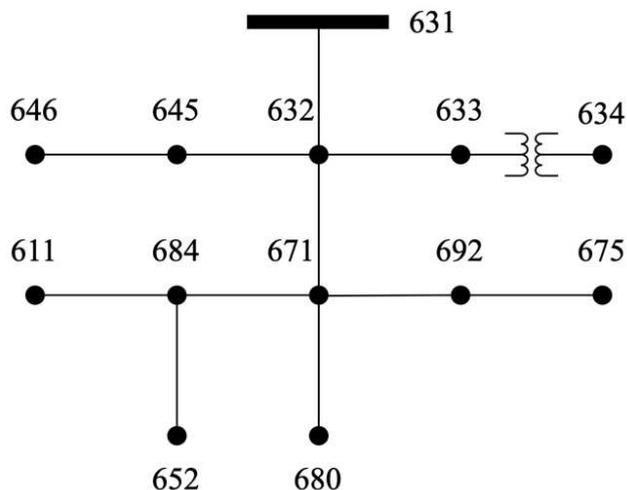
No Capítulo 2, foi enfatizado que uma das motivações do presente trabalho é o desenvolvimento de um algoritmo eficiente de FPP que seja escalável, isto é, que seja apto a simular aplicações de grande porte. Nesse sentido, é importante que o método consiga simular com boa precisão sistemas maiores e, sobretudo, situações nas quais um grande número de variáveis aleatórias são introduzidas. No caso de técnicas baseadas em *clustering*, tal teste é crítico, devido à ‘maldição da dimensionalidade’, a qual sustenta que esses métodos perdem muita precisão à medida que o número de variáveis incertas do problema aumenta. A partir disso, o objetivo do teste 3 é efetuar uma análise crítica a respeito do desempenho do FPMP-CRA diante de diferentes dimensionalidades das condições de entrada.

#### 4.4.1 APRESENTAÇÃO DA TOPOLOGIA IEEE 13 E CONDIÇÕES DE SIMULAÇÃO

Para efetuar o estudo proposto, são realizadas simulações considerando a topologia IEEE 13 (IEEE PES Test Feeders, 2024). Trata-se de um SD desequilibrado, que contém 20 cargas, 10 linhas, e um transformador que abaixa a tensão de 4,16 kV para 0,48 kV. Há linhas de distribuição trifásicas, bifásicas e monofásicas, cargas equilibradas e desequilibradas, em delta e em estrela. Logo, trata-se de uma topologia altamente desequilibrada, própria para simular comportamentos inerentes aos SD. Seu carregamento base total vale  $1175 + j616$  kVA,  $1039 + j665$  kVA e  $1252 + j821$  kVA, nas fases A, B e C, respectivamente. O regulador de tensão e os capacitores originalmente presentes foram desconsiderados para enaltecer os resultados referentes às variáveis aleatórias consideradas no estudo. Nos trabalhos futuros, o seu impacto em estudos probabilísticos será devidamente estudado. Seu diagrama unifilar pode ser visto na Figura 20.

O presente teste contempla dois estudos. O primeiro consiste na simulação do FPMP-CRA considerando diferentes dimensionalidades das variáveis aleatórias de entrada (isto é, diferentes números de variáveis aleatórias e diferentes números de amostragens por variável aleatória). O intuito é averiguar a eficiência do método proposto sob condições diferentes quanto à dimensionalidade do banco de dados de entrada. Já o segundo estudo consiste na formulação de diretrizes para aprimorar ainda mais o desempenho do FPMP-CRA, sobretudo quando há condições de alta dimensionalidade, fundamentais para a simulação de um FPP escalável.

Figura 20 – Topologia IEEE 13.



Fonte: elaborada pelo autor (2025).

Em ambos, as variáveis aleatórias são as cargas, modeladas por distribuições de probabilidade Gaussianas cujas médias são os valores originais da topologia e os desvios padrões valem 10% dos valores médios. As cargas são todas tratadas como independentes, pois no teste 2 foi visto essa configuração resulta no caso mais desafiador sob a perspectiva do *clustering*.

#### 4.4.2 ESTUDO 1: VARIANDO A DIMENSIONALIDADE DO FPMP-CRA

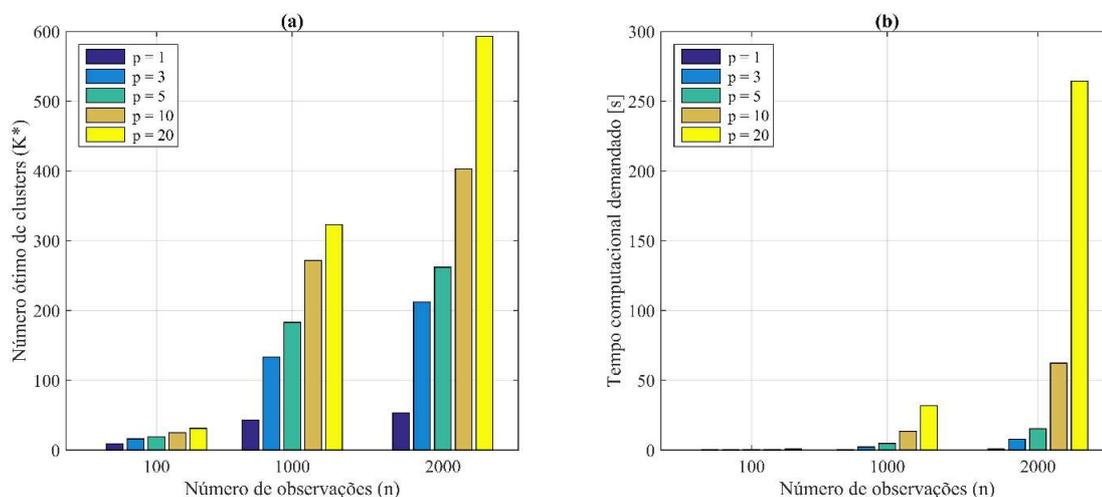
No primeiro estudo avaliado, o FPMP-CRA é simulado utilizando a topologia IEEE 13 sob diferentes condições de entrada. São conduzidas simulações considerando diferentes números de variáveis aleatórias e amostragens por variáveis aleatórias. Conforme mencionado, o objetivo é investigar o comportamento do método proposto mediante diferentes dimensionalidades do banco de dados de entrada.

Como o sistema originalmente possui 20 cargas, foram feitas simulações considerando 1, 3, 5, 10 e 20 variáveis aleatórias. O processo de escolha de quais cargas são assumidas como aleatórias e quais permanecem determinísticas é de acordo com o carregamento médio. Por exemplo, no caso em que há apenas uma carga aleatória, tal carga é aquela cujo carregamento médio é o maior. Analogamente, no caso em que há três variáveis aleatórias, escolhe-se aquelas três cujos carregamentos médios são os maiores. As demais permanecem fixas em seus valores médios. No último caso, todas as cargas são aleatórias e nenhuma é fixa.

No que tange ao número de amostras por variável aleatória, foram feitas simulações considerando 100, 1000 e 2000 amostras aleatórias das PDFs de entrada geradas previamente. Por exemplo, no primeiro caso, simula-se o FPMP-CRA considerando-se apenas uma variável aleatória, que é a carga cujo carregamento médio é o maior, e 100 amostras aleatórias dessa PDF. Logo, o banco de dados de entrada é um vetor de 100 linhas e uma coluna. Este processo é repetido para as demais situações no que se refere ao número de variáveis aleatórias e número de amostras, em que o banco de dados de entrada será sempre uma matriz de  $n$  linhas (referentes ao número de amostras) por  $r$  colunas (referentes ao número de variáveis aleatórias).

A Figura 21 (a) mostra os resultados referentes aos números ótimos de *clusters* definidos para cada caso, calculados por meio do procedimento exposto na Seção 3.2. Já a Figura 21 (b) mostra o tempo computacional demandado (em segundos) para efetuar tal otimização, para cada caso proposto.

Figura 21 – (a) Número ótimo de *clusters* considerando diferentes dimensionalidades do banco de dados no IEEE 13; (b) Tempo computacional demandado, em segundos, de acordo com a dimensionalidade do banco de dados de entrada.



Fonte: elaborada pelo autor (2025).

Nota-se que o número ótimo de *clusters* tende a crescer à medida que a dimensionalidade do banco de dados aumenta. Conforme já citado nos Capítulos 2 e 3, tal fenômeno ocorre devido à ‘maldição da dimensionalidade’ segundo a qual os algoritmos de *clustering* particionais estão submetidos. Como tal algoritmo depende do cálculo de

distâncias, tal métrica perde precisão com o aumento da dimensionalidade dos espaços a serem comprimidos. Consequentemente, as partições tendem a ser ruins. O efeito disso é a necessidade de um maior número de *clusters* para representar estatisticamente o banco de dados original.

Por exemplo, para uma situação em que são consideradas 3 variáveis aleatórias e 1000 amostras por variável (logo, um banco de dados representado por uma matriz de dimensões 1000x3), são necessários 133 *clusters* para obter resultados precisos no FPMP-CRA. Quando se aumenta o número de variáveis para 5, mantendo-se o número de amostras por variável (agora, uma matriz de dimensões 1000x5), o número requerido de *clusters* sobe para 212. Portanto, há um aumento de aproximadamente 60% do número de *clusters* para acomodar apenas duas variáveis a mais. Esse fenômeno é ainda mais agravante quando são simuladas situações de maior dimensionalidade. No caso de 20 variáveis aleatórias, o número ótimo de *clusters* é 31 para 100 amostras, 323 para 1000 amostras e 593 para 2000 amostras. Há uma enorme diferença constatada de acordo com a dimensionalidade. É importante ressaltar que a adoção de um número menor de *clusters* pode levar a resultados imprecisos do FPMP-CRA e esses valores obtidos são as quantidades mínimas requeridas para alcançar-se as PDFs das variáveis aleatórias de saída com precisão suficiente.

Outra implicação desse resultado é o aumento do tempo computacional demandado. De acordo com o procedimento de métodos de FPP baseados em *clustering*, o número de FPDs executados será igual ao número de *clusters* calculados. Dessa maneira, o aumento da dimensionalidade do banco de dados de entrada e o consequente maior número de *clusters* requeridos acarretam um aumento considerável do tempo computacional do método como um todo. Primeiramente devido à necessidade de simulação de um número muito grande de pontos de operação. Mas também devido à dificuldade da própria técnica de *clustering* na convergência de sua partição. Quanto maior a dimensionalidade do conjunto a ser dividido, mais iterações são necessárias para que a técnica convirja. Isso está representado na Figura 21 (b). No caso de apenas uma variável e 100 amostras, a técnica levou aproximadamente 0,24 segundo para convergir. Já no caso de maior dimensionalidade testado no presente estudo, foram necessários 4 minutos e 24 segundos. No entanto, há ganhos computacionais significativos em todos os casos, comparando com a simulação de todos os cenários possíveis do banco de dados de entrada. A Tabela 9 apresenta a porcentagem de ganhos computacionais para cada caso simulado ao utilizar-se o FPMP-CRA. O ganho computacional é calculado por meio de (4.3).

$$GC(\%) = \left( \frac{N_T - K^*}{N_T} \right) 100\% \quad (4.3)$$

Onde GC é o ganho computacional, ou seja, a porcentagem de cenários eliminados pelo processo de *clustering*.  $N_T$  é o número total de amostras do caso (100, 1000 ou 2000) e  $K^*$  é o número ótimo de *clusters* do caso.

É possível perceber que os ganhos são altíssimos quando um número baixo de variáveis aleatórias é empregado, alcançando valores que ultrapassam 90% em alguns casos. Porém, tais vantagens decrescem à medida que o número de variáveis aleatórias aumenta. Mesmo considerando o espaço de maior dimensionalidade estudado, os ganhos ainda chegam a valores por volta de 70%, o que ainda representa uma margem de razoável aplicabilidade.

Tabela 9 – Ganhos computacionais alcançados pelo FPMP-CRA de acordo com a dimensionalidade do banco de dados de entrada.

Número de amostras (n)	Número de variáveis aleatórias (p)				
	1	3	5	10	20
100	91%	84%	81%	75%	69%
1000	95,7%	86,7%	81,7%	72,8%	67,7%
2000	97,3%	89,4%	86,9%	79,8%	70,3%

Fonte: elaborada pelo autor (2025).

Os resultados disponibilizados na Tabela 9 estão em consonância com o estado da arte do FPP: a ‘maldição da dimensionalidade’ é o maior aspecto negativo de métodos baseados em *clustering* e estratégias para mitigar seus efeitos estão em estudo, principalmente para justificar seu uso em aplicações de grande porte sem que haja prejuízos de acurácia dos resultados – mantendo-se sua enorme eficiência computacional, que lhe é característica.

Apesar de o FPMP-CRA reduzir significativamente o número de cenários a serem simulados, é importante formular alternativas para aumentar ainda mais sua eficiência computacional, ampliando seu alcance de aplicabilidade para também ser apto e eficaz para problemas de FPP de grande porte e que possua a escalabilidade do número de variáveis aleatórias. Dessa maneira, o próximo estudo é responsável por propor diretrizes para alterar o FPMP-CRA de maneira a contemplar também esses casos mais complexos.

#### 4.4.3 ESTUDO 2: DIRETRIZES PARA APRIMORAR A EFICIÊNCIA DO FPMP-CRA

Na Seção 4.4.2, foi visto que quanto maior a dimensionalidade do banco de dados de entrada, maior o número de *clusters* requerido para alcançar resultados precisos de acordo com o método proposto de estimação do número ótimo de *clusters*. Consequentemente, maior será o tempo computacional demandado. Logo, torna-se fundamental adotar bancos de dados com a menor dimensionalidade possível, buscando aproveitar os máximos benefícios possíveis do *clustering*. Para isso, são sugeridas *diretrizes* para reduzir tanto o número de amostras quanto o número de variáveis aleatórias e alterações pontuais no FPMP-CRA, que vão proporcionar a eficiência necessária para sua aplicação em situações de larga escala. Essas modificações favorecem a escalabilidade do método, que é um recurso altamente requisitado nas necessidades atuais de análise de SD, e que ainda não foram explorados no que tange aos métodos disponíveis na literatura de FPP desequilibrados, tal como apontado na revisão bibliográfica conduzida no Capítulo 2.

As diretrizes propostas estão listadas a seguir:

1. Ao invés de adotar uma amostragem convencional como no MCS, passa-se a utilizar a técnica LHS. Conforme já discutido no Capítulo 2, ela é mais eficiente em relação ao MCS, fazendo com que sejam requeridas menos amostras para atingir os mesmos resultados. Além disso, devido à sua estrutura própria de amostragem, ela impede a existência de pontos sobrepostos, o que favorece muito as técnicas de *clustering*.
2. Para topologias que apresentem mais do que 5 variáveis aleatórias, a técnica de *clustering* sugerida é uma versão alternativa do K-Medoids chamada CLARA (do inglês: *CLustering for Large Applications*)<sup>16</sup>. A sua utilização justifica-se à medida que é recomendado pela literatura como um algoritmo capaz de lidar com bancos de dados de alta dimensionalidade.
3. Por fim, cargas com baixos desvios padrões (aqui consideradas as cargas com desvio padrão abaixo de 10 kW) são consideradas determinísticas. Essa consideração justifica-se uma vez que suas variações irão produzir baixo impacto devido à presença de cargas de maior variabilidade. No entanto, do ponto de vista de *clustering*, a exclusão de duas ou mais variáveis aleatórias é capaz de acelerar muito o processo de convergência do método, tal como exposto na Tabela 9.

---

<sup>16</sup> A técnica K-Medoids CLARA tem seus principais passos descritos em detalhes no Apêndice C.

As diretrizes acima expostas foram aplicadas na simulação do sistema IEEE 13 sob as condições expressas na seção anterior, considerando todas as cargas como aleatórias. Foram consideradas 100 amostras do método LHS. Nessas condições, o número ótimo de *clusters* calculado, considerando as diretrizes aplicadas, passou a ser 35 *clusters*, em contraste dos 593 obtidos na Seção anterior. Para promover comparações dos resultados e mostrar a precisão do método quando da aplicação das diretrizes propostas, foram simulados também dois algoritmos convencionais de MCS de amostragem fixa, um considerando 2000 amostras e outro considerando 10000 amostras.

A Tabela 10 apresenta os valores esperados para as magnitudes de tensão considerando o FPMP-CRA, o MCS com 2000 amostras e 10000 amostras, respectivamente. Similarmente, a Tabela 11 apresenta os resultados referentes aos desvios padrões, considerando a mesma comparação da Tabela 10.

Tabela 10 – Resultados para IEEE 13: valores esperados.

Nó	FPMP-CRA (proposto)	MCS (2000 amostras)	MCS (10000 amostras)
631-A	1,0000	1,0000	1,0000
631-B	1,0000	1,0000	1,0000
631-C	1,0000	1,0000	1,0000
632-A	0,9493	0,9493	0,9494
632-B	0,9832	0,9829	0,9829
632-C	0,9278	0,9270	0,9271
645-B	0,9735	0,9732	0,9732
645-C	0,9258	0,9250	0,9251
646-B	0,9717	0,9714	0,9714
646-C	0,9237	0,9229	0,9230
671-A	0,9086	0,9087	0,9088
671-B	0,9881	0,9875	0,9875
671-C	0,8679	0,8662	0,8665
684-A	0,9065	0,9066	0,9067
684-C	0,8640	0,8624	0,8627
652-A	0,8994	0,8994	0,8995
611-C	0,8602	0,8585	0,8588

692-A	0,9086	0,9087	0,9088
692-B	0,9881	0,9875	0,9875
692-C	0,8679	0,8662	0,8665
675-A	0,9001	0,9003	0,9003
675-B	0,9893	0,9887	0,9887
675-C	0,8640	0,8622	0,8625
680-A	0,9086	0,9087	0,9088
680-B	0,9881	0,9875	0,9875
680-C	0,8679	0,8662	0,8665
633-A	0,9466	0,9467	0,9467
633-B	0,9807	0,9804	0,9804
633-C	0,9248	0,9240	0,9242
634-A	0,9110	0,9103	0,9105
634-B	0,9390	0,9387	0,9387
634-C	0,9359	0,9359	0,9359

Fonte: elaborada pelo autor (2025).

Tabela 11 – Resultados para IEEE 13: desvios padrões.

Nó	FPMP-CRA (proposto)	MCS (2000 amostras)	MCS (10000 amostras)
631-A	0,0000	0,0000	0,0000
631-B	0,0000	0,0000	0,0000
631-C	0,0000	0,0000	0,0000
632-A	0,0043	0,0042	0,0042
632-B	0,0031	0,0029	0,0029
632-C	0,0034	0,0039	0,0038
645-B	0,0036	0,0033	0,0033
645-C	0,0034	0,0040	0,0039
646-B	0,0037	0,0034	0,0033
646-C	0,0034	0,0041	0,0040
671-A	0,0083	0,0081	0,0080
671-B	0,0056	0,0051	0,0052
671-C	0,0069	0,0073	0,0072

684-A	0,0083	0,0082	0,0081
684-C	0,0071	0,0075	0,0074
652-A	0,0086	0,0085	0,0084
611-C	0,0074	0,0077	0,0075
692-A	0,0083	0,0081	0,0080
692-B	0,0056	0,0051	0,0052
692-C	0,0069	0,0073	0,0072
675-A	0,0092	0,0088	0,0087
675-B	0,0057	0,0052	0,0053
675-C	0,0074	0,0079	0,0077
680-A	0,0083	0,0081	0,0080
680-B	0,0056	0,0051	0,0052
680-C	0,0069	0,0073	0,0072
633-A	0,0044	0,0043	0,0042
633-B	0,0032	0,0029	0,0029
633-C	0,0035	0,0039	0,0038
634-A	0,0047	0,0042	0,0042
634-B	0,0039	0,0032	0,0032
634-C	0,0031	0,0034	0,0032

---

Fonte: elaborada pelo autor (2025).

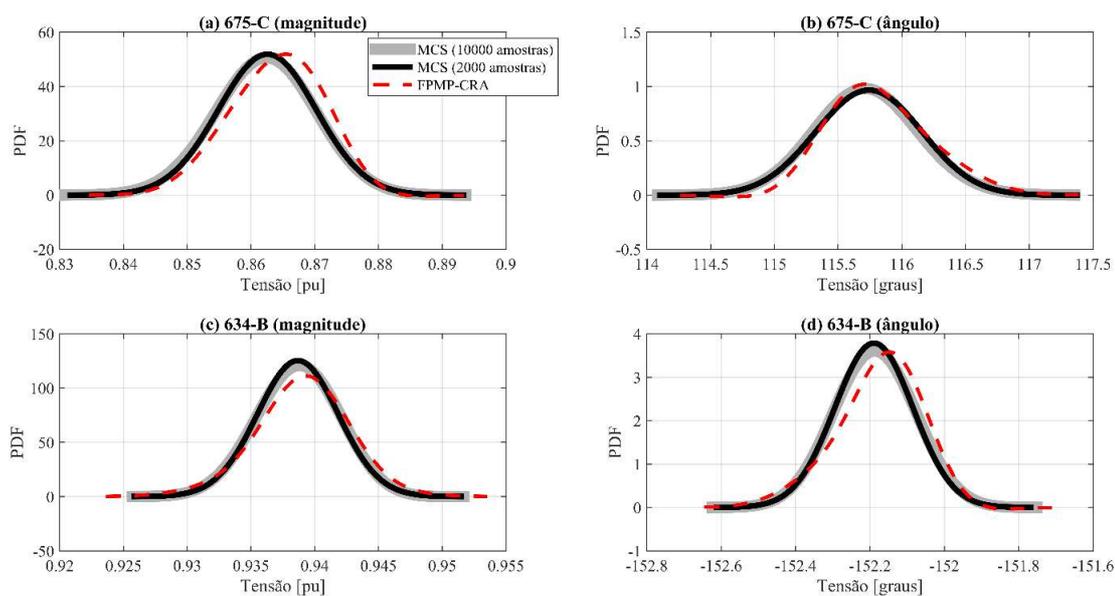
É possível constatar que os erros dos resultados encontrados por meio do FPMP-CRA em relação ao MCS de 10000 amostras são muito pequenos. No caso do valor esperado de magnitude de tensão, o erro médio avaliado considerando todas as posições vale 0,058%. No caso do desvio padrão de magnitude de tensão, o erro médio vale 7,034%. Esses erros são muito pequenos e possuem um baixo impacto na reconstituição analítica das PDFs de tais variáveis aleatórias de saída, via EGC.

As posições que apresentaram maiores erros foram 675-C e 634-B. Na Figura 22, são mostradas suas PDFs recuperadas analiticamente, considerando tanto o FPMP-CRA quanto as abordagens via MCS, referente às magnitudes e ângulos de tensão nas duas posições citadas. Optou-se por representar essas duas posições por motivos de aproveitamento de espaço, mas pode-se extrapolar tais resultados para as demais.

Nota-se que as PDFs são muito similares, corroborando com o resultado destacado anteriormente, referente à precisão do método quanto aos primeiros dois momentos estatísticos.

Em relação ao tempo computacional, destaca-se a enorme diferença do método proposto em relação aos demais simulados. O FPMP-CRA leva apenas 5,42 segundos para convergir, contra 286,948 segundos (aproximadamente 4 minutos e 47 segundos) para o MCS de 2000 amostras, e 1429,176 segundos (aproximadamente 23 minutos e 49 segundos) para o MCS de 10000 amostras. Em outras palavras, o FPMP-CRA é aproximadamente 53 vezes mais rápido do que o MCS de 2000 amostras e cerca de 264 vezes mais rápido em relação ao MCS de 10000 amostras, provando sua enorme eficiência computacional.

Figura 22 – Comparação FPMP-CRA x MCS: tensões elétricas na topologia IEEE 13.



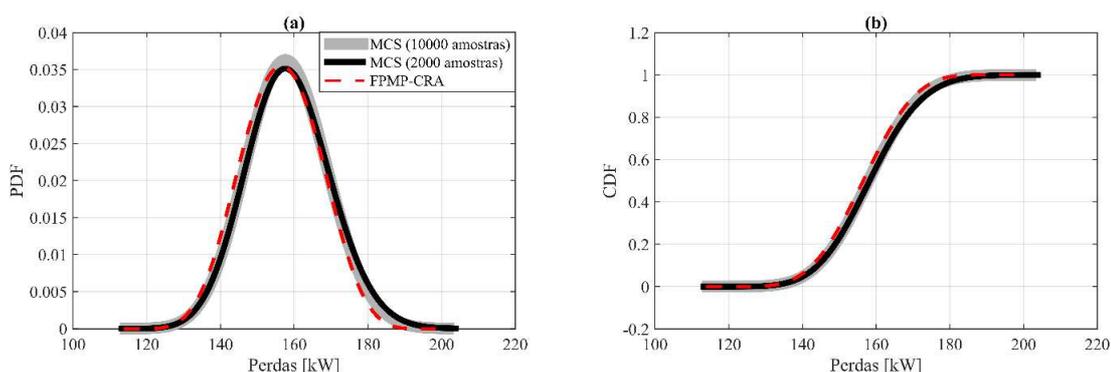
Fonte: elaborada pelo autor (2025).

Essa é uma das principais vantagens do método proposto: consegue alcançar resultados muito precisos em um tempo computacional extremamente favorável. Isso é imprescindível para a aplicação em problemas de planejamento que se baseiam em formulações de otimização que adotam restrições estocásticas, como as equações do FPP. Nesses casos, o problema deve ser avaliado muitas vezes, de maneira que métodos pouco eficientes sob a perspectiva do tempo computacional (tais como o próprio MCS e demais

métodos numéricos) tornam-se inaplicáveis. Uma alternativa é a simplificação do número de variáveis aleatórias, através da consideração da carga completamente correlacionada (Uniyal; Sarangi, 2021). No entanto, já foi visto, no teste 2, que há sérios prejuízos quanto à precisão dos resultados quando tal simplismo é assumido. A vantagem do FPMP-CRA reside no fato de não haver necessidade de simplificação da modelagem do problema. Pode-se considerar quaisquer níveis de correlação entre as cargas e a eficiência computacional se mantém.

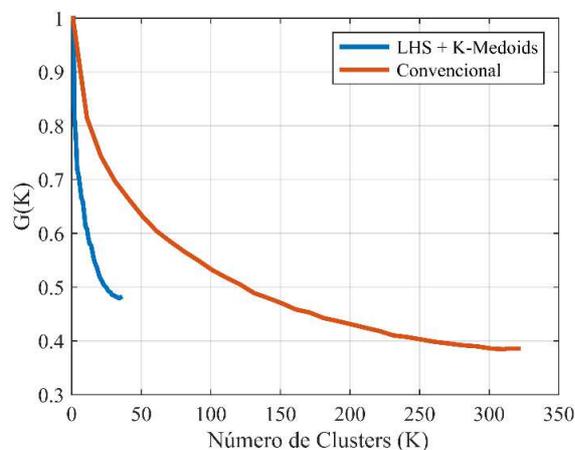
A Figura 23 apresenta a comparação entre FPMP-CRA e o MCS considerando a variável aleatória de perdas elétricas ativas totais no sistema. Esse parâmetro é conveniente para se estabelecer comparações, uma vez que depende de todas as outras variáveis de estado aleatórias, isto é. tensões em magnitude e ângulo.

Figura 23 – Comparação FPMP-CRA x MCS: perdas ativas totais na topologia IEEE 13.



Fonte: elaborada pelo autor (2025).

Figura 24 – Curvas de convergência para estimar número ótimo de *clusters*.



Fonte: elaborada pelo autor (2025).

Novamente, percebe-se uma altíssima precisão do método proposto em relação ao MCS, onde pode-se notar que as curvas são aproximadamente sobrepostas.

Por fim, buscando enfatizar a importância do método de otimização do número de *clusters*, são apresentadas, na Figura 24, as curvas de convergência: em azul, quando da utilização das diretrizes propostas, em vermelho, quando as diretrizes não são aplicadas (isto é, o método convencional). Para facilitar a compreensão visual, as curvas são traçadas até que o ponto ótimo seja atingido.

É possível notar que o método converge muito mais rápido quando as diretrizes propostas são aplicadas (alcançando um número total de 35 *clusters*, conforme já citado anteriormente). Considerando o método convencional, isto é, sem as diretrizes propostas, o número ótimo seria 323 *clusters*, o que demandaria um tempo computacional quase 10 vezes maior.

#### 4.4.4 CONCLUSÕES PARCIAIS DO TESTE 3

Os algoritmos de *clustering* particionais baseados em distâncias perdem sua precisão à medida que a dimensionalidade do banco de dados a ser comprimido aumenta. Isso é chamado de ‘maldição da dimensionalidade’, que também afeta quaisquer métodos de FPP que utilizam essas técnicas no seu desenvolvimento matemático e computacional.

O efeito desse fenômeno no FPP é a necessidade de se adotar um número cada vez maior de *clusters* para que a matriz comprimida seja estatisticamente equivalente à matriz original que contém as amostras das variáveis aleatórias de entrada. Através do método proposto de estimativa do número ótimo de *clusters*, foi possível constatar, por meio de simulações considerando a topologia IEEE 13, que quando o banco de dados possui uma baixa dimensionalidade, a técnica de *clustering* consegue eliminar até 97% de informações redundantes, enaltecendo a enorme eficiência da técnica. No entanto, tal percentual decai para aproximadamente 70% quando dimensionalidades maiores são aplicadas. O caso de maior dimensionalidade estudado contemplou 20 variáveis aleatórias (todas as cargas do sistema) e 2000 amostras das PDFs por variável. Nessa situação, foi possível reduzir em 70,3% a quantidade de cenários a serem simulados. Apesar de ser um bom resultado, é nítida a influência da ‘maldição da dimensionalidade’ nesse estudo.

Logo, buscando o aumento da eficiência do FPMP-CRA sob circunstâncias que impliquem a necessidade de simular problemas de alta dimensionalidade, foram propostas

alterações pontuais no seu algoritmo básico, chamadas *diretrizes para simulação de alta dimensionalidade*. Essas diretrizes promovem o preenchimento de uma lacuna presente no atual estado da arte de métodos de FPP, pois, como apontado no Capítulo 2, situações de alta dimensionalidade ainda estão em estudo, principalmente no que se refere aos métodos de FPP baseados em *clustering*.

Nessa perspectiva, a ideia das diretrizes é explorar a enorme eficiência computacional do FPMP-CRA, porém corrigindo a sua principal desvantagem, que reside na queda de eficiência quando a dimensionalidade do problema aumenta. Portanto, as diretrizes nada mais são do que estratégias para reduzir a dimensão do FPP sem que haja prejuízos significativos na precisão dos resultados das PDFs das variáveis de saída. São, no total, três diretrizes: amostragens conduzidas pela técnica LHS, o uso de uma versão adaptada do *clustering* K-Medoids chamada CLARA e a consideração de cargas cujos desvios padrões são menores do que 10 kW como determinísticas.

Através de simulações computacionais, foi possível constatar o aprimoramento do desempenho do FPMP-CRA por meio das diretrizes propostas: houve um substancial ganho computacional (cerca de 264 vezes mais rápido em relação ao MCS convencional e cerca de 10 vezes mais rápido em relação ao próprio FPMP-CRA sem as diretrizes), sem que houvesse perda significativa na qualidade dos resultados no que se refere à precisão de PDFs das variáveis aleatórias de saída.

A aplicabilidade do método proposto em situações de alta dimensionalidade é uma vantagem em relação a outros métodos de FPP presentes na literatura. Sua precisão, similar a métodos numéricos como o MCS, aliada à sua excelente performance computacional, credencia o mesmo a ser uma importante ferramenta para ser aplicada a problemas de planejamento, sobretudo quando as equações estocásticas do FPP são assumidas como restrições de um problema de otimização. Nesses casos, a velocidade de convergência do método torna-se crítica, e métodos numéricos tradicionais tornam-se inviáveis.

