



ANÁLISE COMPARATIVA DE MÉTODOS PARA ESTIMATIVA DE VALORES  
EXTREMOS DE PROCESSOS ALEATÓRIOS NÃO-GAUSSIANOS

Miguel Alfonso Calderón Ibarra

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Civil.

Orientador: Luís Volnei Sudati Sagrilo

Rio de Janeiro  
Agosto de 2017

ANÁLISE COMPARATIVA DE MÉTODOS PARA ESTIMATIVA DE VALORES  
EXTREMOS DE PROCESSOS ALEATÓRIOS NÃO-GAUSSIANOS

Miguel Alfonso Calderón Ibarra

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

Examinada por:



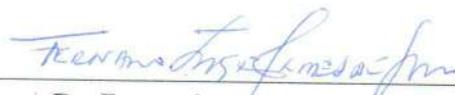
Prof. Luís Volnei Sudati Sagrilo, D.Sc.



Dr. Paulo Mauricio Videiro, Ph.D.



Dr. Allan Carré de Oliveira, D.Sc.



Dr. Fernando Jorge Mendes de Sousa, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

AGOSTO DE 2017

Ibarra, Miguel Alfonso Calderón

Análise comparativa de métodos para estimativa de valores extremos de processos aleatórios não-gaussianos / Miguel Alfonso Calderón Ibarra. – Rio de Janeiro, UFRJ / COPPE, 2017.

XVII, 106 p.: il.; 29,7 cm.

Orientador: Luís Volnei Sudati Sagrilo

Dissertação (Mestrado) – UFRJ / COPPE / Programa de Engenharia Civil, 2017.

Referências Bibliográficas: p. 86-89.

1. Processos não-gaussianos. 2. Predição de Valores Extremos. 3. Métodos de Estimativa Estatística. 4. Projeto de risers de aço. I. Sagrilo, Luís Volnei Sudati. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

*Aos meus pais, Lucha e Toño*  
*Aos meus irmãos, Toño e Nelly*  
*A minha avó, Nélide, in memoriam*

## AGRADECIMENTOS

Meu maior e mais sincero agradecimento ao Doutor Luís Volnei Sudati Sagrilo, não apenas pela sua orientação e pelos ensinamentos, mas também pela sua paciência, suas sugestões e, sobretudo, pela motivação.

Meu obrigado também ao Doutor Fernando Jorge Mendes de Sousa pela sua contribuição nesta pesquisa.

A toda a equipe do Laboratório de Análise e Confiabilidade de Estruturas Offshore (LACEO/COPPE/UFRJ), por tornarem o ambiente de trabalho agradável e produtivo.

A aquelas pessoas que são muito importantes na minha vida: meus pais, meus irmãos, Ariana Tavarez, Airem Vargas e as famílias Sánchez Jáuregui e Torres Ibarra. A todos vocês, meus maiores torcedores, muito obrigado por estar sempre comigo, pelo seu apoio e carinho.

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

## ANÁLISE COMPARATIVA DE MÉTODOS PARA ESTIMATIVA DE VALORES EXTREMOS DE PROCESSOS ALEATÓRIOS NÃO-GAUSSIANOS

Miguel Alfonso Calderón Ibarra

Agosto/2017

Orientador: Luís Volnei Sudati Sagrilo

Programa: Engenharia Civil

Este trabalho faz um estudo comparativo entre quatro métodos para estimativa de valores extremos de séries temporais não gaussianas. Os métodos estudados são: Modelo dos Polinômios de Hermite, Distribuição Lognormal Deslocada Generalizada (SGLD), Taxas Médias de Excedências Condicionadas (ACER) e Modelo da Distribuição de Weibull. Nos dois primeiros o processo é transformado em um gaussiano padrão equivalente utilizando os quatro primeiros momentos estatísticos da série temporal. Os dois últimos procedimentos operam diretamente com os picos observados na série temporal, sendo que somente o ACER leva em conta a dependência estatística entre eles.

Três estudos de caso são apresentados para avaliar o desempenho destas metodologias e a sua precisão em função do tempo de simulação. Em particular, as séries temporais utilizadas são séries do fator de utilização da seção transversal obtidas segundo a norma DnV-OS-F201 para o projeto de risers de aço.

Mostra-se que o modelo de Hermite tem sua aplicabilidade limitada somente a algumas combinações de skewness-kurtosis; a SGLD, devido à sua grande versatilidade para cobrir combinações de skewness-kurtosis, é capaz de modelar qualquer tipo de distribuição com boa precisão; os modelos ACER e Weibull estimam eficientemente os valores extremos, mesmo para tempos de simulação curtos.

Observa-se também que simulações mais longas são desejáveis para reduzir as incertezas das aproximações.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M. Sc.)

COMPARATIVE ANALYSIS OF METHODS FOR EXTREME VALUES  
ESTIMATION OF NON-GAUSSIAN RANDOM PROCESSES

Miguel Alfonso Calderón Ibarra

August/2017

Advisor: Luís Volnei Sudati Sagrilo

Department: Civil Engineering

This work makes a comparative study between four methods for extreme values estimation of non-Gaussian time series. The methods studied are: Hermite Moment Model, Shifted Generalized Lognormal Distribution (SGLD), Average Conditional Exceedance Rates (ACER) and Weibull Distribution Model. In the first two the process is transformed into an equivalent Standard Gaussian using the first four Statistical Moments of the time series. The last two procedures operate directly with the peaks observed in the time series regardless their statistical dependence, with only the ACER taking into account the statistical dependence between the peaks.

Three case studies are presented to evaluate the performance of these methodologies and their accuracy as a function of the simulation time. In particular, the time series used are Utilization Factor series on the cross-section by the DnV-OS-F201 Standard for the steel risers design.

It is shown that the Hermite Model has its applicability limited to some skewness-kurtosis combinations; the SGLD, due to its great versatility to cover the skewness-kurtosis combinations, is able to model any kind of distribution with high precision; the ACER and the Weibull models estimate efficiently the extreme values even for short simulation times.

It is also observed that longer simulations are desirable to reduce the estimated values uncertainties.

## Sumário

<b>CAPÍTULO I INTRODUÇÃO.....</b>	<b>1</b>
I.1. Contexto .....	1
I.2. Motivação.....	2
I.3. Objetivos .....	3
I.4. Organização do texto.....	4
<b>CAPÍTULO II PRINCIPAIS CONCEITOS DE ESTATÍSTICA E PROBABILIDADE .....</b>	<b>5</b>
II.1. Variáveis Aleatórias.....	5
II.2. Valores característicos das variáveis aleatórias .....	7
II.3. Equivalência Estatística: .....	9
II.4. Covariância .....	10
II.5. Coeficiente de correlação de Pearson .....	11
II.6. Distribuições de Probabilidades.....	12
II.6.1. Distribuição de Probabilidades Normal ou Gaussiana .....	12
II.6.2. Distribuição de Probabilidades Lognormal .....	14
II.6.3. Shifted Generalized Lognormal Distribution (SGLD).....	14
II.6.4. Distribuição de Probabilidades de Weibull .....	19
II.6.5. Distribuição de Probabilidades de Rayleigh.....	20
II.6.6. Distribuição de Probabilidades baseada nos Polinômios de Hermite .....	20
<b>CAPÍTULO III ANÁLISE ESTATÍSTICA DE PROCESSOS ALEATÓRIOS....</b>	<b>25</b>
III.1. Estacionariedade dos processos aleatórios .....	27
III.2. Ergodicidade de um processo aleatório .....	28
III.3. Função de auto-covariância de um processo aleatório ergódico .....	29
III.4. Densidade espectral de um processo aleatório .....	30
III.5. Fator de largura de banda .....	31
III.5.1. Processos de banda estreita e de banda larga .....	32
III.6. Distribuições de probabilidades associadas a um processo aleatório.....	33
III.6.1. Distribuição do Processo Aleatório.....	34
III.7. Frequência de cruzamento positivo no nível $y(t)=a$ .....	36

III.8. Distribuição dos Picos de um Processo Gaussiano .....	37
III.9. Estimativa dos Parâmetros Estatísticos Básicos de um Processo Aleatório a partir de uma realização .....	39
<b>CAPÍTULO IV ESTATÍSTICA DE VALORES EXTREMOS .....</b>	<b>41</b>
IV.1. Distribuições teóricas de Valores Extremos (Estatística de Ordem).....	41
IV.2. Distribuições assintóticas de valores extremos .....	44
IV.2.1. Distribuição Tipo I ou de Gumbel (para máximos) .....	45
IV.2.2. Distribuição Tipo II ou de Fréchet (para máximos).....	46
IV.3. Distribuição do Pico Extremo de um Processo Aleatório Gaussiano .....	48
<b>CAPÍTULO V MÉTODOS PARA ESTIMATIVA DE VALORES EXTREMOS DE PROCESSOS NÃO-GAUSSIANOS.....</b>	<b>49</b>
V.1. Métodos de Transformação de Processos Aleatórios .....	50
V.1.1. Método baseado nos Polinômios de Hermite .....	52
V.1.2. Método baseado na SGLD .....	53
V.2. Métodos Baseados na Distribuição dos Picos do Processo .....	53
V.2.1. Método baseado na Distribuição de Weibull .....	54
V.2.2. Método baseado na SGLD .....	56
V.2.3. Average Conditional Exceedance Rate (ACER) .....	57
V.3. Estimativa da incerteza das aproximações .....	60
<b>CAPÍTULO VI ESTUDOS DE CASOS .....</b>	<b>64</b>
VI.1. Análise Estatística das Séries Temporais dos Fatores de Utilização.....	66
VI.2. Análise Estatística dos Picos das Séries Temporais dos Fatores de Utilização ....	72
VI.3. Análise de Valores Extremos das Séries Temporais dos Fatores de Utilização – Valores de Referência.....	75
VI.4. Análise de Valores Extremos das Séries Temporais dos Fatores de Utilização – Resultados dos Métodos Estudados.....	77
<b>CAPÍTULO VII CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS .....</b>	<b>83</b>
VII.1 Conclusões .....	83
VII.2 Sugestões para trabalhos futuros.....	85

<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>86</b>
<b>ANEXO A CRITÉRIO DE PROJETO SEGUNDO A NORMA DNV LRFD .....</b>	<b>90</b>
<b>ANEXO B SOLUÇÕES APROXIMADAS PARA O SISTEMA DE EQUAÇÕES DOS POLINÔMIOS DE HERMITE.....</b>	<b>92</b>
<b>ANEXO C MÉTODO DE NEWTON-RAPHSON PARA CÁLCULO DOS PARÂMETROS (<math>r, \sigma</math>) DA SHIFTED GENERALIZED LOGNORMAL DISTRIBUTION (SGLD) .....</b>	<b>94</b>
<b>ANEXO D PARÂMETROS ESTATÍSTICOS DAS SÉRIES TEMPORAIS DE 3 HORAS.....</b>	<b>95</b>
<b>ANEXO E PARÂMETROS ESTATÍSTICOS DOS PICOS DAS SÉRIES TEMPORAIS DE 3 HORAS .....</b>	<b>101</b>

## LISTA DE FIGURAS

<i>Figura I.1 –Ações ambientais atuantes sobre uma plataforma offshore tipo Spar Buoy</i>	1
<i>Figura II.1 –Definição do valor mais provável de uma variável aleatória</i>	7
<i>Figura II.2 – Significado gráfico do coeficiente de assimetria (SAGRILO, 2011)</i>	9
<i>Figura II.3 – Duas funções cumulativas <math>F_X</math> e <math>F_Y</math>.estatisticamente equivalentes nos pontos <math>X = x</math> e <math>Y = y</math></i>	10
<i>Figura II.4 –Interpretação gráfica do significado do coeficiente de correlação (SAGRILO, 2011)</i>	11
<i>Figura II.5 –Forma Geral da Função Densidade de Probabilidades Normal ou Gaussiana</i>	13
<i>Figura II.6 – Rotação da SGLD com assimetria positiva (em vermelho) e sua equivalente com assimetria negativa (em verde).</i>	17
<i>Figura III.1 –Exemplo de n diferentes realizações de um processo aleatório <math>X(t)</math></i>	25
<i>Figura III.2 – Exemplo de uma realização de um processo aleatório estacionário</i>	27
<i>Figura III.3 – Exemplo da função de auto-covariância de um processo aleatório</i>	30
<i>Figura III.4 – Densidade espectral <math>S(\omega)</math> de um processo aleatório</i>	31
<i>Figura III.5 –Processo de banda estreita (LIMA, 2008)</i>	32
<i>Figura III.6 – Processo de banda larga (LIMA, 2008)</i>	33
<i>Figura III.7 – Densidades espectrais de processos de banda estreita e banda larga</i>	33
<i>Figura III.8 – Definição da largura <math>dt_1</math> do processo <math>y(t)</math> para a faixa entre <math>y</math> e <math>y+dy</math>.</i>	34

<i>Figura III.9 – Discretização do processo <math>y(t)</math> e a probabilidade contida entre <math>y</math> e <math>y+\Delta y</math></i> .....	34
<i>Figura III.10 – Frequência de cruzamentos positivos no nível <math>y(t)=a</math> (LIMA, 2008) ..</i>	36
<i>Figura III.11 – Picos do processo <math>x(t)</math>.....</i>	37
<i>Figura III.12 –Estabilidade da média de uma amostra em relação ao tamanho da mesma</i> .....	40
<i>Figura IV.1 – Funções densidades de probabilidades de valores extremos máximos e mínimos em relação à sua distribuição inicial.....</i>	44
<i>Figura V.1 – Região onde o modelo de Hermite é monotônico .....</i>	52
<i>Figura V.2 – Picos da série temporal <math>X(t)</math>.....</i>	54
<i>Figura V.3 – Ajuste dos picos observados a uma distribuição de Weibull.....</i>	55
<i>Figura V.4 – Exemplo dos gráficos das primeiras 6 funções ACER.....</i>	59
<i>Figura V.5 – Seleção dos valores máximos observados nas <math>N</math> realizações independentes</i> .....	62
<i>Figura VI.1 – Riser na configuração Steel Lazy Wave .....</i>	64
<i>Figura VI.2 – Estudo de Caso 1: Comparação entre a distribuição do processo ajustada a uma SGLD e uma distribuição Normal .....</i>	67
<i>Figura VI.3 – Estudo de Caso 2: Comparação entre a distribuição do processo ajustada a uma SGLD e uma distribuição Normal .....</i>	67
<i>Figura VI.4 – Estudo de Caso 3: Comparação entre a distribuição do processo ajustada a uma SGLD e uma distribuição Normal .....</i>	68

<i>Figura VI.5 – Estudo de Caso 1: Realização #2 do processo com duração de 10800 segundos (3 horas) .....</i>	69
<i>Figura VI.6 – Estudo de Caso 1: Realização #10 do processo com duração de 10800 segundos (3 horas) .....</i>	69
<i>Figura VI.7 – Estudo de Caso 1: Realização #20 do processo com duração de 10800 segundos (3 horas) .....</i>	69
<i>Figura VI.8 – Estudo de Caso 2: Realização #22 do processo com duração de 10800 segundos (3 horas) .....</i>	70
<i>Figura VI.9 – Estudo de Caso 2: Realização #35 do processo com duração de 10800 segundos (3 horas) .....</i>	70
<i>Figura VI.10 – Estudo de Caso 2: Realização #40 do processo com duração de 10800 segundos (3 horas) .....</i>	70
<i>Figura VI.11 – Estudo de Caso 3: Realização #37 do processo com duração de 10800 segundos (3 horas) .....</i>	71
<i>Figura VI.12 – Estudo de Caso 3: Realização #38 do processo com duração de 10800 segundos (3 horas) .....</i>	71
<i>Figura VI.13 – Estudo de Caso 3: Realização #50 do processo com duração de 10800 segundos (3 horas) .....</i>	71
<i>Figura VI.14 – Estudo de Caso 1: Pares de picos consecutivos para a realização #2. Correlação igual a 0.29150 .....</i>	73
<i>Figura VI.15 – Estudo de Caso 2: Pares de picos consecutivos para a realização #22. Correlação igual a 0.44566 .....</i>	73
<i>Figura VI.16 – Estudo de Caso 3: Pares de picos consecutivos para a realização #37. Correlação igual a 0.50813 .....</i>	74

<i>Figura VI.17 – Estudo de Caso 1: Ajuste dos picos máximos observados (normalizados) das 50 séries do FU a uma distribuição Tipo I</i> .....	76
<i>Figura VI.18 – Estudo de Caso 2: Ajuste dos picos máximos observados (normalizados) das 50 séries do FU a uma distribuição Tipo I</i> .....	76
<i>Figura VI.19 – Estudo de Caso 3: Ajuste dos picos máximos observados (normalizados) das 50 séries do FU a uma distribuição Tipo I</i> .....	77
<i>Figura VI.20 – Estudo de Caso 1: Estimativas do valor extremo mais provável de 3 horas (Viés/tendência do estimador)</i> .....	78
<i>Figura VI.21 – Estudo de Caso 1: Coeficientes de variação do estimador do valor extremo mais provável de 3 horas</i> .....	78
<i>Figura VI.22 – Estudo de Caso 2: Estimativas do valor extremo mais provável de 3 horas (Viés/tendência do estimador)</i> .....	79
<i>Figura VI.23 – Estudo de Caso 2: Coeficientes de variação do estimador do valor extremo mais provável de 3 horas</i> .....	79
<i>Figura VI.24 – Estudo de Caso 3: Estimativas do valor extremo mais provável de 3 horas (Viés/tendência do estimador)</i> .....	80
<i>Figura VI.25 – Estudo de Caso 3: Coeficientes de variação do estimador do valor extremo mais provável de 3 horas</i> .....	80

## LISTA DE TABELAS

Tabela 1 . Casos particulares da SGLD .....	18
Tabela 2. Resumo dos principais dados relativos ao Caso de Estudo 1 .....	65
Tabela 3. Resumo dos principais dados relativos ao Caso de Estudo 2 .....	65
Tabela 4. Resumo dos principais dados relativos ao Caso de Estudo 3 .....	65
Tabela 5. Estudo de Caso 1: Parâmetros estatísticos das 50 realizações de 3 horas .....	66
Tabela 6. Estudo de Caso 2: Parâmetros estatísticos das 50 realizações de 3 horas .....	66
Tabela 7. Estudo de Caso 3: Parâmetros estatísticos das 50 realizações de 3 horas .....	66
Tabela 8. Estudo de Caso 1: Parâmetros estatísticos dos picos das 50 realizações de 3h .....	72
Tabela 9. Estudo de Caso 2: Parâmetros estatísticos dos picos das 50 realizações de 3h .....	72
Tabela 10. Estudo de Caso 3: Parâmetros estatísticos dos picos das 50 realizações de 3h .....	72
Tabela 11. Parâmetros da distribuição do Pico Extremo (Tipo 1) para cada Estudo de Caso .....	75
Tabela 12 Estudo de Caso 1: Parâmetros estatísticos das 50 realizações de 3 horas .....	95
Tabela 13 Estudo de Caso 2: Parâmetros estatísticos das 50 realizações de 3 horas .....	97
Tabela 14 Estudo de Caso 3: Parâmetros estatísticos das 50 realizações de 3 horas .....	99
Tabela 15 Estudo de Caso 1: Parâmetros estatísticos dos Picos das 50 realizações de 3h .....	101

Tabela 16 Estudo de Caso 2:Parâmetros estatísticos dos Picos das 50 realizações de 3h  
..... 103

Tabela 17 Estudo de Caso 3:Parâmetros estatísticos dos Picos das 50 realizações de 3h  
..... 105

## NOMENCLATURA

ACER	Average Conditional Exceedance Rate (Taxa Média de Excedências Condicionadas)
BBP	Broad Banded Process (Processo de Banda Larga)
DNV	Det Norske Veritas
FEM	Finite Elements Method (Método dos Elementos Finitos)
FLS	Fatigue Limit State (Estado Limite de Fadiga)
FPSO	Floating, Production, Storage and Offloading (Unidade Flutuante de Produção, Armazenamento e Transferência)
FU	Fator de Utilização segundo o critério LRFD da norma DNV
LRFD	Load and Resistance Factor Design (Fator de projeto baseado em carregamentos e resistências)
LTR	Long Term Response (Resposta de Longo Prazo)
MPV	Most Probable Value (Valor Mais Provável)
NBP	Narrow Banded Process (Processo de Banda Estreita)
NTNU	Norges Teknisk-Naturvitenskapelige Universitet (Universidade Norueguesa de Ciência e Tecnologia)
PCF	Probability Cumulative Function (Função Cumulativa de Probabilidades)
PDF	Probability Density Function (Função Densidade de Probabilidades)
SGLD	Shifted Generalized Lognormal Distribution (Distribuição Lognormal Generalizada Deslocada)
SLS	Serviceability Limit State (Estado de Serviço de Limite)
SLWR	Steel Lazy Wave Riser (Riser de aço na configuração Lazy Wave)
STR	Short Term Response (Resposta de Curto Prazo)
TLP	Tensioned Leg Platform (Plataforma de Pernas Tracionadas)
ULS	Ultimate Limit State (Estado Limite Último)

# CAPÍTULO I

## INTRODUÇÃO

### I.1. Contexto

No projeto de plataformas petrolíferas, os efeitos causados pelos fenômenos ambientais com maior influência sobre estas estruturas variam no tempo e, por esta razão, as respostas (esforços, deslocamentos, forças, etc.) são usualmente obtidas através de simulações numéricas baseadas em análises estruturais dinâmicas. Os principais fenômenos que atuam sobre as estruturas offshore são as ondas, o vento e a corrente marinha (representados na Figura I.1) e, como os efeitos gerados por estes parâmetros são de natureza randômica, é necessário recorrer a análises dinâmicas aleatórias para verificação deste tipo de estruturas.

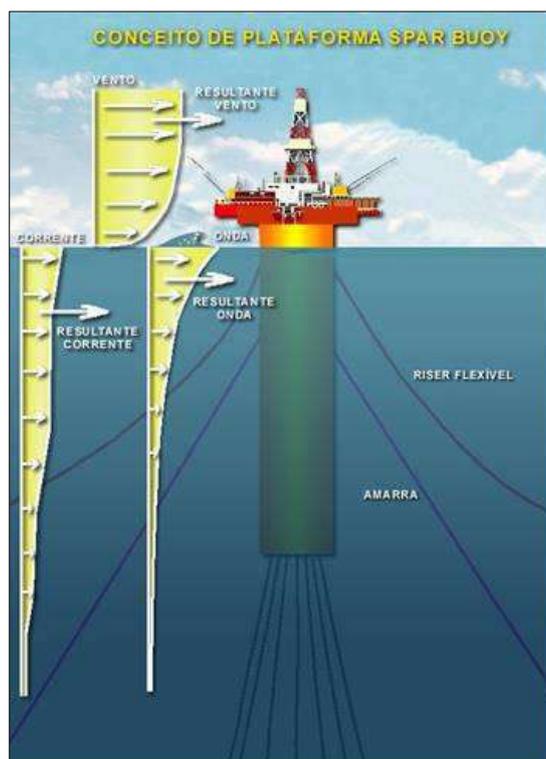


Figura I.1 –Ações ambientais atuantes sobre uma plataforma offshore tipo Spar Buoy

Assim, a avaliação de respostas de estruturas offshore pode ser feita utilizando diversas metodologias de projeto (API RP-2RD, 1998). Dentre elas, destaca-se a denominada tempestade de projeto, que pressupõe a utilização de valores extremos mais prováveis de curto-prazo (usualmente 3-h) de um dado parâmetro de resposta. Neste

contexto, existe a necessidade de se recorrer à métodos probabilísticos utilizados na análise de processos aleatórios para estimar valores extremos.