#### 4.5 TESTE 4: APLICAÇÕES COM GERAÇÃO DISTRIBUÍDA FOTOVOLTAICA

As redes de distribuição cada vez mais estão munidas de equipamentos de geração distribuída. Diz-se, portanto, que os SD estão se tornando ativos, pois as unidades geradoras estão localizadas próximas aos consumidores. Nesse sentido, tais sistemas têm seu comportamento significativamente alterado devido a essa quebra de premissa.

Dentre os empreendimentos mais comuns, destaca-se a GD fotovoltaica. Com o passar dos anos, a tecnologia fotovoltaica foi se tornando mais competitiva no mercado e, aliado a subsídios e incentivos governamentais, vêm sendo uma alternativa muito comumente vista nas redes de distribuição.

Trata-se de uma fonte de energia intermitente, cuja geração depende de condições atmosféricas, sobretudo a irradiação solar (Barik; Das, 2019). Logo, é altamente variável e estocástica e, por isso, é imprescindível que tais elementos sejam modelados com incertezas no problema de FP. As condições atmosféricas podem apresentar correlações estatísticas temporais, isto é, a probabilidade de ocorrência de um evento em determinado horário pode condicionar a ocorrência de um outro evento em um horário distinto. Isso impacta na modelagem do problema de FPP quando GDs fotovoltaicas são introduzidas.

A proliferação das GDs fotovoltaicas leva a diversos desafios. Destacam-se os distúrbios de tensão – mais especificamente sobretensões em horários de maior incidência solar.

Por esses motivos, o presente teste apresenta alguns intuitos: (i) simulação do FPMP-CRA diante da presença de geração distribuída fotovoltaica; (ii) modelagem de correlações temporais e estudo da aplicabilidade do FPMP-CRA sob essas condições de entrada; (iii) cálculo do número ótimo de *clusters* em um problema de diferentes níveis temporais; (iv) precisão dos resultados do FPMP-CRA para entradas modeladas por funções não Gaussianas correlacionadas.

#### 4.5.1 APRESENTAÇÃO DA TOPOLOGIA IEEE 123 E CONDIÇÕES DE SIMULAÇÃO

O sistema empregado para simulação deste teste é o IEEE 123. Trata-se de um sistema mais complexo, com 123 barras e 256 nós, que possui linhas e cargas monofásicas, bifásicas e trifásicas, além de um transformador. Os reguladores e capacitores foram desprezados, para enaltecer os efeitos das variáveis aleatórias. Nos trabalhos futuros, o impacto da inclusão dos reguladores será devidamente estudado sob a perspectiva probabilística. 20 GDs fotovoltaicas foram adicionadas à topologia. Na Figura 25, a topologia IEEE 123 é apresentada. As posições marcadas na cor vermelha são aquelas em que há a inserção de um gerador fotovoltaico. Mais informações sobre a topologia podem ser acessadas em *IEEE PES Test-Feeders* (2024).

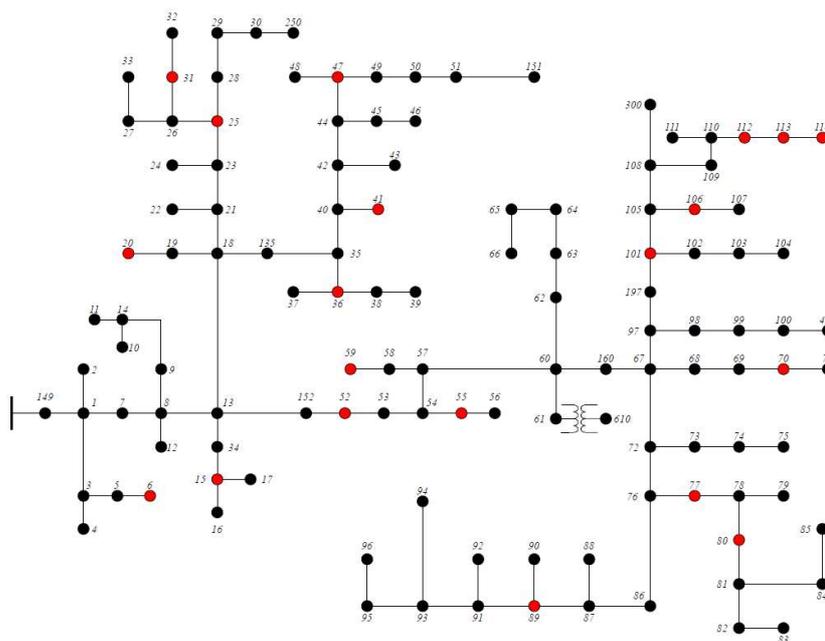
As localizações dos geradores e suas respectivas potências nominais são mostradas na Tabela 12. Esses valores foram retirados de Chai *et al.* (2025), ainda que, nesse trabalho, a topologia seja modificada em relação à proposta pelo IEEE.

Tabela 12 – Dados dos geradores fotovoltaicos

Localização	Potência [MW]
6, 20, 89	0,1
25, 36, 47, 52, 55, 77, 80, 112	0,2
31, 59, 101, 106, 113	0,4
15, 41, 70, 114	0,6

Fonte: elaborada pelo autor, baseada em Chai *et al.* (2025).

Figura 25 – Topologia IEEE 123.



Fonte: elaborada pelo autor (2025).

Neste trabalho, é assumido que os geradores localizados em nós bifásicos ou trifásicos têm suas potências geradas de maneira igual entre as fases.

As variáveis aleatórias consideradas para esse teste são as cargas do sistema e as gerações fotovoltaicas. Como discutido no Capítulo 2, ambas apresentam incertezas inerentes e precisam ser modeladas adequadamente para análises mais completas e relevantes.

Um horizonte de planejamento diário com discretização horária é considerado. Isso significa que há 24 variáveis aleatórias (uma para cada hora do dia) relativas às cargas, e 24 associadas às gerações fotovoltaicas. Todas essas variáveis são correlacionadas temporalmente utilizando a teoria de cópulas estatísticas, cujo procedimento para geração de amostras será esclarecido na Seção 4.5.2.

As cargas são modeladas por PDFs Gaussianas, cujas médias e desvios padrões são calculados diretamente de um banco de dados de medições reais, com o intuito de trazer o problema para condições mais próximas da realidade. As correlações temporais também são calculadas tendo como referência esse banco de dados.

Já as gerações fotovoltaicas são modeladas por PDFs Beta, cujos parâmetros  $\alpha$  e  $\beta$  são calculados diretamente de um banco de dados de medições reais de irradiação solar. As suas correlações temporais também são extraídas desse banco de dados<sup>17</sup>.

Como os dados obtidos de medições de cargas e irradiações são de localizações ligeiramente diferentes, optou-se por desconsiderar as correlações espaciais entre cargas e GDs. Em outras palavras, as cargas e GDs fotovoltaicas são variáveis aleatórias independentes entre si, em um mesmo intervalo de tempo. As dependências simuladas dizem respeito apenas ao quesito temporal, isto é, as cargas e GDs são correlacionadas ao longo do tempo. No entanto, ressalta-se que o modelo aplicado é facilmente adaptável para condições de correlações espaciais, além das dependências temporais, como será visto em detalhes na Seção 4.5.2, desde que dados confiáveis sejam obtidos no que diz respeito a isso.

As gerações fotovoltaicas são modeladas conforme mostrado em Barik e Das (2019): considera-se um modelo simplificado de injeção de potência ativa, diretamente proporcional à irradiação solar no tempo simulado. O fator de proporcionalidade é a potência nominal de cada gerador, expressa na Tabela 12.

Além disso, são consideradas duas simplificações, porém justificáveis:

- As gerações das GDs fotovoltaicas são consideradas perfeitamente correlacionadas espacialmente. Isso é uma simplificação razoável, uma vez que, apesar do sistema IEEE 123 ser complexo em número de nós e presença de elementos elétricos, ele não é tão extenso (possui extensão total de aproximadamente 17 km). Isso significa que as irradiações solares que incidem em todos os pontos do sistema podem ser consideradas aproximadamente as mesmas, o que levaria a condições muito parecidas de geração em todos os lugares do sistema

---

<sup>17</sup> As modelagens de funções densidade de probabilidade Gaussiana e Beta podem ser vistas em detalhes no Apêndice A.

trabalhado. Dessa forma, variações de altitude, condições atmosféricas (nuvens, poluição, etc) e sombreamento (prédios, árvores, etc) são desconsiderados.

- As cargas são consideradas perfeitamente correlacionadas espacialmente. Essa consideração é feita, uma vez que as cargas possuem variabilidade muito menor do que a geração renovável em estudo. Logo, o efeito da independência das cargas é menor do que o próprio efeito de variabilidade da GD.

Por fim, as duas considerações também se justificam por um particular interesse do teste em questão: um dos alvos da presente simulação é avaliar os impactos das correlações temporais no problema de FPP. Para isso, isola-se o fenômeno que se deseja avaliar das demais dependências espaciais que possam existir.

Ademais, busca-se mostrar que o FPMP-CRA é robusto o suficiente para lidar com esse tipo de dependência estatística, ao contrário de muitos métodos da literatura, que não são capazes de englobar tais características, tal como visto na revisão bibliográfica proposta no Capítulo 2.

#### 4.5.2 PROCEDIMENTO PARA GERAÇÃO DE AMOSTRAS CORRELACIONADAS TEMPORALMENTE

O FPMP-CRA é aplicado exatamente da mesma forma como mostrado no Capítulo 3. Em relação aos casos anteriores, a única diferença reside no cálculo das amostras a serem comprimidas, que agora precisam ser correlacionadas temporalmente. Neste teste, elas são geradas a partir da teoria de Cópulas. Para determinar os parâmetros das distribuições de probabilidade marginais, bancos de dados de medições horárias de demandas e irradiações solares são empregados. O procedimento completo pode ser visto no fluxograma exposto na Figura 26.

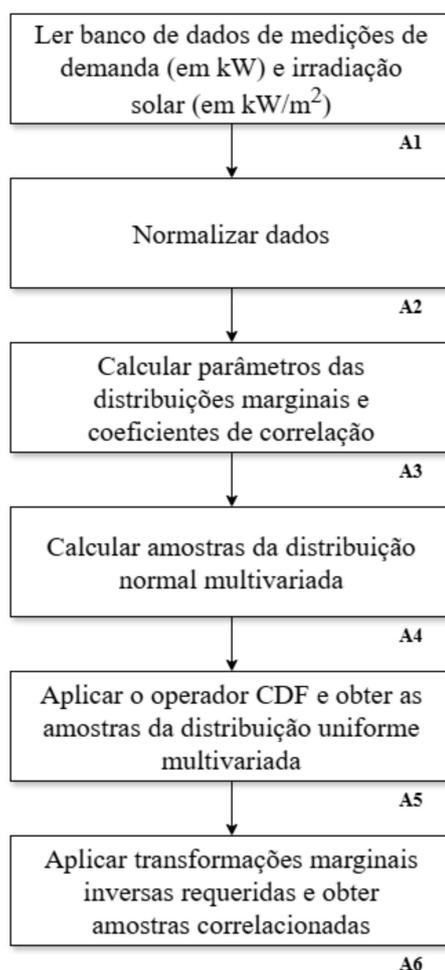
Cada passo do fluxograma será devidamente destacado e seus procedimentos serão elucidados.

*A1) Leitura dos bancos de dados:* o primeiro passo consiste na leitura de bancos de dados referentes ao consumo de potência ativa demandada (em kW), em medições horárias, e à irradiação solar (em kW/m<sup>2</sup>), também em medições horárias. Neste trabalho, foram consideradas medições em determinadas áreas da cidade do Rio de Janeiro. Como os dados obtidos de demandas e irradiações solares não são correspondentes à mesma região da cidade,

optou-se por desconsiderar as correlações estatísticas existentes entre cargas e gerações fotovoltaicas. No entanto, o procedimento é robusto o suficiente para levar em consideração esse aspecto, como será visto adiante, caso o usuário possua mais informações e dados relevantes. Além disso, o intuito, no presente teste, é averiguar os impactos das correlações temporais, que é o alvo primordialmente traçado de análise, e, não verificar impactos de correlações espaciais entre elementos elétricos.

Logo, são obtidas duas matrizes cujas dimensões são iguais a  $[N \times 24]$ , ou seja,  $N$  medições (de potência demandada e irradiação solar) para cada hora do dia.

Figura 26 – Procedimento para obter amostras correlacionadas temporalmente.



Fonte: elaborada pelo autor (2025).

*A2) Normalização*: os dados, nessa etapa, são normalizados, a fim de serem aplicados genericamente como variáveis de entrada de qualquer sistema elétrico. As cargas são

normalizadas pelo seu valor médio, considerando todas as medições, de modo que são obtidos fatores multiplicativos de cargas. A operação conduzida é mostrada em (4.4).

$$fc_i^h = \frac{D_i^h}{D^{med}} \quad (4.4)$$

Onde:  $fc_i^h$  é o fator multiplicativo de carga correspondente à  $i$ -ésima observação da hora  $h$ ;  $D_i^h$  é a  $i$ -ésima medição de demanda na hora  $h$ ;  $D^{med}$  é a demanda média, considerando todos os elementos do banco de dados.

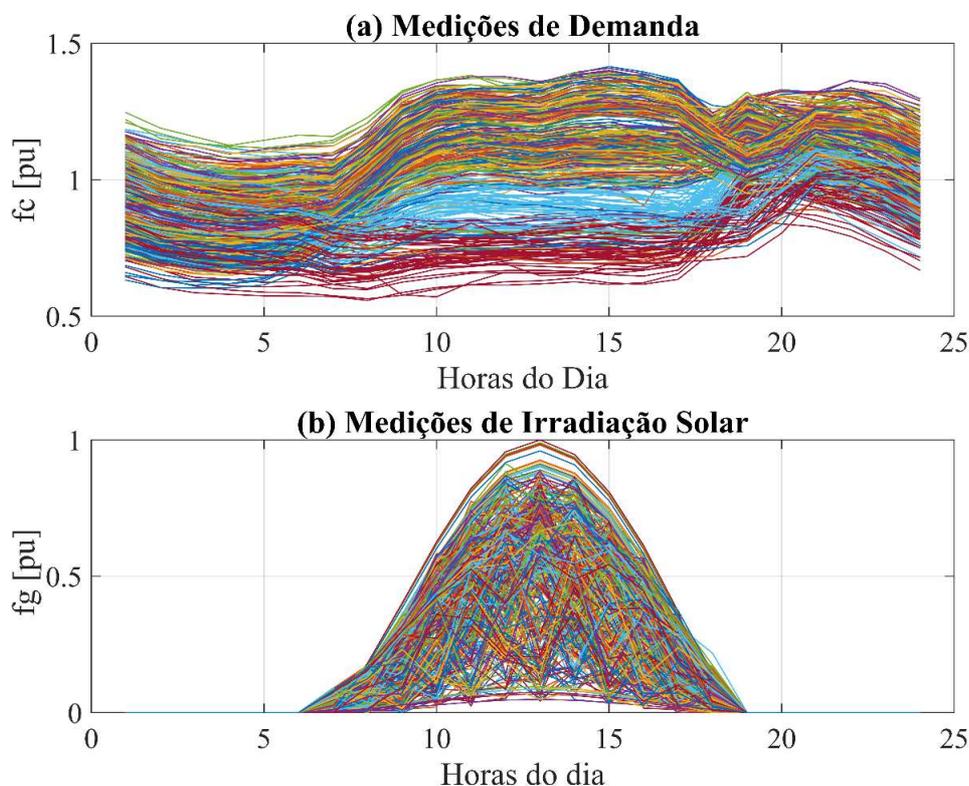
Similarmente, os dados de irradiação solar também são normalizados. Porém, os dados são divididos pela irradiação máxima, uma vez que a distribuição Beta, que melhor modela tal parâmetro, deve estar compreendida no intervalo  $[0,1]$ . Os valores obtidos são fatores multiplicativos de geração. A operação conduzida é mostrada em (4.5).

$$fg_i^h = \frac{H_i^h}{H^{max}} \quad (4.5)$$

Onde  $fg_i^h$  é o  $i$ -ésimo fator de geração, considerando hora  $h$ ;  $H_i^h$  é a  $i$ -ésima medição de irradiação solar para hora  $h$ ;  $H^{max}$  é a máxima irradiação solar, considerando todos os dados do banco de dados.

Após as operações (4.4) e (4.5), são obtidas duas famílias de curvas horárias, uma correspondente à carga do sistema e outra correspondente à geração fotovoltaica. O total de curvas de cada família é igual a  $N$ , o número de medições horárias. Na Figura 27, as curvas já normalizadas podem ser observadas. O número  $N$  de medições horárias adotado foi 365, compreendendo medições horárias de um ano de observações. Esse valor pode ser maior ou menor, dependendo da abrangência dos dados disponíveis. Como visto nos testes anteriores, isso irá interferir no número ótimo de *clusters*. Ressalta-se também que a discretização pode variar (horária, a cada 10 minutos, minuto a minuto, etc), dependendo dos dados disponibilizados e das intenções e objetivos do usuário. Simulações com diferentes discretizações temporais e seus devidos acoplamentos serão efetuadas e analisadas em trabalhos futuros.

Figura 27 – Curvas obtidas através dos dados: (a) Demandas; (b) Irradiação Solar.



Fonte: elaborada pelo autor (2025).

A3) *Calcular parâmetros*: nessa etapa, os parâmetros das distribuições marginais são calculados. Trata-se dos parâmetros que serão introduzidos na última etapa da teoria de Cópulas, isto é, os parâmetros que definirão as distribuições marginais que, por sua vez, serão temporalmente correlacionadas. Conforme, mencionado, as cargas serão modeladas por distribuições Gaussianas (cujos parâmetros são a média e o desvio padrão), e a irradiação solar será modelada pela distribuição Beta (cujos parâmetros são  $\alpha$  e  $\beta$ )<sup>18</sup>.

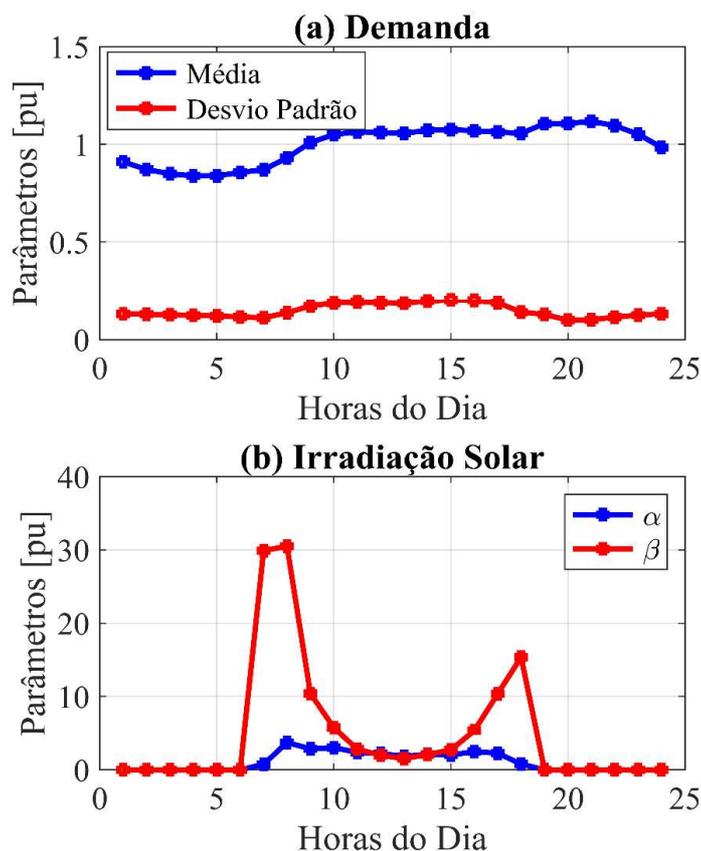
Os parâmetros da distribuição Gaussiana por hora do dia podem ser vistos na Figura 28.

É importante notar que, no caso dos parâmetros da irradiação solar, os valores de  $\beta$  são muito elevados entre 6 e 10 horas, e entre 16 e 18 horas. Trata-se de horários em que há uma alta variabilidade de irradiação ao longo dos dias do ano analisado, devido às diferentes

<sup>18</sup> Considerações acerca das modelagens das distribuições Gaussianas e Beta podem ser vistas no Apêndice A, assim como o cálculo de seus parâmetros.

estações. Nos horários entre 11 e 14 horas, os valores de  $\beta$  são mais baixos, e mais próximos aos de  $\alpha$ , o que indica distribuições mais simétricas.

Figura 28 – Parâmetros das distribuições marginais.



Fonte: elaborada pelo autor (2025).

Outro procedimento feito nessa etapa consiste no cálculo dos coeficientes de correlação das amostras horárias. Nesta simulação, são consideradas correlações temporais tanto de demandas quanto de irradiações solares. A matriz de correlação, denotada por  $\mathbf{R}$  e mostrada em (4.6), é usada para modelar as correlações de variáveis na teoria de Cópulas, tal como explicado no Capítulo 2. Para calcular as correlações, basta empregar a equação (2.2). São calculadas todas as correlações de demandas e irradiações solares, para todo par de horas possível. A teoria de Cópulas irá gerar, portanto, amostras correlacionadas que atendem a todas as correlações horárias calculadas por meio das observações contidas no banco de dados, tomadas duas a duas. Conforme expresso, as correlações entre cargas e irradiações são

desprezadas devido à falta de dados consistentes para tal análise estatística. Portanto, tais variáveis são assumidas como independentes para o presente estudo. A matriz  $\mathbf{R}$  de correlação pode ser definida como uma matriz dividida em quatro blocos, tal como em (4.6).

$$\mathbf{R} = \begin{bmatrix} \mathbf{P}_{DD} & \mathbf{P}_{DI} \\ \mathbf{P}_{ID} & \mathbf{P}_{II} \end{bmatrix} \quad (4.6)$$

Onde:  $\mathbf{P}_{DD}$  é a matriz de correlação temporal das demandas;  $\mathbf{P}_{DI}$  e  $\mathbf{P}_{ID}$  são as matrizes de correlação temporal-espacial entre demandas e irradiações solares;  $\mathbf{P}_{II}$  é a matriz de correlação temporal entre irradiações solares.

Lembra-se que  $\mathbf{P}_{DI} = \mathbf{P}_{ID} = \mathbf{0}$  (matriz nula), pois são desconsideradas as correlações entre demandas e irradiações solares. Além disso, cada matriz tem dimensão [24x24], o número de horas por dia. Logo, expandindo  $\mathbf{R}$ , tem-se que:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{a1a2} & \rho_{a1a3} & \dots & \rho_{a1a24} & 0 & 0 & 0 & \dots & 0 \\ \rho_{a2a1} & 1 & \rho_{a2a3} & \dots & \rho_{a2a24} & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{a24a1} & \rho_{a24a2} & \rho_{a24a3} & \dots & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & \rho_{i1i2} & \rho_{i1i3} & \dots & \rho_{i1i24} \\ 0 & 0 & 0 & \dots & 0 & \rho_{i2i1} & 1 & \rho_{i2i3} & \dots & \rho_{i2i24} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \rho_{i24i1} & \rho_{i24i2} & \rho_{i24i3} & \dots & 1 \end{bmatrix} \quad (4.7)$$

Onde:  $\rho_{axdy}$  e  $\rho_{ixiy}$  são os coeficientes de correlação das demandas e irradiações solares, respectivamente, referentes às horas  $x$  e  $y$ ;  $x, y \in \{0, \dots, 23\}$ . A diagonal principal de  $\mathbf{R}$  engloba apenas valores unitários, pois a covariância entre duas variáveis aleatórias iguais resulta na variância da variável aleatória. Como a matriz de correlação é a matriz de covariância normalizada pelo produto dos desvios padrões, isso resulta na unidade para essas posições da matriz.  $\mathbf{R}$  tem dimensão [48x48].

Logo, as amostras resultantes do processo de Cópula terão os parâmetros das distribuições marginais e as correlações temporais aproximadamente iguais às originalmente presentes nos bancos de dados. Isso confere ao problema simulado verossimilhança com condições reais de operação dos sistemas. Ademais, a teoria de Cópulas permite a geração de mais amostras do que se teria em um banco de dados limitado de medições, permanecendo com suas informações estatísticas inerentes.

*A4) Calcular amostras da distribuição normal multivariada:* os passos A4 a A6 correspondem às etapas padronizadas da teoria de Cópulas estatísticas, mostradas e esclarecidas em detalhes no Capítulo 2, mais especificamente na Figura 4. Inicialmente, calcula-se as amostras da distribuição normal multivariada, com os parâmetros de entrada correspondendo aos previamente avaliados pelo usuário, isto é, médias e a matriz de correlação. Nesse caso, há 48 variáveis (demanda e irradiação solar para cada hora do dia), de maneira que a matriz de correlação possui dimensão [48x48], tal como visto na equação (4.7). Nessa etapa, são obtidas amostras normalmente distribuídas, tal que suas correlações obedecem aos valores informados na matriz  $\mathbf{R}$ . Ressalta-se que, nas simulações presentes, há correlações temporais entre demandas e irradiações solares entre si, mas não correlações espaciais-temporais entre demandas e irradiações.

*A5) Aplicar o operador CDF:* nesta etapa, o operador CDF é aplicado às amostras Gaussianas multivariadas, de modo que cada variável aleatória resultante tenha comportamento uniformemente distribuído.

*A6) Aplicar transformações marginais inversas:* finalmente, transformações inversas são aplicadas, para gerar amostras correlacionadas temporalmente de acordo com as distribuições de probabilidade desejadas pelo usuário. Conforme mencionado, demandas têm suas distribuições marginais modeladas por funções Gaussianas. Já as irradiações solares, funções Beta. Os parâmetros de cada PDF são os valores extraídos dos bancos de dados, isto é, os mostrados na Figura 28. Além disso, o procedimento de Cópulas garante que as correlações temporais são aproximadamente iguais aos valores informados na matriz  $\mathbf{R}$ <sup>19</sup>. Em outras palavras, as variáveis aleatórias produzem amostras que respeitam as funções de probabilidade requeridas pelo usuário, além de estarem temporalmente correlacionadas de acordo com os valores informados pelo mesmo. Isso garante uma representação física muito fiel em relação à realidade, o que aumenta o detalhamento das condições de entrada do FPMP-CRA.

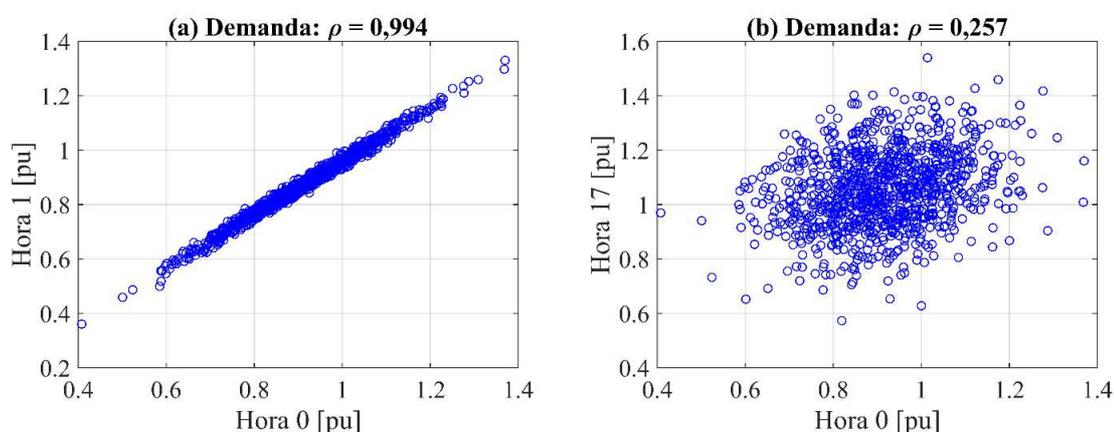
Para as simulações práticas propostas, observou-se que:

- As correlações entre demandas em horas consecutivas são muito fortes (valores acima de 0,99). Elas vão diminuindo à medida que a janela temporal aumenta. Por exemplo: a correlação temporal das demandas entre as horas 0 e 1 vale 0,994. Isso denota que existe enorme previsibilidade entre o consumo de potência entre essas duas horas (são quase perfeitamente correlacionadas). Em outras palavras, uma alteração de potência consumida na

<sup>19</sup> Diz-se que as correlações são aproximadamente iguais, não exatamente iguais, pois elas tendem aos valores informados na Cópula, à medida que o número de amostras tende ao infinito.

hora 0 irá provocar uma alteração quase proporcional no valor da hora 1. Trata-se de uma correlação muito forte. Já a correlação entre as horas 0 e 17 vale 0,257. As dispersões de amostras terão caráter menos correlacionado (mais ‘independente’), por causa disso. Uma ilustração comparativa desses dois casos pode ser vista na Figura 29. De forma geral, demandas de horas consecutivas são fortemente correlacionadas. Essa correlação vai diminuindo à medida que as horas se distanciam.

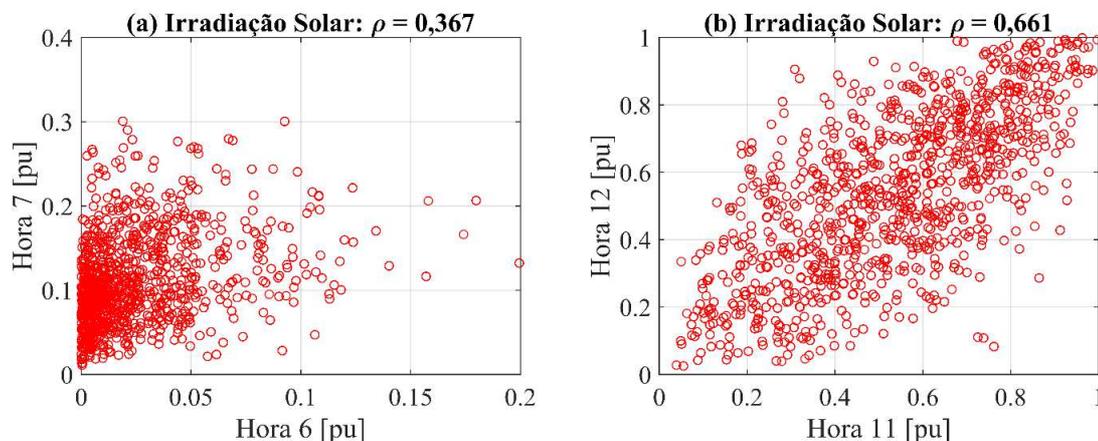
Figura 29 – Amostras temporalmente correlacionadas de demanda.



Fonte: elaborada pelo autor (2025).

- Já as correlações temporais de irradiação solar são mais baixas para as primeiras horas após o Sol nascer, e mais elevadas para as horas consecutivas mais próximas do meio-dia. Por exemplo, a correlação temporal entre a irradiação solar das horas 6 e 7 vale 0,367. Já a correlação temporal entre as horas 12 e 13 vale 0,661. Isso pode ser justificado pela mudança de estações do ano: os dias têm intensidades de irradiações diferentes mediante a estação e, dependendo, o Sol pode nascer mais cedo ou mais tarde, o que aumenta a imprevisibilidade desse fenômeno. Já nas horas mais próximas do meio-dia, a irradiação é mais previsível para essa região da cidade do Rio de Janeiro. No entanto, frisa-se que, mesmo para horas mais próximas do meio-dia, a correlação não é tão forte quanto as vistas para cargas. Isso é esperado, uma vez que as incertezas inerentes às gerações fotovoltaicas são mais expressivas (mais imprevisíveis) do que no caso de demandas, devido ao caráter mais estocástico da irradiação solar, que, por sua vez, depende de condições atmosféricas altamente incertas. Uma ilustração que cobre o ponto destacado pode ser vista na Figura 30.

Figura 30 – Amostras correlacionadas temporalmente de irradiação solar.



Fonte: elaborada pelo autor (2025).

- De qualquer maneira, é importante destacar que as considerações destacadas se referem exclusivamente para os bancos de dados trabalhados. Caso a localização geográfica mude, tais indícios podem ser totalmente diferentes.

Percebe-se que as amostras de demanda, na Figura 29 (a) são concentradas em uma região muito mais limitada em relação à Figura 29 (b). Isso é consequência da correlação forte que existe entre demandas de horas consecutivas.

Analogamente, na Figura 30 (a), as amostras se concentram primordialmente na origem do plano cartesiano. Isso significa que, em muitos cenários, as irradiações solares são nulas em ambas as horas 6 e 7. Existe uma concentração razoável de cenários em que a irradiação é nula para a hora 6, mas diferente de zero para a hora 7. Essa diferença é justificada pelas diferentes estações do ano, que alteram o horário do nascimento do Sol. Na Figura 30 (b), as amostras são mais dispersas.

#### 4.5.3 RESULTADOS DO FPMP-CRA CORRELACIONADO TEMPORALMENTE

Em consonância com o que já foi relatado, para cada cenário obtido pelo *clustering*, a carga do sistema é atualizada segundo (4.8) e (4.9).

$$P_{k,i}^h = f c_i^h \cdot P_{k,nom} \quad (4.8)$$

$$Q_{k,i}^h = fc_i^h \cdot Q_{k,nom} \quad (4.9)$$

Onde:  $P_{k,i}^h$  e  $Q_{k,i}^h$  são as potências ativa e reativa, respectivamente, da  $k$ -ésima carga, considerando  $i$ -ésimo cenário e  $h$ -ésima hora do dia;  $fc_i^h$  é o fator multiplicativo de carga, considerando  $i$ -ésimo cenário e  $h$ -ésima hora do dia;  $P_{k,nom}$  e  $Q_{k,nom}$  são as potências ativa e reativa, respectivamente, da  $k$ -ésima carga, informadas originalmente na topologia. Note que o fator de potência é mantido no mesmo valor originalmente contido na topologia.

Similarmente, as potências das GDs fotovoltaicas são atualizadas de acordo com o cenário a ser computado, de acordo com (4.10).

$$P_{m,i}^h = fg_i^h \cdot P_{m,nom} \quad (4.10)$$

Onde:  $P_{m,i}^h$  é a potência ativa da  $m$ -ésima GD, considerando o  $i$ -ésimo cenário e a  $h$ -ésima hora do dia;  $fg_i^h$  é o fator multiplicativo de geração, correspondente ao  $i$ -ésimo cenário e  $h$ -ésima hora do dia; e  $P_{m,nom}$  é o valor nominal da GD, de acordo com a Tabela 12. Portanto, considera-se uma relação linear entre a geração fotovoltaica e a irradiação solar, tal como sugerido em Barik e Das (2019). A potência reativa é mantida nula, tal como no modelo adotado por esses mesmos autores. Modelos mais completos para modelar a GD e possibilidades de controle de potência reativa serão estudados em trabalhos futuros.

Como já mencionado, os passos para solucionar o FPMP-CRA são precisamente os mesmos discutidos no Capítulo 3, com a única diferença que as amostras a serem comprimidas via *clustering* são, agora, correlacionadas temporalmente, seguindo o procedimento mostrado na Seção 4.5.2.

Como as amostras geradas são correlacionadas temporalmente, optou-se por realizar o *clustering* para cada grupo horário de amostras, tomado de forma individualmente. Em outras palavras, para cada hora do dia, reúnem-se as amostras previamente calculadas de demandas e irradiações solares, calcula-se o número ótimo de *clusters* para esse conjunto e, por fim, a partição é conduzida utilizando-se K-Means. Apesar de serem 48 variáveis aleatórias (demandas e irradiações solares para cada hora do dia), optou-se por utilizar K-Means, pois, ao particionar o problema temporalmente, restam apenas duas variáveis aleatórias por hora simulada. Essa escolha foi motivada por duas razões:

- Como os espaços de amostras são muito diferentes para cada hora do dia, o número ótimo de *clusters* requisitado para representar adequadamente o espaço de amostras de cada hora do dia pode ser significativamente, especialmente entre 0 e 5 horas, e entre 18 e 23 horas, em que há apenas uma variável aleatória (a demanda), pois, a irradiação solar para essas horas é nula e pode ser desconsiderada. Como visto no Teste 3, o número de variáveis aleatórias impacta sensivelmente no número ótimo de *clusters*. Consequentemente, se existem horas que requerem um número menor de *clusters* do que outras, a adoção de um número fixo de *clusters* para todas as horas do dia pode trazer um sobredimensionamento ou subdimensionamento do tempo computacional.
- A partição da matriz completa de amostras é mais desafiadora e mais custosa computacionalmente, além de requerer um número muito maior de *clusters* para representar adequadamente a matriz de entrada, dadas as 48 variáveis aleatórias.

É importante ressaltar que essa escolha não prejudica a modelagem das correlações temporais, pois as amostras originais já estão correlacionadas temporalmente. Logo, se as matrizes comprimidas, após o *clustering*, representam adequadamente as matrizes originais, as correlações continuarão corretamente definidas, seja fazendo-se o *clustering* do horizonte temporal completo ou hora a hora.

O número ótimo de *clusters* para cada hora do dia está presente na Tabela 13. Nota-se que, de fato, o número ótimo de *clusters* difere significativamente de acordo com a hora do dia. O número tende a ser maior entre 6 e 17 horas, uma vez que a irradiação solar tende a ser diferente de zero nesses horários, independentemente da estação do ano, o que aumenta o número de variáveis aleatórias. Além disso, é possível perceber que o número de *clusters* tende a ser igual ou similar em horas consecutivas, duas a duas, o que está em consonância com os pressupostos de que as correlações temporais são mais expressivas em horas consecutivas.

O número total de cenários a serem avaliados é igual a 1410, contra 24000 da matriz original de amostras. Isso configura uma redução de mais de 94% dos cenários. Caso a partição fosse conduzida considerando-se todas as horas conjuntamente, o número ótimo de *clusters* seria 175, totalizando  $175 \cdot 24 = 4200$  cenários a serem avaliados. Apesar de representar uma redução de 82,50% em relação à configuração original, são quase 3 vezes mais cenários em confronto com a estratégia adotada, o que aumentaria substancialmente o tempo computacional.

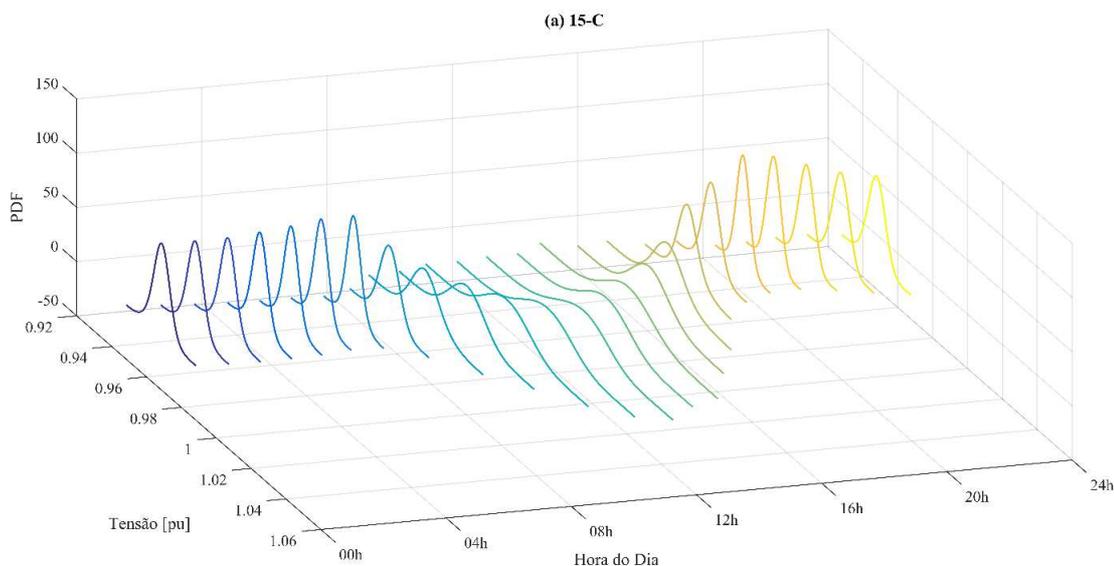
Tabela 13 – Número ótimo de *clusters* por hora do dia.

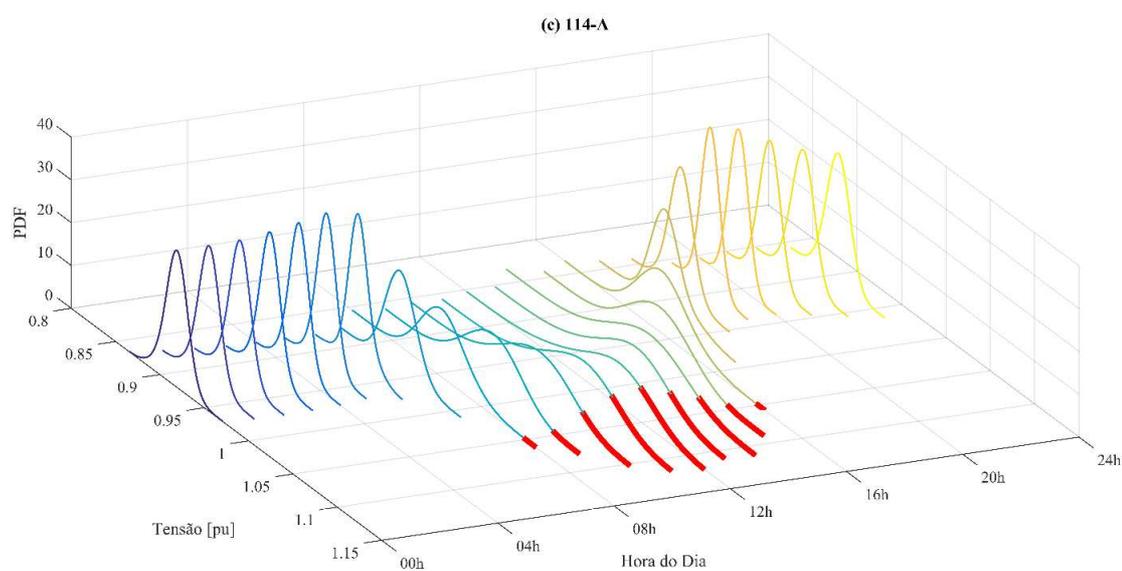
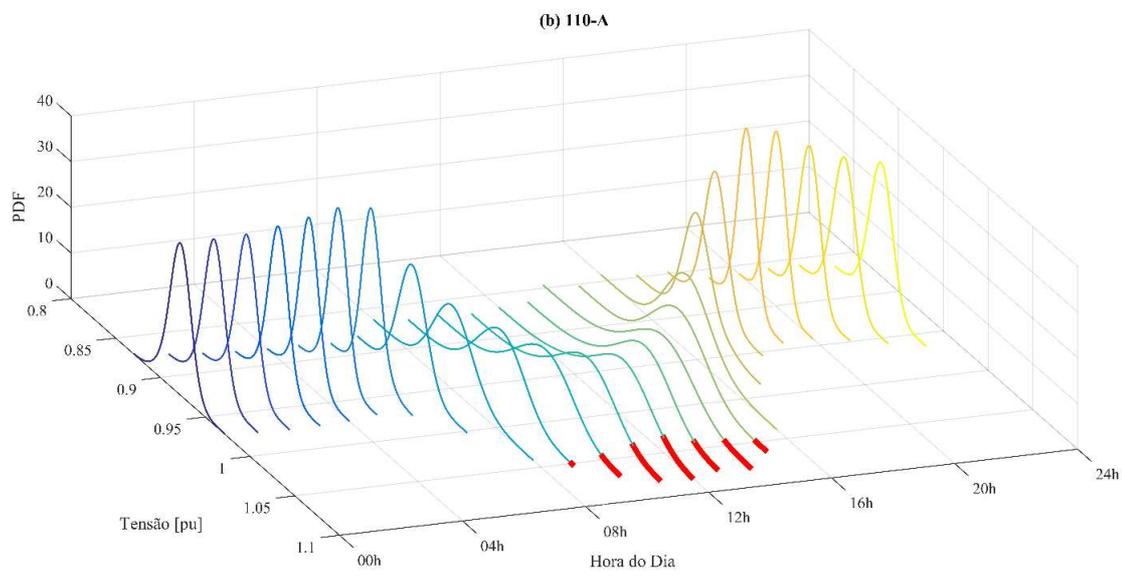
Hora	K	Hora	K	Hora	K	Hora	K
0	34	6	52	12	93	18	34
1	35	7	73	13	72	19	43
2	44	8	93	14	74	20	33
3	32	9	82	15	95	21	35
4	43	10	83	16	83	22	33
5	45	11	83	17	72	23	44

Fonte: elaborada pelo autor (2025).

Após a compressão de cada espaço de amostras para cada hora do dia, e da execução do FPMP-CRA, são obtidas as informações estatísticas de cada variável aleatória de saída a se analisar. Por questões de conveniência de aproveitamento de espaço, foram escolhidas para análise as magnitudes de tensão nodais das seguintes posições: 15-C, 110-A e 114-A, que são as posições que têm maior penetração de potência oriunda de GDs.

Figura 31 – PDFs de magnitude de tensão, de acordo com a hora do dia: (a) 15-C; (b) 70-A; (c) 114-A.





Fonte: elaborada pelo autor (2025).

Na Figura 31 (a), há as PDFs de magnitude de tensão para o nó 15-C, considerando cada uma das 24 horas do dia. É possível perceber que, nas primeiras e nas últimas horas do dia – quando não há irradiação solar – as PDFs têm um pico maior, uma vez que há apenas o impacto da imprevisibilidade da carga. Isso resulta na limitação de seus desvios padrões,

tornando o intervalo de operação mais reduzido e previsível. Já nas horas entre 6h e 17h, as funções são muito mais abrangentes, devido à forte incerteza da geração fotovoltaica.

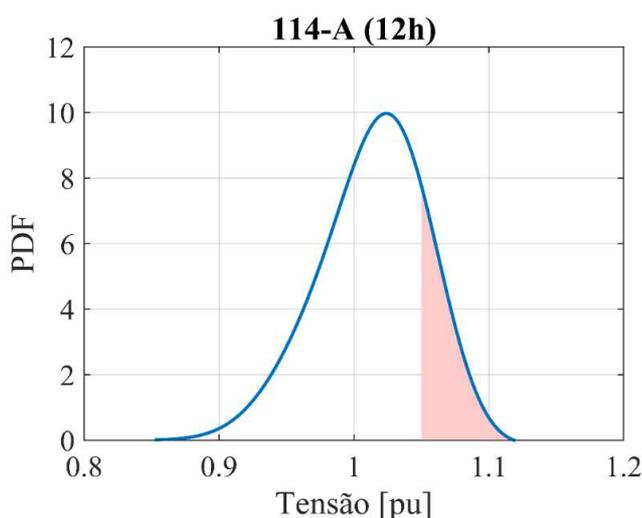
Além disso, o intervalo de operação, nesses horários, é muito mais amplo, existindo preocupações acerca de sobretensão (as tensões são muito mais elevadas nesses horários, devido às GDs). No entanto, os valores de tensão podem assumir valores igualmente muito baixos, devido à incerteza da geração fotovoltaica ao longo das estações do ano. Existem cenários em que a produção de potência é muito baixa, com elevados carregamentos, o que ocasiona elevadas quedas de tensão no alimentador. Ressalta-se que esses cenários são contabilizados apenas por meio de um método probabilístico capaz de lidar com correlações de variáveis aleatórias de entrada, tal como o FPMP-CRA. Uma abordagem determinística, ou até uma probabilística que contém limitações quanto a essa modelagem, não é capaz de gerar resultados confiáveis.

Outro aspecto importante de se perceber é a semelhança que existe nas PDFs de horas consecutivas, o que enaltece a importância de se considerar as correlações temporais na modelagem do problema. Qualitativamente, é possível perceber que há um padrão no comportamento das PDFs para horários antes de 6h e após 17h, quando não há impactos da irradiação solar, e um outro padrão entre tais horários. Isso permite ao usuário ter certa previsibilidade de como se manifestará o comportamento estocástico da tensão nodal, antes mesmo da simulação do FPMP-CRA. Essa característica é proveniente das fortes correlações que existem entre variáveis aleatórias de horários consecutivos, como comentado na Seção 4.5.2.

As conclusões citadas valem para as análises das Figuras 31 (b) e (c), nas quais as posições analisadas são 110-A e 114-A. No entanto, ao contrário da situação comentada anteriormente, para essas últimas posições há probabilidades de sobretensão (isto é, magnitudes de tensão que excedem 1,05 pu, limiar máximo estabelecido pelo PRODIST). As ‘caudas’ das PDFs que denotam regiões em que há probabilidade de sobretensão estão destacadas em vermelho, em ambas as figuras. Note que as probabilidades de sobretensão existem apenas nos horários em que há maior incidência solar, o que é esperado, uma vez que tal condição aumenta a produção de potência por parte das GDs. As probabilidades são ainda mais expressivas no caso da posição 114-A, pois os nós adjacentes (112-A e 113-A) também apresentam penetração de potência fotovoltaica. De forma particular, na Figura 32, é apresentada a PDF para magnitude de tensão da posição 114-A, às 12 horas. A região

marcada em rosa corresponde à probabilidade de violação do limiar de tensão máximo admitido pela ANEEL.

Figura 32 – PDF de tensão para posição 114-A, às 12 horas.



Fonte: elaborada pelo autor (2025).

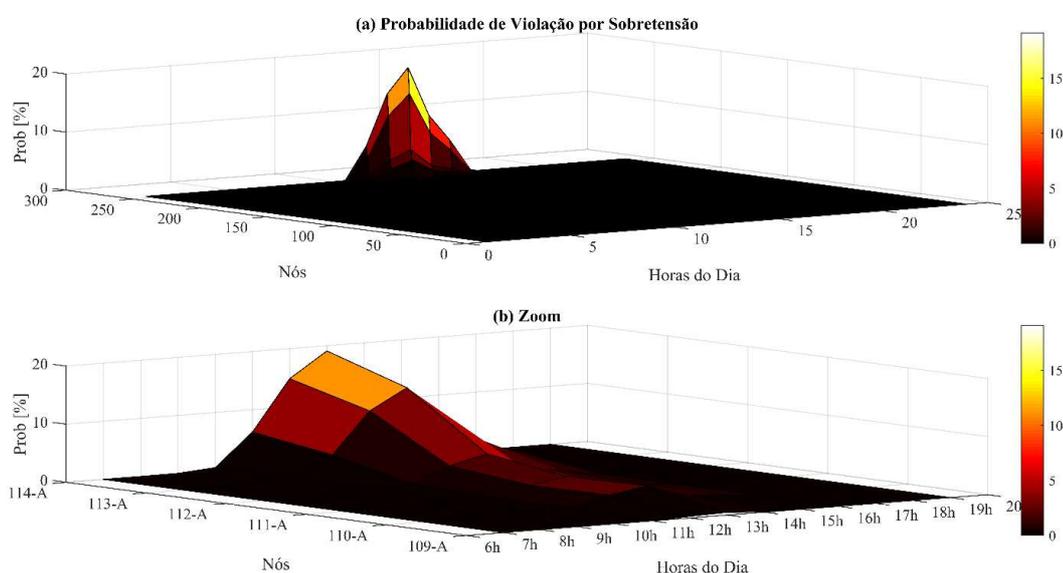
Nesse caso, a probabilidade de violação de tensão é 19,31%, valor este muito expressivo. Além disso, em quase 7% dos cenários, há tensões acima de 1,07 pu, que é uma tensão de atendimento considerada crítica pelo PRODIST (2022). É imprescindível que as concessionárias levem em consideração tais distúrbios em suas redes de distribuição, para subsidiar tomadas de ação relevantes. No entanto, ressalta-se que a constatação de tais fenômenos só é possível quando são utilizadas as ferramentas apropriadas, como o FPMP-CRA. É de interesse estimar a duração das sobretensões, e isso também é uma possibilidade de simulação da ferramenta proposta, desde que os dados que compõem as discretizações temporais sejam precisas o suficiente. Isso será tratado em trabalhos futuros.

De forma a generalizar o resultado anterior, as probabilidades de violação de tensão por sobretensão são apresentadas na Figura 33 (a), para todos os nós do sistema, e para todas as horas do dia. Trata-se, portanto, de um gráfico tridimensional.

É possível perceber que na grande maioria das posições, não há qualquer violação, mesmo considerando os diferentes horários do dia. Isso é destacado no gráfico através do *plateau* na cor preta, que percorre a maioria dos nós e horários do dia. No entanto, há uma região específica do sistema em que há severos riscos de sobretensão. Trata-se da região compreendida pelos nós 109 a 114, sendo nós monofásicos, possuindo apenas a fase A. Além

disso, é uma região com intensa produção fotovoltaica, possuindo GDs nos nós 112, 113 e 114. Dessa forma, a intensa geração nos horários de maior intensidade de irradiação solar faz com que essa região seja a mais crítica do sistema quanto à ocorrência desses distúrbios. Na Figura 33 (b), há um *zoom* para melhor visualização das probabilidades de sobretensão. São destacados tanto a região mais crítica (nós 109 a 114), quanto os horários do dia em que essas violações ocorrem (6h às 18h). Percebe-se que as violações podem chegar próximas a 20% em determinados cenários, o que é bastante preocupante para a operação correta do sistema.

Figura 33 – Probabilidades de violação por sobretensão.

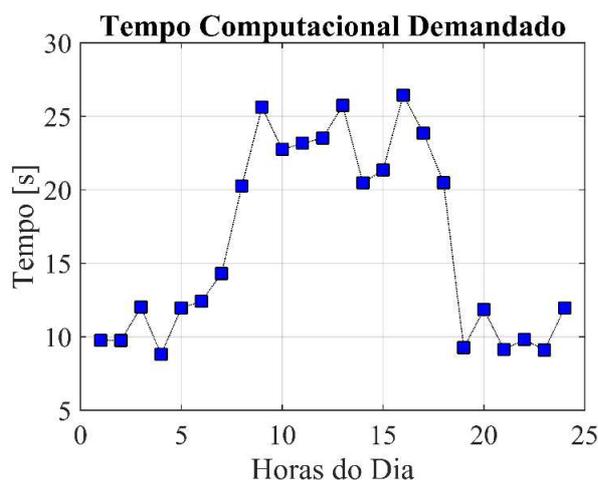


Fonte: elaborada pelo autor (2025).

A identificação de regiões críticas em redes de distribuição ativas é imprescindível para a tomada de ações relevantes, e concessionárias requerem cada vez mais métodos que calculem essas probabilidades de violação de forma precisa, para que as soluções empregadas para mitigar tais distúrbios sejam eficientes. Ademais, a identificação correta dos cenários a serem simulados, assim como das equações probabilísticas do FPMP-CRA podem ser empregadas para subsidiar problemas de máxima capacidade de hospedagem, em que a identificação das regiões críticas é fundamental, sobretudo em se tratando de redes altamente desequilibradas, como a apresentada neste teste. Ressalta-se que a precisão e a relevância das ações tomadas pelos agentes planejadores ou operadores passa por um estudo que considere a abordagem multifásica das redes, para que os resultados das simulações sejam fidedignos.

Por fim, é feita uma análise comparativa acerca do tempo computacional demandado pelo FPMP-CRA. Na Figura 34, estão contidos os tempos demandados referentes à simulação de cada hora do dia.

Figura 34 – Tempo computacional demandado pelo FPMP-CRA correlacionado temporalmente.



Fonte: elaborada pelo autor (2025).

Os tempos variam de acordo com a hora do dia simulada, pois, conforme visto nos testes anteriores, tal parâmetro é diretamente proporcional ao número de *clusters* do FPMP-CRA. Portanto, em consonância com os valores mostrados na Tabela 13, as horas que requerem um número maior de *clusters* também vão demandar um tempo computacional maior. Trata-se das horas mais influenciadas pela irradiação solar (entre 7h e 19h), devido ao aumento do número de variáveis aleatórias e de amostras não-nulas de irradiações solares, o que torna mais complexa a tarefa do *clustering*. Logo, nota-se que a curva do tempo computacional segue aproximadamente o formato da curva de geração fotovoltaica, pois ela que irá condicionar a complexidade computacional do FPMP-CRA.

De qualquer forma, o tempo maior observado foi 26 segundos (às 16h), o que está em completa aplicabilidade para problemas de planejamento ou operação de curto prazo. Mesmo que se queria avaliar o estado probabilístico do sistema na hora seguinte, o FPMP-CRA provê velocidade de convergência suficiente para o usuário ser informado de tais condições a tempo de se promover as ações necessárias nesses dois âmbitos. Ressalta-se que o FPMP-CRA não pode ser aplicado, sob essas condições, para situações de operação em tempo real, mas esse

tempo pode ser ainda acelerado por computação paralela e pela alteração do próprio algoritmo de FP multifásico executado para cada *cluster*, o que será explorado em trabalhos futuros.

Em comparação com o método tradicional, ou seja, simulando-se todos os cenários gerados pela cópula estatística, o FPMP-CRA promove um ganho de aproximadamente 93%. A simulação de todos os cenários faz a convergência ocorrer em 6711 segundos (aproximadamente 1h52min) contra 7 minutos do método proposto, considerando-se todas as 24 horas. Logo, trata-se de um ganho muito expressivo. Em diversas aplicações, precisa-se saber com urgência o estado do sistema em um horizonte diário, ou seja, em se tratando de algumas ou muitas horas adiante. Um exemplo que pode ser citado envolve aplicações de mercado de energia elétrica, em que os preços das gerações são atualizados de acordo com as dinâmicas estocásticas. Uma ferramenta que promova uma simulação rápida, eficiente e precisa é capaz de subsidiar resultados positivos para a liquidação dos custos de uma geração distribuída no âmbito do mercado. Aplicações de otimização do FP também podem ser citadas. Nelas, o custo computacional já é demasiadamente alto, especialmente em problemas que considerem 24 horas em um horizonte de planejamento diário. Em problemas de natureza combinatória, como alocação de recursos energéticos distribuídos, é comum a adoção de algoritmos metaheurísticos e evolucionários. Em alguns casos, a complexidade computacional e combinatória é tão elevada que métodos clássicos são inviáveis. Nesse sentido, métodos probabilísticos rápidos são requisitados, pois o FPP é executado muitas vezes até que se determine a solução ótima. Por isso, métodos tradicionais, como o MCS são inviáveis para essas aplicações. Usuários podem usar o FPMP-CRA correlacionado temporalmente para explorar tais situações.

As Figuras 35 a 37 mostram a comparação dos dois principais momentos das magnitudes de tensões nodais (média e desvio padrão), calculadas pelo método tradicional MCS (marcados em círculos azuis) em relação aos obtidos via FPMP-CRA (marcados por 'x'). Por questões de aproveitamento de espaço, foram definidos os horários de 00h, 12h e 18h para serem demonstrados.

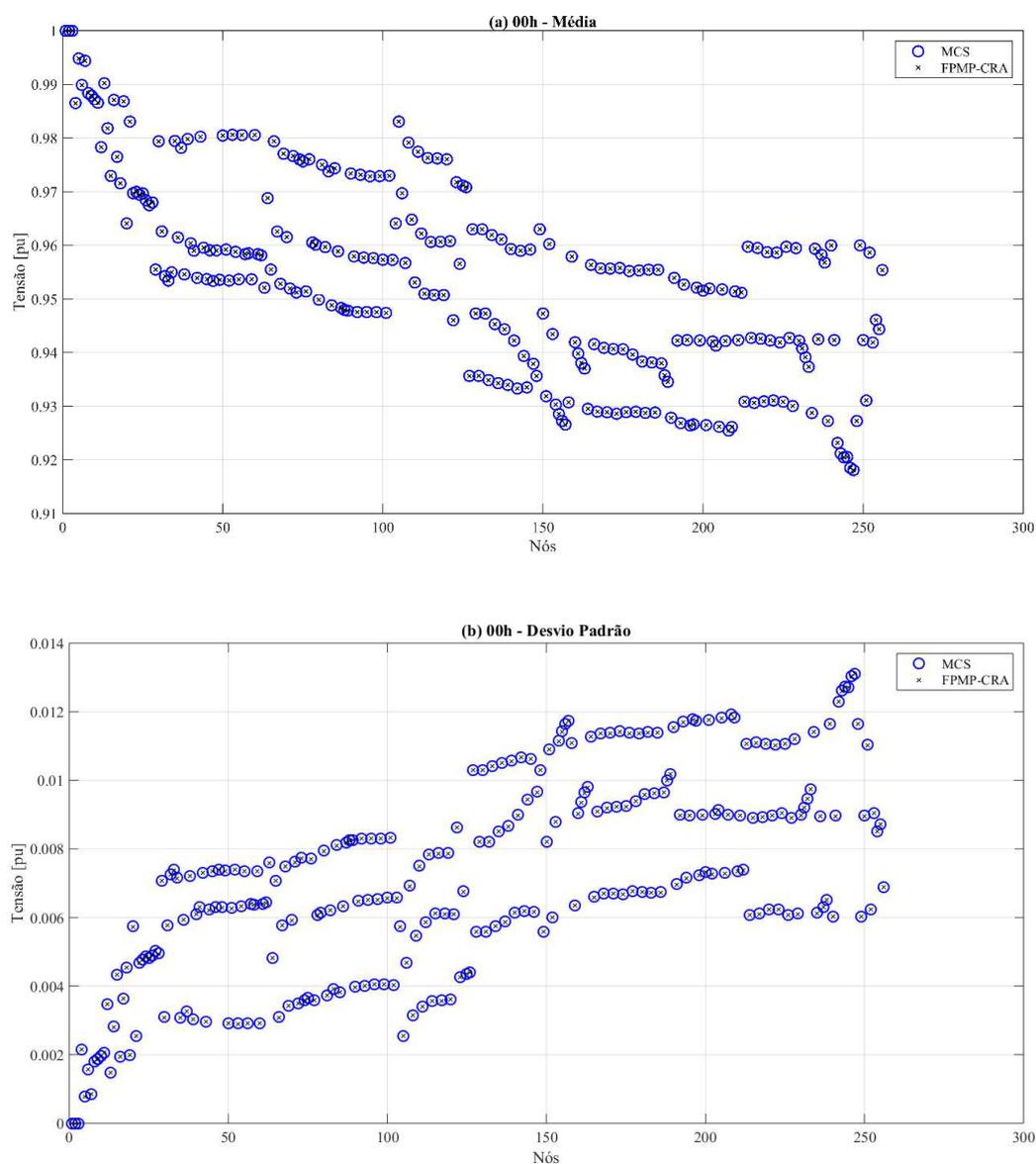
Os valores dos dois métodos são quase coincidentes para todas as posições e todos os horários do dia, o que demonstra que:

- A estratégia de otimização do número de *clusters* foi feita de maneira correta, uma vez que os resultados são estatisticamente equivalentes em ambos os métodos.
- Promover o *clustering* hora a hora, ao invés de solucionar usando as curvas de carga e geração conjuntamente é um procedimento válido. Uma vez que os espaços originais de dados

contêm as correlações instituídas nas cópulas e os espaços comprimidos são estatisticamente equivalentes, tem-se que estes também mantêm as correlações temporais. No entanto, isso garante uma eficiência computacional maior, como já mencionado, visto que existirá um número menor de variáveis aleatórias por hora do dia, o que faz o método requerer um número ótimo menor de *clusters*.