De maneira geral, os processos aleatórios podem ser divididos em dois tipos: gaussianos e não-gaussianos. A teoria sobre processos gaussianos é amplamente difundida na literatura e, neste caso, existem soluções analíticas para vários parâmetros de interesse em engenharia, incluindo os valores extremos. Entretanto, numa análise de uma estrutura offshore, devido às diversas fontes de não-linearidades, muitas vezes os parâmetros de resposta de interesse não podem ser classificados como gaussianos. Nesta situação, devido à não existência de soluções analíticas, deve-se recorrer a métodos numéricos para estimativa de valores extremos de parâmetros de resposta da estrutura.

## **I.2. Motivação**

Na análise estrutural de plataformas petrolíferas, a grande maioria das respostas dinâmicas são de caráter não-gaussiano, devido principalmente à não linearidade do sistema (por exemplo, os originados pela parcela de arrasto na equação semi-empírica de Morison, pela interação entre meio fluido e superfície livre, interação solo-estrutura, etc.). Como não existem soluções analíticas para a estimativa de valores extremos de processos aleatórios não-gaussianos, torna-se de grande relevância a identificação de metodologias numéricas que permitam estimar de forma não tendenciosa estes valores para este tipo de processo aleatório.

Atualmente existem na literatura vários métodos para estimar valores extremos de séries aleatórias não-gaussianas. A base de dados destes métodos varia. Alguns deles utilizam os parâmetros estatísticos da série temporal do processo (média, desvio padrão, coeficiente de assimetria e coeficiente de curtose) para transformá-lo em um processo aleatório gaussiano equivalente. O valor extremo para este último processo é calculado analiticamente e faz-se a transformada inversa para obter o valor corresponde do processo não-gaussiano original.

Por outro lado, existem métodos que se baseiam na distribuição de picos observados na série temporal do processo aleatório. Dentre esta última classe de métodos, boa parte dos métodos disponíveis não leva em conta a possível dependência estatística

entre picos próximos entre si, enquanto alguns poucos permitem levar em conta este aspecto na estimativa do valor extremo do processo aleatório.

### I.3. Objetivos

O principal objetivo deste trabalho é o de efetuar uma análise comparativa das distintas metodologias mais atuais propostas na literatura para estimativa de valores extremos de curto prazo de processos aleatórios não-gaussianos. Em particular, serão analisadas as séries temporais dos fatores de utilização da seção transversal de risers de aço seguindo o critério LRFD da norma DNV (DnV-OS-F201. , 2010).

Os métodos investigados neste trabalho serão:

- Modelo baseado nos Polinômios de Hermite (WINTERSTEIN, 1987).
- Modelo baseado nas Taxas Médias de Excedências Condicionadas (do inglês ACER: *Average Conditional Exceedance Rates*) (KARPA, 2015).
- Ajuste de uma Distribuição Lognormal Deslocada Generalizada (do inglês SGLD: *Shifted Generalized Lognormal Distribution*) em duas alternativas: ao processo completo e aos picos do processo (LOW , 2015).
- Ajuste de uma distribuição de Weibull de três parâmetros aos picos do processo (MADSEN & KRENK, 1986).

É interessante observar que todos estes métodos assumem que o processo aleatório é ergódico e que as séries temporais são amostradas com um certo de passo de tempo  $\Delta t$  com uma duração total  $T$ . Como esta duração é sempre finita, os parâmetros estatísticos que servem de base para os métodos de estimativa de extremos são estimados através de uma amostra também de tamanho finito e, portanto, sujeitos a variabilidades estatísticas associadas ao tamanho da amostra. Assim, neste trabalho também será investigada a variabilidade dos valores extremos estimados em função do tamanho da amostra, que por sua vez dependerá da duração da simulação.

#### **I.4. Organização do texto**

Para facilitar o entendimento dos conceitos e a terminologia usada ao longo desta dissertação, no CAPÍTULO II são apresentados aspectos teóricos sobre Probabilidade e Estatística, com ênfase nas variáveis aleatórias e as principais distribuições de probabilidades que as representam. No CAPÍTULO III são apresentadas as técnicas de tratamento dos processos aleatórios através da Probabilidade e a Estatística.

A essência principal deste trabalho encontra-se nos seguintes três capítulos. No CAPÍTULO IV é explicada a teoria para o tratamento estatístico dos Valores Extremos e no CAPÍTULO V são abordadas, de modo detalhado, as metodologias para a estimativa deste valores. Com o fim de determinar a sua eficiência e viabilidade, estes métodos são avaliados em três estudos de caso. Os detalhes e os resultados destas análises são mostrados no CAPÍTULO VI.

As principais conclusões obtidas dos casos de estudo são discutidas no CAPÍTULO VII e são registradas algumas sugestões para trabalhos futuros envolvendo os métodos avaliados e os resultados obtidos.

Finalmente são listadas as referências bibliográficas utilizadas para a elaboração deste trabalho.

Acompanha esta dissertação uma série de cinco ANEXOS que permitem aprimorar o entendimento do texto.

## CAPÍTULO II

### PRINCIPAIS CONCEITOS DE ESTATÍSTICA E PROBABILIDADE

#### II.1. Variáveis Aleatórias

Em probabilidade e estatística, uma variável aleatória (também chamada de variável randômica ou variável estocástica) é uma função contínua que relaciona números reais com um determinado valor de um espaço amostrado. Assim, para cada evento determinado está relacionada uma probabilidade, definida pela frequência de ocorrência do evento dentre uma grande quantidade de experimentos (ANG & TANG, 1984).

Em geral, quando os resultados de um experimento não são previsíveis, ele pode ser caracterizado através de variáveis aleatórias. Pelo contrário, se os resultados do experimento podem ser previsíveis, sua caracterização deve-se fazer usando variáveis determinísticas. Este trabalho aborda apenas o primeiro tipo de variáveis.

Seguindo as convenções, denominaremos as variáveis aleatórias através de letras maiúsculas (por exemplo,  $X$ ) e os valores específicos da mesma serão representados por letras minúsculas (no caso,  $x$ ). Assim, definiremos a Função Densidade de Probabilidades  $f_X(x)$  (abreviada como PDF, do inglês *Probability Density Function*) da variável contínua  $X$  aquela função que satisfaz à seguinte condição:

$$P(x \leq X \leq x + dx) = f_X(x)dx \quad (\text{II-1})$$

onde  $P(A)$  significa a probabilidade do evento  $A$  acontecer. A probabilidade da variável  $X$  assumir valores entre  $a$  e  $b$  conhecidos é dada pela expressão:

$$P(a \leq x \leq b) = \int_a^b f_X(x)dx \quad (\text{II-2})$$

Graficamente esta probabilidade corresponde à área sob a curva de  $f_X(x)$  contida no intervalo entre  $a$  e  $b$ . É possível demonstrar que toda PDF satisfaz às seguintes duas propriedades:

- a) A probabilidade de ocorrência de qualquer evento é sempre não-negativa.

$$f_X(x) \geq 0 \quad (\text{II-3})$$

b) A soma das probabilidades de todos os possíveis resultados é 1 (ou 100%).

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (\text{II-4})$$

Outra função muito utilizada no tratamento de variáveis aleatórias é a chamada Função Cumulativa de Probabilidades (abreviada como PCF, do inglês *Probability Cumulative Function*)  $F_X(x)$  definida como:

$$F_X(a) = \int_{-\infty}^a f_X(x) dx \quad (\text{II-5})$$

que significa a probabilidade da variável  $X$  assumir valores menores ou iguais a um determinado valor  $x = a$  definido. É possível demonstrar que toda Função Cumulativa de Probabilidade satisfaz às seguintes propriedades:

a) O contradomínio da função  $F_X(x)$  é o conjunto  $[0,1]$

$$0 \leq F_X(x) \leq 1 \quad (\text{II-6})$$

b) A função  $F_X(x)$  tem assíntotas horizontais em 0 e em 1.

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad (\text{II-7})$$

$$\lim_{x \rightarrow \infty} F_X(x) = 1 \quad (\text{II-8})$$

A partir da inversão da Eq (II-5) podem-se relacionar a Função Densidade de Probabilidade com sua respectiva Função Cumulativa de Probabilidade usando a seguinte equação:

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (\text{II-9})$$

Assim, uma distribuição de probabilidades possui sempre uma função densidade de probabilidades  $f_X(x)$  e sua correspondente função cumulativa de probabilidades  $F_X(x)$ .

## II.2. Valores característicos das variáveis aleatórias

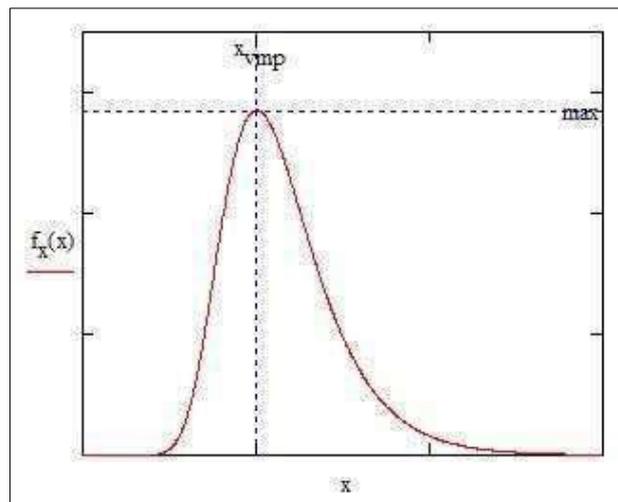
No tratamento de variáveis aleatórias, existem alguns valores característicos que permitem conhecer informações quantitativas sobre a localização, dispersão, forma e tendência da sua distribuição de probabilidades. Primeiramente são apresentadas as medidas de localização: média, moda e mediana.

A média, valor médio ou valor esperado, representada por  $\mu_X$  ou  $E(x)$ , é definida como:

$$\mu_X = E(x) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (\text{II-10})$$

A moda ou valor mais provável (abreviado VMP) de uma variável aleatória contínua é um ponto  $X = x_{vmp}$  em que função densidade de probabilidades  $f_X(x)$  atinge o valor máximo, como mostrado na *Figura II.1*. Assim, o VMP pode ser calculado como aquele valor de  $X$  que satisfaz à seguinte equação:

$$\frac{df_X(x_{vmp})}{dx} = 0 \quad (\text{II-11})$$



*Figura II.1 – Definição do valor mais provável de uma variável aleatória*

A seguir são apresentadas as medidas de dispersão de uma variável aleatória contínua  $X$ : valor médio quadrático, variância, desvio padrão e coeficiente de variação.

O valor médio quadrático, representado por  $E(X^2)$ , é definido como:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx \quad (\text{II-12})$$

A variância, representada por  $\sigma_X^2$  ou  $Var(X)$ , é definida como:

$$\begin{aligned} \sigma_X^2 = Var(X) &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2\mu_X \int_{-\infty}^{\infty} x f_X(x) dx + \mu_X^2 \int_{-\infty}^{\infty} f_X(x) dx \\ &= E(X^2) - \mu_X^2 \end{aligned} \quad (\text{II-13})$$

O desvio padrão, representado por  $\sigma_X$ , é definido como:

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{Var(x)} \quad (\text{II-14})$$

A variância e o desvio padrão são medidas da dispersão dos valores da variável aleatória em relação à sua média. Valores grandes da variância e do desvio padrão indicam que a variável apresenta grande dispersão em torno da média, enquanto que valores pequenos destes parâmetros indicam que os valores da variável encontram-se concentrados próximos da média

O coeficiente de variação, representado por  $CoV_X$  ou  $\delta_X$ , é uma medida adimensional de dispersão de  $X$  definido como a relação entre o desvio padrão e a média:

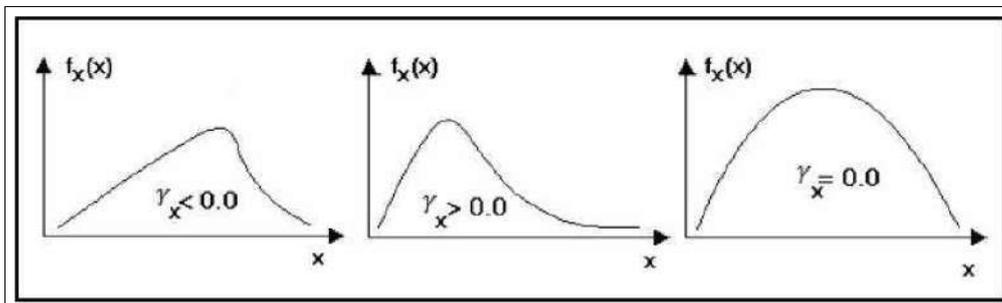
$$CoV_X = \delta_X = \frac{\sigma_X}{\mu_X} \quad (\text{II-15})$$

Além destes valores característicos de uma variável aleatória apresentados anteriormente, existem outros relacionados à forma da distribuição que são os coeficientes de assimetria e de curtose.

O coeficiente de assimetria ou coeficiente de skewness, representado por  $\gamma_X$ , é definido como:

$$\gamma_X = \frac{E(X - \mu_X)^3}{\sigma_X^3} = \frac{\int_{-\infty}^{\infty} (x - \mu_X)^3 f_X(x) dx}{\sigma_X^3} \quad (\text{II-16})$$

O coeficiente de assimetria, como o próprio nome indica, é um parâmetro relacionado à simetria da distribuição em relação à sua média. Um valor igual a zero significa que  $f_X(x)$  é simétrica. Caso este parâmetro seja positivo, isto indica que a cauda da distribuição é mais longa na direção dos valores maiores que a sua média. No caso contrário, se ele for negativo, a cauda da distribuição é mais longa na direção dos valores menores que sua média. Estes valores são indicados na *Figura II.2*.



*Figura II.2 – Significado gráfico do coeficiente de assimetria (SAGRILO, 2011)*

O coeficiente de curtose ou coeficiente de kurtosis, representado por  $\kappa_X$ , é definido como:

$$\kappa_X = \frac{E(X - \mu_X)^4}{\sigma_X^4} = \frac{\int_{-\infty}^{\infty} (x - \mu_X)^4 f_X(x) dx}{\sigma_X^4} \quad (\text{II-17})$$

O coeficiente de curtose é uma medida de esbeltez de  $f_X(x)$  e é usado para comparar uma dada distribuição de probabilidades em relação à distribuição normal, cujo valor teórico para este parâmetro é 3. Uma distribuição com  $\gamma_X \geq 3$  apresenta uma tendência de possuir valores extremos maiores do que uma distribuição normal com a mesma média e desvio padrão.

### **II.3. Equivalência Estatística:**

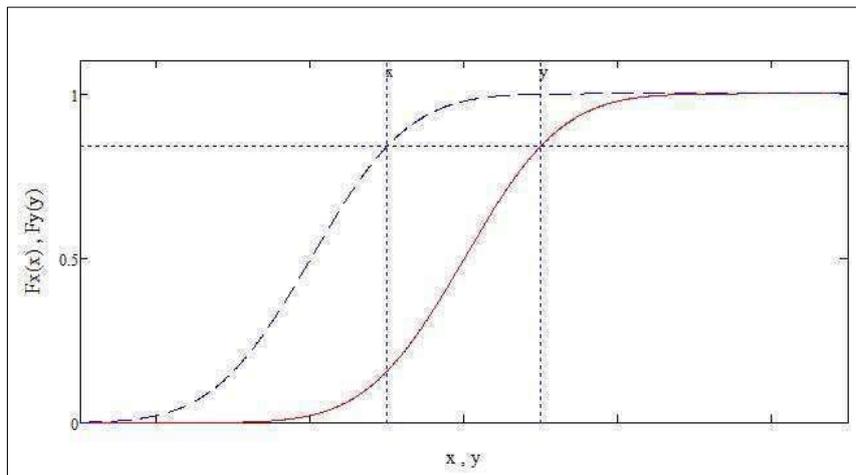
Sendo duas variáveis aleatórias  $X$  e  $Y$ , elas são estatisticamente equivalentes nos pontos  $X = x$  e  $Y = y$ , se as suas funções cumulativas de probabilidades avaliadas nesses pontos também são equivalentes, o que matematicamente pode ser representado através das seguintes igualdades:

$$F_X(x) = F_Y(y) \quad (\text{II-18})$$

$$x = F_X^{-1}(F_Y(y)) \quad (\text{II-19})$$

$$y = F_Y^{-1}(F_X(x)) \quad (\text{II-20})$$

onde  $F_X^{-1}(\cdot)$  e  $F_Y^{-1}(\cdot)$  correspondem às inversas das funções cumulativas de probabilidades das distribuições das variáveis  $X$  e  $Y$ , respectivamente. Esta equivalência é ilustrada na *Figura II.3*



*Figura II.3 – Duas funções cumulativas  $F_X(\cdot)$  e  $F_Y(\cdot)$  estatisticamente equivalentes nos pontos  $X = x$  e  $Y = y$*

Este conceito é muito utilizado já que permite a transformação de uma variável em outra ou a mudança de domínios e contradomínios dos espaços em que elas estão definidas.

#### **II.4. Covariância**

A covariância entre duas variáveis aleatórias  $X$  e  $Y$ , representada por  $COV(X, Y)$ , é definida pela seguinte equação:

$$COV(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = COV(Y, X) \quad (\text{II-21})$$

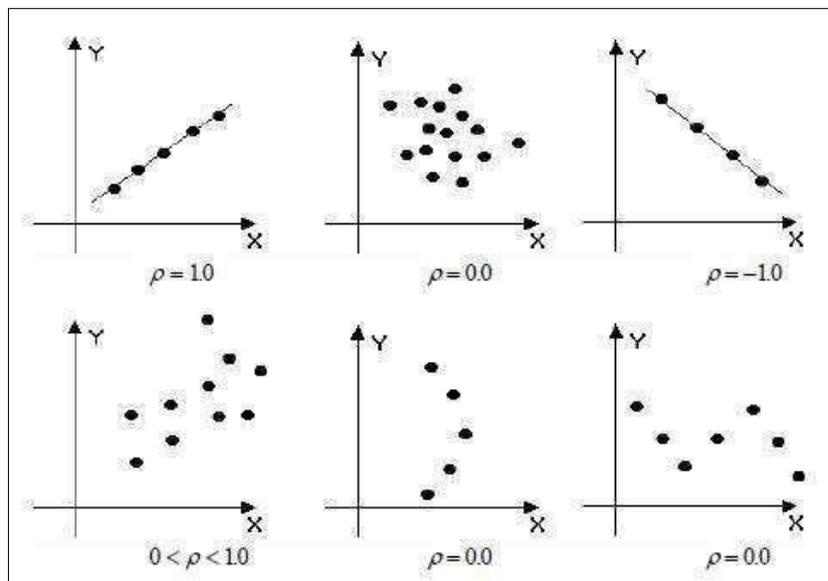
A covariância é uma medida da dependência estatística linear entre as duas variáveis. Se  $X$  e  $Y$  são independentes, então, a sua covariância é nula.

## II.5. Coeficiente de correlação de Pearson

O coeficiente de correlação de Pearson (ou simplesmente coeficiente de correlação) entre as duas variáveis  $X$  e  $Y$  é um coeficiente adimensional, representado por  $\rho_{X,Y}$ , que mede a relação (ou dependência) linear entre os valores das duas variáveis aleatórias. Ele é calculado por meio da seguinte equação:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \rho_{Y,X} \quad (\text{II-22})$$

O coeficiente de correlação pode assumir valores entre  $[-1,1]$ , sendo o sinal um indicativo do sentido da relação. Quando  $\rho_{Y,X} > 0$ , indica que quando  $X$  aumenta,  $Y$  também aumenta. Pelo contrário, quando  $\rho_{Y,X} < 0$ , indica que quando  $X$  aumenta,  $Y$  diminui. Quando  $\rho_{X,Y} = \pm 1$ , indica uma dependência total e direta entre as duas variáveis. Caso  $\rho_{X,Y} = 0$ , não existe uma relação linear entre  $X$  e  $Y$ . A *Figura II.4*, mostra graficamente exemplos do valor do coeficiente de correlação para diferentes conjuntos de valores das variáveis  $X$  e  $Y$ .



*Figura II.4 – Interpretação gráfica do significado do coeficiente de correlação (SAGRILO, 2011)*

## II.6. Distribuições de Probabilidades

Em geral, qualquer função contínua capaz de satisfazer às Eqs. (II-1) até (II-9) pode ser utilizada para o cálculo de probabilidades, porém, na bibliografia, existe uma extensa quantidade de funções previamente definidas. Neste item serão apresentadas apenas algumas delas, com interesse nas mais utilizadas na área de engenharia de estruturas offshore.

A grande variabilidade das distribuições de probabilidades permite inferir que, através delas, é possível modelar o comportamento de um grande número de variáveis aleatórias. Por este motivo, o grande desafio do engenheiro é fazer a escolha certa da distribuição de probabilidades que seja mais adequada para a representação do fenômeno (ou em ocasiões uma parcela do fenômeno) que esteja sendo estudado.

A principal vantagem de utilizar as distribuições disponíveis na literatura é que, além de permitir a universalização das pesquisas, muitas delas representam satisfatoriamente fenômenos comumente estudados e para elas existem propriedades previamente conhecidas ou soluções analíticas para determinados problemas.

### II.6.1. Distribuição de Probabilidades Normal ou Gaussiana

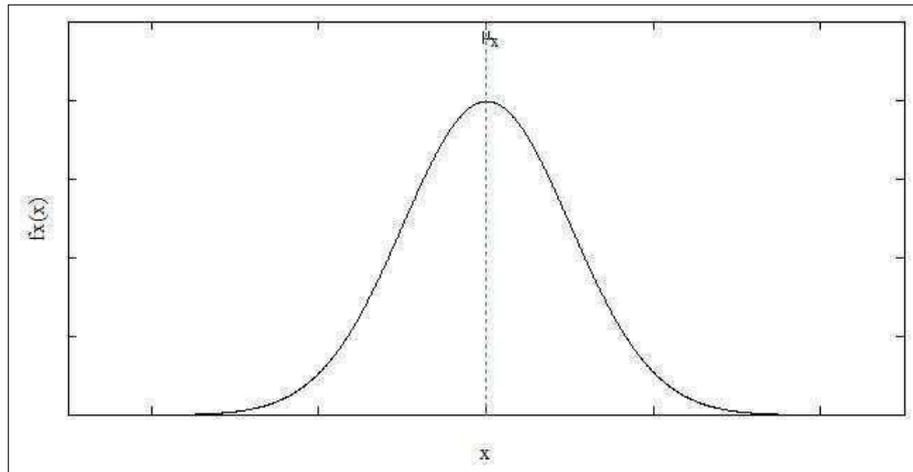
Uma variável  $X$  que segue uma distribuição de probabilidades normal ou gaussiana é conhecida comumente como variável normal ou variável gaussiana, e sua função densidade de probabilidades é dada pela seguinte equação:

$$f_X(x) = \frac{1}{\sqrt{2\pi} \sigma_X} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu_X}{\sigma_X} \right)^2 \right] \quad (\text{II-23})$$

Pode ser observado que esta distribuição depende apenas de dois parâmetros  $\mu_X$  e  $\sigma_X$ , que correspondem, respectivamente, à média e ao desvio padrão da variável  $X$ . Frequentemente se faz referência a esta distribuição utilizando a notação  $N(\mu_X, \sigma_X)$ .

Além disso, deve ser mencionado que não existe uma função analítica para calcular a sua função cumulativa de probabilidades, representada por  $\Phi(x)$ , o que deve ser calculado utilizando métodos numéricos ou tabelas disponíveis na literatura.

A curva da função densidade de probabilidades da distribuição gaussiana definida pela Eq. (II-23) possui duas características particulares: a primeira é que ela é simétrica em relação à sua média, e a segunda é que a forma da curva na região próxima ao valor mais provável é sempre a mesma, como mostra a *Figura II.5*. Estas características se traduzem em um coeficiente de assimetria igual a zero e um coeficiente de curtose igual a três.



*Figura II.5 – Forma Geral da Função Densidade de Probabilidades Normal ou Gaussiana*

Um caso particular desta distribuição é a denominada Distribuição Normal Padrão, que é obtida através da introdução de uma variável auxiliar  $Y$  (também chamada de variável reduzida) dada pela seguinte relação:

$$Y = \frac{X - \mu_X}{\sigma_X} \quad (\text{II-24})$$

Ao se fazer esta transformação da variável gaussiana  $X$  na variável reduzida  $Y$ , é possível demonstrar que a nova variável continua sendo gaussiana e possui média nula e desvio padrão unitário, tendo a sua função densidade de probabilidades reduzida para a seguinte equação:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right] \quad (\text{II-25})$$

Seguindo a convenção, serão utilizadas as notações  $\Phi(y)$  para se referir a função cumulativa de probabilidades de uma variável normal padrão avaliada em  $Y = y$ ,  $\Phi^{-1}(y)$

para se referir a função inversa da anterior, e  $N(0,1)$  para indicar que a variável segue uma Distribuição Normal Padrão.

### II.6.2. Distribuição de Probabilidades Lognormal

Uma variável  $X$  segue uma distribuição de probabilidades lognormal quando a variável definida pelo seu logaritmo natural ( $\ln(x)$ ) é gaussiana. Assim, a sua função densidade de probabilidades é dada pela seguinte equação:

$$f_X(x) = \frac{1}{\sqrt{2\pi} x \xi} \exp \left[ -\frac{1}{2} \left( \frac{\ln(x) - \lambda}{\xi} \right)^2 \right] \quad (\text{II-26})$$

Neste caso, a distribuição depende também de dois parâmetros  $\lambda$  e  $\xi$ . Frequentemente se faz referência a esta distribuição utilizando a notação  $LN(\lambda, \xi)$ . Estes parâmetros guardam as seguintes relações com a média e o desvio padrão de  $X$ :

$$\mu_X = \exp \left[ \lambda + \frac{1}{2} \xi^2 \right] \quad (\text{II-27})$$

$$\sigma_X = \mu_X \sqrt{\exp[\xi^2] - 1} \quad (\text{II-28})$$

### II.6.3. Shifted Generalized Lognormal Distribution (SGLD)

LOW (2012) apresentou uma generalização da distribuição de probabilidades lognormal que foi definida através de quatro parâmetros. Esta função é chamada de “Distribuição Lognormal Generalizada Deslocada”, originalmente batizada como *Shifted Generalized Lognormal Distribution* (abreviada como SGLD), e tem suas PDF e CDF definidas através das seguintes equações:

$$f_X(x) = \frac{\alpha}{x-b} \exp \left( -\frac{1}{r\sigma^r} \left| \ln \left( \frac{x-b}{\theta} \right) \right|^r \right), \quad x > b \quad (\text{II-29})$$

$$F_X(x) = \frac{1}{2} + \frac{1}{2} \operatorname{sgn} \left( \frac{x-b}{\theta} - 1 \right) g \left( \frac{1}{r}, \frac{\left| \ln \left( \frac{x-b}{\theta} \right) \right|^r}{r} \right), \quad x > b \quad (\text{II-30})$$

onde:  $b$  é o parâmetro de locação,  $\theta$  é o parâmetro de escala,  $r > 0$  e  $\sigma > 0$  são os dois parâmetros de forma e  $sgn(\cdot)$  é a função sinal. O parâmetro  $\alpha$  é definido como:

$$\alpha = \frac{1}{2r^{1/r}\sigma\Gamma\left(1 + \frac{1}{r}\right)} \quad (\text{II-31})$$

e a função  $g(v, x)$  é a Função Gamma Incompleta dada por:

$$g(v, x) = \frac{\int_0^x t^{v-1}e^{-t} dt}{\Gamma(v)} \quad (\text{II-32})$$

Para o cálculo dos momentos estatísticos desta distribuição faz-se uso de uma variável auxiliar  $Y$  definida como (LOW , 2012):

$$Y = \frac{X - b}{\theta} \quad (\text{II-33})$$

Não é difícil demonstrar que as seguintes relações entre os quatro primeiros momentos das variáveis  $X$  e  $Y$ :

$$\mu_Y = \frac{\mu_X - b}{\theta} \quad (\text{II-34})$$

$$\sigma_Y = \frac{\sigma_X}{\theta} \quad (\text{II-35})$$

$$\gamma_Y = \gamma_X \quad (\text{II-36})$$

$$\kappa_Y = \kappa_X \quad (\text{II-37})$$

Adicionalmente, Low (2012) mostra que os quatro primeiros momentos de  $Y$  podem ser obtidos por:

$$\mu_Y = \mu_y(r, \sigma) = E[Y^1] \quad (\text{II-38})$$

$$\sigma_Y = \sigma_y(r, \sigma) = \sqrt{E[Y^2] - (E[Y^1])^2} \quad (\text{II-39})$$

$$\gamma_Y = \gamma_y(r, \sigma) = \frac{E[Y^3] - 3(\mu_y(r, \sigma))(\sigma_y(r, \sigma))^2 - (\mu_y(r, \sigma))^3}{(\sigma_y(r, \sigma))^3} \quad (\text{II-40})$$

$$\kappa_Y = \kappa_y(r, \sigma) = \frac{E[Y^4] - 4\mu_y(r, \sigma)E[Y^3] + 6(\mu_y(r, \sigma))^2(\sigma_y(r, \sigma))^2 + 3(\mu_y(r, \sigma))^4}{(\sigma_y(r, \sigma))^4} \quad (\text{II-41})$$

onde:

$$E[Y^k] = \frac{1}{\Gamma\left(\frac{1}{r}\right)} \sum_{n=0}^{\infty} \left[ \frac{(k\sigma)^{2n}}{(2n)!} r^{2n/r} \Gamma\left(\frac{2n+1}{r}\right) \right] ; k = 1, 2, 3, 4 \quad (\text{II-42})$$

Utilizando o método dos momentos os quatro parâmetros da SGLD para representar a variável  $X$  podem ser obtidos, conforme descrito a seguir. Lembrando que  $\sigma$  e  $r$  controlam diretamente a assimetria e a curtose da distribuição, selecionamos o seguinte sistema para obter os valores dos parâmetros  $r$  e  $\sigma$ :

$$\gamma_x = \gamma_y(r, \sigma) \quad (\text{II-43})$$

$$\kappa_x = \kappa_y(r, \sigma) \quad (\text{II-44})$$

A solução deste sistema é feita através de métodos numéricos. O ANEXO B mostra um procedimento de resolução baseado no método de Newton-Raphson.

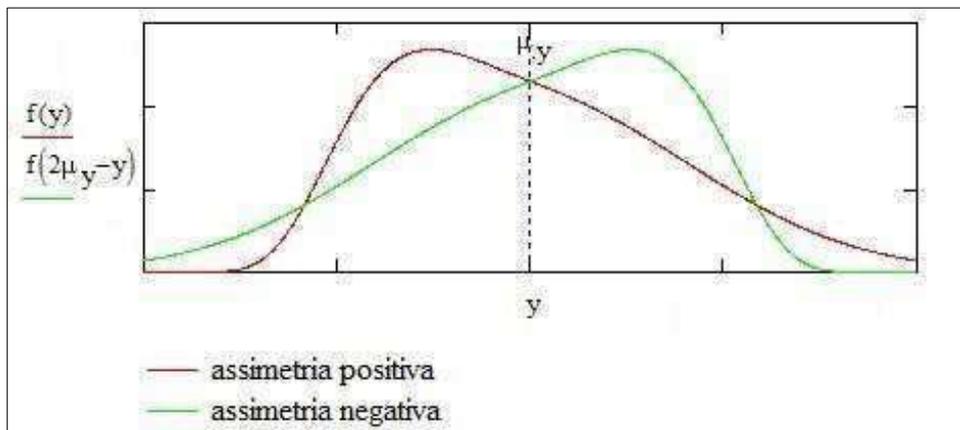
Uma vez calculados os valores de  $r$  e  $\sigma$ , para encontrar os parâmetros  $\theta$  e  $b$  resolvemos as seguintes equações:

$$\theta = \frac{\sigma_X}{\sigma_Y(r, \sigma)} \quad (II-45)$$

$$b = \mu_X - \theta \mu_Y(r, \sigma) \quad (II-46)$$

Conforme mencionado por Low (2012), é importante ressaltar que a determinação dos quatro parâmetros da SGLD conforme descrito acima somente é válida para coeficientes de assimetria positivos, porém isto não representa uma limitação da metodologia.

Em casos de distribuições com assimetria negativa ( $\gamma_X < 0$ ) devem se calcular primeiro os parâmetros considerando o valor absoluto do coeficiente de skewness (positivo) e obtêm-se a sua respectiva função densidade de probabilidades (Eq. (II-29)) com assimetria positiva. Posteriormente, para obter a PDF com assimetria negativa, a função obtida pela Eq. (II-29) é rebatida em relação à média, como mostra a *Figura II.6*. Matematicamente isto se faz substituindo  $x = 2\mu_X - x$  e calculando-se a  $f_X(2\mu_X - x)$ . Esta nova função tem assimetria negativa, mantendo as mesmas média, desvio padrão e coeficiente de curtose.



*Figura II.6 – Rotação da SGLD com assimetria positiva (em vermelho) e sua equivalente com assimetria negativa (em verde).*

Já para o caso de distribuições simétricas ( $\gamma_X = 0$ ), a SGLD recai na distribuição chamada Distribuição Exponencial de Potência (*Exponential Power Distribution*), cuja função densidade de probabilidades é definida por:

$$f_x(x) = \alpha \exp\left(-\frac{1}{r\sigma^r}|x - \mu|^r\right), -\infty < x < \infty \quad (\text{II-47})$$

Na SGLD o papel principal do parâmetro  $r$  é controlar o coeficiente de curtose, enquanto que o do parâmetro  $\sigma$  é controlar o coeficiente de assimetria. Na prática, esta distribuição tem uma capacidade de atuar num espaço bastante amplo de combinações de assimetria e de curtose, fato que atribui a ela uma grande flexibilidade de uso.

Outra particularidade da SGLD, que permite entender a sua grande versatilidade, é que através da manipulação dos seus parâmetros, ela consegue ser equivalente a diversos tipos de distribuições existentes na literatura. A Tabela 2 apresenta alguns casos particulares (ou situações limites) da SGLD.

*Tabela 1 . Casos particulares da SGLD*

<b>PDFs simétricas (<math>\sigma = 0</math>)</b>	<b>PDFs assimétricas (<math>\sigma &gt; 0</math>)</b>
Distribuição Normal ( $r = 2$ )	Distribuição Lognormal ( $r = 2, b = 0$ )
Distribuição de Laplace ( $r = 1$ )	Distribuição Log-Laplace ( $r = 1, b = 0$ )
Distribuição Uniforme ( $r \rightarrow \infty$ )	Distribuição Log-Uniforme ( $r \rightarrow \infty, b = 0$ )

Adicionalmente, conforme será visto mais adiante, é útil conhecer que a função cumulativa de probabilidades inversa da SGLD que para uma  $p \neq 1/2$  é dada por:

$$x = F_x^{-1}(p) = b + \theta \exp\left(\operatorname{sgn}\left(p - \frac{1}{2}\right) \sigma \left( r g^{-1}\left(\frac{1}{r}, \frac{2p - 1}{\operatorname{sgn}\left(p - \frac{1}{2}\right)}\right)\right)^{1/r}\right) \quad (\text{II-48})$$

onde  $p$  é valor da probabilidade acumulada,  $g^{-1}(\cdot, \cdot)$  é a Função Gamma Incompleta Inversa que se deriva da Eq. (II-32) tal que  $z = g^{-1}(s, w)$  corresponde ao valor  $w = g(s, z)$ . Já para  $p = 1/2$  a Eq. (II-48) se reduz a  $x = b + \theta$ .

## II.6.4. Distribuição de Probabilidades de Weibull

Uma variável  $X$  segue uma distribuição de probabilidades de Weibull quando as suas funções densidade e cumulativa de probabilidades são dadas pelas seguintes equações, para  $x \geq u$ :

$$f_X(x) = \frac{(x-u)^{\lambda-1}}{\alpha^\lambda} \lambda \exp \left[ - \left( \frac{x-u}{\alpha} \right)^\lambda \right] \quad (\text{II-49})$$

$$F_X(x) = 1 - \exp \left[ - \left( \frac{x-u}{\alpha} \right)^\lambda \right] \quad (\text{II-50})$$

Pode ser observado que esta distribuição depende de três parâmetros:  $u$  é o parâmetro de locação,  $\alpha$  é o parâmetro de escala e  $\lambda$  é o parâmetro de forma. No presente trabalho será utilizada a notação W3P para se referir a esta distribuição.

Os três parâmetros da distribuição de Weibull estão diretamente relacionados com os valores característicos da variável aleatória  $X$ :  $\mu_X$ ,  $\sigma_X$ ,  $\gamma_X$  e  $\kappa_X$  através das seguintes igualdades:

$$\mu_X = u + \alpha \Gamma \left[ 1 + \frac{1}{\lambda} \right] \quad (\text{II-51})$$

$$\sigma_X = \alpha \sqrt{\Gamma \left[ 1 + \frac{2}{\lambda} \right] - \Gamma^2 \left[ 1 + \frac{1}{\lambda} \right]} \quad (\text{II-52})$$

$$\gamma_X = \frac{\Gamma \left[ 1 + \frac{3}{\lambda} \right] - 3\Gamma \left[ 1 + \frac{1}{\lambda} \right] \Gamma \left[ 1 + \frac{2}{\lambda} \right] + 2\Gamma^3 \left[ 1 + \frac{1}{\lambda} \right]}{\left( \Gamma \left[ 1 + \frac{2}{\lambda} \right] - \Gamma^2 \left[ 1 + \frac{1}{\lambda} \right] \right)^{3/2}} \quad (\text{II-53})$$

$$\kappa_X = \frac{\Gamma \left[ 1 + \frac{4}{\lambda} \right] - 4\Gamma \left[ 1 + \frac{1}{\lambda} \right] \Gamma \left[ 1 + \frac{3}{\lambda} \right] + 6\Gamma^2 \left[ 1 + \frac{1}{\lambda} \right] \Gamma \left[ 1 + \frac{2}{\lambda} \right] - 3\Gamma^4 \left[ 1 + \frac{1}{\lambda} \right]}{\left( \Gamma \left[ 1 + \frac{2}{\lambda} \right] - \Gamma^2 \left[ 1 + \frac{1}{\lambda} \right] \right)^2} \quad (\text{II-54})$$

Em algumas ocasiões, o parâmetro de locação é considerado nulo ( $u = 0$ ), o que reduz a dependência desta distribuição a apenas dois parâmetros ( $\lambda$  e  $\alpha$ ). Esta forma da distribuição será novamente tratada no itens V.1.2. e V.2.2.

### II.6.5. Distribuição de Probabilidades de Rayleigh

Um caso particular da distribuição de Weibull é a chamada distribuição de Rayleigh, que é obtida quando seus três parâmetros adquirem os seguintes valores:  $u = 0$ ,  $\lambda = 2$  e  $\alpha = \sqrt{2}\alpha_R$ . As suas funções densidade de probabilidades e cumulativa de probabilidades são definidas apenas para valores de  $x \geq 0$ , através das seguintes equações:

$$f_X(x) = \frac{x}{\alpha_R^2} \exp\left[-\frac{1}{2}\left(\frac{x}{\alpha_R}\right)^2\right] \quad (\text{II-55})$$

$$F_X(x) = 1 - \exp\left[-\frac{1}{2}\left(\frac{x}{\alpha_R}\right)^2\right] \quad (\text{II-56})$$

A distribuição de Rayleigh depende apenas do parâmetro de escala  $\alpha_R > 0$ .

### II.6.6. Distribuição de Probabilidades baseada nos Polinômios de Hermite

Utilizando os conceitos de equivalência estatística e da variável normal padrão, WINTERSTEIN (1987) apresentou uma nova distribuição de probabilidades baseada nos Polinômios de Hermite, como será explicado a seguir.

Supondo uma variável aleatória  $X$  qualquer, e seus correspondentes valores característicos:  $\mu_X, \sigma_X, \gamma_X$  e  $\kappa_X$ . Definindo uma variável reduzida  $Y$ , através da Eq. (II-24), a mesma terá os seguintes valores característicos:  $\mu_Y = 0, \sigma_Y = 1, \gamma_Y = \gamma_X$  e  $\kappa_Y = \kappa_X$

Uma nova variável aleatória  $U$  normal padrão é introduzida, de maneira tal que ela seja estatisticamente equivalente à variável reduzida, e pela definição dada na Eq. (II-18) podemos escrever:

$$F_Y(y) = F_U(u) = \Phi(u) \quad (\text{II-57})$$

$$y = F_Y^{-1}(F_X(x)) = F_Y^{-1}(\Phi(u)) = g(u) \quad (\text{II-58})$$

A Eq. (II-58) representa uma relação matemática não linear entre as variáveis  $Y$  e  $U$ . Identificamos esta relação como uma função  $g(u)$ . A princípio, esta função é desconhecida, porém permite inferir que ela é capaz de transformar variáveis normais padrão  $U$  em variáveis reduzidas  $Y$  estatisticamente equivalentes.

No método proposto por WINTERSTEIN(1987), a função  $g(u)$  é aproximada pelos polinômios de Hermite, abreviados como  $He_n$ . Assim, sendo  $n$  o grau do polinômio, ele pode ser definido pela seguinte equação:

$$He_n(u) = \exp\left(\frac{1}{2}u^2\right) (-1)^n \frac{d^n}{du^n} \left[ \exp\left(-\frac{1}{2}u^2\right) \right] \quad (\text{II-59})$$

De tal modo, chega-se a uma relação entre  $U$  e  $Y$ , dada pela seguinte expansão em série polinomial:

$$y = k \left[ u + \sum_{n=1}^N c_n He_{n-1}(u) \right] \quad (\text{II-60})$$

Nesta soma:  $N$  é o número de termos considerados,  $k$  é um fator de escala que permite controlar a variância de  $Y$  e  $c_n$  são seus respectivos coeficientes a serem calculados.

Na intenção de viabilizar uma solução não trivial, a série infinita dada pela Eq. (II-60) deve ser truncada num  $N$  definido. Na prática, adota-se uma variável reduzida, de maneira que a variância de  $Y$  seja unitária e, conseqüentemente, se faz com que  $k = 1$ . Por outro lado, para garantir a compatibilidade dos quatro primeiros momentos estatísticos adota-se  $N = 4$ .

Desta forma, os quatro primeiros termos dos polinômios de Hermite (isto é,  $n = 1, 2, 3, 4$ ) são dados pelas seguintes equações:

$$He_0(u) = 1 \quad (\text{II-61})$$

$$He_1(u) = u \quad (\text{II-62})$$

$$He_2(u) = u^2 - 1 \quad (\text{II-63})$$

$$He_3(u) = u^3 - 3u \quad (\text{II-64})$$

Finalmente, a série dada pela Eq. (II-60), é reescrita como:

$$y = g(u) \approx c_1 + c_2u + c_3(u^2 - 1) + c_4(u^3 - 3u) \quad (\text{II-65})$$

Por outro lado, os valores característicos da variável  $Y$  podem ser calculados como:

$$m_n = \int_{-\infty}^{\infty} y^n f_U(u) du = \int_{-\infty}^{\infty} (g(u))^n f_U(u) du \quad \text{para } n = 1, \dots, 4 \quad (\text{II-66})$$

sendo  $f_U(u)$  a função densidade de probabilidades da distribuição normal padrão.

Explicitando a Eq. (II-66) para os quatro primeiros momentos estatísticos de  $Y$  tem-se as seguintes expressões:

$$\mu_Y = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) [c_1 + c_2u + c_3(u^2 - 1) + c_4(u^3 - 3u)] du = 0 \quad (\text{II-67})$$

$$\sigma_Y = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) [c_1 + c_2u + c_3(u^2 - 1) + c_4(u^3 - 3u)]^2 du = 1 \quad (\text{II-68})$$

$$\gamma_Y = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) [c_1 + c_2u + c_3(u^2 - 1) + c_4(u^3 - 3u)]^3 du = \gamma_X \quad (\text{II-69})$$

$$\kappa_Y = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) [c_1 + c_2u + c_3(u^2 - 1) + c_4(u^3 - 3u)]^4 du = \kappa_X \quad (\text{II-70})$$

Ao se resolver as integrais anteriores, obtém-se um sistema de quatro equações não lineares, dadas por:

$$\begin{aligned}
c_1 &= 0 \\
2c_3^2 + c_2^2 + 6c_4^2 &= 1 \\
8c_3^3 + 6c_3c_2^2 + 36c_2c_3c_4 + 1086c_3c_4^2 &= \gamma_x \\
60c_3^4 + 60c_3^2c_2^2 + 576c_3^2c_2c_4 + 2232c_3^2c_4^2 + 252c_4^2c_2^2 + \\
+ 1296c_4^3c_2 + 3348c_4^4 + 24c_2^3c_4 + 3c_2^4 &= \kappa_x
\end{aligned} \tag{II-71}$$

Resolvendo-se o sistema de equações definido acima, é possível conhecer os coeficientes  $c_2, c_3$  e  $c_4$ , uma vez que  $c_1 = 0$ . O sistema de equações anterior pode ser resolvido utilizando o método de Newton-Raphson. O ANEXO B mostra algumas soluções analíticas aproximadas para a solução deste sistema.

Conhecidos os coeficientes dos polinômios de Hermite, eles podem ser inseridos na função  $g(u)$  da Eq. (II-65) para calcular valores estatisticamente equivalentes entre as variáveis aleatórias  $Y$  e  $U$ , o que se traduz na seguinte função:

$$Y(U) = c_2U + c_3(U^2 - 1) + c_4(U^3 - 3U) \tag{II-72}$$

Por conseguinte, definida a função  $g(u)$ , também será possível determinar a sua inversa como uma expressão analítica fechada  $u = g^{-1}(y)$ . Notando que, da Eq. (II-58), a função  $g^{-1}(\cdot)$  representa a função de distribuição de probabilidades normal padrão. Assim, baseado na teoria de probabilidades, a Função Densidade de Probabilidade da variável  $Y$  será escrita como (ANG & TANG 1975):

$$f_Y(y) = f_U(u) \left| \frac{d}{dy} u(y) \right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \left| \frac{d}{dy} u(y) \right| \tag{II-73}$$

Sabendo-se que:  $\frac{d}{dy} u(y) = \frac{1}{\frac{d}{du} y(u)}$  e  $\frac{d}{du} He_n(u) = n He_{n-1}(u)$  tem-se:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(u(y))^2\right) [c_2 + c_3(2(u(y))) + c_4(3(u(y))^2 - 3)] \quad (\text{II-74})$$

Finalmente, para voltar na variável original  $X$ , calcula-se a função inversa da Eq. (II-24), dada por:

$$X = Y\sigma_x + \mu_x \quad (\text{II-75})$$

## CAPÍTULO III

### ANÁLISE ESTATÍSTICA DE PROCESSOS ALEATÓRIOS

Se um processo é imprevisível, ele é chamado de processo aleatório (também chamado de processo estocástico ou processo randômico). O comportamento dos processos aleatórios ao longo do tempo pode ser representado graficamente através de um conjunto de funções do tempo. Cada uma das funções que representam um processo aleatório é chamada de realização do processo ou simplesmente “série temporal”.