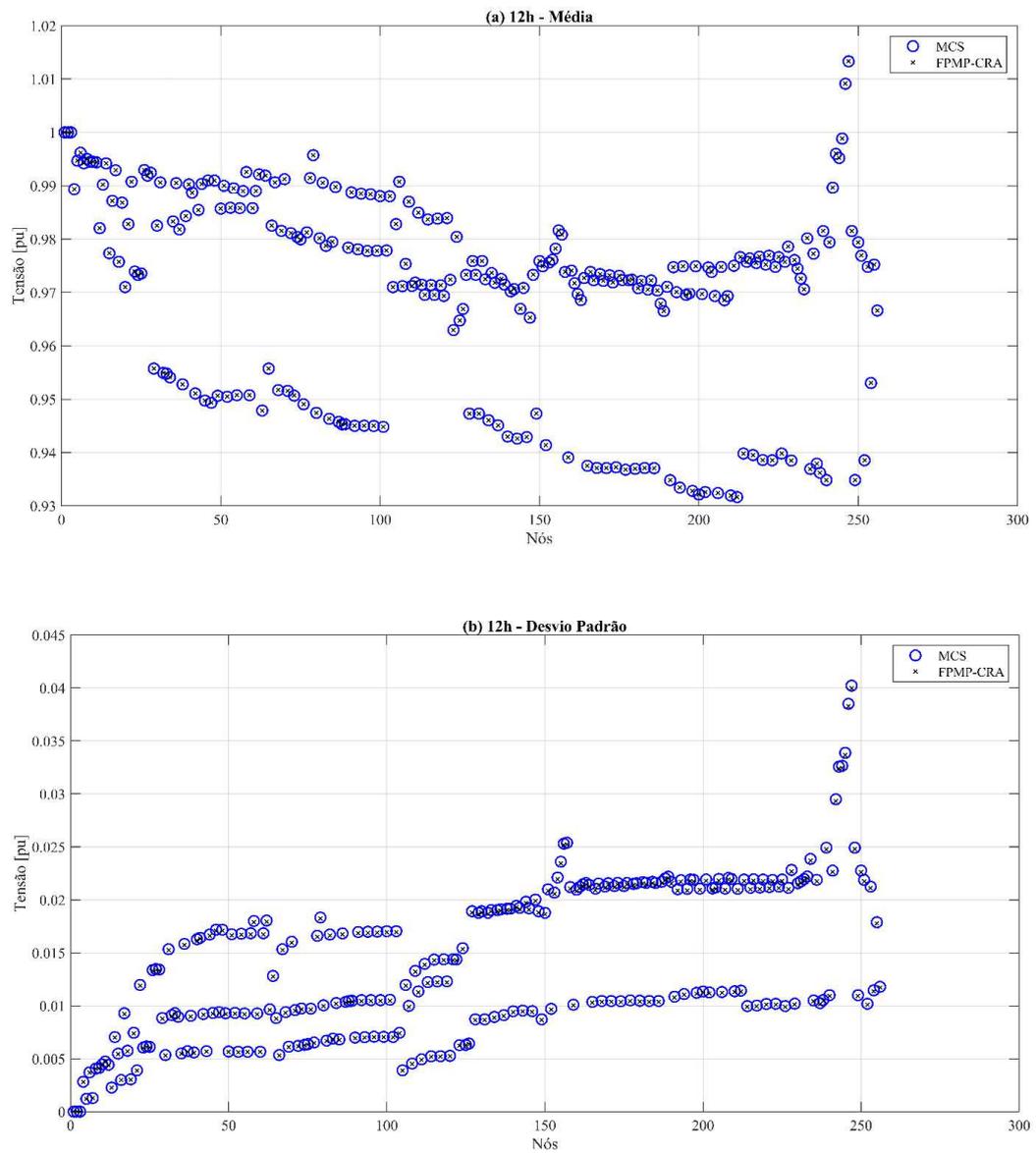
- 

Figura 35 – Momentos estatísticos das magnitudes de tensão: MCS x FPMP-CRA (00h).



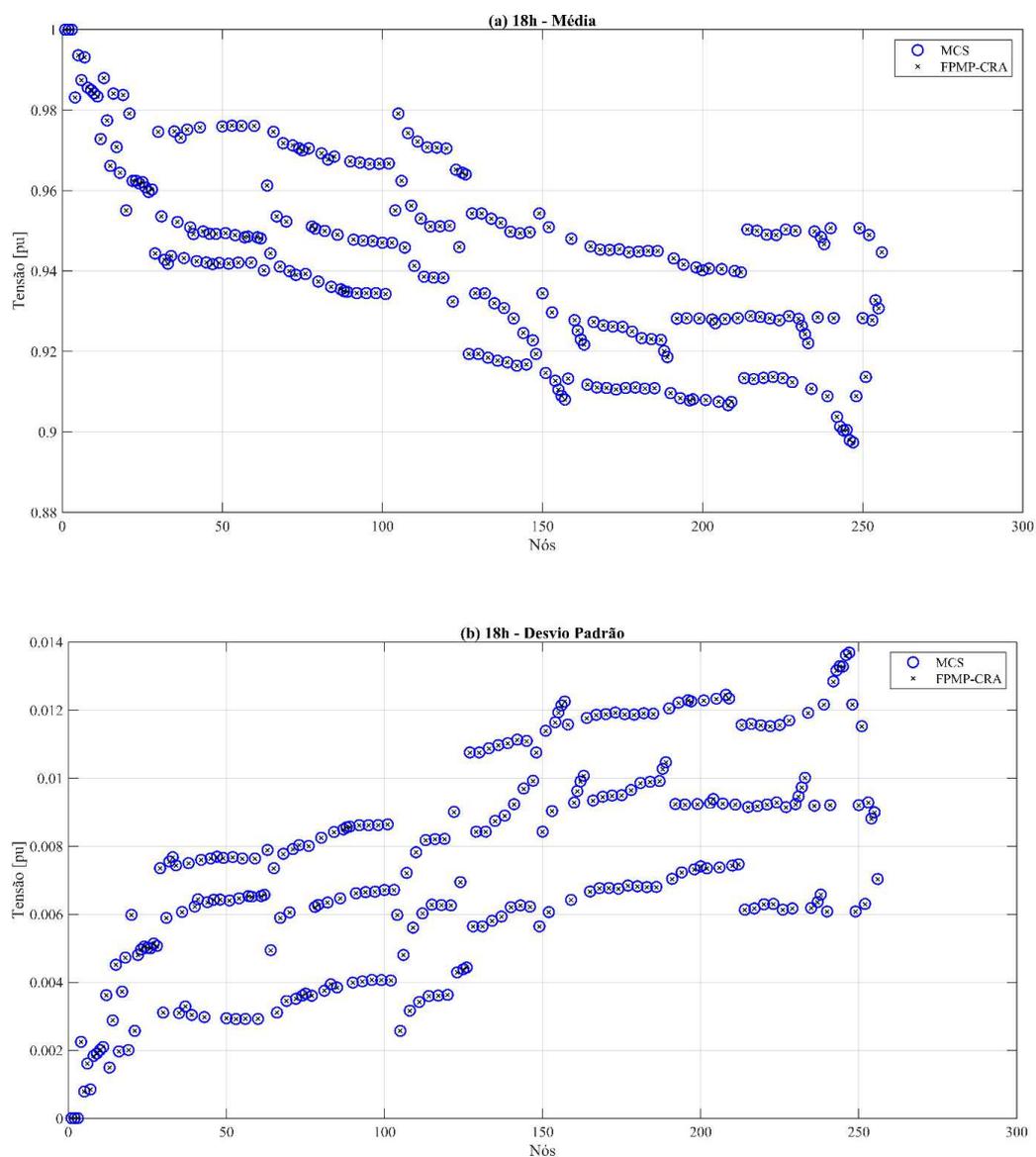
Fonte: elaborada pelo autor (2025).

Figura 36 – Momentos estatísticos das magnitudes de tensão: MCS x FPMP-CRA (12h).



Fonte: elaborada pelo autor (2025).

Figura 37 - Momentos estatísticos das magnitudes de tensão: MCS x FPMP-CRA (18h).



Fonte: elaborada pelo autor (2025).

Para validar matematicamente a enorme precisão do FPMP-CRA em relação ao MCS, são propostos os seguintes índices, denotados por (4.11) e (4.12), que levam em consideração o erro médio para todas as posições e horas do dia.

$$\epsilon_1(\%) = \frac{1}{N \cdot 24} \sum_{j=1}^{24} \sum_{i=1}^N \left( \frac{V_{ij,med}^{FPMP-CRA} - V_{ij,med}^{MCS}}{V_{ij,med}^{MCS}} \right) 100\% \quad (4.11)$$

$$\epsilon_2(\%) = \frac{1}{N \cdot 24} \sum_{j=1}^{24} \sum_{i=1}^N \left( \frac{V_{ij,std}^{FPMP-CRA} - V_{ij,std}^{MCS}}{V_{ij,std}^{MCS}} \right) 100\% \quad (4.12)$$

Onde:  $N$  é o número de nós da topologia, no caso, 256;

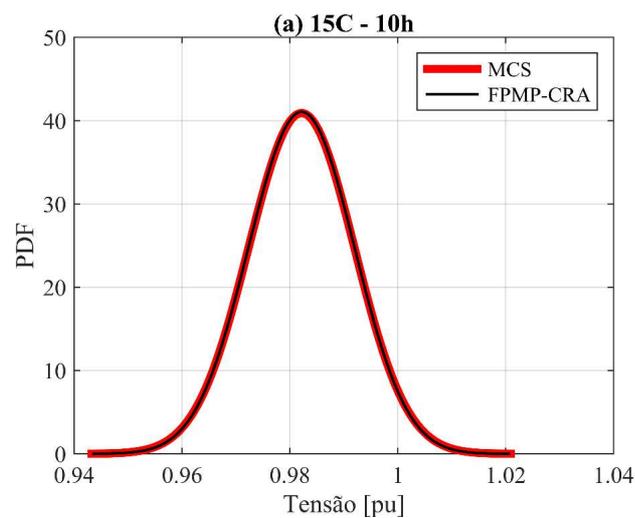
$V_{ij,med}^{FPMP-CRA}$  é a magnitude de tensão média na posição  $i$  e hora  $j$ , calculada pelo FPMP-CRA;

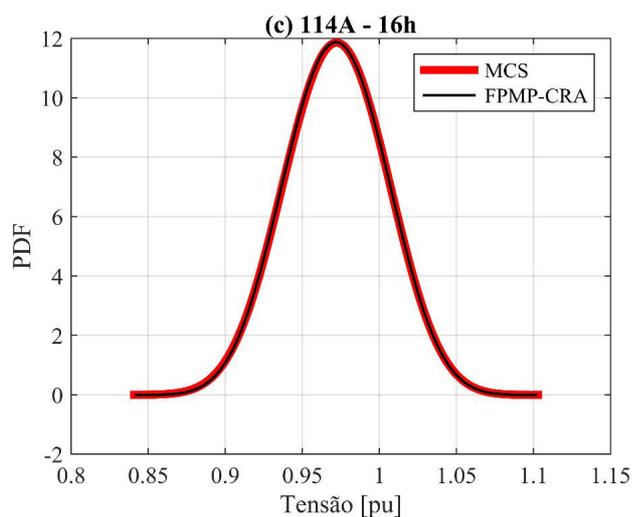
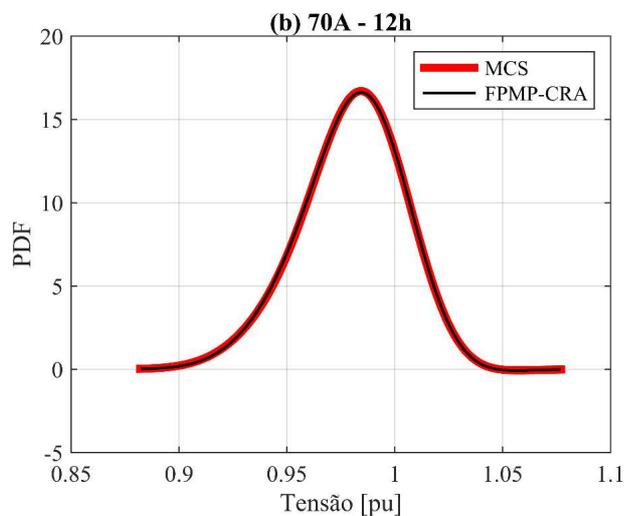
$V_{i,med}^{MCS}$  é a magnitude de tensão média na posição  $i$  e hora  $j$ , calculada pelo MCS de amostragem fixa;

$V_{i,std}^{FPMP-CRA}$  é o desvio padrão da magnitude de tensão na posição  $i$  e hora  $j$ , calculada pelo FPMP-CRA;

$V_{i,std}^{MCS}$  é o desvio padrão da magnitude de tensão na posição  $i$  e hora  $j$ , calculada pelo MCS de amostragem fixa.

Figura 38 – Comparação das PDFs de magnitude de tensão obtidas via MCS e FPMP-CRA.





Fonte: elaborada pelo autor (2025).

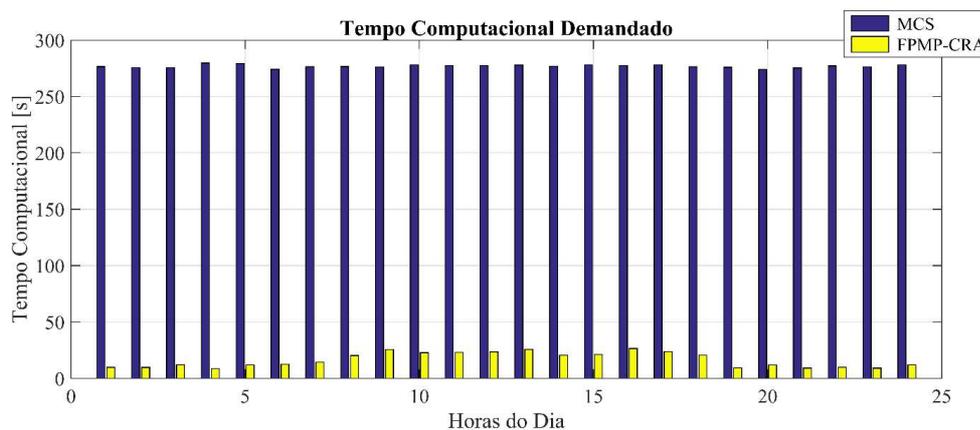
Para a atual simulação, os valores de  $\epsilon_1$  e  $\epsilon_2$  são, respectivamente, 0,0004% e 0,4065%. Tais valores são muito limitados, o que garante uma enorme precisão do método proposto, e, conseqüentemente, uma acurácia muito grande na reconstituição analítica das PDFs das variáveis aleatórias de saída. Para ilustrar isso, as PDFs de magnitude de tensão para as posições 15-C (às 10h), 70-A (às 12h) e 114-A (às 16h) são mostradas na Figura 38, estabelecendo uma comparação entre as obtidas via MCS (em vermelho) e as via FPMP-CRA (em preto).

Em conformidade com os resultados mostrados anteriormente, nota-se que a reconstituição analítica das PDFs das variáveis de saída a partir do FPMP-CRA apresentam

enorme precisão em relação ao MCS. Isso corrobora para a validação do método e sua aplicabilidade em situações de correlação temporal. Isso ocorre para todos os nós e todas as horas do dia, em consonância com os índices de precisão de média e desvio padrão extremamente limitados.

Ressalta-se que o método proposto é aproximadamente 17 vezes mais rápido em relação ao MCS. Uma comparação do tempo computacional para a simulação do FPP para cada hora do dia é mostrada na Figura 39, confrontando-se o MCS com o FPMP-CRA. No caso deste último, as horas do dia com maior irradiação solar tendem a demandar um tempo computacional maior, devido ao número ótimo de *clusters*, que é proporcionalmente maior. Já em relação ao MCS, foi considerada uma amostragem fixa para todas as horas do dia, o que mantém os tempos computacionais mais uniformes. Mesmo assim, a superioridade do método proposto é visível e nítida.

Figura 39 – Comparação do tempo computacional: MCS x FPMP-CRA.



Fonte: elaborada pelo autor (2025).

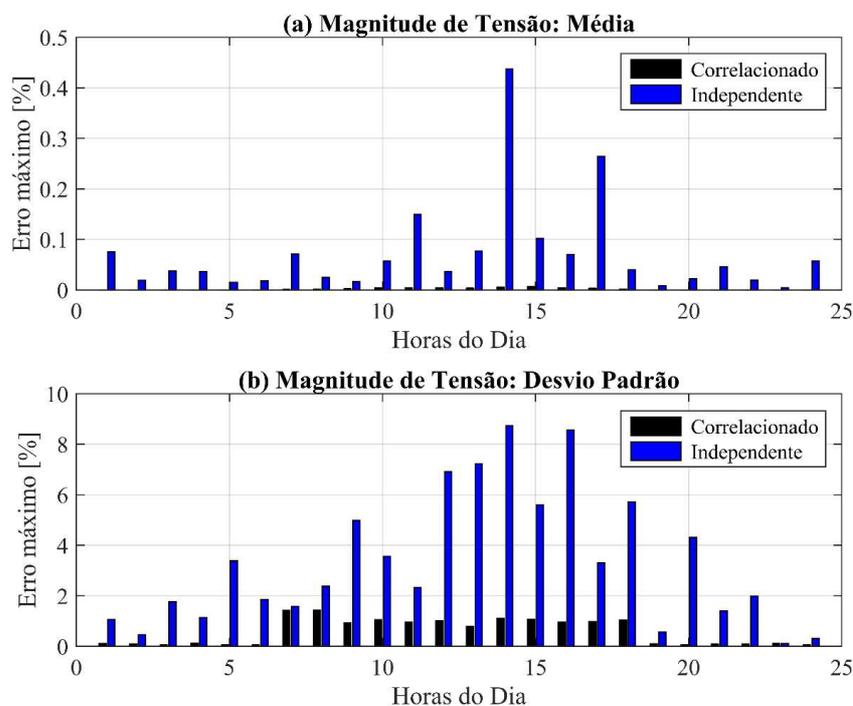
Finalmente, resta mostrar os impactos da adoção de um método de FPP que não considere as correlações temporais inerentes aos fenômenos modelados. Para isso, foram simulados duas vezes o FPMP-CRA: considerando-se as correlações temporais, tal como explicadas na Seção 4.5.2, e considerando todas as variáveis aleatórias do problema como independentes. Nesse último caso, a distribuição de probabilidade de uma variável aleatória em determinado instante de tempo não tem qualquer impacto na distribuição de probabilidade das variáveis aleatórias dos demais horários. Os dois casos são comparados com uma

referência, aqui adotada como o MCS que leva em conta todos os cenários possíveis gerados pela cópula estatística.

Na

Figura 40, são apresentados os erros máximos constatados em ambos os casos em relação à simulação de referência, considerando-se: (a) médias das magnitudes de tensão e (b) desvios padrões desse mesmo parâmetro. Em outras palavras, para cada hora do dia, é relatado o maior erro possível, dentre todas as posições do sistema.

Figura 40 – Erros máximos: FPP correlacionado x independente.



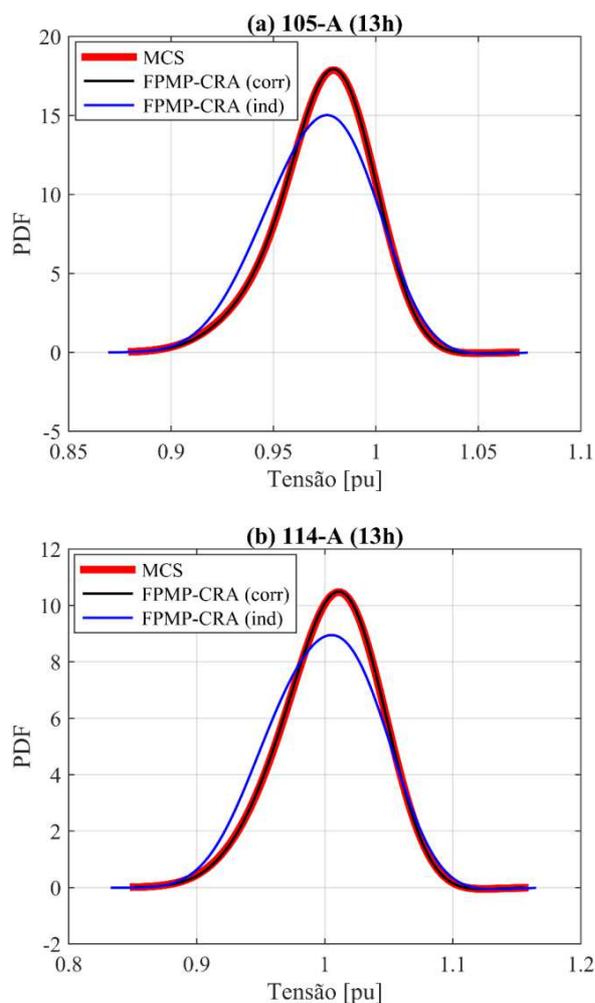
Fonte: elaborada pelo autor (2025).

Percebe-se, inicialmente, que as correlações não impactam significativamente as médias das tensões (isto é, os valores esperados), uma vez que os erros máximos do caso independente não ultrapassam 0,6%. De qualquer forma, verifica-se que os erros são ainda menores quando as correlações são levadas em consideração. Além disso, já foi visto no Capítulo 2 que a análise do sistema por meio da média dos parâmetros aleatórios de saída pode gerar conclusões equivocadas e incompletas.

Quando os desvios padrões são o alvo de análise, é possível verificar que os erros são muito maiores quando uma abordagem independente é empregada, sobretudo nos horários de maior penetração de potência fotovoltaica. Tais desvios podem chegar a quase 10%. Isso pode provocar alterações significativas nas PDFs recuperadas analiticamente.

Para ilustrar esse fenômeno, são mostradas, na Figura 41, as PDFs para a magnitude de tensão das posições 105-A e 114-A usando-se uma abordagem independente (em azul), correlacionada (em preto) e a simulação de referência (em vermelho). É possível constatar que as PDFs dos casos independente e correlacionado são bem diferentes, ainda que suas médias sejam próximas. Isso pode provocar falhas e erros de avaliação das probabilidades, sobretudo em se tratando de eventos raros (avaliados pelas ‘caudas’ das curvas).

Figura 41 – Comparações PDFs: FPP correlacionado x independente.



Fonte: elaborada pelo autor (2025).

Portanto, conclui-se que é fundamental o uso de métodos capazes de levar em conta as correlações, não só espaciais, mas temporais, para simular problemas de redes de distribuição ativas com penetração de GDs renováveis. O FPMP-CRA, por conseguinte, mostra-se robusto por estar apto a incluir essas condições de simulação e configura-se como uma importante ferramenta de simulação probabilística de redes de distribuição ativas.

#### 4.5.4 CONCLUSÕES PARCIAIS DO TESTE 4

Uma requisição importante para métodos de FPP, atualmente, é sua aplicabilidade em se tratando da consideração de variáveis aleatórias correlacionadas temporalmente. Isso é particularmente importante para simulação de redes de distribuição ativas, dotadas da presença de GD renovável. Fontes eólicas e fotovoltaicas são fortemente dependentes de condições atmosféricas que são estocásticas e correlacionadas no tempo. Portanto, visando resultados ainda mais fidedignos, devem ser feitas as corretas premissas para modelar essas variáveis, conforme discutido no Capítulo 2.

No teste 4, o objetivo foi simular o FPMP-CRA em situações que considerem variáveis aleatórias correlacionadas temporalmente, e mostrar que o método proposto é robusto o suficiente para lidar com essas premissas.

O sistema IEEE 123 foi alterado por meio da inserção de geradores fotovoltaicos, cuja geração de potência ativa depende da irradiação solar, que é uma variável aleatória correlacionada no tempo. Além disso, foi considerado que a demanda do sistema também é incerta e correlacionada temporalmente. Para obter as correlações e os parâmetros de cada função de probabilidade, um banco de dados de medições reais foi empregado.

A teoria de cópulas foi usada para gerar amostras de demandas e GDs fotovoltaicas correlacionadas no tempo. Posteriormente, o algoritmo K-Means foi usado para comprimir tais espaços de amostras. Foi visto que o número ótimo de *clusters* varia substancialmente com a hora do dia, a depender da intensidade da irradiação solar.

Através do FPMP-CRA, pode-se estimar as probabilidades de violação de tensão por sobretensão – devido à penetração de potência dos geradores fotovoltaicos – para todos os nós e horas do dia. Foi possível fazer um estudo de sensibilidade para identificar regiões do sistema mais críticas para a ocorrência desse distúrbio.

Foi verificado que o FPMP-CRA é 93% mais rápido do que uma abordagem convencional que simule todos os cenários gerados pela teoria de cópulas. Isso pode ser

vantajoso em estudos de mercado de energia elétrica e em problemas de fluxo de potência ótimo probabilístico. O tempo computacional demandado é consistente com as aplicações de curto prazo de planejamento e operação, não demorando mais do que 30 segundos para convergência de cada FPMP-CRA horário.

Por fim, foi constatado que o FPMP-CRA apresenta uma precisão muito significativa em relação ao *benchmark* adotado, o MCS, contendo erros nos momentos das variáveis aleatórias de saída menores do que 0,5%. O impacto da desconsideração das correlações temporais também foi alvo de estudo, em que se pode averiguar que erros de até 10% podem ocorrer nos desvios padrões das variáveis aleatórias de saída. Os valores esperados não são tão impactados, ainda que o FPMP-CRA correlacionado tenha tido um desempenho muito melhor em relação à sua abordagem independente. Mas, ressalta-se, mais uma vez, que a análise do sistema por meio dos valores esperados apenas pode ocasionar erros e falhas de avaliação, como visto nos testes anteriores e na revisão bibliográfica proposta no Capítulo 2. Nesse sentido, destaca-se a enorme acurácia do FPMP-CRA no que tange aos momentos estatísticos de ordem maior e sua conseqüente reconstituição analítica das PDFs.

Conclui-se, por conseguinte, que o FPMP-CRA é robusto o suficiente para lidar com situações nas quais as variáveis aleatórias são temporalmente correlacionadas e, além disso, se configura como uma importante ferramenta para ser aplicada em diversos âmbitos devido à sua eficiência computacional comprovada.

#### 4.6 TESTE 5: ESTUDOS PROBABILÍSTICOS DE SEGURANÇA DE TENSÃO DE NEUTRO

Conforme visto nos testes anteriores, o perfil de tensão dos SDs são cada vez mais incertos, devido à crescente imprevisibilidade de geração e carga diante das transformações tecnológicas que ocorreram nos últimos anos (Ramadhani *et al.*, 2020). Isso motivou pesquisadores a desenvolverem métodos cada vez mais robustos para averiguar a segurança de tensão, buscando ações relevantes para o controle desse parâmetro. No caso de redes de distribuição multiterradas, há uma preocupação séria acerca da segurança das tensões de neutro-terra (ou tensões *NEV*, do inglês: *neutral-to-earth voltages*, ou simplesmente “tensões de neutro”, como serão denominadas de agora em diante), cujos potenciais elétricos podem surgir devido ao desequilíbrio do carregamento do sistema. Tais fenômenos podem causar

acidentes graves, caso haja o contato não acidental de seres humanos com massas energizadas (Tsai-Hsiang Chen; Wen-Chih Yang, 2001).

Para obter as avaliações corretas das tensões de neutro e, conseqüentemente, estimar as probabilidades de ocorrência de cenários graves, é necessário empregar uma ferramenta probabilística capaz de conter a modelagem de sistemas multifásicos que contemplem as representações explícitas dos condutores de neutro e os próprios aterramentos. Caso contrário, a negligência de uma modelagem apropriada pode levar a erros de avaliação das tensões de neutro e, conseqüentemente, acidentes humanos ou danos em equipamentos podem ocorrer (Penido *et al.*, 2013). Diante disso, métodos capazes de estimar com precisão as PDFs das tensões de neutro tornam-se imprescindíveis para o planejamento de redes de distribuição multiterradas, com sua modelagem intrínseca sendo contemplada.

A previsão do comportamento das tensões de neutro via FPP ainda não foi publicada na literatura, tampouco o estudo de segurança de tensão de neutro utilizando-se essas bases. O objetivo do presente teste é mostrar a aplicabilidade robusta do FPMP-CRA para simular redes de distribuição desequilibradas, multifásicas, com a representação explícita dos condutores de neutros e aterramentos, de maneira que ele pode ser empregado para subsidiar estudos de segurança de tensão – mais especificamente enfocando-se nas tensões de neutro.

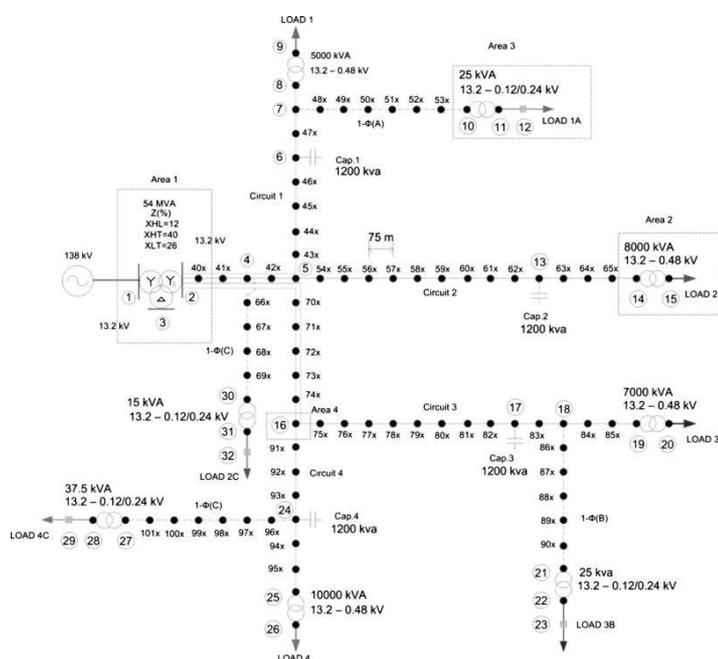
O método para solucionar o FPMP-CRA segue exatamente os mesmos passos vistos no Capítulo 3. A solução individual do FPD para cada agente de *cluster*, via MICN, engloba a representação explícita do condutor neutro e das impedâncias de aterramento. Tal método é ideal para simular redes de distribuição multifásicas e multiterradas (Penido *et al.*, 2013).

As variáveis aleatórias são as resistências de aterramento. São modeladas como cargas do tipo impedância constante e são representadas por variáveis Gaussianas. As resistências dependem de fatores que apresentam incertezas, como a resistividade, tipo e umidade do solo, além da geometria, desgaste e corrosão dos eletrodos. Logo, não há uma previsibilidade determinística acerca do valor das resistências de aterramento, de maneira que a consideração de incertezas pode levar a uma formulação mais robusta e completa do estado probabilístico das tensões de neutro. Além disso, as resistências são assumidas como variáveis aleatórias independentes, pois tais condições incertas podem variar significativamente de um ponto do sistema para outro. Ressalta-se que todas as outras variáveis são mantidas como determinísticas, uma vez que o intuito do presente teste é verificar as PDFs das tensões de neutro, dadas as incertezas das resistências de aterramento.

#### 4.6.1 APRESENTAÇÃO DA TOPOLOGIA IEEE NEV E CONDIÇÕES DE SIMULAÇÃO

O alimentador-teste IEEE NEV, apresentado em IEEE (2024), foi utilizado para demonstrar a eficiência do FPMP-CRA e sua aplicabilidade para lidar com sistemas complexos. Deseja-se mostrar sua viabilidade para estimar as PDFs das magnitudes de tensão de neutro.

Figura 42 – Topologia IEEE NEV.



Fonte: Penido *et al.* (2013).

Segundo Penido *et al.* (2013), este sistema apresenta tanto desafios de ordem computacional quanto de modelagem. A maioria dos métodos não conseguem simular tal sistema, devido à sua enorme complexidade. Para isso, é requerido o desenvolvimento de métodos multifásicos, para representar adequadamente os condutores de neutro, os aterramentos e os cabos de telecomunicações. O sistema é formado por diversos circuitos, com impedâncias mútuas entre condutores de fase, neutro e cabos de telecomunicações. Além disso, apresenta aterramentos e diversos tipos de cargas e transformadores. De acordo com os mesmos autores, a ausência de um rigor maior para a modelagem correta desses equipamentos ocasiona o cálculo de tensões de neutro equivocadas, contendo sérios erros de avaliação. Esses erros podem levar a danos em equipamentos, configurações inexatas de dispositivos de

proteção e até acidentes humanos severos. Eis, portanto, a motivação para mostrar a aplicabilidade da simulação do FPMP-CRA em sistemas tão complexos quanto o apresentado. Sua rede pode ser vista na Figura 42. Mais detalhes acerca da topologia podem ser verificados em IEEE (2024).