Constituem-se exemplos de processos aleatórios na área offshore: a variação da elevação da superfície do mar numa dada locação, a variação da velocidade do vento num dado ponto de observação, os movimentos de uma unidade flutuante submetida às ações ambientais de onda, vento e correnteza, etc. Nos fenômenos anteriores, não é possível prever o valor exato que adquirirão em algum instante de tempo futuro.

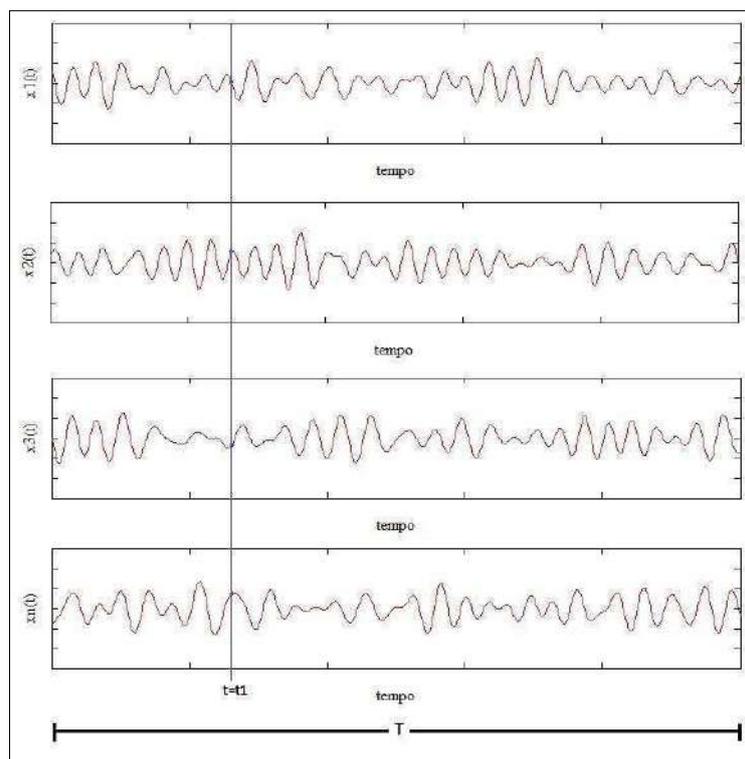


Figura III.1 –Exemplo de  $n$  diferentes realizações de um processo aleatório  $X(t)$

A Figura III.1 ilustra um conjunto (ou coleção) de  $n$  experimentos (ou observações) de um certo processo aleatório  $X(t)$ , para os quais foram registradas suas medições (simultaneamente ou não) durante um período de tempo com duração igual a

$T$ . Pode ser observado que para um instante de tempo  $t = t_1$  qualquer, existe uma nova amostra de valores formada pelas  $n$  medições diferentes de cada uma das séries temporais. É intuitivo imaginar que, mesmo sendo aleatórios, este novo conjunto de valores está distribuído em torno de um valor e sob uma dispersão particular, no início desconhecidas.

Em termos matemáticos, é possível tratar o conjunto de valores das  $n$  medições do experimento no instante  $t = t_1$ , i.e.,  $\{x_1(t_1), x_2(t_1), x_3(t_1), \dots, x_n(t_1)\}$  como sendo uma amostra de uma variável aleatória. A mesma lógica pode ser aplicada para uma quantidade  $m$  de diferentes instantes de tempo  $t = t_1, t_2, t_3 \dots t_m \leq T$ .

Como será visto, estes conjuntos atendem certas relações matemáticas que podem ser utilizadas para extrair informações estatísticas sobre o seu comportamento. Sendo assim, cada uma das  $n$  novas variáveis aleatórias terá seus próprios parâmetros estatísticos. Por exemplo, para dois instantes de tempo  $t_1$  e  $t_2$ , o valor médio, o valor médio quadrático, a variância e a covariância serão dados por:

<p><i>Instante de tempo <math>t_1</math></i></p> $E[x(t_1)] = \mu_1 = \sum_{i=1}^n \frac{x_i(t_1)}{n}$ $E[(x(t_1))^2] = \sum_{i=1}^n \frac{(x_i(t_1))^2}{n}$ $Var[x(t_1)] = \sigma_1^2 = E[(x(t_1))^2] - \mu_1^2$	<p><i>Instante de tempo <math>t_2</math></i></p> $E[x(t_2)] = \mu_2 = \sum_{i=1}^n \frac{x_i(t_2)}{n}$ $E[(x(t_2))^2] = \sum_{i=1}^n \frac{(x_i(t_2))^2}{n}$ $Var[x(t_2)] = \sigma_2^2 = E[(x(t_2))^2] - \mu_2^2$
$COV(x(t_1), x(t_2)) = E[(x(t_1) - \mu_1)(x(t_2) - \mu_2)] = \sum_{i=1}^n \frac{(x_i(t_1))(x_i(t_2))}{n} - (\mu_1)(\mu_2)$	

Esta abordagem, baseada em medições “através” de várias séries temporais, ao invés de fazer as medições “ao longo” de uma única série temporal, permite classificar os processos aleatórios em função das propriedades estatísticas das novas variáveis  $\{x_1(t_i), x_2(t_i), x_3(t_i), \dots, x_n(t_i)\}$  e das distribuições de probabilidades que as definem, sem dependerem (estritamente) da variação do tempo, como será explicado neste capítulo.

Como abreviação da nomenclatura, ao se referir ao processo aleatório  $X(t)$ , poderá ser sobre-entendido que ele é uma função do tempo, de forma tal que será omitida a letra “ $t$ ”, e reescrevendo simplesmente como  $X$ .

### III.1. Estacionariedade dos processos aleatórios

Sabendo-se que cada uma das variáveis aleatórias  $x(t_i)$  segue uma determinada distribuição de probabilidades, caso essa distribuição seja idêntica para qualquer  $t$  o processo será conhecido como processo estacionário. Em consequência, suas propriedades estatísticas serão independentes da variação do tempo, e apenas dependerão da duração da realização considerada.

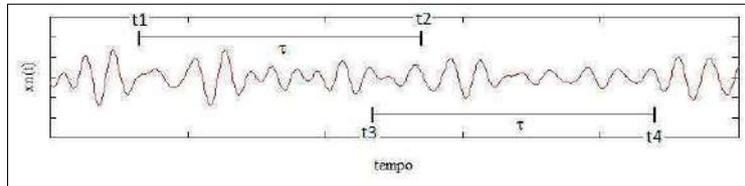


Figura III.2 – Exemplo de uma realização de um processo aleatório estacionário

Em alusão a *Figura III.2*, o processo será estacionário quando o valor esperado em  $t = t_1$  seja igual para outro instante de tempo qualquer  $(t_2, t_3, \dots, t_n)$ , e a covariância dependerá apenas da separação  $\tau = t_2 - t_1 = t_4 - t_3$ . Com isso, a mesma conclusão pode se aplicar para o resto dos valores característicos da variável, obtendo as seguintes propriedades:

$$E[X(t_1)] = E[X(t_2)] = E[X(t_3)] = \dots = E[X(t_n)] = \mu_X \quad (\text{III-1})$$

$$\sigma[X(t_1)] = \sigma[X(t_2)] = \sigma[X(t_3)] = \dots = \sigma[X(t_n)] = \sigma_X \quad (\text{III-2})$$

$$\gamma[X(t_1)] = \gamma[X(t_2)] = \gamma[X(t_3)] = \dots = \gamma[X(t_n)] = \gamma_X \quad (\text{III-3})$$

$$\kappa[X(t_1)] = \kappa[X(t_2)] = \kappa[X(t_3)] = \dots = \kappa[X(t_n)] = \kappa_X \quad (\text{III-4})$$

$$COV[X(t_1), X(t_2)] = COV[X(t_3), X(t_4)] = \dots = COV[X(t), X(t + \tau)] \quad (\text{III-5})$$

Na prática, é usual considerar que processos com durações relativamente curtas são aproximadamente estacionários, ou também, processos com durações longas podem ser divididos em períodos curtos, cada um dos quais pode ser considerado com características de estacionariedade. Assim, entendemos que “o que acontece num tempo  $t_1$  qualquer tem a mesma probabilidade de acontecer em qualquer outro instante de tempo  $t_2$ ”.

Pelo contrário, quando num processo as suas propriedades estatísticas nunca podem ser consideradas constantes, independentemente do intervalo de tempo estudado, o processo será chamado de não-estacionário. Em geral, os processos com durações muito prolongadas são não-estacionários.

Quando um processo não-estacionário é sub-dividido em intervalos menores (1 hora, 2 horas, 3 horas) com características de estacionariedade, estes tempos de duração pequenos são denominadas intervalos de curto prazo. A representação de um tempo de duração consideravelmente grande (1 ano, 10 anos, 100 anos, etc.) através de vários períodos de curto-prazo é conhecida como representação de longo prazo.

### III.2. Ergodicidade de um processo aleatório

Um caso particular de um processo estacionário é o processo denominado ergódico. Um processo aleatório ergódico é um processo aleatório estacionário cujos parâmetros estatísticos são iguais aos parâmetros temporais calculados ao longo de uma única realização. Assim, este processo pode ser caracterizado a partir de apenas uma realização individual (série temporal) observada do mesmo processo (NEWLAND, 1993).

Matematicamente, para um processo ergódico, a média, a variância e a covariância podem ser calculadas aplicando-se as seguintes equações:

$$\mu_X = E[x(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt \quad \text{(III-6)}$$

$$\sigma_X^2 = Var[x(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x(t) - \mu_X]^2 dt \quad \text{(III-7)}$$

$$COV[x(t), x(t + \tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x(t) - \mu_x][x(t + \tau) - \mu_x] dt \quad (III-8)$$

A ergodicidade do processo implica a sua estacionariedade; mas a recíproca não é necessariamente verdadeira.

Na análise estrutural de plataformas petrolíferas, devido à alta dificuldade que implicaria obter uma quantidade suficiente de realizações dos carregamentos ambientais e dos efeitos produzidos por eles, uma prática comum é assumir que eles, além de serem estacionários, são ergódigos. As considerações a serem feitas a seguir assumem que o processo aleatório investigado é estacionário e ergódigo.

### III.3. Função de auto-covariância de um processo aleatório ergódigo

Citando a *Figura III.2*, a função de auto-covariância  $R(\tau)$  permite prever estatisticamente o comportamento da série no instante de tempo  $[t + \tau]$  a partir do valor conhecido no instante de tempo  $t$ . Esta função é definida como o valor esperado do produto  $[x(t)x(t + \tau)]$ , o que equivale à seguinte equação:

$$R(\tau) = E[x(t)x(t + \tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x(t)x(t + \tau)] dt \quad (III-9)$$

Em particular, quando  $\tau = 0$ , a equação anterior é equivalente à Eq. (II-13), que é propriamente a definição da variância, o que permite escrever:

$$R(0) = \sigma_x^2 \quad (III-10)$$

Devido à aleatoriedade do processo, é possível imaginar que na medida em que cresce o valor de  $\tau$ , torna-se mais difícil prever o seu comportamento e, conseqüentemente a auto-covariância tende a diminuir. A *Figura III.3* apresenta a forma típica de uma função de auto-covariância.

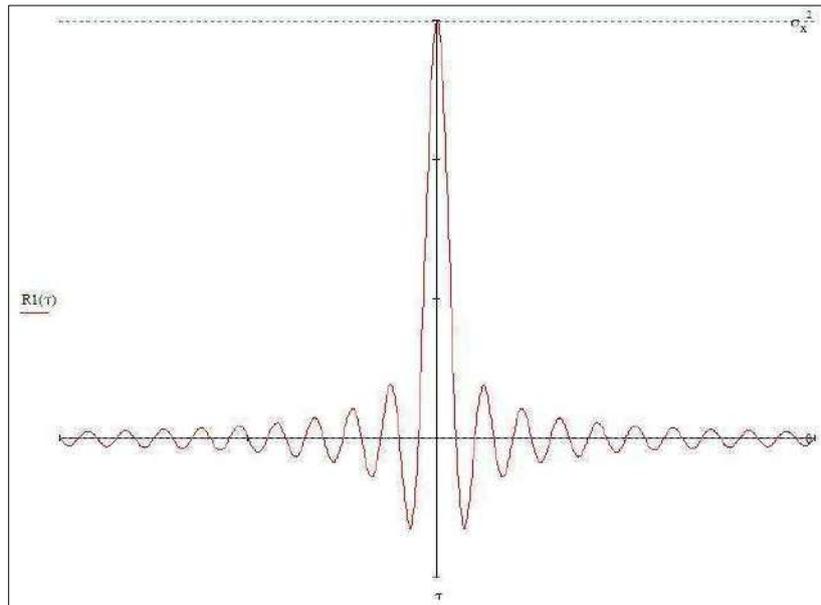


Figura III.3 – Exemplo da função de auto-covariância de um processo aleatório

#### III.4. Densidade espectral de um processo aleatório

Fazendo uso da definição da função de auto-covariância, definimos a densidade espectral (às vezes chamada simplesmente de espectro) de um processo aleatório como uma função que, através da identificação das periodicidades do sinal, permite levar o processo do domínio do tempo até o domínio da frequência. A densidade espectral corresponde à Transformada de Fourier da função de auto-covariância e é calculada através da seguinte equação:

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau} d\tau \quad (\text{III-11})$$

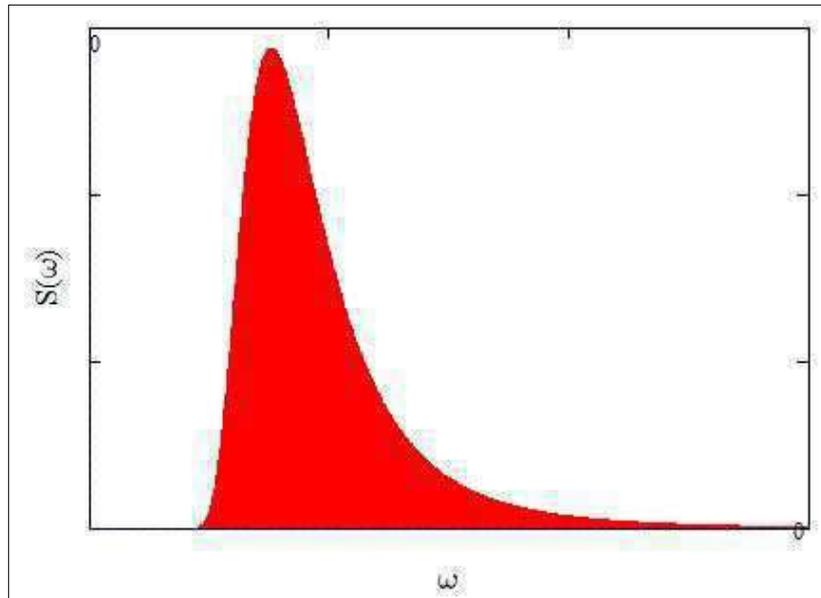
onde  $\omega$  indica a frequência angular, medida em rad/s.

A relação inversa é dada por:

$$R(\tau) = \int_0^{\infty} S(\omega) e^{i\omega\tau} d\omega \quad (\text{III-12})$$

A partir desta expressão é possível observar que considerando  $\tau = 0$ , a variância do processo é igual a área sob a curva da sua densidade espectral (*Figura III.4*), i.e.:

$$\sigma_X^2 = \int_0^{\infty} S(\omega) d\omega \quad (\text{III-13})$$



*Figura III.4 – Densidade espectral  $S(\omega)$  de um processo aleatório*

A propriedade anterior tem um grande significado, devido à que permite obter uma relação matemática direta entre as frequências e os parâmetros estatísticos da variável que representam.

### **III.5. Fator de largura de banda**

O fator de largura de banda é uma medida quantitativa que permite observar a forma que a energia do sinal ( $S(\omega)$ ) é distribuída em relação às suas faixas de frequências ( $\omega$ ). Este fator é representado por  $\varepsilon$ , que pode assumir valores entre 0 e 1, e calcula-se pela seguinte equação:

$$\varepsilon = \sqrt{1 - \frac{(m_2)^2}{(m_0)(m_4)}} \quad (\text{III-14})$$

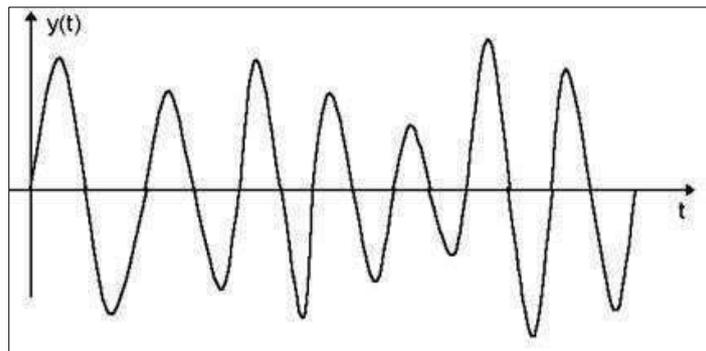
onde  $m_n$  indica o momento espectral de ordem  $n$ , que é calculado através da seguinte equação:

$$m_n = \int_0^{\infty} S(\omega)(\omega^n)d\omega \quad (\text{III-15})$$

O fator de largura de banda permite classificar os processos em dois tipos: processos de banda estreita e processos de banda larga.

### III.5.1. Processos de banda estreita e de banda larga

Em um processo de banda estreita, a maior parte da densidade espectral está concentrada numa pequena faixa de frequências. Isto implica em que para cada cruzamento positivo pelo seu valor médio corresponde um único máximo local, como mostra a *Figura III.5*. Neste tipo de processo, o fator de largura de banda adquire valores próximos ao zero ( $\varepsilon \rightarrow 0$ ).



*Figura III.5 –Processo de banda estreita (LIMA, 2008)*

Por outro lado, em processos de banda larga, a densidade espectral está distribuída numa faixa ampla de frequências. Uma série temporal de um processo de banda larga é caracterizado por apresentar vários máximos locais entre dois cruzamentos com ascendência positiva pelo seu valor médio, como mostra a *Figura III.6*. Neste tipo de processos, o fator de largura de banda assume valores maiores que zero ( $\varepsilon > 0$ ).

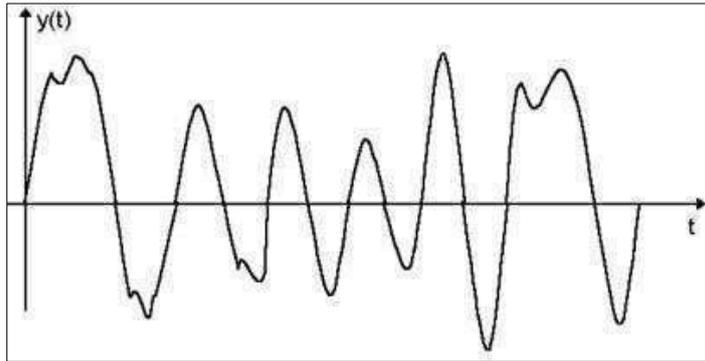


Figura III.6 – Processo de banda larga (LIMA, 2008)

A Figura III.7 ilustra genericamente os espectros destes dois tipos de processos mencionados anteriormente.

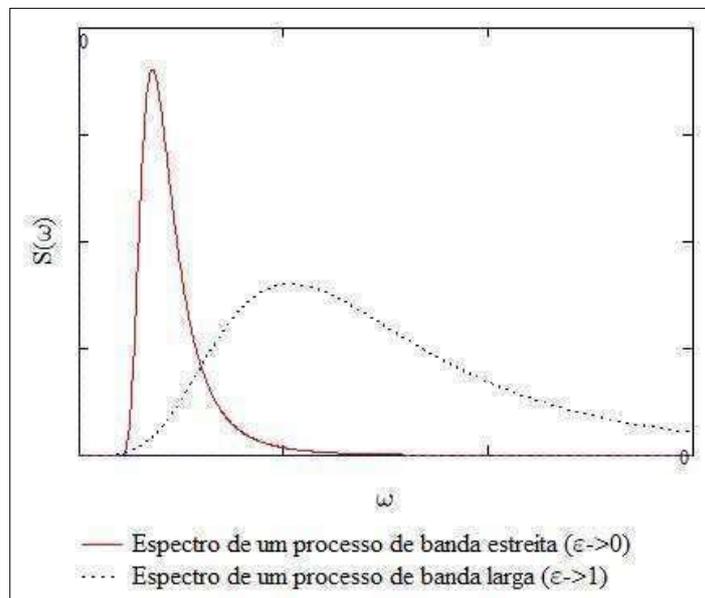


Figura III.7 – Densidades espectrais de processos de banda estreita e banda larga

### III.6. Distribuições de probabilidades associadas a um processo aleatório

Com o intuito de facilitar o tratamento estatístico dos processos aleatórios, e assumindo sua ergodicidade, recorre-se a associar a série temporal que o representa com distribuições de probabilidades, como será explicado na sequência.

Na intenção de simplificar os cálculos, utilizaremos a chamada variável reduzida, definida na Eq. (II-24).

### III.6.1. Distribuição do Processo Aleatório

Utilizando a definição da função densidade de probabilidades dada na Eq. (II-2), a probabilidade de um processo  $X(t)$  estar contido entre dois valores  $x_a$  e  $x_b$  é dada por:

$$P(x_a \leq x(t) \leq x_b) = \int_{x_a}^{x_b} f_X(x(t)) dx \quad (\text{III-16})$$

Assim, a determinação da probabilidade do processo  $y(t)$  estar contido entre dois valores  $y$  e  $y + dy$ , consistirá na acumulação de todas as larguras  $dt$ . Isto é exemplificado na Figura III.8, pela definição da Eq. (II-2), é dada por:

$$P(y \leq y(t) \leq y + dy) = \int_y^{y+dy} f_Y(y(t)) dy = f_Y(y(t)) dy = \sum_{i=1}^N \frac{dt_i}{T} \quad (\text{III-17})$$

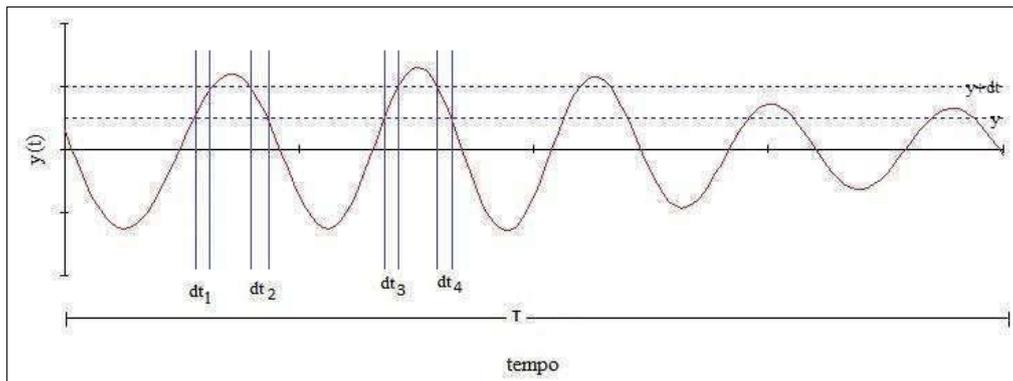


Figura III.8 – Definição da largura  $dt_1$  do processo  $y(t)$  para a faixa entre  $y$  e  $y+dy$

Na prática, os processos usualmente são digitalizados através de  $N$  pontos discretos utilizando um intervalo de amostragem ( $\Delta t$ ) constante com uma duração total  $T = N\Delta t$  conforme ilustra a Figura III.9

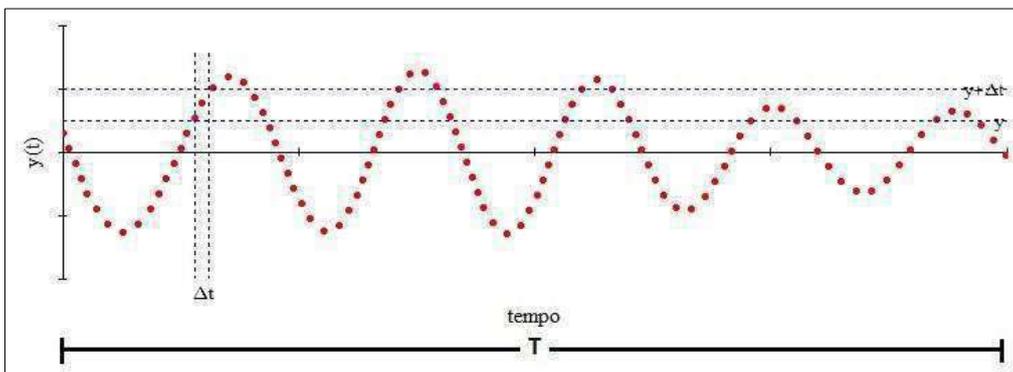


Figura III.9 – Discretização do processo  $y(t)$  e a probabilidade contida entre  $y$  e  $y+\Delta y$

Com o processo na sua forma discreta em um número finito de pontos, a probabilidade de se obter valores no intervalo  $[y; y + \Delta y]$  consiste simplesmente na contagem de todos os pontos que caem dentro desta faixa e a divisão pelo número total de pontos  $N$ . Desta maneira, a Eq. (III-17) poderá ser reescrita de forma discreta como:

$$P(y \leq y(t) \leq y + \Delta y) = f_Y(y)\Delta y = \sum_{i=1}^N \frac{\Delta t_i}{T} = \frac{n_i}{N} \quad (\text{III-18})$$

$$f_Y(y) = \frac{n_i}{N\Delta y}$$

onde:

$n_i$	número de pontos contidos no intervalo $[y; y + \Delta y]$
$N = \frac{T}{\Delta t}$	número total de pontos do processo $y(t)$ discretizado
$T$	tempo total da duração do processo $y(t)$
$\Delta t$	passo de tempo considerado na discretização

Assim, a função  $f_Y(y)$  é a função densidade de probabilidades da distribuição de probabilidades que descreve (ou caracteriza) o processo aleatório. No caso de um processo aleatório ser caracterizado por uma distribuição Normal, i.e., a distribuição de Gauss, é a que melhor se ajusta à função empírica calculada na Eq. (III-18), o processo aleatório é chamado de processo aleatório Gaussiano. Neste caso, a função densidade de probabilidades é definida como:

$$f_Y(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{y}{\sigma_Y} \right)^2 \right] = \frac{1}{\sqrt{m_0} \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{y}{\sqrt{m_0}} \right)^2 \right] \quad (\text{III-19})$$

sendo  $m_0$  a área do respectivo espectro do processo aleatório.

Em situações em que  $f_Y(y)$  não pode ser caracterizada por uma distribuição de Gauss o processo recai na classe de processos Não-Gaussianos.

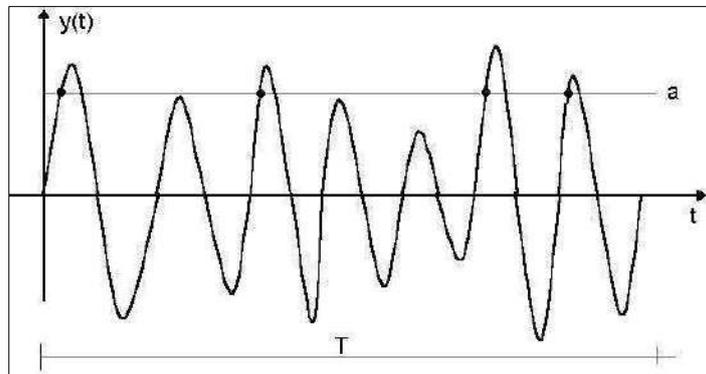
### III.7. Frequência de cruzamento positivo no nível $y(t)=a$

No caso de um processo Gaussiano, a taxa de cruzamentos do processo num nível  $y(t) = a$  com ascendência positiva (vide *Figura III.10.*) é calculada como:

$$v_a = \frac{N_a}{T} = \frac{1}{2\pi} \sqrt{\frac{m_2}{m_0}} \exp\left[-\frac{1}{2}\left(\frac{a^2}{m_0}\right)\right] \quad (\text{III-20})$$

onde:

$a$	nível fixo de $y(t)$
$T$	duração considerada do processo $y(t)$
$N_a$	número de vezes que a função $y(t)$ cruza o nível horizontal $y(t) = a$ na direção crescente
$m_0$ e $m_2$	momentos espectrais de ordem zero e de ordem dois do espectro, respectivamente, dados pela Eq. (III-15)



*Figura III.10 – Frequência de cruzamentos positivos no nível  $y(t)=a$  (LIMA, 2008)*

Desta forma, a frequência de cruzamento pelo nível zero  $v_0$  (ou pela média da variável reduzida) de um processo aleatório gaussiano é dada por:

$$v_0 = \frac{N_0}{T} = \frac{1}{2\pi} \sqrt{\frac{m_2}{m_0}} \quad (\text{III-21})$$

onde  $N_0$  é o número de vezes que a função  $y(t)$  cruza o eixo horizontal na direção ascendente.

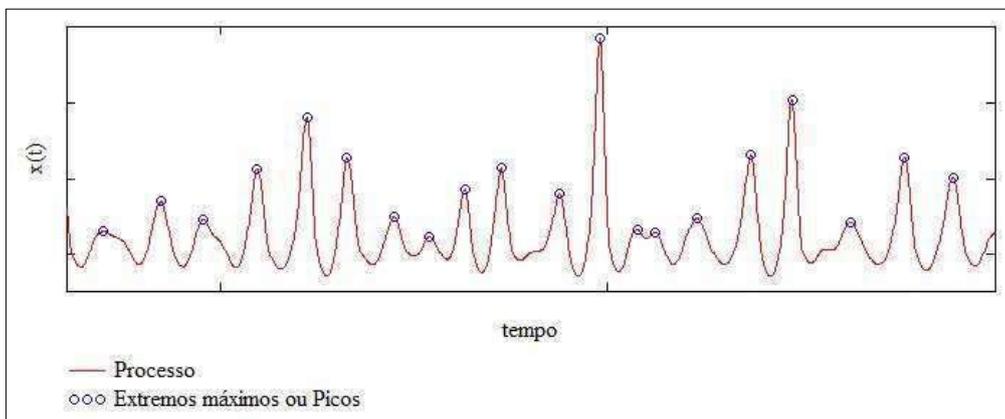
Para processos não-Gaussianos a estimativa da frequência de cruzamentos zero é feita a partir da sua definição, i.e., contando-se os cruzamentos e dividindo-se pela duração da série temporal.

### III.8. Distribuição dos Picos de um Processo Gaussiano

Os picos de um processo aleatório, identificados na *Figura III.11*, possuem uma distribuição de probabilidades própria. No caso específico do processo ser Gaussiano, a distribuição dos picos segue uma distribuição conhecida como distribuição de Rice (RICE, 1945), cuja função densidade de probabilidades é dada pela seguinte equação:

$$f_P(p) = \frac{\varepsilon}{\sqrt{m_0}\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{p^2}{m_0\varepsilon^2}\right] + \frac{p}{m_0} \sqrt{1-\varepsilon^2} \exp\left[-\frac{1}{2} \frac{p^2}{m_0}\right] \Phi\left[\frac{p}{\sqrt{m_0}\varepsilon} \sqrt{1-\varepsilon^2}\right] \quad (\text{III-22})$$

onde  $\varepsilon$  é o fator de largura de banda definido na Eq. (III-14),  $m_0$  é o momento espectral de ordem zero definido na Eq. (III-15) e  $\Phi(\cdot)$  é a função cumulativa da distribuição normal padrão.



*Figura III.11 – Picos do processo  $x(t)$*

Outro caso ainda mais particular acontece quando o processo, além de ser Gaussiano, é também de banda estreita ( $\varepsilon = 0$ ). Neste caso a distribuição dos picos se reduz ao modelo de Rayleigh (item II.6.5.) com o parâmetro  $\alpha_R^2 = m_0$ , i.e., as funções densidade de probabilidades e cumulativa de probabilidades são obtidas através das seguintes equações:

$$f_p(p) = \frac{p}{m_0} \exp\left[-\frac{1}{2}\left(\frac{p^2}{m_0}\right)\right] \quad (\text{III-23})$$

$$F_p(p) = 1 - \exp\left[-\frac{1}{2}\left(\frac{p^2}{m_0}\right)\right] \quad (\text{III-24})$$

No outro extremo, se o processo é gaussiano, porém, de banda larga com  $\varepsilon = 1$ , a distribuição dos picos é também definida por uma distribuição Gaussiana com os parâmetros  $\mu = 0$  e  $\sigma = \sqrt{m_0}$ , i.e.:

$$f_p(p) = \frac{1}{\sqrt{m_0}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{p^2}{m_0}\right)\right] \quad (\text{III-25})$$

No caso de um processo Gaussiano a taxa de picos ou frequência de picos é definida por:

$$\nu_p = \frac{N_p}{T} = \frac{1}{2\pi} \sqrt{\frac{m_4}{m_2}} \quad (\text{III-26})$$

onde  $N_p$  é o número de picos identificados na série temporal aleatória de duração total  $T$ .

Deve ser observado que no caso de processos Gaussianos, basta conhecermos a função densidade espectral para obter estimativas de  $\nu_0$  e  $\nu_p$ , assim como, a distribuição do processo e a distribuição dos seus picos.

Nos casos onde os processos não são Gaussianos, não existe uma solução analítica tanto para descrever a distribuição dos picos quanto para obter a taxa de picos. Para tais casos se faz necessário outro tipo de abordagem. A taxa de picos é estimada pela sua definição, ou seja, basta contar os picos e dividir pela duração do sinal. No CAPÍTULO V, serão investigadas várias metodologias para representar a distribuição dos picos de processos não-Gaussianos.

### III.9. Estimativa dos Parâmetros Estatísticos Básicos de um Processo Aleatório a partir de uma realização

Assumindo-se que o processo aleatório é ergódico, os seus parâmetros estatísticos podem ser estimados a partir de uma única realização. Tomando-se uma série digitalizada (vide *Figura III.9*) composta de  $N$  pontos discretos a cada passo de tempo  $\Delta t$ , cujo tempo total é  $T = N\Delta t$ , a média do processo pode ser estimada como:

$$m_X = \frac{1}{T} \int_0^T x(t) dt = \frac{1}{N\Delta t} \sum_{i=1}^N x_i \Delta t = \frac{1}{N} \sum_{i=1}^N x_i \quad (\text{III-27})$$

ou seja, a média é calculada como se estivéssemos tratando de uma amostra de um novo conjunto finito de  $N$  valores:  $\{x_1, x_2, x_3, \dots, x_N\}$ . Isto é válido também para os demais parâmetros estatísticos. Assim, estimadores para a média, variância, coeficiente de assimetria e de curtose de um processo aleatório  $X(t)$  são dados por:

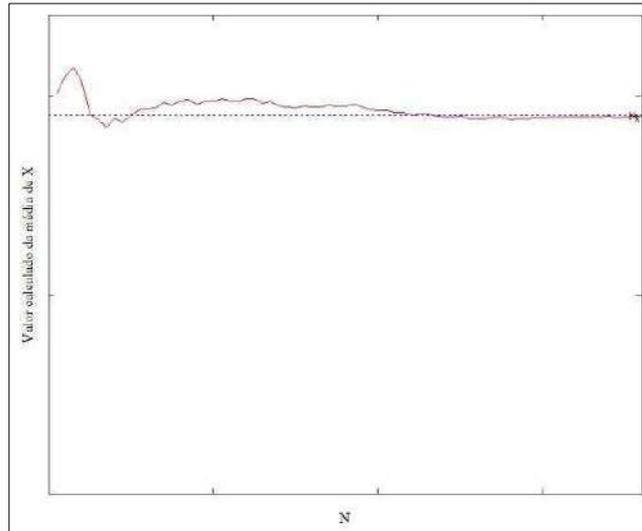
$$m_X = \frac{1}{N} \sum_{i=1}^N x_i \quad (\text{III-28})$$

$$s_X^2 = \frac{1}{N} \sum_{i=1}^N (x_i - m_X)^2 \quad (\text{III-29})$$

$$g_X = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - m_X)^3}{s_X^3} \quad (\text{III-30})$$

$$k_X = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - m_X)^4}{s_X^4} \quad (\text{III-31})$$

Um aspecto importante é observar a estabilidade destes parâmetros (definição de ergodicidade) em função do tamanho da série temporal. A *Figura III.12* ilustra a estabilidade da média em função do tamanho da série temporal. Observa-se que no final o estimador da média está convergindo para um valor fixo. Em outras palavras, a série tem que ser suficientemente longa para garantir a estabilidade dos parâmetros estatísticos.



*Figura III.12 –Estabilidade da média de uma amostra em relação ao tamanho da mesma*

A mesma amostra  $\{x_1, x_2, x_3, \dots, x_N\}$  é utilizada para se buscar uma distribuição de probabilidades teórica, por uma técnica de ajuste de distribuição (ANG & TANG, 1984), que descreva o processo aleatório adequadamente. Como já dito antes, se esta distribuição for a Gaussiana o processo é chamado de processo aleatório Gaussiano.

## CAPÍTULO IV

### ESTADÍSTICA DE VALORES EXTREMOS

Em muitos problemas da Engenharia o principal intuito é identificar os efeitos provocados pelas condições mais adversas. No caso particular da Engenharia Estrutural, o interesse principal concerne aos maiores valores dos efeitos provocados pelos carregamentos externos atuantes e/ou aos menores valores de resistência interna da estrutura. Nestas situações os interesses recaem num conjunto específico das variáveis aleatórias que são os seus valores extremos máximos ou mínimos, dependendo do caso.

Os valores máximos (ou mínimos) observados de uma variável aleatória dependem de cada realização e, portanto, eles constituem também uma variável aleatória. Neste capítulo aborda-se a chamada Estatística de Valores Extremos, que é o ramo da estatística devotado ao tratamento de valores máximos e mínimos de uma variável aleatória. Como o interesse deste trabalho recai na estatística do pico extremo (máximo valor) de um processo aleatório, maior ênfase será dada para a estatística de valores extremos máximos.

#### IV.1. Distribuições teóricas de Valores Extremos (Estatística de Ordem)

Devido à sua excepcionalidade, naturalmente os valores extremos de uma variável aleatória (por exemplo: o nível máximo anual de um rio numa dada localidade) constituem-se de amostras com poucos elementos. Caso a amostra seja grande o suficiente, a distribuição do valor extremo constitui-se na distribuição que melhor se ajusta a este conjunto de dados. Entretanto, na grande maioria dos problemas práticos de interesse, por exemplo, eventos centenários, é difícil se dispor de uma amostra suficiente de dados para fazer tal ajuste. Em função disto, foi desenvolvida uma teoria denominada Estatística de Ordem (ANG & TANG, 1984) que permite a obtenção da distribuição de valores extremos de interesse a partir da distribuição inicial da variável aleatória e de algumas hipóteses simplificadoras.

Uma amostra genérica de tamanho  $n$  de uma variável aleatória  $X$  pode ser representada como:

$$X = \{X_1, X_2, \dots, X_n\} \quad (\text{IV-1})$$

onde  $X_i$  é uma variável aleatória que representa a variabilidade da  $i$ -ésima observação de  $X$ . Como todas estas variáveis pertencem a mesma população, suas distribuições de probabilidade são idênticas a da variável  $X$ , i.e.:

$$F_{X_1}(x) = \dots = F_{X_n}(x) = F_X(x) \quad (\text{IV-2})$$

Chamando de  $Y_n$  o valor extremo máximo de  $X$  numa amostra de tamanho  $n$  tem-se a seguinte definição:

$$Y_n = \text{máx}\{X_1, X_2, \dots, X_n\} \quad (\text{IV-3})$$

Assumindo-se que um valor específico  $X = y$  pertence a população de valores extremos de  $X$ , tem-se a seguinte condição:

$$P(Y_n \leq y) = P\left(X_1 \leq y \bigcap X_2 \leq y \dots \bigcap X_n \leq y\right) \quad (\text{IV-4})$$

Assumindo-se que os valores observados de  $X$  são estatisticamente independentes uns dos outros a Eq. (IV-4) pode ser escrita como:

$$P(Y_n \leq y) = P(X_1 \leq y)P(X_2 \leq y) \dots P(X_n \leq y) \quad (\text{IV-5})$$

Considerando que  $P(X \leq x) = F_X(x)$  e que as observações são identicamente distribuídas, tem-se que a função cumulativa da distribuição do valor máximo extremo de  $X$  é dada por:

$$F_{Y_n}(y) = [F_X(y)]^n \quad (\text{IV-6})$$

e conseqüentemente sua função densidade de probabilidades é dada por:

$$f_{Y_n}(y) = \frac{dF_{Y_n}(y)}{dy} = n[F_X(y)]^{n-1}f_X(y) \quad (\text{IV-7})$$

onde  $F_X(\cdot)$  e  $f_X(\cdot)$  correspondem à função densidade e a função cumulativa de probabilidades da variável original  $X$ , que na Estatística de Ordem é denominada de Distribuição Inicial.

Definindo-se  $Y_1$  o valor extremo mínimo de  $X$  numa amostra de tamanho  $n$ , tem-se a seguinte definição:

$$Y_1 = \min\{X_1, X_2, \dots, X_n\} \quad (\text{IV-8})$$

Aplicando-se raciocínio similar ao que foi usada acima chega-se aos seguintes resultados para a distribuição do valor mínimo extremo de  $X$ :

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n \quad (\text{IV-9})$$

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1} f_X(y) \quad (\text{IV-10})$$

Desta forma demonstra-se que, conhecendo-se a distribuição de probabilidades da variável  $X$  (Distribuição Inicial), é possível extrair suficientes informações sobre ela para determinar as distribuições dos seus valores extremos máximos (resolvendo-se a Eq. (IV-7)) ou mínimos (resolvendo-se a Eq. (IV-10)), de acordo com o interesse da aplicação. A *Figura IV.1* ilustra genericamente esta situação. Como pode ser observado nesta figura, as distribuições de valores extremos (máximos e mínimos) são completamente dependentes das caudas (direta e esquerda, respectivamente) da distribuição inicial. A seguir serão abordados somente aspectos relacionados à distribuição de valores máximos.

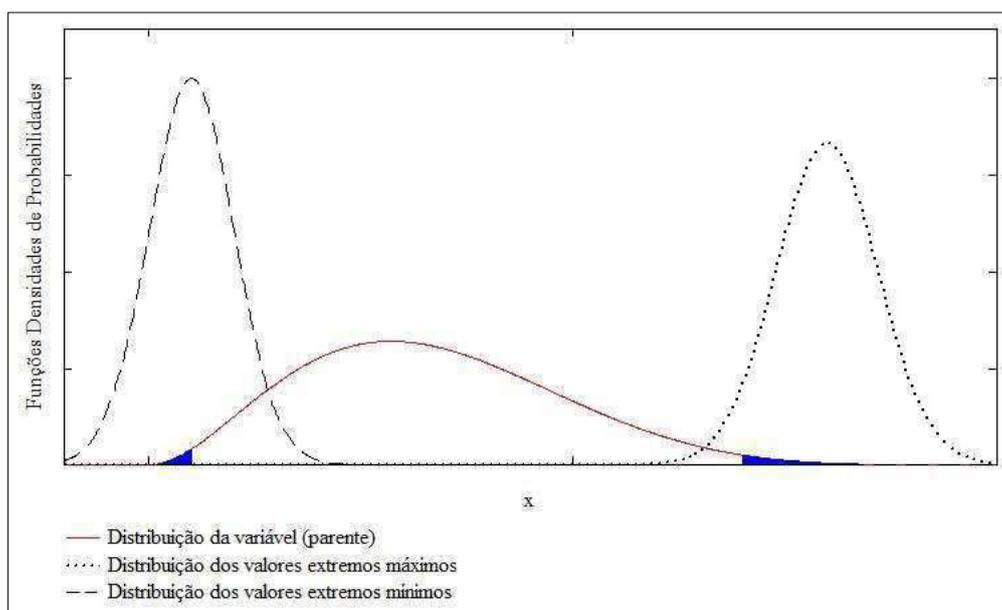


Figura IV.1 – Funções densidades de probabilidades de valores extremos máximos e mínimos em relação à sua distribuição inicial

## IV.2. Distribuições assintóticas de valores extremos

Nas equações (IV-3) a (IV-10) o parâmetro  $n$  representa o número de observações da variável  $X$ , por exemplo, num dado intervalo de tempo de interesse. Vários pesquisadores observaram que as distribuições de extremos teóricas definidas no item anterior convergem para formas assintóticas para valores grandes de  $n$ , ou melhor dizendo, para  $n \rightarrow \infty$ . A forma assintótica de convergência é dependente do decaimento da cauda da distribuição inicial na direção do extremo considerado (máximo ou mínimo). As formas assintóticas identificadas na literatura são as seguintes (ANG & TANG, 1984):

- Tipo I (ou de Gumbel)
- Tipo II (ou de Fréchet)
- Tipo III (ou de Weibull)

A forma Tipo III é aplicável para distribuições iniciais truncadas na direção do valor extremo de interesse. Como neste trabalho o foco não é em distribuições truncadas, esta forma não será abordada. As duas outras formas serão descritas mais detalhadamente nas seções seguintes.

### IV.2.1. Distribuição Tipo I ou de Gumbel (para máximos)

Variáveis que apresentam decaimento exponencial convergem para uma distribuição assintótica Tipo I ou de Gumbel. Uma variável aleatória  $X$  possui decaimento exponencial se a seguinte relação for atendida:

$$\frac{f_X(x)}{1 - F_X(x)} = -\frac{f'_X(x)}{f_X(x)} \quad (\text{IV-11})$$

Dentre as mais conhecidas, as distribuições de probabilidade que atendem o requisito acima na direção do valor extremo máximo são a Normal, a Weibull e a própria distribuição de Gumbel quando usada como distribuição inicial. Estas distribuições irão assintoticamente (em função de  $n$ ) convergir para a distribuição de Gumbel que é dada por:

$$f_X(x) = \alpha \exp[-\alpha(x - u) - \exp[-\alpha(x - u)]] \quad (\text{IV-12})$$

$$F_X(x) = \exp[-\exp[-\alpha(x - u)]] \quad (\text{IV-13})$$

onde  $u$  é o parâmetro de locação e  $\alpha$  é o parâmetro de escala. Além disso, a média e o desvio padrão tem as seguintes relações com estes parâmetros:

$$\mu_X = u + \frac{0.57722}{\alpha} \quad (\text{IV-14})$$

$$\sigma_X = \frac{\pi}{\alpha\sqrt{6}} \quad (\text{IV-15})$$

O mais importante desta distribuição é que o parâmetro  $u$  corresponde ao valor mais provável (ou moda) de  $X$ , valor que será muito utilizado neste trabalho. É interessante observar que este parâmetro tem a seguinte relação com a distribuição inicial (ANG & TANG, 1984):

$$F_X(u) = 1 - \frac{1}{n} \quad (\text{IV-16})$$

No caso de uma variável inicial  $X$  seguir a distribuição Normal com média  $\mu$  e desvio padrão  $\sigma$ , os parâmetros da distribuição assintótica são dados por:

$$u = \sigma \left( \sqrt{2\ln(n)} - \frac{\ln(\ln(n)) + \ln(4\pi)}{2\sqrt{2\ln(n)}} \right) + \mu \quad (\text{IV-17})$$

$$\alpha = \frac{\sqrt{2\ln(n)}}{\sigma} \quad (\text{IV-18})$$

No caso de uma variável inicial  $X$  possuir uma distribuição de Weibull com dois parâmetros  $\alpha_w$  e  $\lambda_w$ , os parâmetros da distribuição assintótica (Gumbel) são dados por:

$$u = \alpha_w (\ln(n))^{\frac{1}{\lambda_w}} \quad (\text{IV-19})$$

$$\alpha = \frac{\lambda_w}{\alpha_w} (\ln(n))^{\frac{\lambda_w-1}{\lambda_w}} \quad (\text{IV-20})$$

Se a distribuição inicial da variável aleatória  $X$  for também uma distribuição de Gumbel com parâmetros  $u_i$  e  $\alpha_i$ , os parâmetros da distribuição assintótica (Gumbel) são dados por:

$$u = u_i + \frac{\ln(n)}{\alpha_i} \quad (\text{IV-21})$$

$$\alpha = \alpha_i \quad (\text{IV-22})$$

#### IV.2.2. Distribuição Tipo II ou de Fréchet (para máximos)

Variáveis aleatórias que apresentam decaimento polinomial na direção do extremo de interesse (máximo) convergem para uma distribuição assintótica Tipo II ou de Fréchet. Uma variável aleatória  $X$  possui decaimento polinomial se a seguinte relação for atendida:

$$\lim_{x \rightarrow \infty} x^k [1 - F_X(x)] = a = cte \quad (IV-23)$$

Dentre as mais conhecidas, a distribuição de probabilidade que atende o requisito acima na direção do valor extremo máximo é a distribuição Log-normal.