Em consonância com a codificação apresentada em Penido *et al.* (2013), adota-se a convenção para a identificação dos nós do sistema através de um código de cinco dígitos expresso por **xxx.c.f**, tal que:

- **xxx** identifica a barra;
- **c** identifica o circuito, que pode variar de 0 a 4. Os dígitos de 1 a 4 identifica o circuito. Já o algarismo 0 está relacionado aos condutores que são comuns a mais do que um circuito, como os condutores de neutro e os cabos de telecomunicações;
- **f** identifica a fase, que pode ser A, B ou C (para os condutores de fase), N (para o condutor de neutro) ou M (para os cabos de telecomunicações).

As variáveis aleatórias para esse teste são as resistências de aterramento. São modeladas por cargas do tipo impedância constante e suas incertezas advêm de amostras de funções de probabilidade Gaussianas independentes. Seus parâmetros são:

- Médias iguais aos valores originalmente informados na topologia.
- Variância igual a 20% da média, o que garante uma boa variabilidade das amostras e permite a simulação de cenários qualitativa e quantitativamente distintos em termos de aterramentos.

No total, há 90 aterramentos e, por conseguinte, 90 variáveis aleatórias. Todos os demais parâmetros do FPP são mantidos determinísticos. O objetivo deste teste é estimar as PDFs das magnitudes de tensão de neutro e efetuar uma análise mais profunda e completa do estado probabilístico dessa rede complexa, que possui tantas características peculiares.

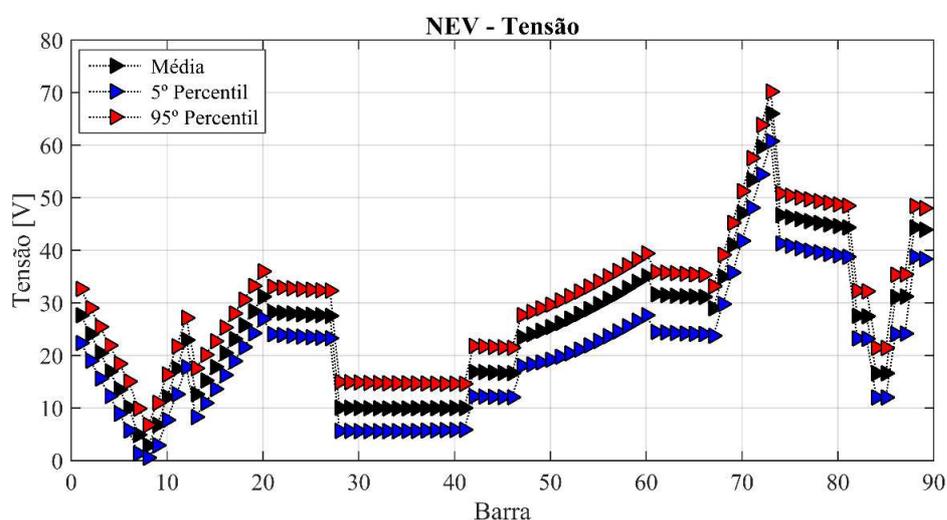
#### 4.6.2 RESULTADOS DO FPMP-CRA CONSIDERANDO REDES MULTIATERRADAS

As amostras são coletadas através da técnica LHS. É adotado um número total de 100 amostras por variável aleatória. É adotada a técnica K-Means para efetuar o *clustering*. O número ótimo de *clusters* é igual a 30. As variáveis aleatórias de saída são recuperadas através do cálculo de seus momentos estatísticos, tendo como base as respostas individuais de cada problema determinístico, para cada ponto operativo representado por seu respectivo centro de *cluster*, e posteriormente, utilizando-se as recuperações analíticas via EGC, tal como

explicado no Capítulo 3. Neste teste, as variáveis aleatórias de saída de interesse referem-se às magnitudes das tensões de neutro.

Na Figura 43, há o perfil de tensão das tensões de neutro, considerando as suas médias (em preto), os seus 5<sup>os</sup> percentis (em azul) e os 95<sup>os</sup> percentis (em vermelho). Em consonância com os estudos anteriores, a análise da média e desses percentis visa demonstrar, respectivamente, a tendência central de cada variável (ou valor esperado), e um intervalo de confiança para essas tensões, o que pode favorecer a previsibilidade de estratégias de planejamento. Tais parâmetros são obtidos diretamente das PDFs recuperadas via EGC.

Figura 43 – Comportamento probabilístico das magnitudes de tensões de neutro.



Fonte: elaborada pelo autor (2025).

Algumas conclusões podem ser inferidas, com base na análise da Figura 43. Primeiramente, é possível notar que as tensões de neutro podem chegar a aproximadamente 70 V em algumas posições do sistema, o que indica um alto risco para acidentes humanos fatais ou danos em equipamentos. Devem ser tomadas atitudes de planejamento e operação para evitar valores tão elevados, priorizando sempre a segurança humana.

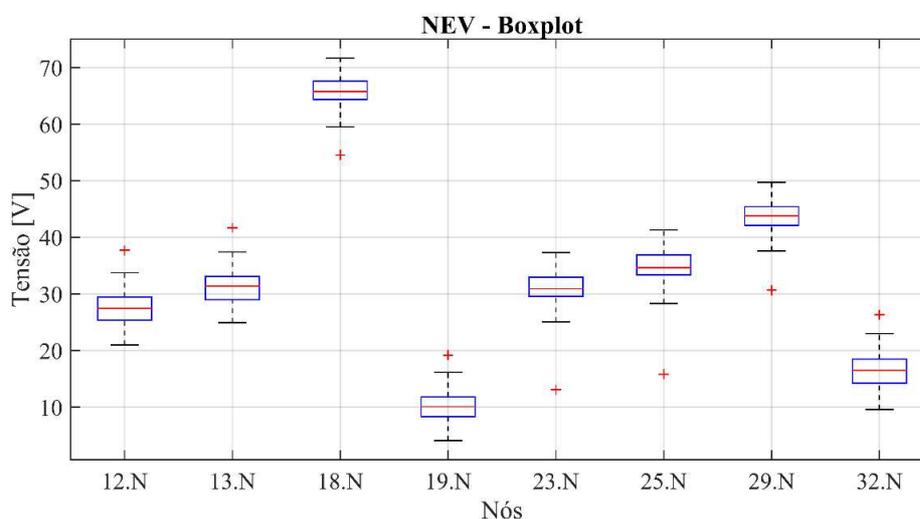
Em segundo lugar, há uma diferença considerável entre os valores de 5<sup>o</sup> e 95<sup>o</sup> percentil para as tensões de neutro. Essa diferença pode variar de 10 a 15 V, o que é algo significativo no que tange à segurança de tensão. Em outras palavras, a tensão de neutro pode oscilar notavelmente, dependendo das condições combinatórias entre todas as resistências de aterramento do sistema. Essa avaliação somente é possível tendo em posse uma ferramenta capaz de representar cada um desses aterramentos como uma variável aleatória independente.

Em terceiro lugar, nota-se que, na maioria das posições, as médias não estão equidistantes dos dois percentis analisados. Isso indica que as PDFs das tensões de neutro não são simétricas. O correto cálculo dos coeficientes de assimetria e de curtose para essas variáveis aleatórias permite a correta avaliação das probabilidades de ocorrências de eventos raros. Isso só é possível de se constatar porque o método proposto pressupõe a recuperação analítica das PDFs das variáveis de saída, o que enaltece sua robustez e aplicabilidade.

Por fim, ressalta-se a diferença substancial que existe entre os valores médios, que poderiam ser obtidos por meio de ferramentas determinísticas (como as contidas nos métodos desenvolvidos na literatura) e aqueles observados por meio de abordagens probabilísticas via FPMP-CRA. A análise probabilística é muito mais completa e abrangente, o que permite ações mais seguras e eficazes de planejamento e operação. O erro percentual pode chegar a aproximadamente 54%, o que pode induzir os engenheiros a ações equivocadas.

Na Figura 44, há uma avaliação probabilística por meio de *boxplot*<sup>20</sup>, enfocando-se nas posições nas quais há cargas desequilibradas conectadas. Os nós estão em consonância com os mostrados na Figura 42.

Figura 44 – Tensões de neutro: *boxplots*.



Fonte: elaborada pelo autor (2025).

<sup>20</sup> O gráfico *boxplot* é em forma de “caixa”. As extremidades da caixa azul representam o primeiro e o terceiro quartil (isto é, 25º e 75º percentis). O comprimento da caixa é, portanto, o “intervalo interquartil” (IIQ). A linha vermelha refere-se à mediana. Fora da caixa, os traços pretos são indicativos dos valores máximos e mínimos, calculados por  $Q3 \pm 1,5 \cdot \text{IIQ}$ . Por fim, as cruces em vermelho são os *outliers*. A análise desse gráfico, portanto, traz muitas informações estatísticas relevantes de forma visual e simples.

Além de demonstrar o intervalo de confiança, os gráficos em *boxplot* permitem a visualização dos *outliers*. Dessa forma, é possível rastrear eventos raros de probabilidade baixa, mas que devem ser levados em consideração para o planejamento das redes de distribuição multiaterradas. Na posição 19.N, a tensão de neutro mínima avaliada é próxima do valor nulo, o que seria a situação operativa desejada. No entanto, nessa mesma posição, há mais de 25% de chance de a tensão ser maior do que 15 V, o que já colocaria em risco a saúde humana e a vida útil de equipamentos. Além disso, existem probabilidades remotas de essa tensão chegar em valores próximos a 20 V, o que é ainda mais drástico. Uma análise sem rigor, que desconsidere ferramentas estatísticas, pode levar a um entendimento errado acerca das condições fidedignas de operação para essa rede, o que poderia levar à negligência dos fenômenos descritos.

Sublinha-se, também, que a diferença do valor médio para as tensões de eventos raros, cobertos pela representação de *outliers*, pode ultrapassar 15 V, como nas posições 23.N e 25.N. Novamente, ressalta-se que a análise das condições operativas do sistema tendo como referência apenas as médias (os valores esperados ou determinísticos) pode ser insuficiente para a real compreensão, entendimento e espírito crítico da operação da rede.

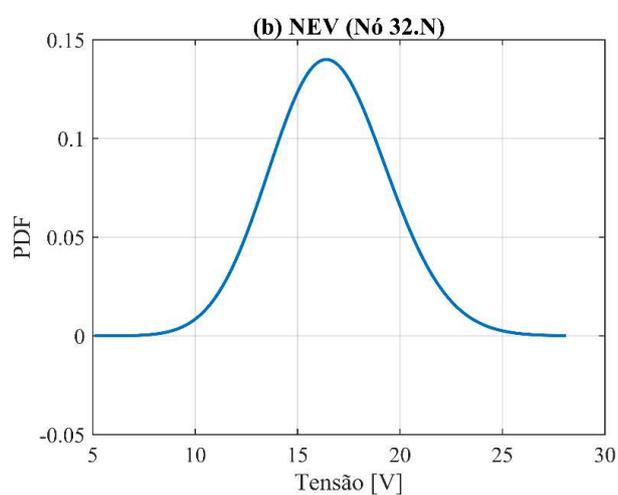
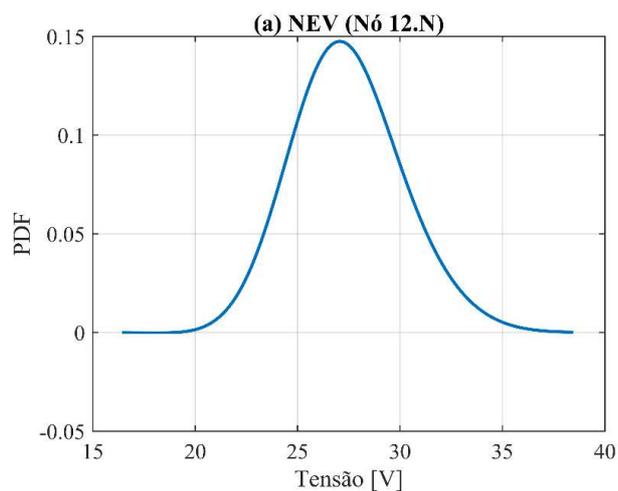
Recorda-se que a identificação de *outliers* operativos, obtidos pela consideração combinatória das diferentes resistências de aterramento do sistema, só pode ser feita mediante a otimização do número de *clusters*. Um número muito reduzido não será capaz de englobar estatisticamente o espaço de amostras original em sua plenitude. Logo, essa etapa do FPMP-CRA é crucial para o correto mapeamento de eventos raros. Conforme visto no Capítulo 3, isso é garantido pela minimização das distâncias *intra-cluster*. Se houver um *outlier*, tal distância permanecerá alta até que ele seja contemplado por um *cluster* próximo. Isso permite a correta identificação dos principais pontos operativos.

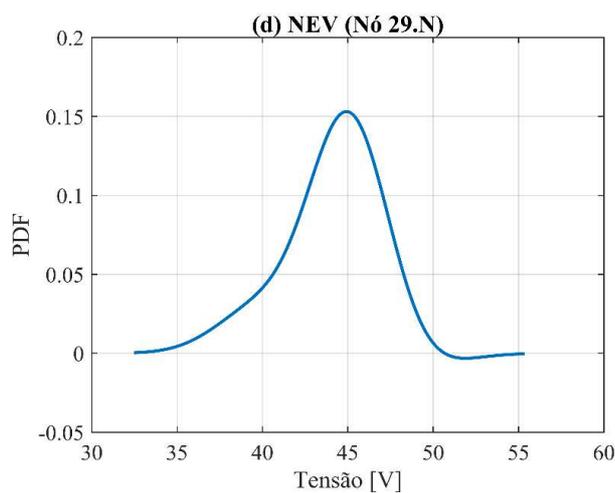
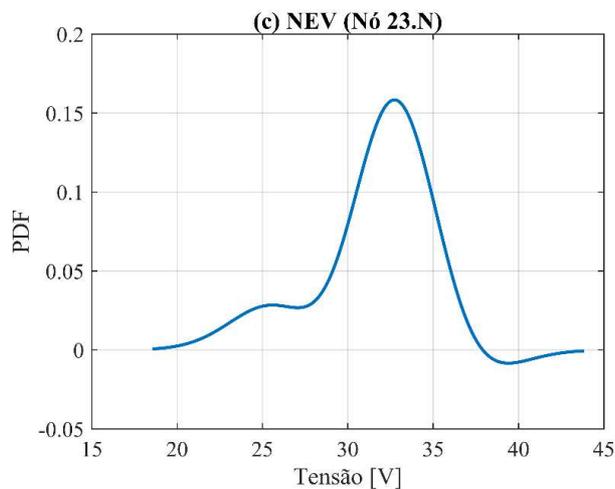
As PDFs recuperadas analiticamente para as tensões de neutro das posições 12.N, 32.N, 23.N e 29.N podem ser vistas na Figura 45.

Em consonância com os resultados demonstrados anteriormente, as PDFs não necessariamente são simétricas, tampouco unimodais. Isso gera uma complexidade significativa para a avaliação de redes multiaterradas, de maneira que a modelagem multifásica, a otimização do número de *clusters* e, sobretudo, a recuperação analítica das PDFs tornam-se fatores cruciais para a avaliação correta do estado probabilístico operativo. Diante disso, fica enaltecida a enorme robustez do FPMP-CRA em se tratando de sua aplicabilidade para subsidiar estudos de segurança de tensão de neutro. Ressalta-se a

aplicabilidade do presente método em estimar a probabilidade de a tensão de neutro exceder algum valor considerado seguro segundo normas internacionais, o que pode vir a ser uma importante ferramenta para estudos de planejamento de redes multiterradas.

Figura 45 – PDFs das tensões de neutro, recuperadas analiticamente usando EGC.





Fonte: elaborada pelo autor (2025).

#### 4.6.3 CONCLUSÕES PARCIAIS DO TESTE 5

O estudo de redes de distribuição multiterradas requer uma modelagem mais completa e abrangente dos elementos elétricos para que seus resultados sejam fidedignos e relevantes. Além da inerente natureza multifásica, é imprescindível a representação explícita dos condutores de neutro e dos aterramentos. O estudo de segurança de tensão de neutro necessita de tais pressupostos e a implementação de uma abordagem que empregue métodos probabilísticos para esse fim ainda não foi conduzida na literatura.

Nesse sentido, o teste 5 direciona-se ao uso do FPMP-CRA para avaliar as PDFs das magnitudes de tensão de neutro em sistemas multiterrados, em que as resistências de aterramento são variáveis aleatórias, modeladas por funções Gaussianas independentes. O sistema-teste empregado para a avaliação foi o IEEE NEV, que possui características extremamente singulares como: presença de transformador de três enrolamentos; modelagem explícita dos condutores de neutro e aterramentos; representação de cabos de telecomunicações; cargas trifásicas desequilibradas; cargas monofásicas; transformadores de tape central. Logo, trata-se de uma topologia desafiadora em termos de modelagem de equipamentos e sob o aspecto computacional.

Foi possível constatar que existem tensões de neutro que podem alcançar valores próximos de 70 V em algumas posições, o que coloca em risco a saúde humana e vida útil de equipamentos. Além disso, a diferença entre o 5º e 95º percentis é significativa, o que mostra que a tensão de neutro pode variar substancialmente apenas com a permutação combinatória das resistências de aterramento ao longo do sistema. Essa diferença pode exceder 15 V, o que não pode ser negligenciado. Ademais, a diferença entre valores médios e os valores limites de cenários simulados (incluindo eventos raros representados por *outliers* estatísticos, modelados por centros de *clusters* longínquos) pode ser maior do que 20 V. Infere-se, portanto, que a abordagem determinística pode levar a erros e equívocos de planejamento e operação de redes multiterradas.

Por fim, destaca-se que as PDFs das tensões de neutro não são necessariamente simétricas ou unimodais. Isso enaltece a robustez do FPMP-CRA, especialmente no que tange às recuperações analíticas dessas funções.

Logo, conclui-se que o FPMP-CRA pode ser uma importante ferramenta para conduzir estudos probabilísticos de segurança de tensão de neutro, uma vez que possui aplicabilidade para modelar corretamente sistemas multifásicos e multiterrados. Além disso, permite a correta representação dos aterramentos, modelando-os como variáveis aleatórias, o que viabiliza resultados amplos, completos e mais relevantes em comparação com uma abordagem determinística tradicional. Sublinha-se a fundamental importância da sua viabilidade no que se refere à: (i) modelagem multifásica, imprescindível para obter resultados fidedignos; (ii) otimização do número de *clusters*, essencial para obter *outliers* operativos e modelagem de eventos raros; (iii) recuperação analítica das PDFs, importante para englobar tanto os aspectos assimétricos quando multimodais inerentes.

#### 4.7 TESTE 6: APLICAÇÕES EM SISTEMAS DE GRANDE PORTE

Os estudos de escalabilidade são cada vez mais requisitados para averiguar a robustez dos métodos desenvolvidos. Além disso, permitem sua aplicação em situações que se aproximam mais da realidade, conferindo-lhe aplicabilidade prática. Conforme visto no Capítulo 2, métodos de FPP que sejam rápidos, precisos, flexíveis e escaláveis estão ainda em desenvolvimento. Logo, o objetivo do presente teste é verificar a performance do FPMP-CRA em aplicações de grande porte.

No caso do FPP, a escalabilidade pode ser estudada sob dois aspectos: mediante a complexidade do sistema (sistemas de grande porte) e mediante o número de variáveis aleatórias. O primeiro torna-se importante à medida que certos problemas de convergência computacional somente surgem quando sistemas muito complexos são simulados. Por isso, a robustez do método é o alvo de teste. Já o segundo possui uma particular importância para métodos de FPP. Conforme indicado no Capítulo 2, os métodos probabilísticos tendem a piorar seu desempenho quando um número muito alto de variáveis aleatórias é introduzido. No caso dos métodos numéricos, um número maior de iterações será necessário para alcançar resultados precisos. É o que também geralmente ocorre nos métodos por aproximações. Muitas vezes, as vantagens presentes nesses últimos são minimizadas devido ao grande número requerido de amostras. É o caso do MEP, do LHS e do MCS, por exemplo. Métodos analíticos, muitas vezes, são intratáveis em situações de escalabilidade. Na revisão bibliográfica proposta no Capítulo 2, é notável a ausência da simulação de casos em que há um número alto de variáveis aleatórias. Frequentemente, são adotadas simplificações para tratar esses aspectos, como a redução do número de variáveis ou da dimensão do sistema.

Nesse sentido, é de suma importância desenvolver métodos que sejam escaláveis, isto é, que não apresentem uma significativa perda de precisão ou de eficiência computacional quando as duas situações descritas são levadas em consideração.

##### 4.7.1 APRESENTAÇÃO DA TOPOLOGIA IEEE 8500 E CONDIÇÕES DE SIMULAÇÃO

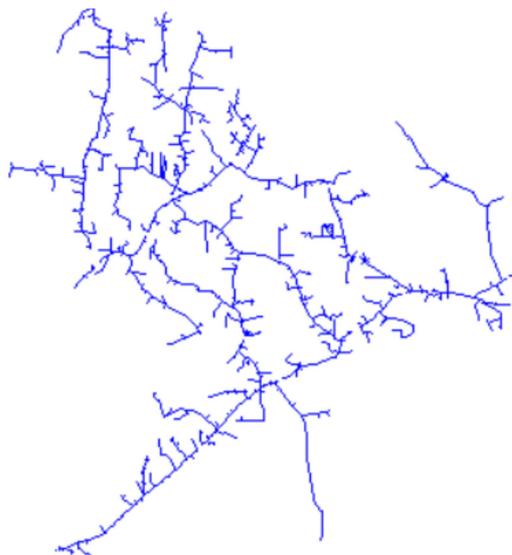
A topologia IEEE 8500 (Arritt; Dugan, 2010) é um sistema-teste de distribuição de altíssimo porte, usado para verificar a escalabilidade de métodos direcionados a SD. Possui 8531 nós em níveis de média e baixa tensão. Apresenta diversos aspectos que o aproxima de uma topologia real de distribuição, como, por exemplo, nós monofásicos, bifásicos e

trifásicos, cargas monofásicas desequilibradas conectadas nos secundários de transformadores de tape central. A solução do FP para essa topologia tem caráter desafiador, e pesquisadores podem verificar a qualidade de seus métodos quanto a sua convergência considerando tal sistema. Sua rede unifilar pode ser constatada na Figura 46, e mais informações sobre a topologia podem ser acessadas em Arrit e Dugan (2010).

Diferentemente de todos os outros testes conduzidos na presente tese, a simulação do sistema IEEE 8500 foi realizada utilizando-se o *software* aberto de simulação de SD chamado *OpenDSS* (do inglês: *Open Distribution System Simulator*), devido à complexidade do sistema e a presença sua opção de simulação já incluída nas bibliotecas internas do *software*. A integração do *OpenDSS* com o código do FPMP-CRA no Matlab® é possível através da interface COM.

Nesse estudo, todas as 2354 cargas foram consideradas aleatórias, sendo modeladas por distribuições Gaussianas, cujas médias e desvios padrões são os valores originais informados na topologia e 10% dos valores médios, respectivamente. Os fatores de potência foram mantidos constantes de acordo com os valores originais e as cargas foram consideradas independentes, para dificultar o processo de *clustering*. Capacitores e reguladores foram desconsiderados. O estudo dos seus impactos probabilísticos será conduzidos em estudos futuros.

Figura 46 – Topologia IEEE 8500.



Fonte: elaborada pelo autor (2025), gerada no *OpenDSS*.

No que tange às configurações do FPMP-CRA, as diretrizes de redução de dimensionalidade foram assumidas, com exceção da diretriz número 3 (que estabelece a redução do número de variáveis aleatórias), pois é alvo do presente estudo averiguar a escalabilidade do número de variáveis aleatórias de entrada. Dessa forma, a amostragem foi conduzida utilizando-se o método LHS (utilizando-se 100 amostras) e o algoritmo de *clustering* adotado foi K-Medoids CLARA.

O presente teste foi dividido em dois estudos. O primeiro consiste na determinação do número ótimo de *clusters* por vias exaustivas. Em outras palavras, simula-se o mesmo problema de FPMP-CRA considerando-se diferentes números de *clusters* até que haja uma estagnação dos resultados estatísticos das variáveis de saída. A partir disso, é possível inferir se o método proposto de estimação do número de *clusters* é apropriado também para aplicações de grande porte e, sobretudo, aplicações de escalabilidade do número de variáveis aleatórias de entrada.

Já o segundo estudo tem por objetivo comparar o desempenho do FPMP-CRA com outros métodos presentes na literatura. A saber: MEP 2m, LHS, MCS, e FPP via K-Means<sup>21</sup>. Tais comparações são efetuadas em três frentes:

1. Comparação da precisão do método: o FPP considerando o IEEE 8500 sob as condições previamente explicadas é simulado considerando o MEP 2m, LHS e FPP via K-Means. Por fim, os erros percentuais das variáveis de saída de cada um deles em relação ao MCS (que é empregado como método de referência) são avaliados.
2. Comparação das funções densidade de probabilidade das variáveis aleatórias de saída: são traçadas as PDFs de variáveis aleatórias de interesse considerando-se cada um dos métodos citados e compara-se a sua precisão em relação ao histograma das observações obtidas através do MCS.
3. Comparação do tempo computacional: é feita uma análise comparativa de todos os métodos citados considerando-se o tempo computacional requerido para a convergência do FPP.

Logo, tendo em vista estas três análises, é possível averiguar criticamente o desempenho do FPMP-CRA sob a perspectiva de simulações de escalabilidade, principalmente no que se refere à escalabilidade do número de variáveis aleatórias, fator crítico para aplicações em sistemas reais.

---

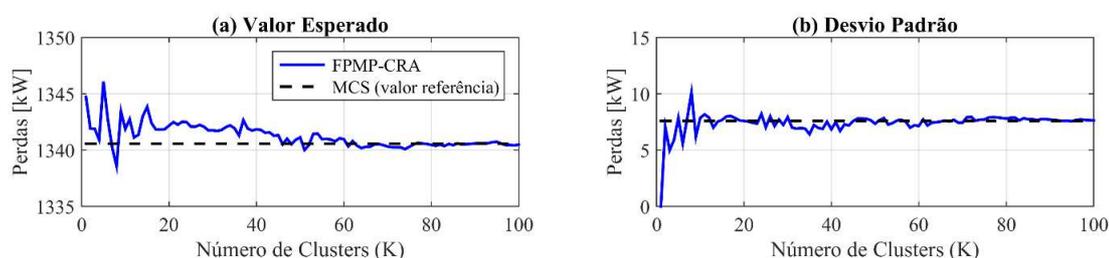
<sup>21</sup> Os desenvolvimentos matemáticos e procedimentos de cada um desses métodos podem ser vistos em detalhes no Apêndice B.

#### 4.7.2 ESTUDO 1: ESTIMATIVA DO NÚMERO ÓTIMO DE *CLUSTERS*

No primeiro estudo realizado, o objetivo é averiguar o número ótimo de *clusters* para aplicações de grande porte e dado um elevado número de variáveis aleatórias. Utilizando-se o algoritmo proposto para otimizar tal parâmetro, o valor encontrado foi 50 *clusters*. Para validar que esse número é adequado para as aplicações de grande porte, o FPMP-CRA foi simulado considerando diferentes números de *clusters* e os resultados foram comparados com a simulação de um MCS de amostragem fixa (10000 amostras aleatórias). Mais especificamente, o FPMP-CRA foi simulado considerando 1, 2, 3, ... até 100 *clusters*, que é o número total de amostras do LHS. A variável aleatória de saída armazenada foram as perdas ativas totais. Conforme já mencionado nos outros estudos, trata-se de um bom parâmetro para efetuar comparações já que depende de todos os valores de magnitudes e ângulos de tensões elétricas, acumulando os erros embutidos em todas as essas variáveis. Em outras palavras, uma alta precisão no valor de perdas indica que há uma tendência na precisão das demais variáveis, pois tal parâmetro depende de todos os outros. A precisão do FPMP-CRA considerando-se diversas outras variáveis aleatórias de saída será alvo de estudo da Seção 4.7.3.

Os resultados encontrados estão dispostos na Figura 47. Na Figura 47 (a), há a variação do valor esperado de perdas ativas totais à medida que o número de *clusters* aumenta. Já na Figura 47 (b), há a variação do desvio padrão. Ambos estão traçados em azul, que indica os valores do FPMP-CRA. Em preto (tracejado), o valor obtido para ambos os parâmetros quando o MCS de amostragem fixa é adotado.

Figura 47 – Estagnação dos resultados estatísticos com a variação do número de *clusters* para IEEE 8500.



Fonte: elaborada pelo autor (2025).

Através da análise da Figura 47, é possível perceber que há uma tendência de estagnação tanto do valor esperado quanto do desvio padrão das perdas elétricas ativas totais a partir do número de *clusters* igual a 50. Além disso, essa estagnação ocorre justamente em direção ao valor encontrado no MCS, o que torna válida a estimativa feita por meio do método proposto. Note que valores menores de *clusters*, como 30 ou 40, por exemplo, acarretam erros pequenos, porém visíveis. No entanto, valores maiores do que 50 não impactam significativamente na qualidade dos resultados estatísticos. Isso corrobora para a conclusão de que o valor de 50 *clusters* é, de fato, a quantidade ideal para esse sistema.

Uma análise similar pode ser conduzida considerando-se as magnitudes de tensões elétricas do sistema. É possível calcular o erro percentual médio levando em conta as 8531 posições, tanto quanto ao valor esperado quanto em relação ao desvio padrão. É o que está indicado pelas equações (4.13) e (4.14).

$$\epsilon_1(\%) = \frac{1}{N} \sum_{i=1}^N \left( \frac{V_{i,med}^{FPMP-CRA} - V_{i,med}^{MCS}}{V_{i,med}^{MCS}} \right) 100\% \quad (4.13)$$

$$\epsilon_2(\%) = \frac{1}{N} \sum_{i=1}^N \left( \frac{V_{i,std}^{FPMP-CRA} - V_{i,std}^{MCS}}{V_{i,std}^{MCS}} \right) 100\% \quad (4.14)$$

Onde: N é o número de nós da topologia, no caso, 8531;

$V_{i,med}^{FPMP-CRA}$  é a magnitude de tensão média na posição *i*, calculada pelo FPMP-CRA;

$V_{i,med}^{MCS}$  é a magnitude de tensão média na posição *i*, calculada pelo MCS de amostragem fixa;

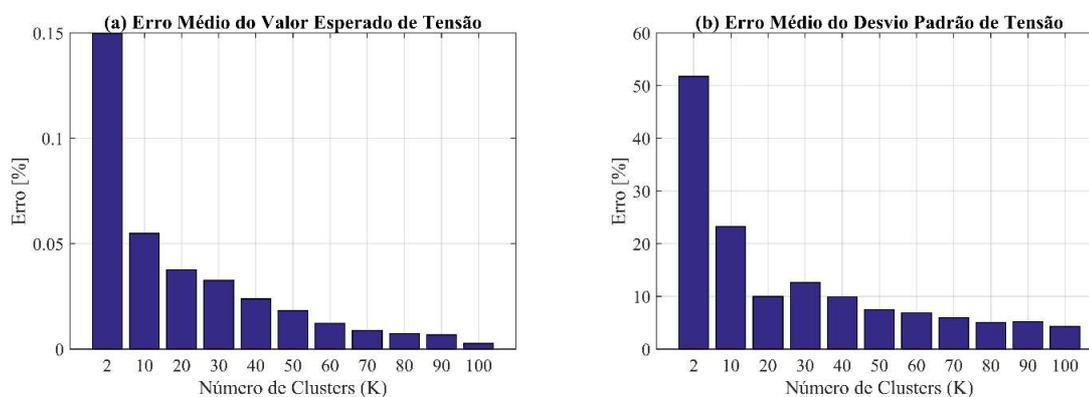
$V_{i,std}^{FPMP-CRA}$  é o desvio padrão da magnitude de tensão na posição *i*, calculada pelo FPMP-CRA;

$V_{i,std}^{MCS}$  é o desvio padrão da magnitude de tensão na posição *i*, calculada pelo MCS de amostragem fixa.