A distribuição de probabilidades de Fréchet é definida pelas seguintes duas funções densidade e cumulativa de probabilidades:

$$f_X(x) = \left(\frac{k}{v}\right) \left(\frac{v}{x}\right)^{k+1} \exp\left[-\left(\frac{v}{x}\right)^k\right] \quad (IV-24)$$

$$F_X(x) = \exp\left[-\left(\frac{v}{x}\right)^k\right] \quad (IV-25)$$

Esta distribuição depende de dois parâmetros:  $v$  é o parâmetro de escala e  $k$  é o parâmetro de forma. Além disso, a média e o desvio padrão da variável têm as seguintes relações com estes parâmetros:

$$\mu_X = v \Gamma\left[1 - \frac{1}{k}\right] \quad (IV-26)$$

$$\sigma_X = v \left[ \Gamma\left[1 - \frac{2}{k}\right] - \Gamma^2\left[1 - \frac{1}{k}\right] \right]^{1/2} \quad (IV-27)$$

onde  $\Gamma(\cdot)$  é a Função Gamma.

Então, se a distribuição inicial da variável aleatória  $X$  for uma distribuição lognormal com parâmetros  $\lambda$  e  $\xi$ , os parâmetros da distribuição assintótica (Fréchet) são dados por:

$$v = \exp\left(\xi \left(\sqrt{2 \ln(n)} - \frac{\ln(\ln(n)) + \ln(4\pi)}{2\sqrt{2 \ln(n)}}\right) + \lambda\right) \quad (IV-28)$$

$$k = \frac{\sqrt{2 \ln(n)}}{\xi} \quad (IV-29)$$

### IV.3. Distribuição do Pico Extremo de um Processo Aleatório Gaussiano

No estudo de processos aleatórios muitas vezes o interesse recai no valor extremo do mesmo num intervalo (ou período) de tempo  $T$ . O valor extremo do processo corresponde ao valor máximo extremo de sua distribuição de picos. No caso de um processo aleatório Gaussiano a distribuição de picos é modelada pela distribuição de Rice (vide Seção III.8. ). Assumindo-se que os picos são estatisticamente independentes é possível chegar a distribuição do pico extremo através da estatística de ordem descrita anteriormente. Além disto é possível demonstrar (MADSEN et. al., 1986) que a distribuição do pico extremo converge para uma distribuição de Gumbel com os parâmetros definidos por:

$$u = \sqrt{m_0} \sqrt{2 \ln(\nu_0 T)} \quad (\text{IV-30})$$

$$\alpha = \frac{\sqrt{2 \ln(\nu_0 T)}}{\sqrt{m_0}} \quad (\text{IV-31})$$

onde  $\nu_0$  é a frequência de cruzamento zero do processo,  $m_0$  é o momento espectral de ordem zero (equivalente à variância do processo) e  $T$  é o período de tempo (em segundos) considerado na análise do extremo de curto prazo ( $T = 3600s, 7200s, 10800s, etc.$ ).

É interessante observar que a distribuição do pico extremo de um processo Gaussiano depende somente de parâmetros que podem ser obtidos a partir da sua densidade espectral e do período de tempo considerado. Adicionalmente, observa-se que a distribuição do valor extremo (e seus parâmetros) independem do fator de largura de banda do espectro.

No caso de processos aleatórios não-Gaussianos não existe uma solução analítica disponível para a obtenção da estatística do seu valor extremo. Por este motivo, são utilizadas técnicas numéricas para tal finalidade. Uma destas técnicas é, a partir de uma série temporal do processo, obter uma amostra dos picos, ajustar uma distribuição de probabilidades para esta amostra e usar a Estatística de Ordem para obter a distribuição do valor extremo. Este tópico, por ser o foco deste trabalho, será discutido em maiores detalhes mais adiante.

## CAPÍTULO V

### MÉTODOS PARA ESTIMATIVA DE VALORES EXTREMOS DE PROCESSOS NÃO-GAUSSIANOS

Como foi comentado, para o estudo dos valores extremos de processos gaussianos, existem soluções analíticas. Porém, quando o processo não segue uma distribuição gaussiana, devem-se aplicar outros métodos para estimar estes valores. Neste capítulo serão apresentadas algumas metodologias para estimativa de valores extremos voltados para processos não-gaussianos.

Existem na literatura vários procedimentos disponíveis para estimativa de valores extremos de processos não-gaussianos. Estes procedimentos podem ser classificados em:

- a) Métodos de transformação do processo
- b) Métodos baseados na frequência de cruzamentos
- c) Métodos baseados na distribuição dos picos

Os métodos de transformação utilizam o conceito de equivalência estatística (vide seção II.3. ) para transformar um processo não-Gaussiano num processo Gaussiano. Assim, as estimativas são feitas para o processo Gaussiano equivalente e depois, utilizando a transformada inversa retorna-se para o processo original. O método mais conhecido nesta linha é o Modelo baseado nos polinômios de Hermite proposto por WINTERSTEIN (1987). No presente trabalho, será apresentada também uma metodologia alternativa baseada na aplicação da SGLD, descrita no item II.6.3. .

Os métodos baseados na frequência de cruzamentos do processo aleatório assumem que para os níveis mais elevados, os cruzamentos podem ser modelados através de um processo de Poisson (NAESS & MOAN, 2012). Procedimentos baseados na frequência de cruzamentos não foram utilizados neste trabalho e maiores informações podem ser encontradas em (NAESS, 2001), (PAIVA JUNIOR, 2010), (NAESS & MOAN, 2012), entre outros.

Métodos baseados na distribuição dos picos são métodos que fazem um ajuste de uma distribuição do conjunto formado apenas pelos picos de uma única realização do processo e posteriormente, utilizando a Estatística de Ordem, obtém-se a distribuição dos valores extremos. Uma das distribuições mais usadas para representar os picos de um processo aleatório não-Gaussiano é a distribuição de Weibull (vide seção II.6.4. ). Porém, existem diversas maneiras de se fazer o ajuste, conforme pode ser visto em (NASCIMENTO, 2009).

Recentemente, um novo método denominado ACER (do inglês: *Average Conditional Exceedance Rate*) foi proposto por (NAESS & GAIDAI, 2013). A principal vantagem deste método é a consideração automática de uma possível dependência estatística entre os picos do processo, fato que é desprezado nos demais procedimentos. Neste trabalho, serão investigadas a distribuição de Weibull de 3 parâmetros, o método ACER e também o uso da distribuição SGLD para modelar os picos do processo aleatório.

A seguir será apresentada uma breve descrição de todos os métodos de estimativa de extremos para processos não-Gaussianos utilizados neste trabalho.

## V.1. Métodos de Transformação de Processos Aleatórios

A ideia fundamental dos métodos de transformação é a de transformar um processo aleatório  $X(t)$  num processo aleatório gaussiano equivalente  $U(t)$  com média zero e desvio padrão unitário. Esta transformação é obtida através da seguinte relação:

$$F_X(x) = \Phi(u) \quad (\text{V-1})$$

$$x = F_X^{-1}(\Phi(u)) \quad (\text{V-2})$$

$$u = \Phi^{-1}(F_X(x)) \quad (\text{V-3})$$

onde  $F_X(x)$  é função cumulativa de probabilidades  $X(t)$ ,  $\Phi(\cdot)$  é a função cumulativa de uma variável normal padrão e  $F_X^{-1}[\cdot]$  e  $\Phi^{-1}[\cdot]$  correspondem às suas inversas, respectivamente.

A importância da equivalência da Eq. (V-1) decorre na possibilidade de se estimar a estatística dos valores extremos (valor mais provável, distribuição dos picos, etc.) para o processo Gaussiano  $U(t)$  equivalente (Eq. (V-2)) e posteriormente fazer a transformação inversa (Eq. (V-3)) para obter os correspondentes valores para a variável original  $X(t)$ .

Como a transformação definida acima é monotônica, a frequência de cruzamentos pelo nível zero  $\nu_0$  de  $X(t)$  é igual à frequência de cruzamentos pela média  $\nu_m$  do processo  $U(t)$ . Assim, por exemplo, o valor extremo mais provável de  $X(t)$ , i.e.,  $x_{MPV}$ , é obtido por:

$$x_{MPV} = F_X^{-1}[\Phi(u_{MPV})] \quad (V-4)$$

Como consequência do anterior, é obtida uma variável Normal Padrão equivalente e, lembrando que para ela a Distribuição do Pico Extremo converge para a Distribuição Tipo I, o valor extremo mais provável de um processo Gaussiano (vide Eq. (IV-30))  $u_{MPV}$  para um período de tempo  $T$  é dado por:

$$u_{MPV} = \sqrt{m_0} \sqrt{2 \ln(\nu_0 T)} = \sqrt{2 \ln(\nu_0 T)} \quad (V-5)$$

onde  $\nu_0$  é a frequência de cruzamento zero dado na Eq. (III-18),  $\sqrt{m_0} = 1$  devido à que a variância de  $U(t)$  é unitária, e  $T$  é o período de tempo (em segundos) considerado na análise do extremo de curto prazo ( $T = 3600s, 7200s, 10800s$ ).

Como a distribuição de probabilidades de  $X(t)$  não é conhecida, neste trabalho foram investigados dois modelos de para representá-la: modelo baseado no polinômios de Hermite e um modelo baseado na SGLD. Estes modelos serão descritos a seguir.

### V.1.1. Método baseado nos Polinômios de Hermite

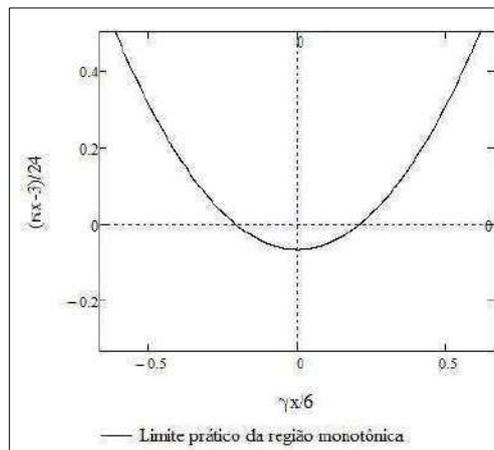
O método baseado nos Polinômios de Hermite usa o procedimento descrito na Seção II.6.6. para obter um modelo de transformação entre os dois espaços  $U(t)$  e  $X(t)$ . A definição dos coeficientes dos polinômios é obtida através do sistema de equações não-lineares, indicado na Eq. (II-71), usando a média, o desvio padrão, o coeficiente de assimetria e o coeficiente de curtose de uma amostra do processo aleatório  $X(t)$ .

Devemos mencionar que este modelo tem sido utilizado durante muitos anos e, inclusive, ele é recomendado por algumas normas de projeto de plataformas auto-elevatórias (também chamadas de “jack-up”), como por exemplo a (ABS , 2004) e a (ISO, 2006) na análise dos extremos de respostas dinâmicas. Porém, conforme ilustrado por (WINTERSTEIN,1987), existem valores de assimetria e curtose que este modelo não consegue representar, ou seja, a transformação torna-se não monotônica. Conforme (WINTERSTEIN,1987) o domínio de validade do modelo se restringe a casos em que a combinação de assimetria e curtose recai na região monotônica.

A região monotônica é definida pela Eq. (V-6), a seguir:

$$\left(\frac{\gamma_x}{6}\right)^2 \leq 3 \left(\frac{\kappa_x - 3}{24}\right) \left[1 - 3 \left(\frac{\kappa_x - 3}{24}\right)\right] \quad (V-6)$$

onde  $\gamma_x$  e  $\kappa_x$  correspondem, respectivamente, ao coeficiente de assimetria e à curtose de  $X(t)$ . Esta condição é ilustrada na *Figura V.1*.



*Figura V.1 – Região onde o modelo de Hermite é monotônico*

### V.1.2. Método baseado na SGLD

O método de transformação baseado na SGLD é bastante similar ao anterior, porém, neste caso usa-se a distribuição SGLD para representar  $F_X(x)$ . O ajuste dos parâmetros  $\theta, b, r$  e  $\sigma$  é feito através do método dos momentos. Em outras palavras, utiliza-se a média, o desvio padrão, o coeficiente de assimetria e a curtose de uma amostra do processo aleatório  $X(t)$  para, através da metodologia descrita no item II.6.3. , obter os parâmetros da SGLD. Uma vez obtidos os parâmetros desta distribuição, a sua função inversa, necessária para fazer a transformada inversa de  $U(t)$  para  $X(t)$ , é calculada através da Eq. (V-4).

Este método está sendo avaliado pela primeira vez neste trabalho. Deve ser enfatizado que ambos os métodos utilizam os quatro primeiros momentos da série temporal completa, pois  $F_X(x)$  representa neste caso a distribuição de probabilidade do processo aleatório e não a distribuição dos seus picos. Assim, neste trabalho o presente método será identificado como *SGLD<sub>Processo\_Completo</sub>*.

### V.2. Métodos Baseados na Distribuição dos Picos do Processo

Os métodos baseados na distribuição dos picos do processo aleatório consistem basicamente em ajustar uma distribuição  $F_{X_p}(x)$  aos picos observados (vide *Figura V.2*) numa série temporal do respectivo processo  $X(t)$  e utilizar a estatística de ordem para obter a distribuição de probabilidades do pico extremo para um período de tempo T, i.e.:

$$F_{X_e}(x) = \left[ F_{X_p}(x) \right]^{\nu_p T} \quad (\text{V-7})$$

onde  $\nu_p$  é a frequência de picos do processo aleatório.

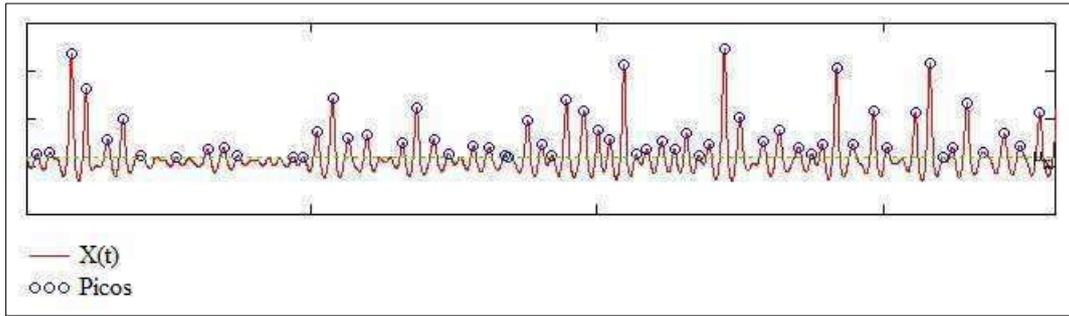


Figura V.2 – Picos da série temporal  $X(t)$

Dependendo do tipo de distribuição utilizada para ajustar a  $F_{X_p}(x)$ , a distribuição de extremos  $F_{X_e}(x)$  pode ser representada por sua forma assintótica conforme descrito no item IV.2. Porém, na grande maioria das vezes o interesse recai no valor extremo mais provável  $x_{MPV}$  e este pode ser obtido através da solução da seguinte equação:

$$F_{X_p}(x_{MPV}) = 1 - \frac{1}{v_p T} \quad (V-8)$$

Neste trabalho foram investigados os usos de uma distribuição de Weibull e de uma distribuição SGLD para representar o comportamento dos picos do processo aleatório  $X(t)$ . É interessante observar que a Eq. (V-8) assume a independência estatística entre os picos do processo aleatório. Por este motivo, adicionalmente, também foi investigado um método proposto recentemente, o ACER, que leva em conta uma possível dependência estatística entre os picos do processo aleatório investigado.

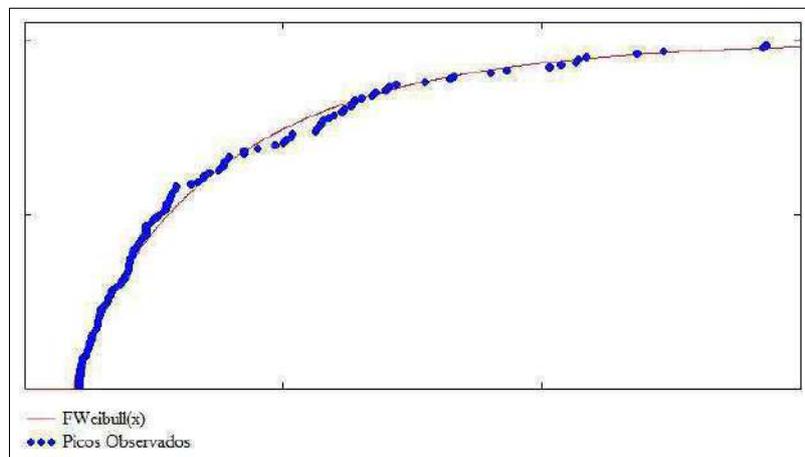
### V.2.1. Método baseado na Distribuição de Weibull

Neste procedimento, assume-se que os picos do processo aleatório podem ser representados por uma distribuição  $F_{X_p}(x)$  definida pela distribuição de Weibull. Entretanto, existem várias maneiras de obter os parâmetros desta distribuição incluindo o método dos momentos, regressão linear, regressão linear mais voltada para a cauda superior dos dados (*Weibull-Tail*, *Weibull-Peaks-over-Threshold*), etc.

Neste trabalho utilizou-se a distribuição de Weibull de três parâmetros onde estes são obtidos pelo método dos momentos. Dispondo-se da média, desvio padrão, coeficiente de assimetria e curtose dos picos de  $X(t)$ , conforme o item II.6.4., dispõe-se de quatro equações (II-51) até (II-54) e três incógnitas para definição dos parâmetros da

distribuição. Como o coeficiente de assimetria e a curtose dependem unicamente do mesmo parâmetro  $\lambda$ , a definição dos parâmetros da distribuição pode ser feita de duas maneiras: utilizando o coeficiente de assimetria ou o coeficiente de curtose para o cálculo de  $\lambda$ . Os outros dois parâmetros ( $u$  e  $\alpha$ ) são obtidos em função da média e do desvio padrão e do próprio valor calculado de  $\lambda$ .

Baseado no trabalho de (NASCIMENTO, 2009), nesta pesquisa somente foi utilizado o coeficiente de assimetria para a definição dos parâmetros da distribuição de Weibull. Para avaliar a razoabilidade do ajuste, a função cumulativa empírica da amostra dos picos observados é comparada com a distribuição de Weibull ajustada, conforme *Figura V.3*.



*Figura V.3 – Ajuste dos picos observados a uma distribuição de Weibull*

Neste caso específico, a distribuição de extremos converge para uma distribuição de Gumbel com os parâmetros de locação  $u_G$  e de escala  $\alpha_G$  dados, respectivamente, por:

$$u_G = u + \alpha [\ln(v_p T)]^{1/\lambda} \quad (V-9)$$

$$\alpha_G = u + \frac{\lambda}{\alpha} [\ln(v_p T)]^{\frac{\lambda-1}{\lambda}} \quad (V-10)$$

Deve ser observado que o valor extremo mais provável do processo aleatório coincide com o parâmetro de locação  $u_G$ , i.e.,

$$x_{MPV} = u_G \quad (V-11)$$

## V.2.2. Método baseado na SGLD

Nesta metodologia assume-se que a distribuição de probabilidades dos picos do processo aleatório  $F_{X_p}(x)$  pode ser representada por uma SGLD (*Shifted Generalized Lognormal Distribution*). Da mesma forma que para a distribuição de Weibull, existem várias técnicas que podem ser utilizadas para obter os parâmetros desta distribuição. Neste trabalho foram utilizados dois procedimentos, descritos a seguir.

O primeiro procedimento, denominado  $SGLD_{Momentos\_dos\_picos}$ , utiliza a técnica dos momentos, conforme descrito na Seção II.6.3. , para calcular os parâmetros da distribuição a partir da média, desvio padrão, coeficiente de assimetria e curtose da amostra dos picos de uma realização (série temporal) do processo aleatório  $X(t)$ .

No segundo procedimento, denominado  $SGLD_{Função\_cumulativa\_dos\_picos}$ , os parâmetros da SGLD são calculados de forma que a função cumulativa desta distribuição teórica (SGLD) se ajuste ao máximo possível à função cumulativa empírica dos dados. Deve ser observado que na análise de extremos o interesse recai na cauda superior da distribuição e por este motivo, e através de alguns testes, fez-se o ajuste para somente um conjunto limitado de 12 valores de picos, i.e.,  $\{x_i\}, i = 1, 2, \dots, 12$ ; associados aos seguintes níveis (probabilidades) da função cumulativa empírica:  $P_i = \{0.70; 0.75; 0.80; 0.85; 0.90; 0.95; 0.975; 0.98; 0.99; 0.995; 0.9975; 0.999\}$ . Nestes pontos, o ajuste deve ser o melhor possível e isto pode ser feito através da solução de um problema de otimização em que os parâmetros da SGLD, i.e.,  $b, \theta, r$  e  $\sigma$  sejam aqueles que minimizem a função de erro quadrático definida abaixo:

$$E(b, \theta, r, \sigma) = \frac{1}{N} \sum_{i=1}^N \left( F_{X_p}(x_i, b, \theta, r, \sigma) - P_i \right)^2 ; N = 12 \quad (V-12)$$

onde  $F_{X_p}(\cdot)$  é a função cumulativa da SGLD.

Como ainda não existe uma distribuição assintótica de extremos para quando a distribuição inicial é uma SGLD, a distribuição de extremos pode ser obtida através da Eq. (IV-7) para os dois procedimentos baseados na SGLD. Da mesma forma, o valor

extremo mais provável, i.e., o  $x_{MPV}$  do processo aleatório  $X(t)$  pode ser estimado através das Eqs. (V-4) e (V-5). Deve ser observado que neste último caso é necessário utilizar a função inversa da SGLD definida na Eq. (II-48).

### V.2.3. Average Conditional Exceedance Rate (ACER)

Podemos observar que os dois métodos de ajuste anteriores (Weibull e SGLD) são baseados nas propriedades estatísticas das medições dos picos (média, desvio padrão, skewness e kurtosis) sem levar em consideração uma possível dependência estatística entre eles nem sua ordem de aparecimento. Em outras palavras, os métodos anteriores assumem que os picos do processo aleatório são estatisticamente independentes.

Na intenção de desenvolver uma metodologia que permitisse capturar informações da dependência estatística entre os picos da série temporal, pesquisadores da Universidade Norueguesa de Ciência e Tecnologia (*Norges Teknisk-Naturvitenskapelige Universitet*, NTNU) (NAESS 2001, NAESS 2009A, NAESS 2009B, NAESS 2013, KARPA, 2015) desenvolveram um método conhecido como “Taxa Média de Excedências Condicionadas”, originalmente batizado em inglês como: *Average Conditional Exceedance Rate* (abreviada: ACER). Esta técnica utiliza uma cascata de dependências estatísticas entre picos sucessivos para criar uma função de extrapolação da cauda da distribuição de extremos, a qual é capaz de estimar o valor extremo mais provável levando em conta a dependência entre os picos do processo. Assim, o caso de picos estatisticamente independentes torna-se um caso particular da metodologia.

O método consiste em identificar a sequência dos  $N$  picos de uma amostra do processo aleatório  $X(t)$  e calcular a função ACER para um determinado pico observado  $X_j$  e um nível de excedências  $\eta$ , representada por  $\varepsilon_{kj}(\eta)$ , através da seguinte equação:

$$\varepsilon_{kj}(\eta) = Prob(X_j > \eta | X_{j-1} \leq \eta, \dots, X_{j-k+1} \leq \eta), 1 \leq k \leq j \leq N \quad (V-13)$$

que pode ser descrita como: **a probabilidade de ter um “j-ésimo” pico  $X_j$  maior do que o valor “ $\eta$ ”, dado que todos os “ $j - 1$ ” picos anteriores não foram maiores do que “ $\eta$ ”.**

Nas funções ACER, o índice  $k$  é um inteiro positivo que indica o número de excedências condicionadas consideradas entre picos próximos. Assim, os índices  $k$  permitem calcular cada uma das distintas “ $k$ -ésimas” funções ACER. Por exemplo, as primeiras quatro funções ACER são dadas pelas seguintes expressões:

$$\varepsilon_{1j}(\eta) = Prob(X_j > \eta) ; k = 1 \quad (V-14)$$

$$\varepsilon_{2j}(\eta) = Prob(X_j > \eta | X_{j-1} \leq \eta) ; k = 2 \quad (V-15)$$

$$\varepsilon_{3j}(\eta) = Prob\left(X_j > \eta | X_{j-1} \leq \eta \bigcap X_{j-2} \leq \eta\right) ; k = 3 \quad (V-16)$$

$$\varepsilon_{4j}(\eta) = Prob\left(X_j > \eta | X_{j-1} \leq \eta \bigcap X_{j-2} \leq \eta \bigcap X_{j-3} \leq \eta\right) ; k = 4 \quad (V-17)$$

Note-se que cada uma das Funções ACER são funções de probabilidades de excedências, devido à que elas estão relacionadas às ocorrências acima do nível  $\eta$ . Empiricamente, esta distribuição é estimada contando-se o número de ocorrências condicionadas e dividindo-se este número pelo número total de picos.

A *Figura V.4* apresenta os comportamentos das seis primeiras funções ACER ( $k = 1, 2, 3, 4, 5, 6$ ) numa escala vertical logarítmica. É razoável assumir que as funções diminuem na medida em que aumenta a quantidade de dependências estatísticas (índice  $k$ ) consideradas e/ou aumenta o nível de excedência  $\eta$ . Outra coisa a salientar é que o índice  $k$ , em geral, é pequeno devido à que para valores maiores a função tende ao zero.

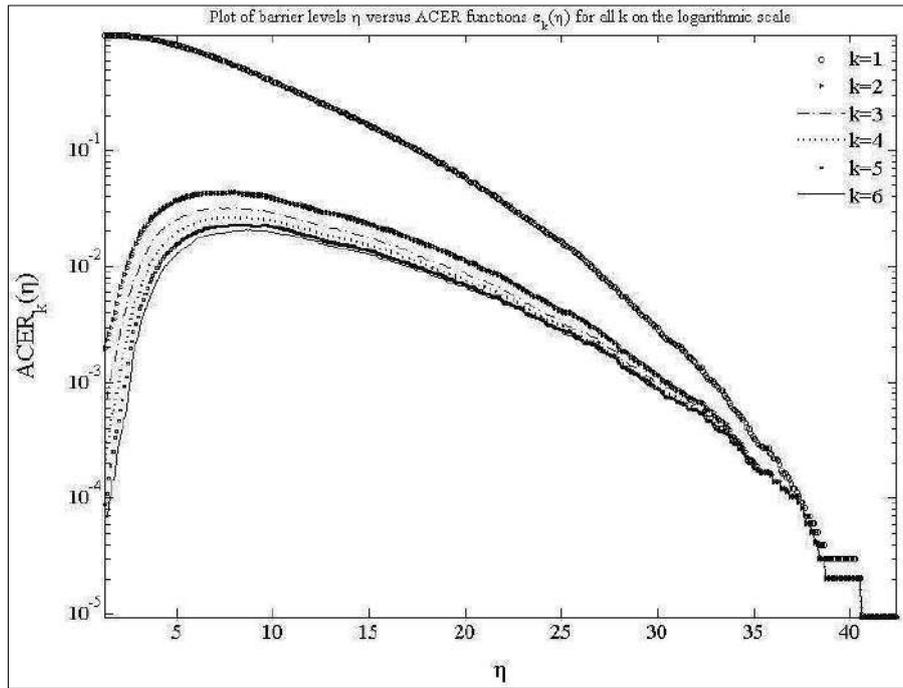


Figura V.4 – Exemplo dos gráficos das primeiras 6 funções ACER

Isto posto, o segundo passo é obter funções contínuas  $\varepsilon_{k_j}(\eta)$  que permitam extrapolar o comportamento da mesma para valores mais elevados do nível  $\eta$ . Conforme exposto por (NAESS, 2012) foi observado que o comportamento da cauda de uma função ACER é tipicamente dominado por uma função exponencial natural ( $e^{-p}$ ). Assim, (NAESS, 2013) propôs o seguinte modelo para representar uma função ACER:

$$\varepsilon_k(\eta) = q_k \exp(-a_k(\eta - b_k)^{c_k}) \quad (V-18)$$

onde  $\eta$  é o nível de excedências e as constantes  $a_k, b_k, c_k$  e  $q_k$  são aquelas que melhor ajustam (minimizam) o comportamento da distribuição empírica focando em níveis mais elevados de  $\eta$ . Estes parâmetros devem ser determinados utilizando métodos de otimização, por exemplo, no trabalho original (NAESS, 2013) foi utilizado o algoritmo de Levenberg–Marquardt.

Esta cascata de dependências tem um significado prático devido ao fato que implicitamente é assumido que existe um  $k_{exato} \ll N$ , tal que avaliado na função  $F_{k_{exato}}(\eta)$  indicará que existe um único valor maior do que ele, o que garante que a  $F_{k_{exato}}(\eta)$  é equivalente a distribuição exata do valor extremo.

Finalmente, a função cumulativa do valor do pico extremo é dada pela seguinte distribuição:

$$F_k(\eta) = \exp \left[ - \left( N_p(\varepsilon_k(\eta)) \right) \right] \quad (\text{V-19})$$

onde  $N_p$  é o número total de picos para o tempo de duração considerado para estimativa do valor extremo. Assim, sabendo-se que a  $F_k(\eta)$  representa a distribuição dos extremos e que ela é contínua e inversível, o  $x_{\text{MPV}}$  é calculado através da resolução da Eq. (V-8).

Na prática, o nível  $\eta$  a ser considerado é aquele que torna estável a função  $\varepsilon_{kj}(\eta)$  e as correspondentes para maiores valores de  $k$ . O nível  $k + 1$  pode ser visto como o número de picos que apresentam alguma dependência estatística entre si. No presente trabalho foi utilizado um programa desenvolvido por (KARPA, 2015) (disponível em <https://folk.ntnu.no/arvidn/ACER/>) para efetuar as análises pelo método ACER e foram consideradas as três primeiras dependências ( $k = 1, 2, 3$ ) para avaliar a influência de considerar distintas dependências estáticas entre os picos na análise de extremos.

### V.3. Estimativa da incerteza das aproximações

Todos os métodos de estimativa de extremos descritos nas seções V.1. e V.2. são baseados em realizações (séries temporais) de duração finita e, conseqüentemente, as estimativas relacionadas às suas propriedades estatísticas dependem do tempo total considerado na simulação do processo (tamanho da série). Assim sendo, na medida em que se aumenta a duração da simulação, maior será a quantidade de informações extraídas a partir dela e conseqüentemente será menor a incerteza nos parâmetros estatísticos estimados.

Por exemplo, nos métodos baseados na amostra de picos são utilizados parâmetros estatísticos, tais como: média, desvio padrão, coeficiente de assimetria e curtose, que são estimados a partir desta amostra. Como estes parâmetros apresentam uma variabilidade em função do tamanho da amostra, os valores extremos estimados a partir deles também apresentarão alguma incerteza estatística.

Por outro lado, os métodos utilizados não são exatos e, portanto, mesmo que não exista nenhuma incerteza nos parâmetros estatísticos da série temporal, os valores extremos podem apresentar alguma tendência com relação ao valor exato. A incerteza final nos valores extremos estimados é composta então de duas fontes: incerteza estatística dos parâmetros da série temporal e a incerteza do modelo utilizado. Porém, este aspecto é bastante difícil de ser calculado analiticamente.

Na intenção de avaliar a tendência dos valores extremos estimados (calculados pelos diferentes métodos investigados) e também a incerteza dos mesmos, neste trabalho foi feita uma investigação numérica com relação ao valor extremo mais provável. A melhor possibilidade para obter a distribuição de valores extremos “exata” de um processo aleatório é obter  $N$  realizações distintas do mesmo (vide *Figura V.5*), todas elas com tamanho igual ao tempo de interesse, e constituir uma amostra de valores extremos que é representada pelos  $N$  máximos valores individuais observados em cada uma delas.

A partir da amostra dos valores extremos ajusta-se uma distribuição de probabilidades e estimam-se todos os parâmetros estatísticos de interesse, entre eles, por exemplo, o valor extremo mais provável  $x_{MPV}$ . Como na Teoria de Extremos a distribuição Tipo I é a distribuição assintótica de extremos para a grande maioria das distribuições iniciais, é comum se tentar representar a amostra de extremos por uma distribuição Tipo I ou de Gumbel (vide item IV.2.1. ).

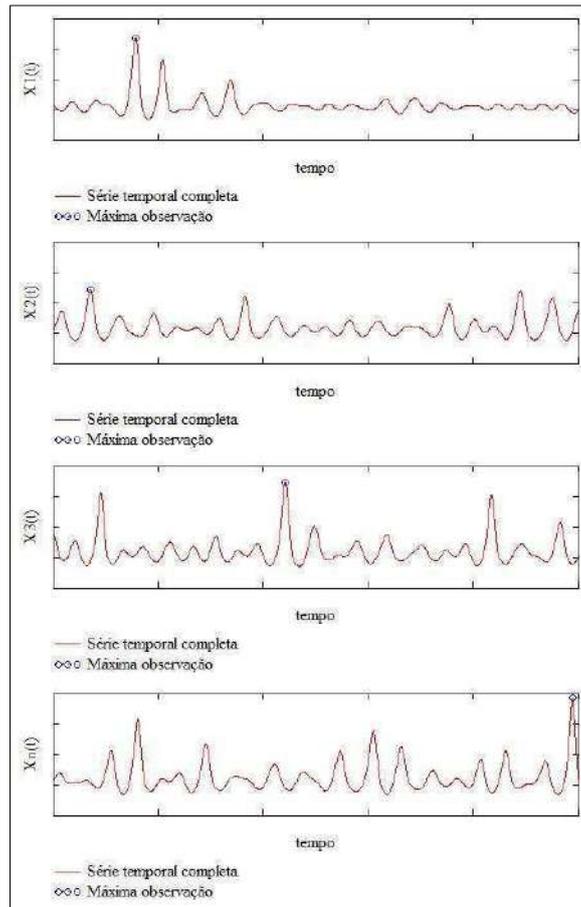


Figura V.5 – Seleção dos valores máximos observados nas  $N$  realizações independentes

Utilizando-se o método dos momentos, os parâmetros  $u$  e  $\alpha$  da distribuição da distribuição Tipo I são dados por:

$$\alpha = \frac{\pi}{\sigma_m \sqrt{6}} \quad (V-20)$$

$$u = \mu_m - \frac{0.57722\sqrt{6}}{\pi} \sigma_m \quad (V-21)$$

onde  $\mu_m$  e  $\sigma_m$  são, respectivamente, a média e o desvio padrão da amostra de valores extremos. É interessante ressaltar que para esta distribuição o valor mais provável é um dos parâmetros da distribuição, i.e.,  $x_{MPV} = u$ .

De posse das  $N$  realizações distintas do processo aleatório é possível utilizá-las individualmente para se fazer as estimativas, por exemplo, do valor extremo mais provável através de cada um dos métodos descritos anteriormente. Adicionalmente, os

comprimentos das séries temporais utilizadas nestas estimativas podem ser maiores, iguais ou menores ao tempo de interesse da análise.

Assim sendo, para cada método e para cada tamanho de simulação podem ser obtidas  $N$  estimativas do valor extremo mais provável. Através destas estimativas pode ser calculada a média  $\mu_{MPV_{T_{sim}}^{Met}}$  e o desvio padrão  $\sigma_{MPV_{T_{sim}}^{Met}}$  do conjunto formado pelos valores extremos estimados por um dado método e para um dado tempo de simulação do processo  $T_{sim}$ . Deve ser observado que o valor extremo mais provável “exato” é sempre estimado para o tempo de interesse  $T$  (no presente trabalho,  $T = 3h = 10800s$ ).

De posse de tais estimativas pode-se estimar a variabilidade na estimativa de um dado método em função do comprimento da série, através do coeficiente de variação ( $CoV$ ):

$$CoV_{MPV_{T_{sim}}^{Met}} = \frac{\sigma_{MPV_{T_{sim}}^{Met}}}{\mu_{MPV_{T_{sim}}^{Met}}} \quad (V-22)$$

e a tendência (ou enviesabilidade) de cada método pode ser estimada através da seguinte expressão:

$$B_{MPV_{T_{sim}}^{Met}} = \frac{\mu_{MPV_{T_{sim}}^{Met}}}{u} \quad (V-23)$$

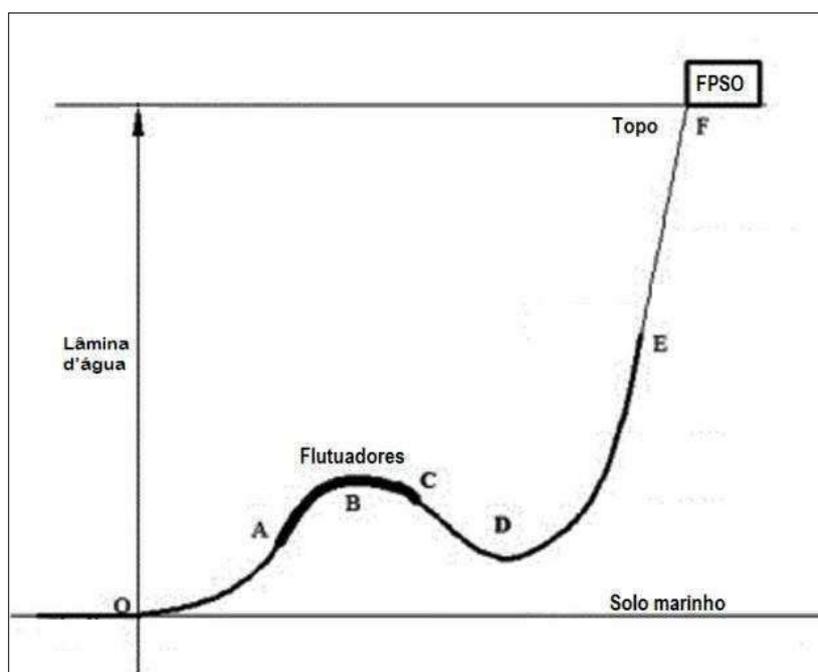
Idealmente, procura-se por um método não tendencioso ( $B_{MPV_{T_{sim}}^{Met}} = 1$ ) e com a menor variabilidade possível ( $\delta_{MPV_{T_{sim}}^{Met}} \rightarrow 0$ ).

## CAPÍTULO VI

### ESTUDOS DE CASOS

Neste capítulo são apresentados três estudos de caso que foram utilizados para avaliar a aplicabilidade dos métodos para estimativa de valores extremos apresentados no CAPÍTULO V. Em particular, calculam-se valores extremos mais prováveis para séries temporais de 3 horas de Fatores de Utilização (FU), que são utilizados no projeto de risers metálicos. Neste trabalho, optou-se por calcular estes fatores utilizando o critério LRFD da norma DNV (DnV-OS-F201. 2010). Uma descrição geral deste critério de projeto é apresentada no ANEXO A.

Todos os três casos de estudo envolvem risers na configuração Lazy-Wave (abreviado SLWR, do inglês *Steel Lazy Wave Riser*) conectado a unidades flutuantes do tipo FPSO (*Floating, Production, Storage and Offloading*), vide *Figura VI.1*. O primeiro caso analisado se refere a uma seção na região dos flutuadores (ponto B da *Figura VI.1*) de um SLWR de 8” conectado a um FPSO numa lâmina d’água de 914 metros; o segundo caso está relacionado à seção do topo (ponto F da *Figura VI.1*) de um SLWR de 10” conectado a um FPSO numa lâmina d’água de 1400 metros; e o terceiro caso corresponde à um ponto na região dos flutuadores (ponto B da *Figura VI.1*) deste mesmo último riser.



*Figura VI.1 – Riser na configuração Steel Lazy Wave*

Para avaliar a acurácia das estimativas, as séries temporais dos Fatores de Utilização foram geradas com diferentes durações (tempos de simulação). Para cada tempo de simulação foram geradas 50 realizações (simulações) distintas, e para cada uma destas realizações foram aplicados todos os métodos para estimativa de valores extremos. As Tabelas 2 a 4 apresentam um resumo dos casos analisados.

*Tabela 2. Resumo dos principais dados relativos ao Caso de Estudo 1*

<b>Descrição</b>	<b>Dados</b>
Altura Significativa de Onda (Hs)	8.20 m
Período de Cruzamento Zero (Tz)	10.00 s
Lâmina d'água (LDA)	914.00 m
Configuração do riser	Lazy Wave
Tipo de Plataforma	FPSO
Seção do riser analisada	Região dos flutuadores
Tempo para análise de valor extremo	10800 s = 3 horas
Passo da simulação ( $\Delta t$ )	0.25 s
Número de realizações distintas (n)	50
Tempos de simulação das séries ( $T_{sim}$ )	3600, 5400, 7200, 10800, 14400, 18000 e 21600 segundos

*Tabela 3. Resumo dos principais dados relativos ao Caso de Estudo 2*

<b>Descrição</b>	<b>Dados</b>
Altura Significativa de Onda (Hs)	5.20 m
Período de Cruzamento Zero (Tz)	8.28 s
Lâmina d'água (LDA)	1400.00 m
Configuração do riser	Lazy Wave
Tipo de Plataforma	FPSO
Seção do riser analisada	Topo do riser
Tempo para análise de valor extremo (T)	10800 s = 3 horas
Passo da simulação ( $\Delta t$ )	0.25 s
Número de realizações distintas (n)	50
Tempos de simulação das séries ( $T_{sim}$ )	3600, 5400, 7200 e 10800 segundos

*Tabela 4. Resumo dos principais dados relativos ao Caso de Estudo 3*

<b>Descrição</b>	<b>Dados</b>
Altura Significativa de Onda (Hs)	5.20 m
Período de Cruzamento Zero (Tz)	8.28 s
Lâmina d'água (LDA)	1400.00 m
Configuração do riser	Lazy Wave
Tipo de Plataforma	FPSO
Seção do riser analisada	Região dos flutuadores
Tempo para análise de valor extremo (T)	10800 s = 3 horas
Passo da simulação ( $\Delta t$ )	0.25 s
Número de realizações distintas (n)	50
Tempos de simulação das séries ( $T_{sim}$ )	3600, 5400, 7200 e 10800 segundos

## VI.1. Análise Estatística das Séries Temporais dos Fatores de Utilização

Em conformidade com as três tabelas anteriores, foram geradas séries temporais com diferentes comprimentos (tempos de simulação). Porém, deve-se fazer uma especial ênfase nas séries com durações iguais a 3 horas (10800s) devido à que é para este tempo que estão sendo calculados os valores extremos mais prováveis.

Para cada uma das 50 séries temporais individuais de comprimento igual a 3 horas (10800s) dos Fatores de Utilização são calculados os principais parâmetros estatísticos (média, desvio padrão, coeficiente de assimetria, coeficiente de curtose e máximos observados). Estes resultados são mostrados em detalhes no ANEXO D. De maneira resumida, são apresentadas nas Tabela 5 a 7 as médias, os desvios padrões e os coeficientes de variação dos 50 valores obtidos para os principais parâmetros estatísticos destas séries temporais.