Na Figura 48, são mostrados o erro médio do valor esperado de tensão, denotado por  $\epsilon_1$ , e o erro médio do desvio padrão de tensão, denotado por  $\epsilon_2$  para o FPMP-CRA, considerando diferentes números de *clusters*.

É possível perceber que quando um número muito baixo de *clusters* é adotado, como  $K = 2$ , o erro é muito alto, sobretudo o desvio padrão, que chega a patamares de 50%. Logo, a precisão das PDFs das variáveis de saída podem ser altamente prejudicadas devido ao baixo número considerado. Os erros tendem a diminuir a partir de 10 *clusters*, porém a diminuição é cada vez mais sutil em  $K = 50$  adiante. Novamente, a quantidade indicada pelo processo de otimização proposto é validada. A partir de 50 *clusters*, o erro médio de desvio padrão das magnitudes de tensão é inferior a 10%. Note que para 100 *clusters* ainda há um erro diferente de zero. Isso é devido ao erro inerente que existe entre o próprio LHS em relação ao MCS. Na Seção 4.7.3, será mostrado que tais erros são muito pequenos quando se trata da recomposição analítica das PDFs das variáveis aleatórias de saída. Isso também justifica o fato de a quantidade de 50 *clusters* não diferir tanto em precisão quando confrontado com a situação na qual 100 *clusters* são empregados, o que indica que a primeira é, de fato, uma boa estimativa para essa simulação.

Figura 48 – Erros estatísticos para magnitudes de tensão, considerando-se a topologia IEEE 8500 e diferentes números de *clusters* no FPMP-CRA.



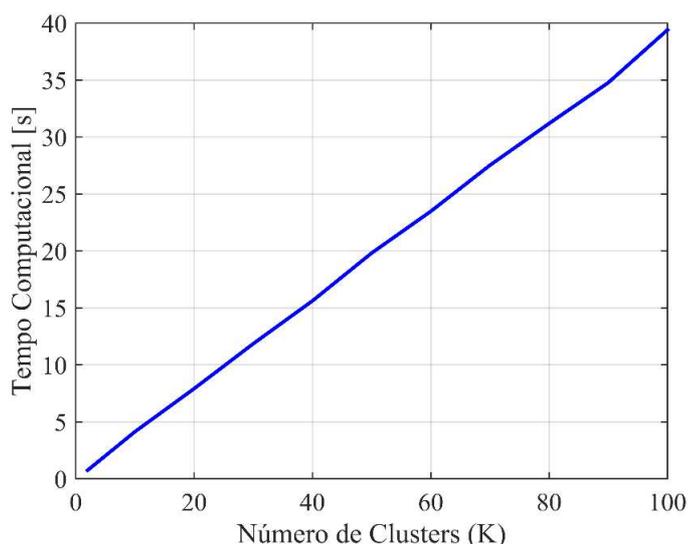
Fonte: elaborada pelo autor (2025).

Em suma, através de análises considerando tanto perdas elétricas quanto magnitudes de tensão, foi comprovado que, de fato, a melhor estimativa quanto ao número de *clusters* quando se almeja alta precisão dos resultados estatísticos é por volta de 50 *clusters*, exatamente o número apontado pelo algoritmo proposto. Por fim, a Figura 49 aponta a variação do tempo computacional requerido para cada uma das simulações que consideram diferentes números de *clusters*.

Trata-se de uma variação aproximadamente linear, o que é absolutamente esperado, uma vez que o tempo computacional demandado pelo FPMP-CRA é proporcional ao número de *clusters*, pois o número adotado será igual ao número de FPDs necessários para a convergência. Nesse sentido, uma quantidade maior do que 50 *clusters* não impacta sensivelmente na precisão dos resultados estatísticos, como já discutido, porém há o aumento proporcional no tempo computacional demandado para sua convergência. Dessa maneira, o aumento do número de *clusters* acarreta um aumento desnecessário da eficiência computacional do método, pois não haverá quaisquer ganhos no que tange à acurácia das PDFs. Isso reforça ainda mais a importância de se saber, em antecedência, uma quantidade estimada de *clusters* necessários para se alcançar determinada precisão dos resultados, providenciada pelo método de estimativa do número ótimo de *clusters* no FPMP-CRA.

Portanto, considerando-se tanto a precisão dos resultados estatísticos quanto o tempo computacional, infere-se que o método de estimativa do número de *clusters* é eficiente também para aplicações de grande porte, como a simulação probabilística do sistema IEEE 8500, e não apresenta quaisquer prejuízos quando um número muito alto de variáveis aleatórias é empregado, comprovando a possibilidade de sua escalabilidade para problemas mais complexos e mais próximos da realidade.

Figura 49 – Variação do tempo computacional do FPMP-CRA de acordo com o número de *clusters*.



Fonte: elaborada pelo autor (2025).

#### 4.7.3 ESTUDO 2: COMPARAÇÃO DO FPMP-CRA COM OUTROS MÉTODOS DE FPP PRESENTES NA LITERATURA

No presente estudo, o objetivo é comparar o desempenho do FPMP-CRA com outros métodos de FPP presentes na literatura. Os métodos escolhidos para efetuar tal validação foram<sup>22</sup>:

1. MEP 2m: o Método de Estimativa de Pontos é uma das abordagens mais comuns para solucionar o FPP. Trata-se de um método por aproximações que utiliza os primeiros momentos estatísticos das variáveis aleatórias de entrada para estimar analiticamente pontos de operação, que serão empregados para a solução determinística do FP. Posteriormente, os principais momentos (média e desvio padrão) das variáveis aleatórias de saída são calculados, a partir dos quais pode-se traçar suas PDFs. O esquema 2m foi escolhido por ser o mais rápido dos esquemas de Hong (1998) e por garantir solução para um número muito alto de variáveis. Esquemas mais desenvolvidos resultam em fatores de peso negativos ou complexos (Che *et al.*, 2019), o que impede sua aplicabilidade nesses casos de escalabilidade.
2. LHS: a simulação do FPP via LHS é fundamental para averiguar o sucesso do FPMP-CRA, dado que é justamente esse algoritmo responsável pela amostragem do banco de dados a ser comprimido pelo *clustering*. Nesse sentido, a comparação do FPMP-CRA com o LHS revela a aptidão do método proposto no que tange à redução de cenários e o erro cometido nesse processo.
3. FPMP-CRA utilizando K-Means: a ideia é simular um problema de FPP via *clustering* tal como é mais comum na literatura. Ou seja, comprimir os cenários gerados utilizando K-Means (que é a técnica amplamente mais utilizada) ao invés do K-Medoids CLARA, que é a abordagem sugerida no presente trabalho, para aplicações escaláveis. O número adotado de *clusters*, para essa formulação, será mantido em 50, para justificar uma comparação sob bases iguais em ambos os métodos. Logo, o intuito é averiguar a importância das diretrizes abordadas na Seção 4.4.3 na mitigação da chamada ‘maldição da dimensionalidade’.
4. MCS: buscando adotar um método de referência, o MCS de amostragem fixa foi adotado para tal fim. Trata-se de uma estratégia amplamente disseminada na literatura para verificar a precisão de algoritmos de FPP. Nesse caso, foi mantido o número de amostras de

---

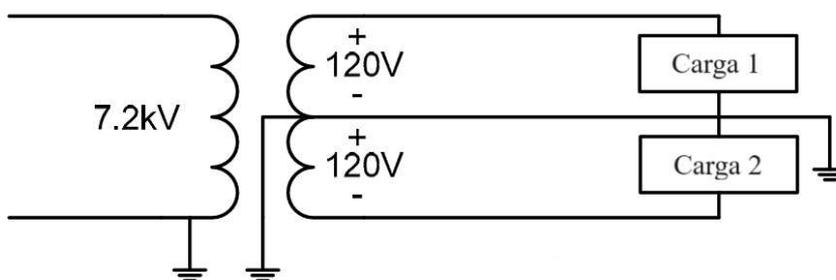
<sup>22</sup> Os desenvolvimentos matemáticos e principais aspectos computacionais de cada método podem ser vistos em maiores detalhes no Apêndice B. Ressalta-se que as condições de modelagem de rede, os equipamentos elétricos e as modelagens de variáveis aleatórias são sempre mantidas as mesmas para simulação de diferentes métodos, para justificar comparações sob as mesmas bases e sob os mesmos pressupostos metodológicos. Altera-se apenas a forma de solução do modelo.

maneira similar à Seção 4.7.2, ou seja, 10000 amostras. Este método pode ser usado como referencial comparativo devido à enorme quantidade de simulações determinísticas que serão conduzidas no seu processo de resolução.

A primeira comparação efetuada diz respeito aos resultados estatísticos de magnitudes e ângulos de tensão, dadas diferentes posições do sistema e escolhidas de forma aleatória. O resultado do método de referência, o MCS, é apresentado por meio de um histograma, que contém a distribuição dos resultados individuais de cada FPD conduzido para cada amostra. Enfatiza-se novamente que a quantidade de *clusters* definida tanto para o FPMP-CRA, quanto para o FPP via K-Means, é igual a 50.

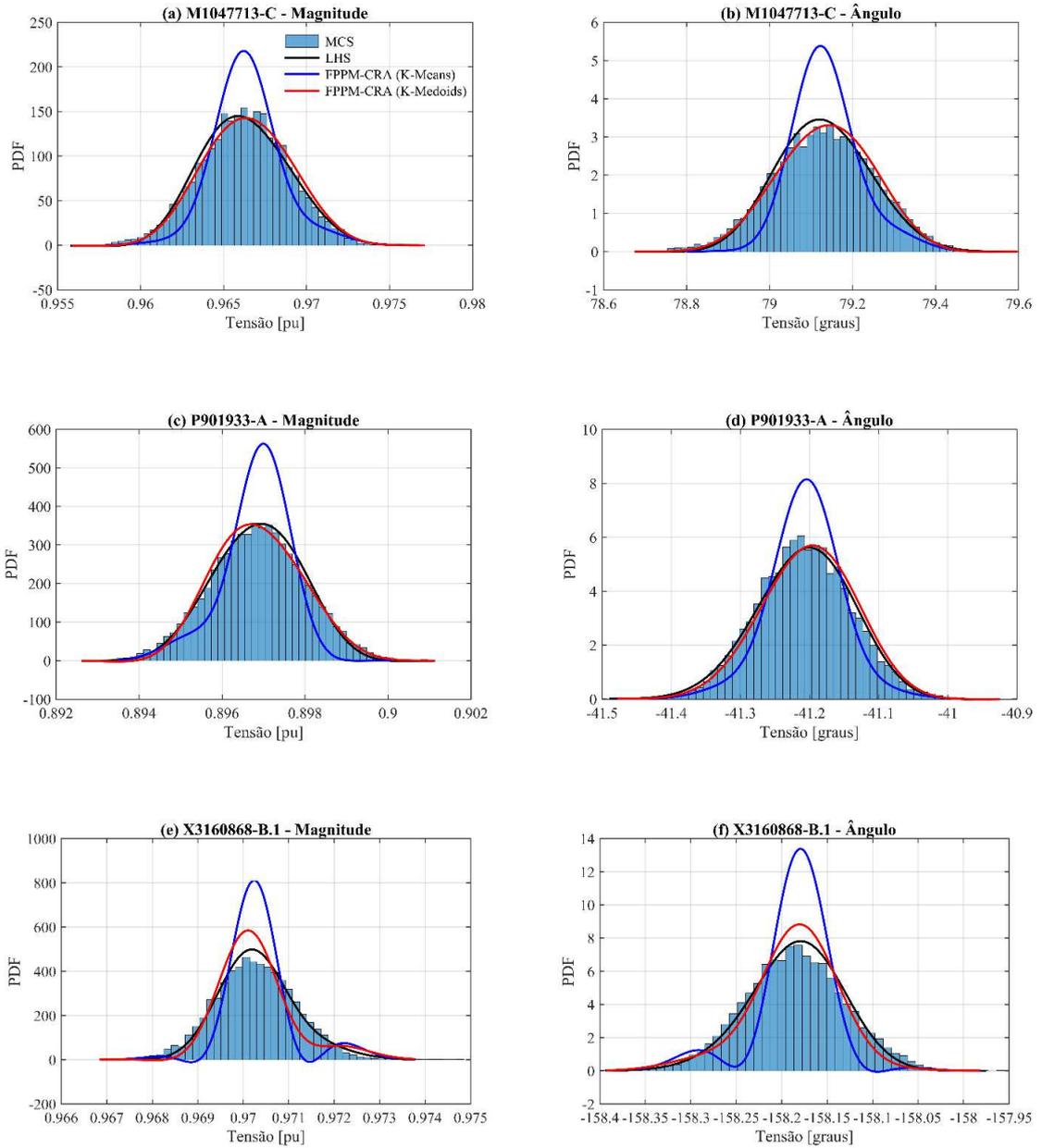
Essa comparação é apresentada na Figura 51. Nela, as PDFs obtidas via LHS estão na cor preta. Já as obtidas via FPMP-CRA utilizando K-Means estão na cor azul. Já as obtidas via FPMP-CRA utilizando K-Medoids estão na cor vermelha. As posições referentes aos resultados mostrados estão seguindo a seguinte padronização: há um código que indica a barra (por exemplo, M1047713), posteriormente a fase correspondente precedida de um sinal hífen (-A, -B ou -C). No caso das barras de carga, ainda é informado o terminal de conexão da carga em relação ao transformador de tape central, designados por 1 ou 2, conforme o esquema da Figura 50.

Figura 50 – Conexões das cargas na topologia IEEE 8500.



Fonte: adaptada de Arrit e Dugan (2010).

Figura 51 – Resultados estatísticos para tensões do IEEE 8500, considerando diferentes métodos probabilísticos.



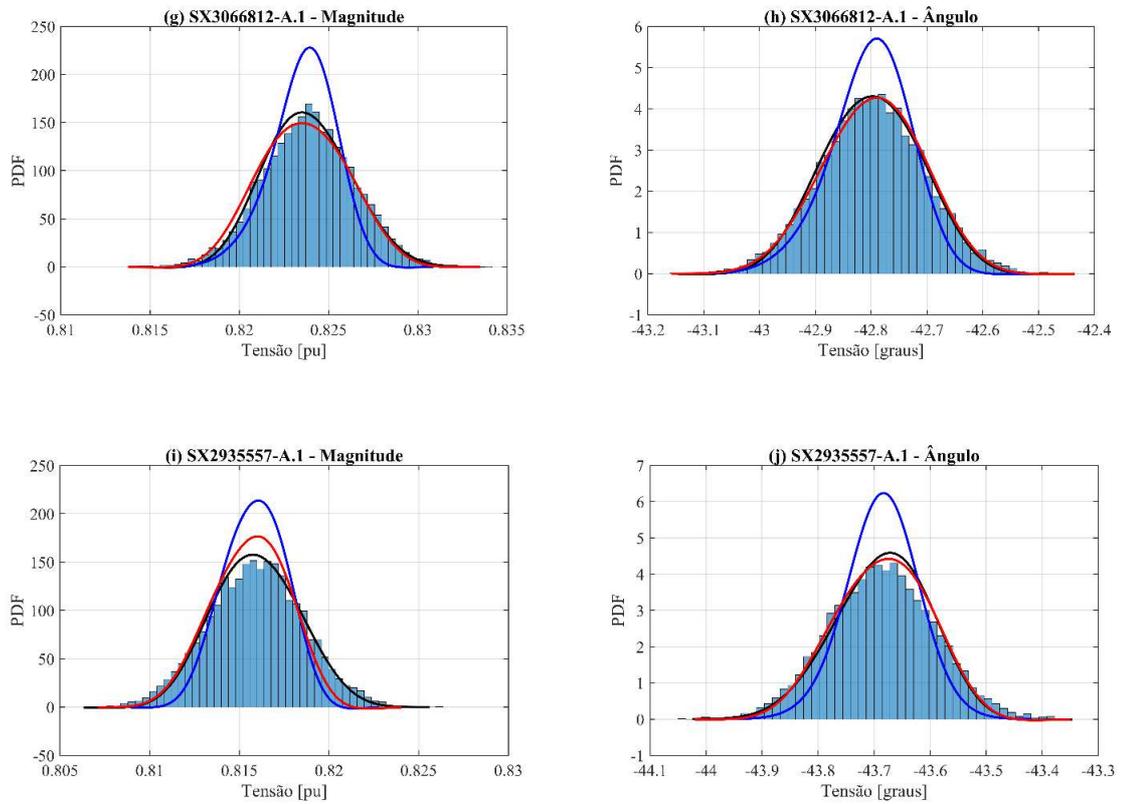
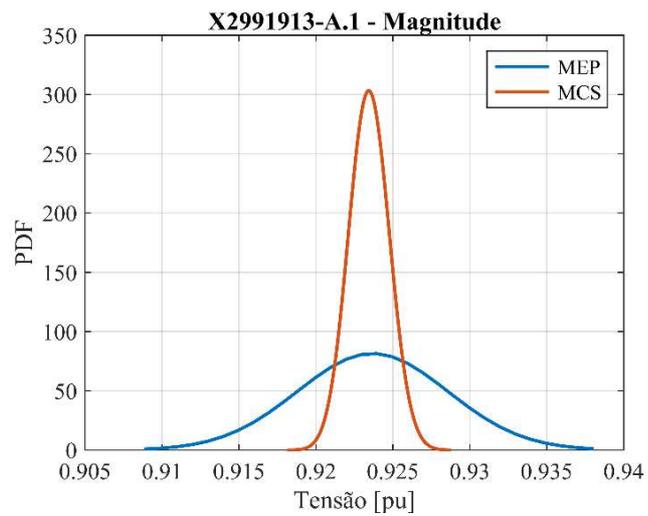


Figura 52 – Comparação MEP x MCS para IEEE 8500.



Os resultados referentes ao MEP não foram incluídos nesta primeira comparação gráfica. Isso é devido à enorme discrepância existente entre este e os demais métodos, o que prejudicaria a compreensão visual proposta. Isso pode ser constatado na Figura 52, em que há apenas as PDFs obtidas pelo MEP e o MCS, para magnitude de tensão na posição X2991913-A.1.

Da análise da Figura 51, infere-se que os resultados obtidos pelo LHS são muito próximos aos calculados via MCS, o que indica que, de fato, tal amostragem é mais eficiente. No caso do FPMP-CRA, tem-se que as PDFs são muito mais precisas quando o algoritmo K-Medoids CLARA é empregado, sendo os seus comportamentos muito semelhantes aos constatados via LHS e MCS. Já os gráficos obtidos pelo FPMP-CRA utilizando-se K-Means provaram ser a pior opção, devido aos grandes erros observados em todas as posições mostradas. As conclusões citadas podem ser generalizadas para todas as outras posições do sistema.

A imprecisão do algoritmo K-Means deve-se à citada ‘maldição da dimensionalidade’ e justifica a utilização de técnicas mais robustas para efetuar o *clustering*, tal como o K-Medoids proposto. Isso impacta significativamente na acurácia da representação das funções de probabilidade das variáveis aleatórias de saída e pode prejudicar estudos de planejamento caso um número muito alto de variáveis seja introduzido na modelagem.

Para caracterizar matematicamente os erros cometidos por cada método, são calculados os erros percentuais médios em relação ao MCS, que é a referência adotada, similarmente ao procedimento adotado na Seção 4.7.2. Tais erros são dispostos nas Tabela 14 e 15, a primeira referente às magnitudes de tensão e a segunda aos ângulos.

Tabela 14 – Erros percentuais médios em relação aos valores esperados e desvios padrões das magnitudes de tensão para topologia IEEE 8500.

Método	$\epsilon_1$ [%]	$\epsilon_2$ [%]
MEP	0,0617	272,8300
LHS	0,0035	4,2406
K-Means	0,0101	22,3048
FPMP-CRA (K-Medoids)	0,0118	6,6994

Fonte: elaborada pelo autor (2025).

Tabela 15 – Erros percentuais médios em relação aos valores esperados e desvios padrões dos ângulos de tensão para topologia IEEE 8500.

Método	$\epsilon_1$ [%]	$\epsilon_2$ [%]
MEP	0,0585	415,2629
LHS	0,0010	5,0928
K-Means	0,0055	19,0112
FPMP-CRA (K-Medoids)	0,0112	6,1601

Fonte: elaborada pelo autor (2025).

Dos dados apresentados, pode-se perceber que os valores esperados são perfeitamente estimáveis utilizando-se qualquer um dos métodos, tanto para magnitudes quanto para ângulos de tensões, uma vez que todos os erros foram baixos (menores do que 0,1% em todos os casos). No entanto, quando se analisa o desvio padrão, verifica-se que existe uma grande diferença quantitativa.

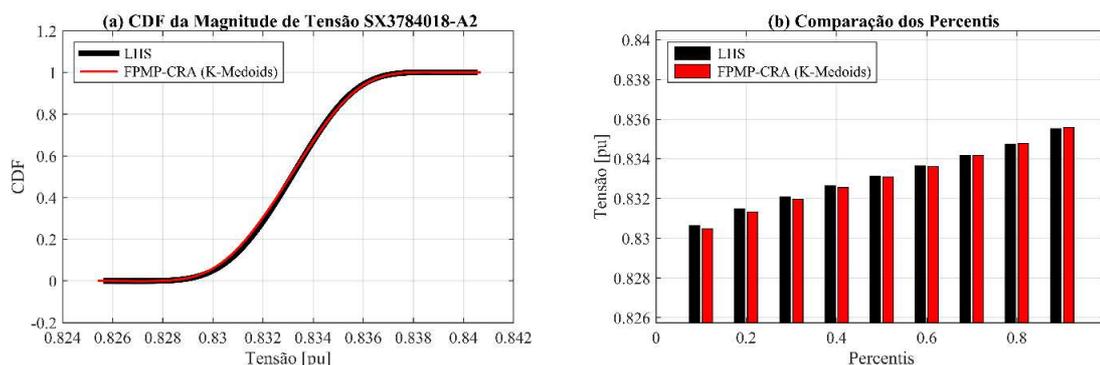
O MEP não é indicado para aplicações de grande porte, com erros significativamente elevados tanto para magnitudes quanto para ângulos. Isso pode ser constatado visualmente por meio da Figura 52, em que as médias de ambas as curvas são muito próximas, porém com desvios padrões absolutamente díspares.

Em consonância com os resultados mostrados na Figura 51, a precisão quanto ao desvio padrão é prejudicada quando o algoritmo K-Means é empregado, devido à alta dimensionalidade do banco de dados que contém as amostras das variáveis de entrada. No entanto, o desvio padrão apresenta precisão muito grande para os métodos LHS e para o FPMP-CRA utilizando K-Medoids CLARA. O erro do primeiro é ainda menor, pois há um erro de compressão do LHS em relação ao MCS conjuntamente com o erro do próprio K-Medoids após o processo de compressão. Nesse sentido é de suma importância mostrar que os erros cometidos pelo *clustering* ao reduzir o banco de dados não interferem substancialmente na recuperação analítica de suas PDFs.

Para isso, é proposta uma comparação acerca das CDFs reconstituídas analiticamente via EGC. Na Figura 53 (a), há a comparação da CDF para magnitude de tensão, confrontando a curva obtida pelo LHS (em cor preta) com aquela via FPMP-CRA (utilizando *clustering* K-Medoids CLARA, em vermelho). Na Figura 53 (b), há uma comparação dos percentis, sendo alvo de análise o décimo, vigésimo, trigésimo, assim por diante, até o nonagésimo percentil. Barras pretas indicam os percentis obtidos via LHS e vermelhos os calculados via FPMP-CRA. Valores parecidos de percentis indicam que o método proposto é eficiente na

eliminação dos cenários redundantes pelo processo de *clustering*. Logo, as CDFs podem ser vistas como extremamente similares. A posição escolhida para efetuar a comparação é a SX3784018-A.2.

Figura 53 – Comparação LHS x FPMP-CRA para topologia IEEE 8500 (magnitude de tensão).



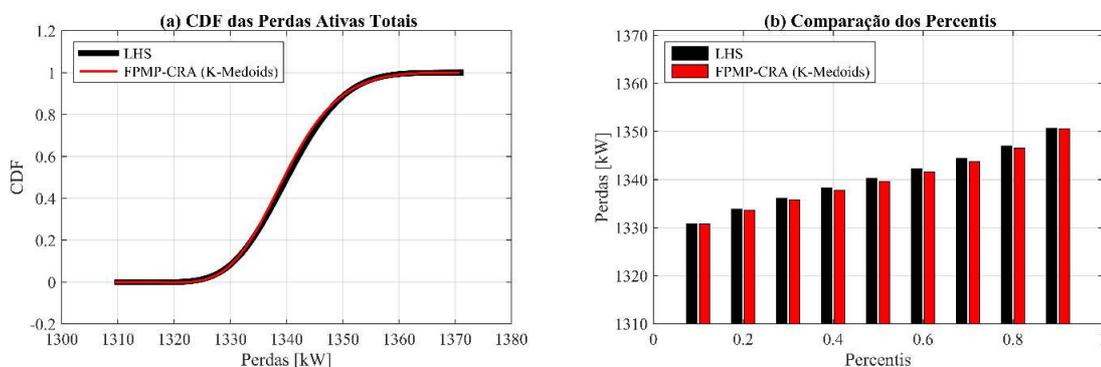
Fonte: elaborada pelo autor (2025).

Em ambas as figuras, pode-se perceber que as CDFs em ambos os métodos são muito próximas. Além disso, os percentis são aproximadamente idênticos, o que permite uma condução precisa para problemas de planejamento. O maior erro constatado vale 0,0196%, no décimo percentil, o que é muito limitado. A mesma conclusão pode ser inferida para todas as outras posições. Isso corrobora com o que foi indicado anteriormente: apesar de existir um erro estatístico em relação ao MCS, tal erro tem baixíssimo impacto na reconstrução analítica das PDFs e CDFs e, conseqüentemente, dos seus percentis.

Nesse sentido, conclui-se que o método proposto é altamente eficiente em reproduzir resultados estatísticos das variáveis de saída, ao contrário de outros métodos apresentados na literatura, que não são indicados para aplicações de grande porte, como o MEP e o FPP via K-Means.

Para ampliar as conclusões destacadas, o mesmo processo foi conduzido, porém tendo como alvo de análise o comportamento estatístico das perdas ativas totais da topologia. Os gráficos são apresentados na Figura 54.

Figura 54 – Comparação LHS x FPMP-CRA para topologia IEEE 8500 (perdas ativas totais).



Fonte: elaborada pelo autor (2025).

Novamente, percebe-se a enorme precisão do FPMP-CRA. Além das CDFs serem aproximadamente idênticas, o maior erro constatado relativo aos percentis vale 0,0474%, no sexagésimo percentil.

Em outras palavras, o método proposto apresenta erros muito pequenos comparados ao MCS e erros também muito limitados em relação ao LHS. Nesse último caso, os resultados são aproximadamente idênticos.

Resta demonstrar o comportamento do FPMP-CRA em relação ao tempo computacional demandado e confrontá-lo com o requisitado pelos outros métodos. Tais informações estão contidas na Tabela 16.

No caso do MCS, seu alto custo computacional já é conhecido na literatura (Ramadhani *et al.*, 2020), devido ao grande número requisitado de amostras solicitado para que resultados precisos sejam atingidos. Porém, trata-se de um método de referência para validação dos demais.

O MEP, por sua vez, apresenta uma grande desvantagem: o número de simulações determinísticas é diretamente proporcional ao número de variáveis aleatórias de entrada. Logo, caso o problema a ser solucionado seja de grande porte, tal método pode se tornar tão ineficiente quanto o próprio MCS. Note que o esquema escolhido foi aquele que necessita de menos pontos de concentração para alcançar a convergência (o esquema 2m) e, mesmo assim, foram necessárias 4708 simulações determinísticas do FP, o que levou a um tempo computacional extremamente grande.

O método LHS se mostrou mais eficiente em relação aos predecessores. Mas o FPMP-CRA consegue ser ainda mais rápido, demorando por volta de 18 segundos para alcançar a

convergência. Portanto, ele é aproximadamente 209 vezes mais rápido do que o MCS, 98 vezes mais rápido do que o MEP e quase duas vezes mais rápido do que o próprio LHS.

Tabela 16 – Tempo computacional demandado por método considerando IEEE 8500.

Método	Número de Simulações	Tempo computacional [s]
MCS	10000	3890,1156 (~1h5min)
MEP	4708	1835,8101 (~30min36s)
LHS	100	35,0224
FPMP-CRA (K-Means)	50	18,2535
FPMP-CRA (K-Medoids)	50	18,5734

Fonte: elaborada pelo autor (2025).

Ressalta-se que o MCS e o LHS podem ser acelerados adotando-se um critério de estagnação. No entanto, um menor número de amostras para esses métodos numéricos irá traduzir em um menor número de *clusters* necessário para o FPMP-CRA, acelerando-o na mesma proporção. Isso foi visto nas diretrizes para redução de dimensionalidade. Os tempos podem ser acelerados por vias de processamento paralelo, ou considerando outros métodos para avaliar cada FP individual.

Em resumo, o método proposto possui um desempenho computacional muito superior aos demais e, devido à sua enorme precisão, pode ser visto como uma importante ferramenta para simulação de problemas probabilísticos direcionados a sistemas desequilibrados de grande porte. No caso de problemas de otimização, em que o FPP precisa ser avaliado inúmeras vezes até que a configuração ótima seja atingida, o tempo computacional demandado torna-se fator de risco. Nesse caso, o FPMP-CRA é uma melhor opção, sobretudo em se tratando de aplicações de escalabilidade das variáveis aleatórias.

#### 4.7.4 CONCLUSÕES PARCIAIS DO TESTE 6

Conforme mencionado no Capítulo 2, poucos trabalhos na literatura se dedicam a aplicações de grande porte. Quando SD desequilibrados são alvo de estudo, essa aplicabilidade é inexistente. Além disso, a escalabilidade de variáveis aleatórias também é um fator negligenciado. O estudo de sistemas de grande porte é fundamental para averiguar as

características do método desenvolvido sob circunstâncias mais próximas da realidade, além de representar situações mais desafiadoras na perspectiva computacional.

No presente teste, o sistema IEEE 8500 foi adotado para efetuar análises de escalabilidade. Nele, cada uma das 2354 cargas foi assumida como aleatória, modelada por PDFs Gaussianas independentes. O procedimento para estimar o número ótimo de *clusters* indicou a necessidade de 50 grupos para providenciar resultados precisos, utilizando-se para essa finalidade o algoritmo K-Medoids CLARA, mais indicado para aplicações de grande dimensionalidade.

A quantidade ideal de *clusters* foi comprovada através de comparações com métodos exaustivos, em que foi visto que o número obtido é, de fato, aquele que promove bons resultados a um tempo computacional aceitável. Menos *clusters* podem resultar em precisão ruim das variáveis aleatórias de saída. Mais *clusters* sobrecarregam o tempo computacional demandado pelo FPP.

Posteriormente, o FPMP-CRA foi confrontado com outros algoritmos presentes na literatura para simular o FPP: MCS, MEP, LHS e FPP via K-Means. O método proposto provou ser muito mais eficiente, apresentando baixos erros percentuais médios para todas as variáveis aleatórias de saída estudadas em relação ao MCS, considerando momentos estatísticos de primeira e segunda ordem. Outros métodos por aproximação não obtiveram sucesso em estimar as PDFs das variáveis de saída de forma correta. O MEP, devido à sua baixa quantidade de pontos de concentração por variável, resultou em erros extremamente severos, alcançando erro superior a 200% no segundo momento estatístico. Já o FPP via K-Means sofre com a ‘maldição da dimensionalidade’, perdendo muita precisão com o aumento das variáveis. Logo, o método proposto alcançou resultados muito superiores para aplicações de grande porte. Ademais, o FPMP-CRA é aproximadamente 209 vezes mais rápido em relação ao MCS, 98 vezes em relação ao MEP e aproximadamente duas vezes mais rápido do que o LHS.

Por isso, devido à precisão comparável ao MCS e ao seu excelente tempo computacional, o FPMP-CRA pode ser visto como uma ferramenta importante para solucionar o FPP para aplicações de grande porte, sobretudo para sistemas desequilibrados dotados de um número muito grande de variáveis aleatórias de entrada.

## 5 CONSIDERAÇÕES FINAIS

O presente capítulo destina-se a apresentar as conclusões finais da tese, assim como os desdobramentos de futuros desenvolvimentos que podem ser conduzidos a partir do conteúdo demonstrado neste trabalho.

### 5.1 CONCLUSÕES

O objetivo primordial do trabalho consistiu no desenvolvimento de uma nova ferramenta para simular problemas de FPP, destinados especificamente a SD desequilibrados. Seu nome é Fluxo de Potência Multifásico Probabilístico baseado em *Clustering* e Recuperações Analíticas (FPMP-CRA). Seu desenvolvimento envolve cinco principais passos: primeiramente, são definidas as variáveis aleatórias de entrada. Em seguida, é conduzido um procedimento de amostragem para identificar os pontos de operação do problema. Logo, uma técnica de *clustering* é aplicada para comprimir essas amostras. O número ótimo de *clusters* é calculado através de um procedimento que leva em consideração as distâncias *intra-cluster* e *inter-cluster*, e é uma contribuição da presente tese. Posteriormente, um problema de FP multifásico determinístico é conduzido para cada condição de entrada cujo cenário é modelado por um centro de *cluster*, armazenando-se as suas respectivas variáveis de saída. Por fim, os principais momentos estatísticos de cada variável aleatória de saída são calculados tendo como base os resultados dos FP. A partir disso, um método analítico chamado Expansão Gram-Charlier é utilizado para recuperar analiticamente as PDFs de cada uma das variáveis de saída.

Foram conduzidos seis testes, que englobaram diversos estudos para averiguar a aplicabilidade do método, tanto em questões técnicas (a respeito da execução do método mediante diferentes condições de entrada) quanto questões computacionais (a respeito do desempenho do tempo computacional do método). As principais conclusões estão listadas a seguir.

- A modelagem de rede foi alvo de investigação do teste 1. Foi visto que a inclusão do desequilíbrio é imprescindível para alcançar resultados precisos em se tratando de SD. Quanto maior o desequilíbrio do sistema, maiores são os erros obtidos nas médias e nos desvios padrões das variáveis aleatórias de saída quando modelagens equilibradas de redes são aplicadas. Nesse sentido, infere-se que, devido ao inerente desequilíbrio de SD, a

consideração da modelagem multifásica é um aspecto obrigatório para simular problemas probabilísticos relacionados a esse tipo de sistema.

- A modelagem probabilística da carga foi estudada no teste 2. Foram comparadas quatro situações de correlação diferentes: cargas fracamente correlacionadas, cargas moderadamente correlacionadas, cargas fortemente correlacionadas e cargas correlacionadas de maneira diferente de acordo com a fase de conexão. A enorme flexibilidade do FPMP-CRA foi logo destacada, pois permite a simulação dessas diferentes situações sem a necessidade de alterações pontuais do método. Foi visto que a correta consideração da correlação da carga é de suma importância para que se possa obter resultados precisos. O desvio padrão é, sobretudo, o principal aspecto afetado quando diferentes correlações são assumidas. Logo, as PDFs de tensões e correntes elétricas diferem significativamente com a variação dessas condições. Além disso, foi constatado que a situação na qual são introduzidas diferentes correlações por fase pode causar maior probabilidade de subtensão em relação ao caso de correlação fraca, o que contraria o que estava previamente estabelecido na literatura, que geralmente considera o caso de cargas independentes como o ‘mais conservador’. Isso evidencia a importância da simulação de sistemas multifásicos e de suas exatas correlações entre as cargas.

- Uma vez que o FPMP-CRA é baseado em *clustering*, é imprescindível adotar estratégias para permitir sua aplicabilidade em situações de alta dimensionalidade, objetivando mitigar a chamada ‘maldição da dimensionalidade’ a que esses algoritmos estão submetidos. Logo, foi visto no teste 3 que a manifestação prática dessa questão é a necessidade de um número cada vez maior de *clusters* para atender às situações com grande número de variáveis aleatórias, o que sobrecarrega o tempo computacional do método. Mesmo assim, por volta de 70% dos cenários previamente amostrados foram eliminados pelo *clustering* nos piores casos de dimensionalidade simulados. Porém, com o intuito de aprimorar ainda mais o método, três diretrizes foram sugeridas para lidar com esses casos. A primeira é a respeito da amostragem, que deve ser feita através do LHS, um método numérico bastante preciso e ideal para a posterior aplicação de *clustering*, pois as amostras tendem a não ficar sobrepostas. A segunda diretriz consiste na utilização de um algoritmo de *clustering* chamado K-Medoids CLARA, que é indicado para comprimir dados de alta dimensionalidade. Por fim, a redução do número de variáveis aleatórias das cargas é sugerida na terceira diretriz. Coloca-se como referência a utilização de cargas com desvios padrões inferiores a 10 kW como determinísticas. Tendo como base tais diretrizes, o mesmo problema

anterior foi simulado em um sistema teste que possui 20 cargas aleatórias. O percentual de cenários eliminados pelo *clustering* passou a ser até cerca de 99%, aumentando consideravelmente a eficiência do FPMP-CRA, mantendo-se sua precisão.

- Devido às constantes transformações das redes de distribuição, é cada vez mais requerida a consideração de inserção de GDs para a simulação de redes ativas. Como os empreendimentos são oriundos sobretudo de tecnologias renováveis, a inclusão de incertezas é imprescindível. No teste 4, foi simulada a rede IEEE 123 com GDs fotovoltaicas, tais que a produção de potência é dependente da irradiação solar, que é uma variável aleatória. Foram consideradas correlações temporais das irradiações solares e das demandas, o que enaltece a robustez do FPMP-CRA. As dependências estatísticas foram modeladas pela teoria de cópulas e o processo de *clustering* foi efetivado hora a hora de um planejamento diário, para aumentar a eficiência computacional do método. Foi constatado que o método consegue estimar com elevada precisão as PDFs das tensões nodais, permitindo cálculo minucioso das probabilidades de violação de tensão por sobretensão, devido à inserção de potência fotovoltaica em horários de pico de irradiação solar. Outra aplicação importantíssima providenciada pelo FPMP-CRA é a possibilidade de identificação regiões geográficas críticas para distúrbios elétricos, que necessitariam de maior atenção por parte de operadores e concessionárias. Foi feita uma comparação acerca do tempo computacional em relação a métodos convencionais e também no caso em que as correlações temporais são negligenciadas, em que foi possível perceber que a modelagem correta dos dispositivos e dos fenômenos considerados devem apresentar rigor estatístico suficiente para atingir resultados relevantes e fidedignos. O FPMP-CRA é de fundamental importância, pois consegue alcançar esses critérios e engloba um enorme número de possibilidades de simulação.

- Para os estudos de redes de distribuição multiterradas, é imprescindível a modelagem explícita do condutor neutro e dos aterramentos. O FPMP-CRA permite tais inclusões, pois seu processo de resolução depende das equações do MICN, que é ideal para tratar a modelagem multifásica. Nesse contexto, o FPMP-CRA é empregado, no teste 5, para subsidiar estudos de segurança de tensão de neutro. Utilizando-se a topologia IEEE NEV, o objetivo foi determinar as PDFs das tensões de neutro e verificar um intervalo de confiança para suas características operativas. As variáveis aleatórias foram as resistências de aterramento do sistema, que dependem de fatores estocásticos como a resistividade e umidade do solo. Os resultados indicaram que existe uma significativa discrepância dos resultados médios em relação a eventos raros que podem se manifestar – e só podem ser detectados por

meio de ferramentas probabilísticas robustas como o FPMP-CRA. Tais diferenças podem chegar a 20 V, o que pode colocar em risco a saúde humana e a vida útil de equipamentos. Em alguns pontos do sistema, a tensão de neutro pode inclusive alcançar 70 V, representando uma operação perigosa e crítica. Graças à eficiência do método proposto de otimização do número de *clusters*, foi possível identificar eventos raros por meio da inclusão de *outliers* que são rastreados pela consideração da otimização da parcela referente às distâncias *intra-clusters*. Isso é de extrema importância pois, apesar de esses cenários serem raros e pouco prováveis, os agentes responsáveis pelo planejamento das redes devem ter em consideração esses eventos para mitigar possíveis acidentes humanos. Foi visto que a diferença entre os cenários médios e esses eventos raros podem chegar a patamares de 20 V. Infere-se que o FPMP-CRA pode ser utilizado para estudos de avaliação de segurança de tensão de neutro, podendo auxiliar engenheiros de planejamento a tomarem ações mitigadoras que visem maior segurança do sistema.

- No teste 6, o objetivo foi simular o FPMP-CRA em uma aplicação de grande porte e de escalabilidade do número de variáveis aleatórias. Esse estudo é cada vez mais requisitado para avaliar situações mais próximas da realidade e impor desafios maiores para os métodos desenvolvidos. Para isso, foi considerada a topologia IEEE 8500 com 2354 cargas aleatórias. O FPMP-CRA demonstrou enorme precisão em relação ao tradicional MCS, sendo aproximadamente 209 vezes mais rápido. Seu desempenho foi confrontado com outros métodos de FPP, como o MEP, o LHS e o FPP via K-Means, apresentando precisão similar ou superior a todos eles. O MEP e FPP via K-Means não são recomendáveis para simulações de grande porte. No caso de FPP via *clustering*, a abordagem via K-Medoids CLARA provou ser muito superior.

O FPMP-CRA se apresentou como uma ferramenta: (i) flexível, pois é capaz de simular diferentes condições de entrada do FPP, incluindo diferentes correlações de variáveis aleatórias de entrada, sem necessidade de modificação prévia do algoritmo; (ii) simples, pois sua implementação matemática e computacional depende sobretudo do cálculo determinístico do FP para estimar os momentos estatísticos das variáveis de saída; (iii) preciso, pois consegue alcançar resultados muito próximos aos do MCS, que é um método confiável de referência, amplamente utilizado na literatura para esse fim; (iv) rápido, pois demonstra excelente tempo computacional em relação aos métodos mais comuns da literatura; (v) escalável, pois pode ser usado em aplicações de grande porte e de um alto número de variáveis aleatórias, o que permite a simulação de problemas mais próximos da realidade.

Tendo em vista todos esses pontos, infere-se que o FPMP-CRA é um método importante para o estado da arte de métodos de FPP para sistemas desequilibrados, sobretudo quando aplicações de grande porte são requeridas ou para aplicações que a velocidade de convergência do método é um aspecto crucial, como em problemas de planejamento baseados em otimização, por exemplo.

## 5.2 TRABALHOS FUTUROS

A seguir, são enumerados possíveis desdobramentos do trabalho apresentado. Ou seja, estudos que podem vir a ser conduzidos futuramente, com a presente tese sendo um ponto de partida.

1. Aplicação do FPMP-CRA em problemas de otimização: a ideia é envolver o método desenvolvido em um problema de otimização cujas restrições de rede são probabilísticas.
2. Utilização de técnicas alternativas de FP para acelerar ainda mais o FPMP-CRA: além de otimizar o número de *clusters* para promover aceleração do tempo computacional, pode-se empregar técnicas mais avançadas e rápidas de FP, como o FP tensorial, para torná-lo ainda mais eficiente.
3. Utilização de processamento paralelo no FPMP-CRA, para acelerar sua convergência ainda mais: uma vez que as execuções determinísticas são independentes, pode-se adotar o processamento paralelo para executá-los e, só depois, estimar os momentos estatísticos.
4. Modelagem de outras variáveis aleatórias de entrada, como gerações eólicas distribuídas e baterias.
5. Representação de cenários correlacionados utilizando outras técnicas além das copulas empregadas no presente trabalho.
6. Modelagens de cenários correlacionados espacialmente e temporalmente ao mesmo tempo.
7. Estudar outras técnicas de *clustering* e verificar eficiência do método.
8. Estudar outras técnicas de recuperação analítica, como Expansão de Cornish-Fischer.
9. Estudar o impacto de dispositivos de controles e tratamentos de limites sob a perspectiva probabilística, como controles por capacitores, reguladores, inversores inteligentes, STATCOMs, entre outros.

10. Análise do impacto probabilístico nas PDFs das variáveis aleatórias de saída devido à inclusão dos reguladores de tensão. Incluir análises que incluem o tempo de resposta do regulador.
11. Simular problemas acoplados temporalmente, mas em discretizações temporais diferentes, como simulações minuto a minuto, incluindo análises do tempo de duração de sobretensões e demais distúrbios elétricos.
12. Análise de sistemas multiterrados com presença de GDs fotovoltaicas ou eólicas.
13. Variar estocasticamente a posição das GDs nos sistemas analisados e conduzir simulação de problemas de máxima capacidade de hospedagem usando o equacionamento probabilístico.
14. Simulações para averiguar probabilidades de fluxo reverso em redes de distribuição ativas.
15. Incluir modelos mais completos para representar as GDs fotovoltaicas, de forma a possibilitar a representação da temperatura como variável aleatória correlacionada com a irradiação solar.
16. Simulação de sistemas reais brasileiros, para demonstrar aplicabilidade prática do método no contexto nacional.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABBASI, A. R.; MOHAMMADI, M. Probabilistic load flow in distribution networks: An updated and comprehensive review with a new classification proposal. **Electric Power Systems Research**, [s. l.], v. 222, p. 109497, 2023.
- ABUD, T. P. *et al.* State of the Art Monte Carlo Method Applied to Power System Analysis with Distributed Generation. **Energies**, [s. l.], v. 16, n. 1, p. 394, 2022.
- ANEEL. **Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional**. [S. l.]: Agência Nacional de Energia Elétrica, 2022. Disponível em: <https://www.gov.br/aneel/pt-br/centrais-de-contudos/procedimentos-regulatorios/prodist>.
- ARAUJO, L. R. D. *et al.* Comparisons between the three-phase current injection method and the forward/backward sweep method. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 32, n. 7, p. 825–833, 2010.
- ARORA, P.; DEEPALI; VARSHNEY, S. Analysis of K-Means and K-Medoids Algorithm For Big Data. **Procedia Computer Science**, [s. l.], v. 78, p. 507–512, 2016.
- ARRITT, R. F.; DUGAN, R. C. The IEEE 8500-node test feeder. *In: IEEE PES T&D 2010, 2010, New Orleans, LA, USA. IEEE PES T&D 2010*. New Orleans, LA, USA: IEEE, 2010. p. 1–6. Disponível em: <http://ieeexplore.ieee.org/document/5484381/>. Acesso em: 17 jul. 2024.
- BAGHAEE, H. R. *et al.* Application of RBF neural networks and unscented transformation in probabilistic power-flow of microgrids including correlated wind/PV units and plug-in hybrid electric vehicles. **Simulation Modelling Practice and Theory**, [s. l.], v. 72, p. 51–68, 2017.
- BALAMURUGAN, K.; SRINIVASAN, D. Review of power flow studies on distribution network with distributed generation. *In: 2011 IEEE NINTH INTERNATIONAL CONFERENCE ON POWER ELECTRONICS AND DRIVE SYSTEMS (PEDS 2011), 2011, Singapore, Singapore. 2011 IEEE Ninth International Conference on Power Electronics and Drive Systems*. Singapore, Singapore: IEEE, 2011. p. 411–417. Disponível em: <http://ieeexplore.ieee.org/document/6147281/>. Acesso em: 13 jul. 2024.
- BARAN, M. E.; WU, F. F. Network reconfiguration in distribution systems for loss reduction and load balancing. **IEEE Transactions on Power Delivery**, [s. l.], v. 4, n. 2, p. 1401–1407, 1989.
- BARIK, S.; DAS, D. Impact of FFC distributed generations in a DNR in the presence of renewable and load uncertainties by mixed-discrete particle swarm-based point estimation method. **IET Renewable Power Generation**, [s. l.], v. 13, n. 9, p. 1431–1445, 2019.
- BARUTCU, I. C.; KARATEPE, E.; BOZTEPE, M. Impact of harmonic limits on PV penetration levels in unbalanced distribution networks considering load and irradiance uncertainty. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 118, p. 105780, 2020.

BERG, R.; HAWKINS, E.; PLEINES, W. Mechanized Calculation of Unbalanced Load Flow on Radial Distribution Circuits. **IEEE Transactions on Power Apparatus and Systems**, [s. l.], v. PAS-86, n. 4, p. 415–421, 1967.

BHAT, N. G.; PRUSTY, B. R.; JENA, D. Cumulant-based correlated probabilistic load flow considering photovoltaic generation and electric vehicle charging demand. **Frontiers in Energy**, [s. l.], v. 11, n. 2, p. 184–196, 2017.

BORKOWSKA, B. Probabilistic Load Flow. **IEEE Transactions on Power Apparatus and Systems**, [s. l.], v. PAS-93, n. 3, p. 752–759, 1974.

CAI, D.; SHI, D.; CHEN, J. Probabilistic load flow with correlated input random variables using uniform design sampling. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 63, p. 105–112, 2014.

CHAI, Y. *et al.* Key node identification and network simplification modelling method for optimal power flow analysis of active distribution network. **Electric Power Systems Research**, [s. l.], v. 238, p. 111066, 2025.

CHE, Y. *et al.* Probabilistic Load Flow Using An Improved Point Estimate Method Considering Wind Generation. *In: 2019 IEEE INNOVATIVE SMART GRID TECHNOLOGIES - ASIA (ISGT ASIA)*, 2019, Chengdu, China. **2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)**. Chengdu, China: IEEE, 2019. p. 4080–4085. Disponível em: <https://ieeexplore.ieee.org/document/8881186/>. Acesso em: 17 jul. 2024.

CHEVALIER, S.; SCHENATO, L.; DANIEL, L. Accelerated Probabilistic Power Flow in Electrical Distribution Networks via Model Order Reduction and Neumann Series Expansion. **IEEE Transactions on Power Systems**, [s. l.], v. 37, n. 3, p. 2151–2163, 2022.

CIRIC, R. M.; FELTRIN, A. P.; OCHOA, L. F. Power flow in four-wire distribution networks-general approach. **IEEE Transactions on Power Systems**, [s. l.], v. 18, n. 4, p. 1283–1290, 2003.

CONSTANTE-FLORES, G. E.; ILLINDALA, M. S. Data-Driven Probabilistic Power Flow Analysis for a Distribution System With Renewable Energy Sources Using Monte Carlo Simulation. **IEEE Transactions on Industry Applications**, [s. l.], v. 55, n. 1, p. 174–181, 2019.

DING, Q. *et al.* Day-Ahead Probabilistic Load Flow Analysis Considering Wind Power Forecast Error Correlation. **Energy and Power Engineering**, [s. l.], v. 09, n. 04, p. 292–299, 2017.

GALIANA, F. D.; BANAKAR, M. Approximation Formulae for Dependent Load Flow Variables. **IEEE Transactions on Power Apparatus and Systems**, [s. l.], v. PAS-100, n. 3, p. 1128–1137, 1981.

GALVANI, S. *et al.* A new approach for probabilistic harmonic load flow in distribution systems based on data clustering. **Electric Power Systems Research**, [s. l.], v. 176, p. 105977, 2019.

GARCIA, P. A. N. *et al.* Three-phase power flow calculations using the current injection method. **IEEE Transactions on Power Systems**, [s. l.], v. 15, n. 2, p. 508–514, 2000.

- GRUOSSO, G. *et al.* Joined Probabilistic Load Flow and Sensitivity Analysis of Distribution Networks Based on Polynomial Chaos Method. **IEEE Transactions on Power Systems**, [*s. l.*], v. 35, n. 1, p. 618–627, 2020.
- HAQUE, M. H. Load flow solution of distribution systems with voltage dependent load models. **Electric Power Systems Research**, [*s. l.*], v. 36, n. 3, p. 151–156, 1996.
- HASAN, K. N.; PREECE, R.; MILANOVIC, J. V. Priority Ranking of Critical Uncertainties Affecting Small-Disturbance Stability Using Sensitivity Analysis Techniques. **IEEE Transactions on Power Systems**, [*s. l.*], v. 32, n. 4, p. 2629–2639, 2017.
- HEIDARIFAR, M.; ANDRIANESIS, P.; CARAMANIS, M. Efficient Load Flow Techniques Based on Holomorphic Embedding for Distribution Networks. *In: 2019 IEEE POWER & ENERGY SOCIETY GENERAL MEETING (PESGM), 2019, Atlanta, GA, USA. 2019 IEEE Power & Energy Society General Meeting (PESGM)*. Atlanta, GA, USA: IEEE, 2019. p. 1–5. Disponível em: <https://ieeexplore.ieee.org/document/8973882/>. Acesso em: 13 jul. 2024.
- HERNÁNDEZ, J. C.; RUIZ-RODRIGUEZ, F. J.; JURADO, F. Modelling and assessment of the combined technical impact of electric vehicles and photovoltaic generation in radial distribution systems. **Energy**, [*s. l.*], v. 141, p. 316–332, 2017.
- HOESE, A.; GARCÉS, F. Stochastic correlated simulation: an extension of the cumulant method to include time-dependent energy sources. **International Journal of Electrical Power & Energy Systems**, [*s. l.*], v. 21, n. 1, p. 13–22, 1999.
- HONG, H. P. An efficient point estimate method for probabilistic analysis. **Reliability Engineering & System Safety**, [*s. l.*], v. 59, n. 3, p. 261–267, 1998.
- IEEE PES TEST FEEDERS. *In: 17 jul. 2024*. Disponível em: <https://cmte.ieee.org/pes-testfeeders/resources/>. Acesso em: 17 jul. 2024.
- KEIHAN ASL, D.; MOHAMMADI, M.; REZA SEIFI, A. Holomorphic embedding load flow for unbalanced radial distribution networks with DFIG and tap-changer modelling. **IET Generation, Transmission & Distribution**, [*s. l.*], v. 13, n. 19, p. 4263–4273, 2019.
- LI, Y. *et al.* Data-Driven Nonparametric Probabilistic Optimal Power Flow: An Integrated Probabilistic Forecasting and Analysis Methodology. **IEEE Transactions on Power Systems**, [*s. l.*], v. 38, n. 6, p. 5820–5833, 2023.
- LLOYD, S. Least squares quantization in PCM. **IEEE transactions on information theory**, [*s. l.*], v. 28, n. 2, p. 129–137, 1982.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics**, [*s. l.*], v. 5, 1967.
- MARTINS, A. S. C.; ARAUJO, L. R. D.; PENIDO, D. R. R. K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow. **International Journal of Electrical Power & Energy Systems**, [*s. l.*], v. 157, p. 109861, 2024.

- MARTINS, A. S. C.; DE ARAUJO, L. R.; PENIDO, D. R. R. An active multiphase probabilistic power flow based on a clustering approach. **Electric Power Systems Research**, [s. l.], v. 230, p. 110264, 2024.
- MORALES, Juan. M.; PEREZ-RUIZ, Juan. Point Estimate Schemes to Solve the Probabilistic Power Flow. **IEEE Transactions on Power Systems**, [s. l.], v. 22, n. 4, p. 1594–1601, 2007.
- MORI, H.; WENJUN JIANG. A New Probabilistic Load Flow Method Using MCMC in Consideration of Nodal Load Correlation. *In*: 2009 15TH INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEM APPLICATIONS TO POWER SYSTEMS (ISAP), 2009, Curitiba. **2009 15th International Conference on Intelligent System Applications to Power Systems**. Curitiba: IEEE, 2009. p. 1–6. Disponível em: <http://ieeexplore.ieee.org/document/5352840/>. Acesso em: 15 jul. 2024.
- MÜLLER, H. H.; RIDER, M. J.; CASTRO, C. A. Artificial neural networks for load flow and external equivalents studies. **Electric Power Systems Research**, [s. l.], v. 80, n. 9, p. 1033–1041, 2010.
- PALAHALLI, H.; MAFFEZZONI, P.; GRUOSSO, G. Gaussian Copula Methodology to Model Photovoltaic Generation Uncertainty Correlation in Power Distribution Networks. **Energies**, [s. l.], v. 14, n. 9, p. 2349, 2021.
- PENIDO, D. R. R. *et al.* A new tool for multiphase electrical systems analysis based on current injection method. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 44, n. 1, p. 410–420, 2013.
- PENIDO, D. R. R. *et al.* Three-Phase Power Flow Based on Four-Conductor Current Injection Method for Unbalanced Distribution Networks. **IEEE Transactions on Power Systems**, [s. l.], v. 23, n. 2, p. 494–503, 2008.
- PUTRA, J. T.; SARJIYA; SETYONEGORO, M. I. B. Modeling of high uncertainty photovoltaic generation in quasi dynamic power flow on distribution systems: A case study in Java Island, Indonesia. **Results in Engineering**, [s. l.], v. 21, p. 101747, 2024.
- RAMADHANI, U. H. *et al.* Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 120, p. 106003, 2020.
- REZAEIAN-MARJANI, S. *et al.* Optimal allocation of D-STATCOM in distribution networks including correlated renewable energy sources. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 122, p. 106178, 2020.
- ROUHANI, M.; MOHAMMADI, M.; AIELLO, M. Soft clustering based probabilistic power flow with correlated inter temporal events. **Electric Power Systems Research**, [s. l.], v. 204, p. 107677, 2022.
- RUIZ-RODRIGUEZ, F. J.; HERNÁNDEZ, J. C.; JURADO, F. Voltage behaviour in radial distribution systems under the uncertainties of photovoltaic systems and electric vehicle charging loads. **International Transactions on Electrical Energy Systems**, [s. l.], v. 28, n. 2, p. e2490, 2018.

- SADEGHIAN, O. *et al.* A robust data clustering method for probabilistic load flow in wind integrated radial distribution networks. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 115, p. 105392, 2020.
- SALEHI, M.; REZAEI, M. M. An improved probabilistic load flow in distribution networks based on clustering and Point estimate methods. **Energy and AI**, [s. l.], v. 14, p. 100272, 2023.
- SANDOVAL, N.; GONG, Y.; CHUNG, C. Y. Three-Phase Second-Order Analytic Probabilistic Load Flow With Voltage-Dependent Load. **IEEE Transactions on Power Systems**, [s. l.], v. 38, n. 1, p. 229–241, 2023.
- SCHUBERT, E. Stop using the elbow criterion for k-means and how to choose the number of clusters instead. **ACM SIGKDD Explorations Newsletter**, [s. l.], v. 25, n. 1, p. 36–42, 2023.
- SHAMSELDEIN, M. A Fast Holomorphic Embedding Power Flow Approach for Meshed Distribution Networks. **International Transactions on Electrical Energy Systems**, [s. l.], v. 2022, p. 1–11, 2022.
- SHIRMOHAMMADI, D. *et al.* A compensation-based power flow method for weakly meshed distribution and transmission networks. **IEEE Transactions on Power Systems**, [s. l.], v. 3, n. 2, p. 753–762, 1988.
- SINGH, V.; MOGER, T.; JENA, D. Probabilistic Load Flow Approach Combining Cumulant Method and K-Means Clustering to Handle Large Fluctuations of Stochastic Variables. **IEEE Transactions on Industry Applications**, [s. l.], v. 59, n. 3, p. 2832–2841, 2023.
- SINGH, V.; MOGER, T.; JENA, D. Uncertainty handling techniques in power systems: A critical review. **Electric Power Systems Research**, [s. l.], v. 203, p. 107633, 2022.
- SUN, L. *et al.* Holomorphic Embedding Load Flow Modeling of the Three-phase Active Distribution Network. *In*: 2018 INTERNATIONAL CONFERENCE ON POWER SYSTEM TECHNOLOGY (POWERCON), 2018, Guangzhou, China. **2018 International Conference on Power System Technology (POWERCON)**. Guangzhou, China: IEEE, 2018. p. 488–495. Disponível em: <https://ieeexplore.ieee.org/document/8602144/>. Acesso em: 13 jul. 2024.
- SUN, Y. *et al.* Probabilistic load flow calculation of AC/DC hybrid system based on cumulant method. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 139, p. 107998, 2022.
- TING, T. O.; WONG, K. P.; CHUNG, C. Y. Hybrid constrained genetic algorithm/particle swarm optimisation load flow algorithm. **IET Generation, Transmission & Distribution**, [s. l.], v. 2, n. 6, p. 800, 2008.
- TSAI-HSIANG CHEN; WEN-CHIH YANG. Analysis of multi-grounded four-wire distribution systems considering the neutral grounding. **IEEE Transactions on Power Delivery**, [s. l.], v. 16, n. 4, p. 710–717, 2001.
- UNIYAL, A.; SARANGI, S. Optimal network reconfiguration and DG allocation using adaptive modified whale optimization algorithm considering probabilistic load flow. **Electric Power Systems Research**, [s. l.], v. 192, p. 106909, 2021.

USAOLA, J. Probabilistic load flow with wind production uncertainty using cumulants and Cornish–Fisher expansion. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 31, n. 9, p. 474–481, 2009.

WAN, S. An Improved k-means Algorithm based on BIC Score and Density Radius. *In: 2022 GLOBAL RELIABILITY AND PROGNOSTICS AND HEALTH MANAGEMENT (PHM-YANTAI)*, 2022, Yantai, China. **2022 Global Reliability and Prognostics and Health Management (PHM-Yantai)**. Yantai, China: IEEE, 2022. p. 1–7. Disponível em: <https://ieeexplore.ieee.org/document/9942182/>. Acesso em: 15 jul. 2024.

WANG, S. *et al.* Joint planning of active distribution networks considering renewable power uncertainty. **International Journal of Electrical Power & Energy Systems**, [s. l.], v. 110, p. 696–704, 2019.

WONG, W. C. *et al.* Quasi-Monte Carlo Based Probabilistic Small Signal Stability Analysis for Power Systems With Plug-In Electric Vehicle and Wind Power Integration. **IEEE Transactions on Power Systems**, [s. l.], v. 28, n. 3, p. 3335–3343, 2013.

WONG, K. P.; LI, A.; LAW, M. Y. Development of constrained-genetic-algorithm load-flow method. **IEE Proceedings - Generation, Transmission and Distribution**, [s. l.], v. 144, n. 2, p. 91, 1997.

YANG, G. *et al.* Optimal scheduling the wind-solar-storage hybrid generation system considering wind-solar correlation. *In: 2013 IEEE PES ASIA-PACIFIC POWER AND ENERGY ENGINEERING CONFERENCE (APPEEC)*, 2013, Kowloon, Hong Kong. **2013 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)**. Kowloon, Hong Kong: IEEE, 2013. p. 1–6. Disponível em: <http://ieeexplore.ieee.org/document/6837181/>. Acesso em: 13 jul. 2024.

YU, H. *et al.* Probabilistic Load Flow Evaluation With Hybrid Latin Hypercube Sampling and Cholesky Decomposition. **IEEE Transactions on Power Systems**, [s. l.], v. 24, n. 2, p. 661–667, 2009.

ZHU, Y.; TOMSOVIC, K. Adaptive power flow method for distribution systems with dispersed generation. **IEEE Transactions on Power Delivery**, [s. l.], v. 17, n. 3, p. 822–827, 2002.

## APÊNDICE A – NOÇÕES DE ESTATÍSTICA

Este apêndice busca complementar os temas apresentados na tese através da exposição de conceitos de estatística que foram utilizados no desenvolvimento do método proposto. Primeiramente, serão abordadas as definições de variável aleatória, funções de probabilidade e momentos estatísticos. Em seguida, maiores detalhes sobre a Expansão Gram-Charlier são mostrados. Por fim, complementos são expostos acerca da teoria de cópulas estatísticas.

### A.1 VARIÁVEIS ALEATÓRIAS, FUNÇÕES DE PROBABILIDADE E MOMENTOS ESTATÍSTICOS

Uma *variável aleatória* é aquela cujo valor depende de um resultado de um experimento. Define-se, portanto, *probabilidades* desses valores, que consistem na frequência relativa de ocorrência desses resultados. Uma variável aleatória que é caracterizada por um conjunto numerável de valores possíveis é chamada *discreta*. Já as que possuem valores incontáveis ou infinitos são chamadas *contínuas*.

A *função densidade de probabilidade* é uma função cujo valor em qualquer amostra (ou ponto) no espaço amostral (o conjunto de valores possíveis tomados pela variável aleatória) informa a probabilidade relativa de que o valor da variável aleatória seja igual a essa amostra. Já a probabilidade de a variável aleatória estar contida em um intervalo particular é dada pela integral da densidade dessa variável sobre tal intervalo. A integral da função densidade de probabilidade é chamada *função distribuição acumulada*.

Variáveis aleatórias podem ser caracterizadas de acordo com suas distribuições de probabilidade. Uma das mais comuns é a *variável aleatória normal* ou *Gaussiana*. Uma variável aleatória é dita normalmente distribuída se a sua função densidade de probabilidade é dada pela expressão (A.1), onde  $\mu$  e  $\sigma$  são, respectivamente, a média e o desvio padrão dessa variável aleatória.

$$f(x; \mu; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{A.1})$$

Outra variável aleatória muito empregada é a chamada *beta*. Ela é definida no domínio  $[0,1]$  e depende de dois parâmetros  $\alpha$  e  $\beta$ . Sua função densidade de probabilidade é dada por (A.2).

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad (\text{A.2})$$

Onde o operador  $\Gamma(\cdot)$  é dado por (A.3).

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \quad (\text{A.3})$$

Segundo Barik e Das (2019), os parâmetros  $\alpha$  e  $\beta$  podem ser calculados tendo a média e o desvio padrão da variável aleatória conhecidos. São as expressões (A.4) e (A.5).

$$\beta = (1 - \mu) \left( \frac{\mu(1 + \mu)}{\sigma^2} - 1 \right) \quad (\text{A.4})$$

$$\alpha = \frac{\mu\beta}{(1 - \mu)} \quad (\text{A.5})$$

Para um conjunto de variáveis aleatórias  $X_1, X_2, \dots, X_n$ , é possível definir uma função densidade de probabilidade relacionada ao conjunto todo. No caso de duas variáveis, tal função se chama *função densidade de probabilidade bivariada*. Para um número maior de variáveis, *função densidade de probabilidade multivariada*.

Uma das mais importantes distribuições conjuntas é a *normal bivariada* ou *Gaussiana bivariada*, dada por (A.6), considerando duas variáveis aleatórias  $X$  e  $Y$ .

$$f(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)} \quad (\text{A.6})$$

Onde:

$\mu_X, \mu_Y$  são as médias de  $X$  e  $Y$ , respectivamente.

$\sigma_X, \sigma_Y$  são os desvios padrões de  $X$  e  $Y$ , respectivamente.

$\rho$  é o coeficiente de correlação entre  $X$  e  $Y$ .

Os *momentos estatísticos* são quantidades mensuráveis que determinam o comportamento das funções densidade de probabilidade. Isto é, indicam como será o formato de seus gráficos.

O momento bruto de  $n$ -ésima ordem é dado por (A.7). O momento bruto de primeira ordem, ou seja, quando  $n = 1$ , tem particular importância e é chamado *média aritmética*, que indica o valor esperado da variável aleatória, ou tendência central (Usaola, 2009).

$$M_n = \int_{-\infty}^{\infty} x^n f(x) dx \quad (\text{A.7})$$

Onde  $f(x)$  é a função densidade de probabilidade da variável aleatória  $X$ .

O momento central de  $n$ -ésima ordem é dado por (A.8). O momento central de segunda ordem tem particular importância e é chamado *variância*, que indica a dispersão em torno da tendência central da variável aleatória (Usaola, 2009). A raiz quadrada da variância é chamada *desvio padrão*.

$$m_n = \int_{-\infty}^{\infty} (x - M_1)^n f(x) dx \quad (\text{A.8})$$

O momento central normalizado de  $n$ -ésima ordem é dado por (A.9). O momento central normalizado de terceira ordem é chamado *coeficiente de assimetria* e mensura o quanto a função densidade de probabilidade é assimétrica em relação à sua tendência central. Coeficientes de assimetria negativos indicam assimetria pela esquerda, enquanto coeficientes de assimetria positivos indicam assimetria pela direita. Já coeficientes de assimetria nulos indicam distribuições simétricas, como a Gaussiana, por exemplo.

O momento central normalizado de quarta ordem é chamado *coeficiente de curtose* e mede a deformação, ou distorção do formato da função densidade de probabilidade. De acordo com a teoria dos cumulantes, a distribuição normal possui coeficiente de curtose igual a 3 (Usaola, 2009).

$$\mu_n = \frac{M_n}{\sigma^n} \quad (\text{A.9})$$

Na presente tese, esses quatro momentos particulares são utilizados para calcular a Expansão Gram-Charlier, que será o enfoque da próxima seção.

## A.2 EXPANSÃO GRAM-CHARLIER

A Expansão Gram-Charlier (EGC) é uma série infinita que aproxima uma distribuição de probabilidade em termos de seus cumulantes (ou momentos estatísticos). Na teoria de probabilidades, chama-se *função característica* a transformação de Fourier de uma função densidade de probabilidade. Logo, a EGC se baseia em escrever a função característica da distribuição a ser aproximada em função de uma função característica conhecida e, posteriormente, recuperá-la usando a transformação de Fourier inversa.

Dessa maneira, é possível aproximar uma PDF desconhecida por uma PDF Gaussiana, dado que a função a ser recuperada analiticamente tem comportamento similar (porém não igual) a essa PDF conhecida, chamada também de *kernel*.

Logo, seja  $f$  a PDF desconhecida, e  $\hat{f}$  sua função característica, e sejam  $k_r$  seus cumulantes, para todo  $r \in \{1, \dots, \infty\}$ . Sejam também  $\psi$  e  $\hat{\psi}$  a PDF conhecida e sua função característica, respectivamente.

Da propriedade de cumulantes (Usaola, 2009), sabe-se que:

$$\hat{f}(\omega) = \exp \left\{ \sum_{r=1}^{\infty} k_r \frac{(i\omega)^r}{r!} \right\} \quad (\text{A.10})$$

De maneira análoga, pode-se escrever a função característica de  $\psi$  em função de seus cumulantes, denotados por  $\gamma_r$ .

$$\hat{\psi}(\omega) = \exp \left\{ \sum_{r=1}^{\infty} \gamma_r \frac{(i\omega)^r}{r!} \right\} \quad (\text{A.11})$$

Dividindo-se (A.10) por (A.11), obtém-se a expressão em (A.12)

$$\hat{f}(\omega) = \exp \left\{ \sum_{r=1}^{\infty} k_r \frac{(i\omega)^r}{r!} - \gamma_r \frac{(i\omega)^r}{r!} \right\} \hat{\psi}(\omega) \quad (\text{A.12})$$

Após manipulações algébricas e tomando a transformação inversa de Fourier, chega-se à formulação conhecida da EGC. Convertendo-se os cumulantes em momentos, finalmente, obtém-se a expressão em (A.13).

$$f(x) = 1 + \left[ \sum_{k=3}^{\infty} c_k(x, \boldsymbol{\mu}) H_k(x, \mu, \sigma) \right] \phi(x; \mu, \sigma) \quad (\text{A.13})$$

Onde  $x$  é o domínio da variável aleatória,  $c_k$  são os coeficientes de Gram-Charlier de ordem  $k$  (que dependem dos momentos estatísticos até a ordem  $k$ , denotados por  $\boldsymbol{\mu}$ ), e são dados por (A.14).  $H_k$  são os polinômios de Chebycheff-Hermite de ordem  $k$ , dados por (A.15) e  $\phi$  é a função densidade de probabilidade Gaussiana, de média  $\mu$  e desvio padrão  $\sigma$ .

$$c_k(x, \boldsymbol{\mu}) = \int_{-\infty}^{\infty} H_k(x, \mu, \sigma) f(x) dx \quad (\text{A.14})$$

$$H_k(x, \mu, \sigma) = (-1)^k \phi^{-1}(x, \mu, \sigma) D^k [\phi(x, \mu, \sigma)] \quad (\text{A.15})$$

Tanto os coeficientes de Gram-Charlier quanto os polinômios de Chebycheff-Hermite possuem fórmulas fixas após as manipulações algébricas oriundas das equações (A.14) e (A.15). Elas estão mostradas em (A.16) a (A.19) até a quarta ordem. Note que o coeficiente de Gram-Charlier de ordem  $k$  depende do momento estatístico de ordem  $k$ . Além disso, a ordem do polinômio de Chebycheff-Hermite também será  $k$ . Dessa forma, pode-se dizer que cada uma das parcelas da série está atrelada a um momento estatístico.

$$c_1 = 0 \quad H_1(x) = x \quad (\text{A.16})$$

$$c_2 = 0 \quad H_2(x) = x^2 - 1 \quad (\text{A.17})$$

$$c_3 = \frac{1}{3!} \mu_3 \quad H_3(x) = x^3 - 3x \quad (\text{A.18})$$

$$c_4 = \frac{1}{4!} (\mu_4 - 3) \quad H_4(x) = x^4 - 6x^2 + 3 \quad (\text{A.19})$$

Finalmente, a série truncada até a quarta ordem pode ser escrita por (A.20).

$$f(x) \cong \left[ 1 + \frac{1}{3!} \mu_3 (x^3 - 3x) + \frac{1}{4!} (\mu_4 - 3)(x^4 - 6x^2 + 3) \right] \phi(x) \quad (\text{A.20})$$

Note que, para o caso particular em que  $f(x)$  é exatamente igual a uma PDF Gaussiana, então  $\mu_3 = 0$  e  $\mu_4 = 3$ , o que leva precisamente ao resultado  $f(x) = \phi(x)$ . No caso em que esses momentos são diferentes, existe uma distorção da forma da PDF devido a eles. Nesse caso, a aproximação feita torna a representação da variável aleatória mais completa e rigorosa, pois seu comportamento não será igual a de uma variável aleatória normalmente distribuída, podendo apresentar distorções, bimodalidades e assimetrias consideráveis.

### A.3 TEORIA DE CÓPULAS ESTATÍSTICAS

Segundo Palahalli, Maffezzoni e Gruosso (2021), a cópula é a entidade ou função que contém informações de dependência das variáveis aleatórias, obtida mediante distribuições marginais conhecidas. Cópulas são usadas para modelar a dependência entre duas ou mais variáveis aleatórias, onde as distribuições marginais de variáveis aleatórias são combinadas para chegar a uma distribuição correlacionada. As relações de dependência entre as funções marginais são investigadas em um domínio comum, e uma vez obtida a distribuição conjunta, as variáveis aleatórias são posteriormente rastreadas até suas formas originais usando CDFs sem qualquer perda da informação associada (Palahalli; Maffezzoni; Gruosso, 2021).

Sejam duas variáveis aleatórias  $X$  e  $Y$ , com distribuições marginais  $f(x)$  e  $f(y)$ , cujas CDFs são denotadas por  $F_X$  e  $F_Y$ , respectivamente. A correlação existente entre as variáveis aleatórias pode ser determinada pelo coeficiente de correlação, dado por (A.21).

$$\rho_{XY} = \frac{1}{N-1} \sum_{i=1}^N \left( \frac{X_i - \mu_X}{\sigma_X} \right) \left( \frac{Y_i - \mu_Y}{\sigma_Y} \right) \quad (\text{A.21})$$

Esse valor está inserido em um intervalo de -1 a +1. Correlações são ditas *positivas* quando as variáveis aleatórias variam conjuntamente. Elas são ditas *negativas* quando seu comportamento de variação é inverso. Valor de correlação nulo indica variáveis independentes. Quanto mais próximo o coeficiente de correlação for de zero, mais “independente” serão os comportamentos das respectivas variáveis aleatórias. Nesse caso, diz-se que a correlação é *fraca*. Quanto mais próximo o coeficiente de correlação for de -1 ou

+1, mais *forte* será a dependência. Quando  $\rho_{XY} = 1$ , as variáveis variam conjuntamente de forma idêntica.

Os passos da Cópula são informados no Capítulo 2. De forma a retomá-los, recorda-se que eles são divididos em três: (i) considere uma distribuição Gaussiana multivariada com parâmetros conhecidos e matriz de correlação definida inicialmente; (ii) aplique o operador *função densidade acumulada* em cada distribuição marginal separadamente, obtendo distribuições uniformes; (iii) aplique distribuições inversas desejadas a cada uma das distribuições uniformes geradas.

O resultado é uma distribuição conjunta cujas distribuições marginais são as informadas no passo (iii) e com os coeficientes de correlação informados no passo (i).

Segue a demonstração do passo (ii). Seja uma variável aleatória denotada por  $X$  normalmente distribuída. Define-se uma variável aleatória  $Y$  tal que  $Y = \Phi_X(X)$ , onde  $\Phi_X(\cdot)$  é o operador relacionado à função distribuição acumulada da variável  $X$ . Logo, por definição, a distribuição acumulada de  $Y$ , denotada por  $\Phi_Y$ , é dada por (A.22).

$$\Phi_Y(t) = \Pr(Y \leq t) \quad (\text{A.22})$$

Mas  $Y = \Phi_X(X)$ , então:

$$\Phi_Y(t) = \Pr(\Phi_X(X) \leq t) \quad (\text{A.23})$$

Aplicando a distribuição inversa  $\Phi_X^{-1}$  em ambos os lados da inequação, obtém-se:

$$\Phi_Y(t) = \Pr(X \leq \Phi_X^{-1}(t)) \quad (\text{A.24})$$

Pois  $\Phi_X^{-1}(\Phi_X(X)) = X$ . Ora, a probabilidade de  $X$  ser menor ou igual a um valor é exatamente a distribuição acumulada aplicada a esse valor. Por conseguinte:

$$\Phi_Y(t) = \Phi_X(\Phi_X^{-1}(t)) \quad (\text{A.25})$$

Sabe-se que  $\Phi_X(\Phi_X^{-1}(t))$  é igual ao próprio argumento  $t$ . Logo:

$$\Phi_Y(t) = t \quad (\text{A.26})$$

A PDF de  $Y$  é dada pela derivada de sua CDF. Logo, derivando em relação a  $t$ :

$$f_Y(t) = 1 \quad (\text{A.27})$$

Logo, a PDF de  $Y$  é uniformemente distribuída, como se desejava mostrar inicialmente.

Segue a demonstração do passo (iii). Seja  $X$  uma variável aleatória de uma família contínua específica e definida *a priori*. Seja  $U$  uma variável aleatória uniformemente distribuída. O objetivo é determinar uma transformação  $T$ , tal que  $T(U) = X$ .

Seja  $F(x)$  a função distribuição acumulada de  $X$  aplicada em  $x$ :

$$F(x) = \Pr(X \leq x) \quad (\text{A.28})$$

Mas  $T(U) = X$ . Então:

$$F(x) = \Pr(T(U) \leq x) \quad (\text{A.29})$$

Aplicando a transformação inversa  $T^{-1}$  em ambos os lados da inequação. Obtém-se:

$$F(x) = \Pr(U \leq T^{-1}(x)) \quad (\text{A.30})$$

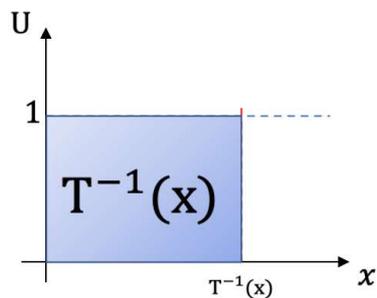
No entanto, o gráfico da distribuição uniforme aplicada a um ponto qualquer  $T^{-1}(x)$  é mostrado na Figura 55. Logo, a expressão  $\Pr(U \leq T^{-1}(x))$ , que é equivalente à CDF de  $U$  no ponto  $T^{-1}(x)$ , é igual à área de tal gráfico. Logo,  $\Pr(U \leq T^{-1}(x)) = T^{-1}(x)$ . Substituindo em (A.30), é obtido o seguinte resultado.

$$F(x) = T^{-1}(x) \quad (\text{A.31})$$

Aplicando a função acumulada inversa:

$$T(x) = F^{-1}(x) \quad (\text{A.32})$$

Figura 55 – Comportamento da distribuição uniforme.



Fonte: elaborada pelo autor (2025).

Em conclusão: a transformação desejada  $T$  será igual à distribuição inversa para a qual se quer transformar a distribuição uniforme. Por exemplo, se se deseja transformar uma variável aleatória uniformemente distribuída em uma normalmente distribuída, basta aplicar a distribuição normal inversa. Esse resultado está consoante com o passo (iii) da teoria de cópulas.

## APÊNDICE B – MÉTODOS DE FPP

No presente apêndice, serão evidenciados os desenvolvimentos matemáticos e computacionais dos métodos de FPP utilizados na tese como forma de validação dos resultados.

### B.1 MÉTODO DA ESTIMATIVA DOS PONTOS

Segundo Morales e Perez-Ruiz (2007), o Método da Estimativa dos Pontos (MEP) concentra as informações estatísticas providenciadas pelos primeiros momentos estatísticos das variáveis aleatórias de entrada, estimando pontos para cada variável, chamados *concentrações*. A função denotada por  $F$  é a que relaciona o FPD aplicado na concentração. Uma concentração compreende um par ordenado formado por uma *localização* e um *fator de peso*. A localização, denotada por  $p_l$ , consiste no ponto sobre o qual será executado o FPD. Já o fator de peso, denotado por  $\omega_l$ , contabiliza a frequência relativa e seu impacto nas variáveis de saída (Morales; Perez-Ruiz, 2007).

Na execução do MEP, são executados FPD para todas as localizações. No caso de um sistema multivariado (com mais de uma variável aleatória), cada concentração é igual ao ponto correspondente à localização calculada, com as demais variáveis mantidas em seus valores médios, denotados por  $\mu_{pi}$ , para toda variável  $i \in \{1, \dots, m\}$ , onde  $m$  é o número de variáveis aleatórias. Matematicamente, pode-se exprimir o FPD para cada concentração por (B.1).

$$Z(l, k) = F(\mu_{p1}, \mu_{p2}, \dots, p_{lk}, \dots, \mu_{pm}) \quad (\text{B.1})$$

O número de pontos a serem avaliados depende do *esquema* utilizado. Os mais comuns são os desenvolvidos por Hong (1998). São eles: 2m, 2m+1, 3m, 4m+1. O nome do esquema indica o número de pontos a serem avaliados, onde  $m$  é o número de variáveis aleatórias. Cada esquema possui um cálculo analítico distinto, cujos desenvolvimentos podem ser vistos em Morales e Perez-Ruiz (2007).

O esquema empregado na presente tese foi o chamado 2m. Nele, cada variável é aproximada por duas concentrações.

O ponto de concentração da  $l$ -ésima variável aleatória é dado por (B.2), para  $k = \{1, 2\}$ .

$$p_{lk} = \mu_{pl} + \xi_{lk}\sigma_{pl} \quad (\text{B.2})$$

Onde  $\mu_{pl}$  e  $\sigma_{pl}$  são a média e o desvio padrão da variável em questão.  $\xi_{lk}$  é chamado *localização padrão*. No caso do esquema 2m, ele é dado por (B.3).

$$\xi_{lk} = \frac{\lambda_{l3}}{2} \pm \sqrt{m + \left(\frac{\lambda_{l3}}{2}\right)^2} \quad \forall k = \{1,2\} \quad (\text{B.3})$$

Onde  $\lambda_{l3}$  é o coeficiente de assimetria da variável. Já os fatores de peso são dados por (B.4) e (B.5).

$$\omega_{l1} = -\frac{1}{m} \frac{\xi_{l2}}{\xi_{l1} - \xi_{l2}} \quad (\text{B.4})$$

$$\omega_{l2} = \frac{1}{m} \frac{\xi_{l1}}{\xi_{l1} - \xi_{l2}} \quad (\text{B.5})$$

Após o cálculo da função  $Z$  para todos os pontos de concentração, a média e o desvio padrão das variáveis de saída podem ser estimadas por (B.6) e (B.7), respectivamente.

$$M_1 = \sum_{l=1}^m \sum_{k=1}^2 \omega_{lk}(Z(l, k)) \quad (\text{B.6})$$

$$m_2 = \sum_{l=1}^m \sum_{k=1}^2 \omega_{lk}(Z(l, k))^2 \quad (\text{B.7})$$

## B.2 MÉTODO DE MONTE CARLO

O método de Monte Carlo, denotado por MCS, compreende os seguintes passos (Abud *et al.*, 2022): (i) defina o modelo computacional e as funções de probabilidade de cada variável aleatória de entrada; (ii) para cada iteração, sorteie uma amostra das funções definidas; (iii) execute numericamente o modelo computacional com a atualização da amostra; (iv) os resultados são armazenados e análises estatísticas podem ser conduzidas.

O valor esperado de uma variável aleatória de saída, denotada por  $H$ , é dada por (B.8).

$$M_1 = \frac{1}{n} \sum_{i=1}^n H(\mathbf{X}_i) \quad (\text{B.8})$$

Onde  $X_i$  é a  $i$ -ésima amostra do vetor de variáveis aleatórias de entrada denotado por  $\mathbf{X}$ , e  $n$  o número total de amostras.

Já o desvio padrão pode ser determinado por (B.9).

$$m_2 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (H(X_i) - M_1)^2} \quad (\text{B.9})$$

Quando a amostragem é conduzida de forma aleatória por meio de amostras *pseudo-randômicas*, esse método pode ser chamado de *Método de Monte Carlo Cru* (Abud *et al.*, 2022). Métodos mais sofisticados foram elaborados para aprimorar o processo de amostragem e, assim, tornando o método mais eficiente, como *Importance Sampling*, MCS via Cadeias de *Markov* e o *Quasi-MCS*. Suas definições podem ser vistas em Abud *et al.* (Abud *et al.*, 2022). Segundo os mesmos autores, uma abordagem considerada *Quasi-MCS* é a técnica LHS, que será abordada na próxima seção.

### B.3 LATIN HYPERCUBE SAMPLING

Segundo Yu *et al.* (2009), a técnica LHS pode ser definida por dois passos principais: *amostragem e permutação*.

O objetivo da amostragem é gerar cenários convenientemente representativos para cada variável aleatória de entrada. Seja  $X_1, X_2, X_3, \dots, X_K$  as  $K$  variáveis aleatórias de entrada. A função de probabilidade acumulada de cada uma delas, denotadas pelo subíndice  $k$ , é dada por (B.10).

$$Y_k = F_k(X_k) \quad (\text{B.10})$$

Dividindo o intervalo de  $Y_k$  por um número  $N$ , relacionado ao número máximo adotado de amostras, tem-se que o tamanho de cada intervalo será igual a  $1/N$ . A cada intervalo gerado é relacionada uma amostra através da escolha de um *ponto central*. Então, o valor da amostra para esse intervalo é calculado pela função de probabilidade acumulada inversa, tal como em (B.11), onde  $x_{kn}$  é a  $n$ -ésima amostra considerando a  $k$ -ésima variável aleatória.

$$x_{kn} = F_k^{-1} \left( \frac{n - 0,5}{N} \right) \quad (\text{B.11})$$

Note que a subtração  $n - 0,5$  indica uma posição intermediária do intervalo da  $n$ -ésima amostra. Os valores obtidos são direcionados a uma matriz de amostragem e esse processo é repetido até que todas as amostras sejam calculadas. Dessa forma, garante-se uma maior abrangência e mitiga pontos sobrepostos ou quase sobrepostos das amostras, o que aumenta a eficiência do método.

O processo de permutação é importante quando as variáveis do problema tendem a ser independentes. Dessa forma, as suas correlações são mitigadas e esse comportamento das amostras pode ser mais bem aproximado. Existem muitas formas de se exercer a permutação e elas são listadas em Yu *et al.* (2009). Na presente tese, o valor amostrado de cada variável de entrada é escolhido aleatoriamente para a linha correspondente na matriz de amostras, sem substituição, o que, segundo os mesmos autores, é o processo mais simples de implementação do LHS.

#### B.4 FPP VIA K-MEANS

Segundo Abud *et al.* (2022), se um método possui um tratamento de amostras via alguma técnica de redução de cenários, então ele não deve ser visto como um método de Monte Carlo tradicional ou cru. Logo, infere-se que ele deve estar acoplado em outra categoria, apesar de continuar a ser um método numérico. É o caso das técnicas de *clustering*, que são algoritmos utilizados para agrupar os dados de acordo com a finalidade de seus elementos, proporcionando, por sua vez, redução do número de cenários a serem avaliados e, conseqüentemente, do tempo computacional demandado.

De acordo com Galvani *et al.* (2019), os passos principais do FPP via *clustering* (mais especificamente, K-Means) são: (i) implementar modelo computacional e definir as variáveis

aleatórias de entrada; (ii) informar o número de *clusters*; (iii) partição do banco de dados que informa as condições das variáveis de entrada (podem ser dados reais ou amostras de funções de probabilidade); (iv) executar o problema determinístico para cada agente de *cluster*; (v) cálculos estatísticos finais.

Sadeghian *et al.* (2020) institui o cálculo de uma função de probabilidade discreta para cada agente de *cluster*, dada pela divisão do número de elementos de cada *cluster* pelo número total de elementos a serem agrupados. Com isso, pode-se estimar com mais precisão a média e o desvio padrão das variáveis aleatórias de saída.

## APÊNDICE C – ALGORITMOS DE *CLUSTERING*

Neste apêndice, são apresentadas as particularidades de cada um dos algoritmos de *clustering* utilizados na presente tese. A saber: K-Means, K-Medoids e K-Medoids CLARA.

### C.1 K-MEANS

Segundo Galvani *et al.* (2019), os métodos de *clustering* são uma das melhores abordagens para se trabalhar com dados estatísticos. Os mesmos autores definem *clustering* como o processo de separar dados em subclasses chamadas *clusters*. Dentro de um mesmo *cluster*, os dados tendem a ser similares. Já dados de *clusters* diferentes tendem a ser distintos. Diferentes critérios de similaridade podem ser empregados. O mais comum é por meio de distâncias (Galvani *et al.*, 2019).

K-Means é considerado um método particional. Elaborado por Lloyd (1982) e formalizado por MacQueen (1967) é, desde então, considerado uma das mais simples e mais populares técnicas. Os seus passos podem ser resumidos nos seguintes (Galvani *et al.*, 2019; Sadeghian *et al.*, 2020):

(Passo 1) defina um banco de dados a ser comprimido;

(Passo 2) determine o número  $K$  de *clusters*;

(Passo 3) selecione uma partição inicial com  $K$  centroides, de forma aleatória ou utilizando heurísticas específicas;

(Passo 4) relacione cada dado com um *cluster*, de acordo com os centroides mais próximos (utilizando alguma métrica pré-definida);

(Passo 5) recalcule os centroides, sendo um ponto médio de todos os elementos contidos nos *clusters* recém definidos;

(Passo 6) retorne ao Passo 4 até que a variação dos centroides seja menor do que uma tolerância pré-especificada.

Matematicamente, o algoritmo K-Means objetiva determinar os centroides, denotados por  $c_j, j \in \{1, \dots, K\}$ , tal que a distância entre todos os pontos do banco de dados original e seus respectivos centroides mais próximos seja mínima. Logo, trata-se de um problema de otimização local, altamente dependente da escolha da partição inicial, cuja função objetivo (FOB) é dada por (C.1) (Sadeghian *et al.*, 2020). Nessa equação, distâncias euclidianas quadráticas são adotadas, as mesmas utilizadas ao longo da presente tese.

$$FOB = \min \sum_{j=1}^K \sum_{i=1}^{a_j} \|x_{ij} - c_j\|^2 \quad (C.1)$$

Onde  $K$  é o número de *clusters*,  $a_j$  é o número de elementos do  $j$ -ésimo *cluster*;  $x_{ij}$  é o  $i$ -ésimo elemento do  $j$ -ésimo *cluster*.

## C.2 K-MEDOIDS

O método K-Medoids é um algoritmo de *clustering* particional similar ao K-Means. Em ambas as técnicas, o objetivo é dividir um banco de dados em  $K$  *clusters*, minimizando as distâncias entre cada elemento e seu centro de *cluster* mais próximo. A diferença primordial reside na forma de calcular os centros de *clusters*. Enquanto na abordagem K-Means, assume-se valores médios dos elementos contidos no *cluster* (chamados centroides), no K-Medoids esses centros de *clusters* são dados reais presentes no banco original, chamados *medoides* (Martins; Araujo; Penido, 2024). Analogamente ao K-Means, pode-se, então, definir uma FOB para o algoritmo K-Medoids tal como vista em (C.2).

$$FOB = \min \sum_{j=1}^K \sum_{i=1}^{a_j} \|x_{ij} - X_j\|^2 \quad (C.2)$$

Onde  $X_j$  é o  $j$ -ésimo centro de *cluster*, que consiste em um dado real presente no banco de dados original.

Segundo Martins, Araujo e Penido (2024), K-Medoids possui algumas vantagens em relação ao K-Means, principalmente no que tange sua aplicação no FPP: (i) é um algoritmo mais robusto para particionar dados de alta dimensionalidade; (ii) os medoides, por serem dados que realmente existem no banco de dados original, proporciona a avaliação do FP usando dados reais, e não médias, o que aumenta sua aplicabilidade prática; (iii) apresenta desempenho melhor do que K-Means quando os dados possuem muitos *outliers* ou ruídos.

### C.3 K-MEDOIDS CLARA

O método K-Medoids CLARA (do inglês: *Clustering for Large Applications*) é uma variação do K-Medoids clássico, baseado em uma busca exaustiva para se definir os melhores medoides, considerando subconjuntos aleatórios do banco de dados original. Os medoides que proporcionarem a menor distância entre todos os pontos e seus respectivos medoides são os escolhidos. O algoritmo engloba os seguintes passos (Martins; Araujo; Penido, 2024):

(*Passo 1*) informe o banco de dados a ser comprimido;

(*Passo 2*) informe o número  $K$  de *clusters*;

(*Passo 3*) selecione um subconjunto aleatório de pontos do banco de dados original;

(*Passo 4*) divida esse subconjunto em  $K$  *clusters*, usando uma partição inicial aleatória ou métodos mais sofisticados;

(*Passo 5*) para cada *cluster* obtido, teste cada um de seus elementos como um medoide em potencial e calcule a distância total entre os pontos do *cluster* e os medoides em potencial; os medoides definitivos serão aqueles que proporcionam as menores distâncias *intra-cluster*.

(*Passo 6*) após obter os melhores medoides, relacione os pontos do subconjunto aleatório a cada *cluster*, de acordo com os medoides mais próximos. Volte ao Passo 5 até que os medoides não se alterem mais.

(*Passo 7*) relacione os pontos restantes (os pontos que não estão contidos no subconjunto aleatório) aos *clusters* definidos no Passo 6.

(*Passo 8*) selecione um novo subconjunto aleatório e volte para o Passo 4. Isso é feito até que os medoides não mudem de uma iteração para outra, isto é, de um subconjunto aleatório para outro.

Uma vez que os resultados são altamente dependentes da partição inicial, recomenda-se simular diversas vezes o K-Medoids CLARA com várias partições iniciais diferentes (chamadas *replicantes*), tomando-se o resultado final como a partição que promove a menor distância *intra-cluster* total, o que também é uma forma de refinar a partição inicial.