*Tabela 5. Estudo de Caso 1: Parâmetros estatísticos das 50 realizações de 3 horas*

	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Máx. Observ.
Média	0.120400	0.042666	4.471900	43.630000	0.853500
Desvio Padrão	0.000333	0.002209	0.906400	18.523000	0.151800
Coef. Var.	0.002765	0.051772	0.202690	0.424540	0.173570

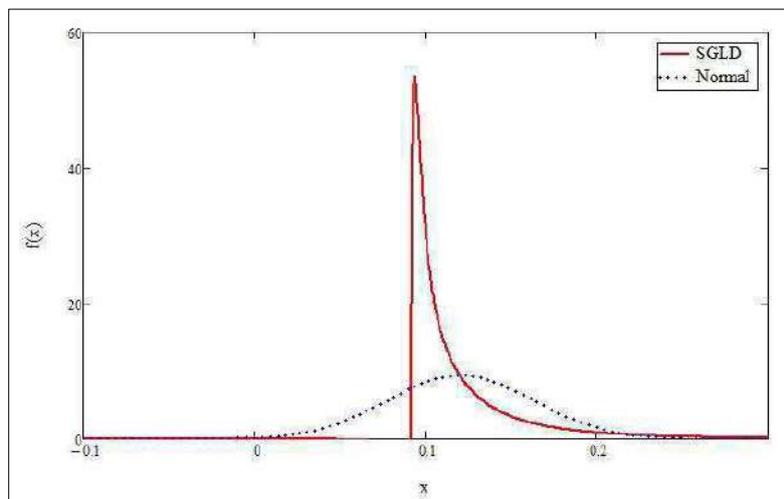
*Tabela 6. Estudo de Caso 2: Parâmetros estatísticos das 50 realizações de 3 horas*

	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Máx. Observ.
Média	0.271690	0.067382	0.060876	2.961300	0.543050
Desvio Padrão	0.000571	0.000356	0.030314	0.079027	0.025737
Coef. Var.	0.002103	0.005283	0.497970	0.026686	0.047394

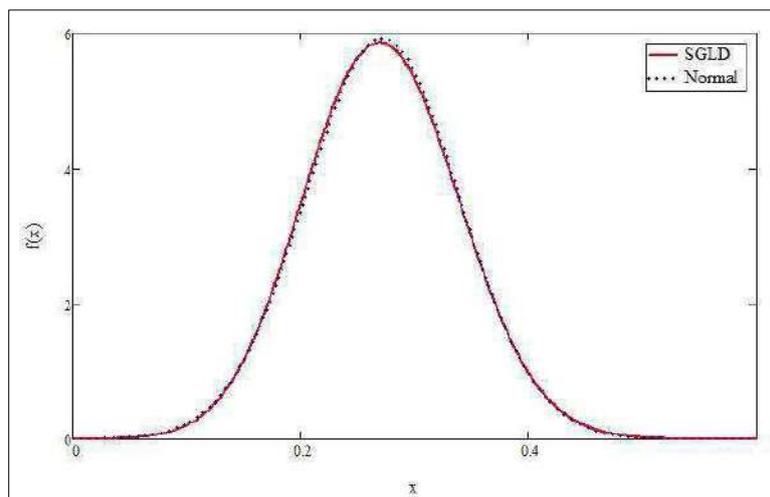
*Tabela 7. Estudo de Caso 3: Parâmetros estatísticos das 50 realizações de 3 horas*

	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Máx. Observ.
Média	0.111170	0.012224	1.078200	6.016600	0.207610
Desvio Padrão	0.000026	0.000123	0.117050	0.977270	0.016536
Coef. Var.	0.000232	0.010039	0.108560	0.162430	0.079649

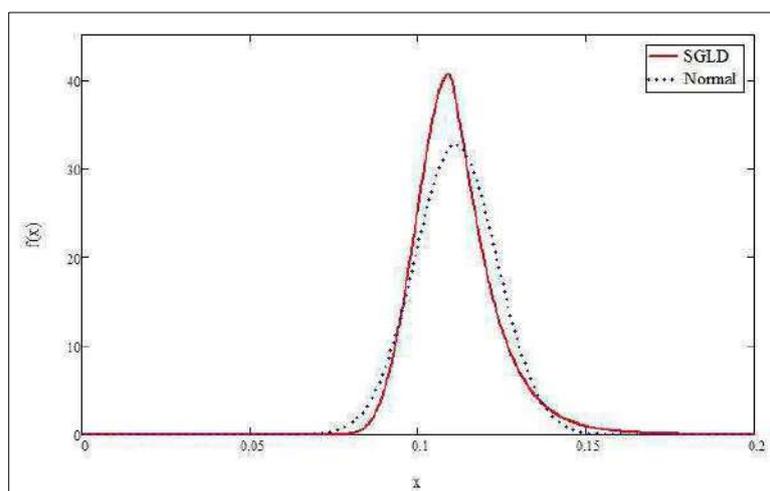
Conforme mencionado no CAPÍTULO II a SGLD possui uma grande versatilidade e pode representar vários tipos de distribuição. Aqui neste ponto ela foi usada para ilustrar comparativamente em relação à distribuição gaussiana o comportamento dos três processos aleatórios investigados. Tomou-se as médias dos parâmetros estatísticos das 50 séries de 3 horas (média, desvio padrão, coeficientes de assimetria e curtose nas Tabela 5 a 7) para ajustar esta distribuição, conforme descrito no item II.6.3. As *Figura VI.2 a VI.4.* ilustram esses resultados. Estas figuras apenas corroboram as observações descritas acima a respeito dos comportamentos (Gaussiano ou não-Gaussiano) dos três casos investigados.



*Figura VI.2 – Estudo de Caso 1: Comparação entre a distribuição do processo ajustada a uma SGLD e uma distribuição Normal*



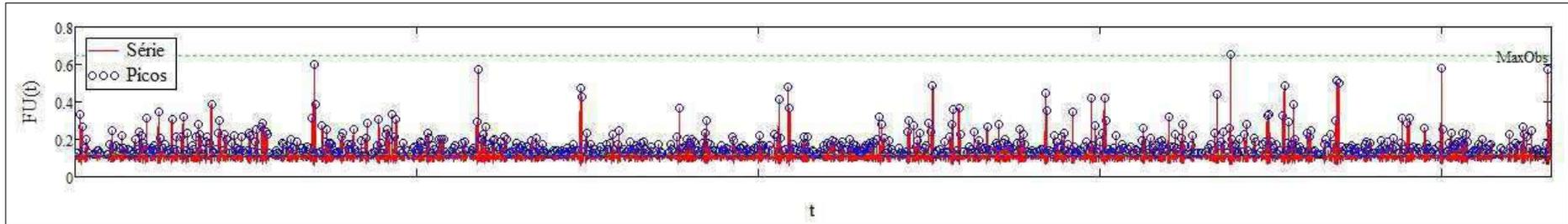
*Figura VI.3 – Estudo de Caso 2: Comparação entre a distribuição do processo ajustada a uma SGLD e uma distribuição Normal*



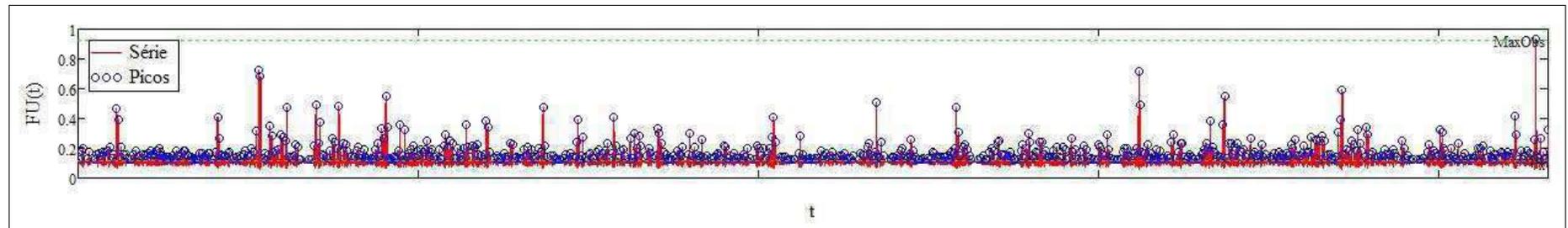
*Figura VI.4 – Estudo de Caso 3: Comparação entre a distribuição do processo ajustada a uma SGLD e uma distribuição Normal*

Como pode ser observado, o Caso de Estudo 1 apresenta valores de coeficiente de assimetria e curtose bastante distantes de um processo Gaussiano. O comportamento não-Gaussiano também é observado no Caso de Estudo 3. O Caso de Estudo 2 é o que apresenta parâmetros estatísticos relativos à assimetria e curtose mais próximos do caso Gaussiano (0 e 3, respectivamente).

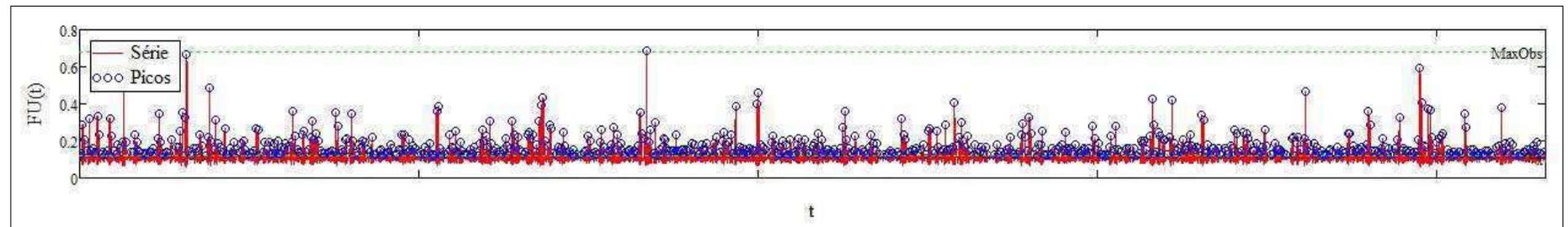
Para ilustrar melhor o que foi dito anteriormente, nas seguintes *Figuras* são apresentadas, para cada Caso de Estudo, três realizações de 3h das séries de Fatores de Utilização. Aproveitando estas figuras, são destacados também os Picos maiores da média e a máxima observação de cada série. Principalmente nos Casos de Estudo 1 e 3, observa-se o comportamento acentuadamente assimétrico das séries temporais, fato que corrobora para identificar o comportamento não-Gaussiano das mesmas.



*Figura VI.5 – Estudo de Caso 1: Realização #2 do processo com duração de 10800 segundos (3 horas)*



*Figura VI.6 – Estudo de Caso 1: Realização #10 do processo com duração de 10800 segundos (3 horas)*



*Figura VI.7 – Estudo de Caso 1: Realização #20 do processo com duração de 10800 segundos (3 horas)*

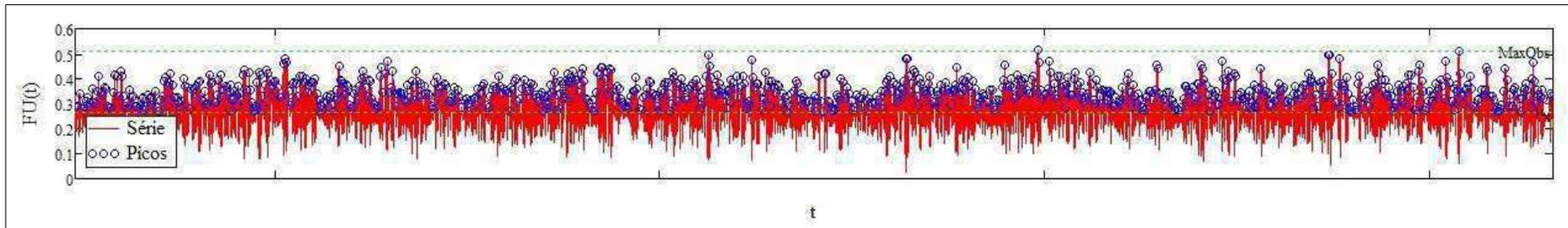


Figura VI.8 – Estudo de Caso 2: Realização #22 do processo com duração de 10800 segundos (3 horas)

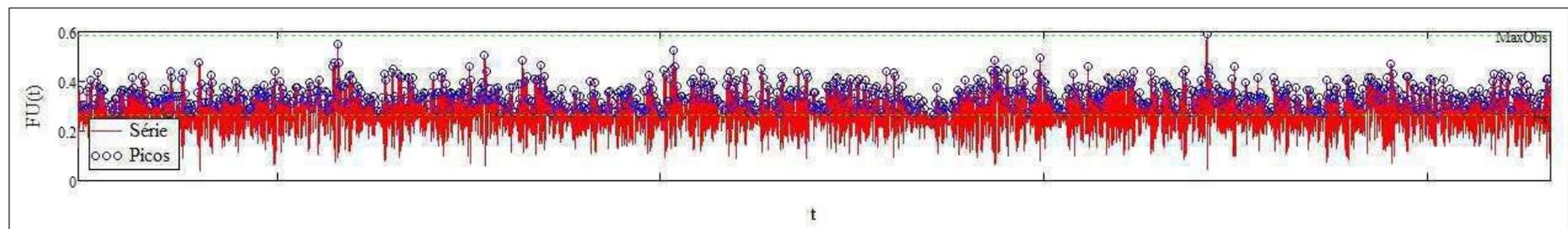


Figura VI.9 – Estudo de Caso 2: Realização #35 do processo com duração de 10800 segundos (3 horas)

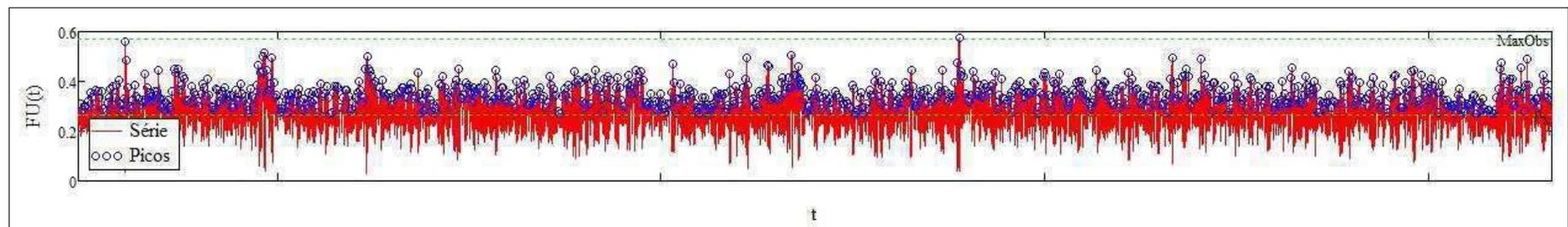


Figura VI.10 – Estudo de Caso 2: Realização #40 do processo com duração de 10800 segundos (3 horas)

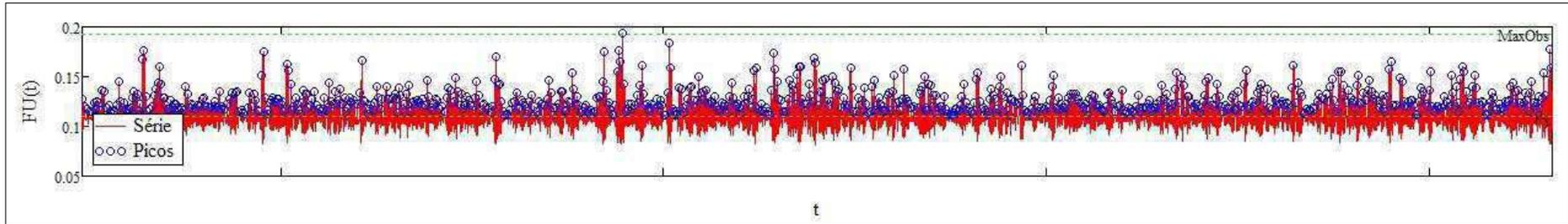


Figura VI.11 – Estudo de Caso 3: Realização #37 do processo com duração de 10800 segundos (3 horas)

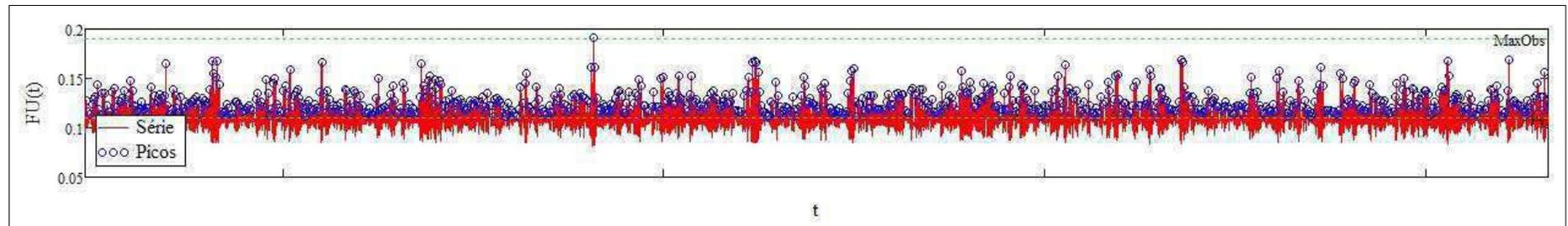


Figura VI.12 – Estudo de Caso 3: Realização #38 do processo com duração de 10800 segundos (3 horas)

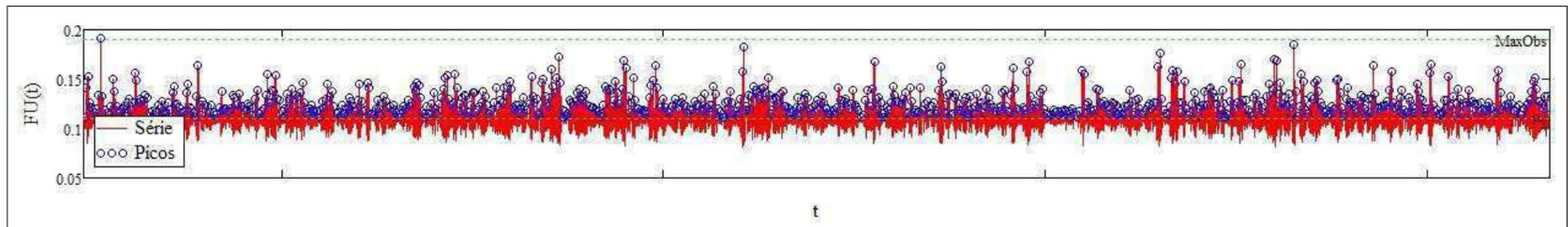


Figura VI.13 – Estudo de Caso 3: Realização #50 do processo com duração de 10800 segundos (3 horas)

## VI.2. Análise Estatística dos Picos das Séries Temporais dos Fatores de Utilização

Consoante com o que foi explicado no item V.2. , uma abordagem na análise de extremos consiste na geração de uma amostra selecionando apenas os picos do processo (vide *Figura V.2*) e posteriormente utilizar a Estatística de Ordem para obter a distribuição do Pico Extremo.

Da mesma maneira, para cada uma das 50 séries temporais individuais de comprimento igual a 3 horas (10800s) dos Fatores de Utilização são calculados os principais parâmetros estatísticos dos Picos (média, desvio padrão, coeficiente de assimetria, coeficiente de curtose e coeficiente de correlação entre picos consecutivos). Estes resultados são mostrados em detalhes no ANEXO E, e de maneira resumida são apresentadas as médias, os desvios padrões e os coeficientes de variação destes parâmetros nas Tabelas Tabela 8 a 10.

*Tabela 8. Estudo de Caso 1: Parâmetros estatísticos dos picos das 50 realizações de 3h*

	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Correl. Picos Consec.
Média	0.17559	0.07584	3.87540	26.29900	0.30260
Desvio Padrão	0.00210	0.00762	0.79855	10.51100	0.04156
Coef. Var.	0.01196	0.10047	0.20606	0.39967	0.13734

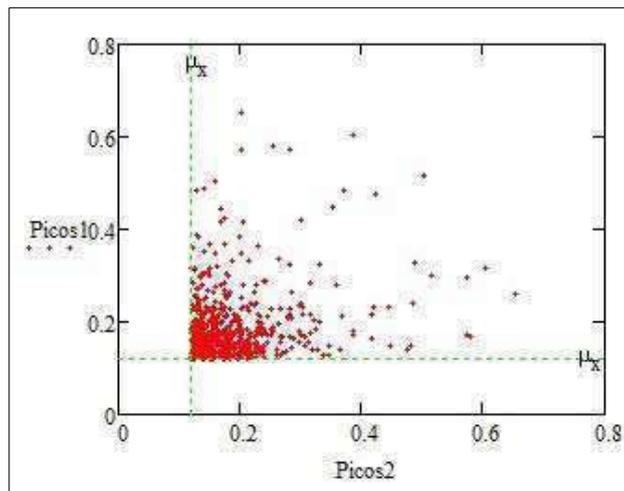
*Tabela 9. Estudo de Caso 2: Parâmetros estatísticos dos picos das 50 realizações de 3h*

	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Correl. Picos Consec.
Média	0.34975	0.04666	0.67029	3.32620	0.44083
Desvio Padrão	0.00117	0.00095	0.08354	0.31094	0.02363
Coef. Var.	0.00336	0.02036	0.12463	0.09348	0.05359

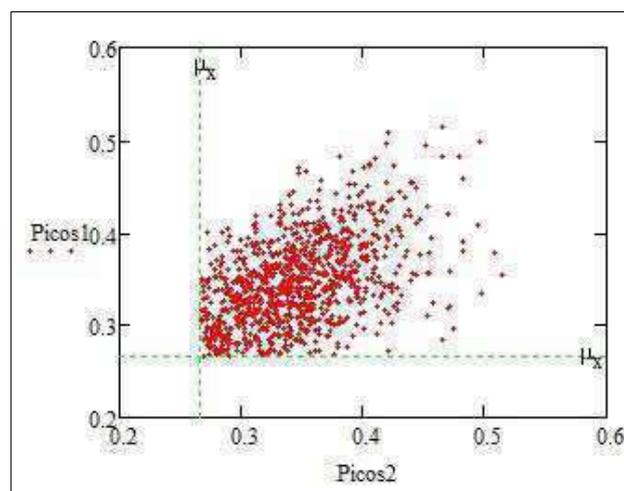
*Tabela 10. Estudo de Caso 3: Parâmetros estatísticos dos picos das 50 realizações de 3h*

	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Correl. Picos Consec.
Média	0.12572	0.01238	1.90180	8.89470	0.50339
Desvio Padrão	0.00029	0.00044	0.28260	2.47060	0.02765
Coef. Var.	0.00230	0.03589	0.14860	0.27776	0.05493

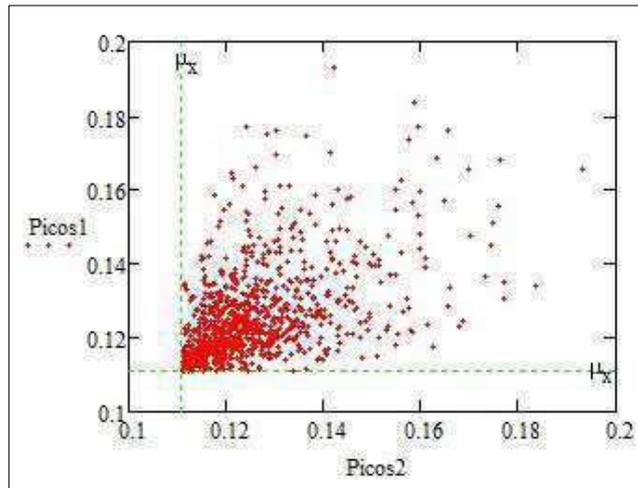
Para o cálculo do coeficiente de correlação entre picos consecutivos, são criados dois conjuntos diferentes com os picos da mesma realização. O primeiro conjunto (Picos1) contendo desde o primeiro até o penúltimo picos e o segundo conjunto (Picos2) contendo desde o segundo até o último picos. Assim, a Eq. (II-22) é aplicada para os dois conjuntos e as *Figura VI.14 a Figura VI.16* ilustram esses resultados para um exemplo particular de uma série temporal de cada caso de estudo.



*Figura VI.14 – Estudo de Caso 1: Pares de picos consecutivos para a realização #2.  
Correlação igual a 0.29150*



*Figura VI.15 – Estudo de Caso 2: Pares de picos consecutivos para a realização #22.  
Correlação igual a 0.44566*



*Figura VI.16 – Estudo de Caso 3: Pares de picos consecutivos para a realização #37. Correlação igual a 0.50813*

Observamos que os valores dos coeficientes de correlação entre os picos consecutivos são relativamente baixos (entre 0.25 e 0.55) o que aparentemente poderia ser traduzido em que não existe uma dependência entre eles. Porém, como será visto, o método ACER mostra que algumas informações podem ser extraídas das dependências estatísticas entre eles para o cálculo da distribuição do pico extremo.

### VI.3. Análise de Valores Extremos das Séries Temporais dos Fatores de Utilização – Valores de Referência

Conforme descrito no item V.3. um valor de referência teoricamente “exato” para o valor extremo mais provável de cada processo analisado pode ser obtido tomando-se os valores máximos observados em cada uma das 50 realizações de 3 horas e depois ajustando uma distribuição Tipo I (Gumbel) a esta amostra. O parâmetro  $u$  desta distribuição corresponde ao valor extremo mais provável que será usado como valor de referência para avaliação da precisão dos demais métodos investigados. As Figuras VI.17 a VI.19 apresentam o ajuste, usando o método dos momentos, da distribuição Tipo I à amostra das máximas observações dos três casos de estudo. Como pode ser observado, o ajuste é muito bom para os três casos, o que corrobora com a hipótese assumida de que os valores extremos seguem uma distribuição Tipo I.

A Tabela 11 apresenta os parâmetros da distribuição de Gumbel (Tipo I) obtidos para os três casos de estudo.

*Tabela 11. Parâmetros da distribuição do Pico Extremo (Tipo I) para cada Estudo de Caso*

Estudo de Caso	$\alpha$	$u$
1	8.450200	0.78522
2	49.832200	0.52820
3	77.561200	0.19897

Somente para reforçar a ideia, o parâmetro  $u$  correspondente a cada caso de estudo apresentado na Tabela 11, e ele é o valor de referência para comparação dos resultados obtidos pelos métodos investigados no trabalho.

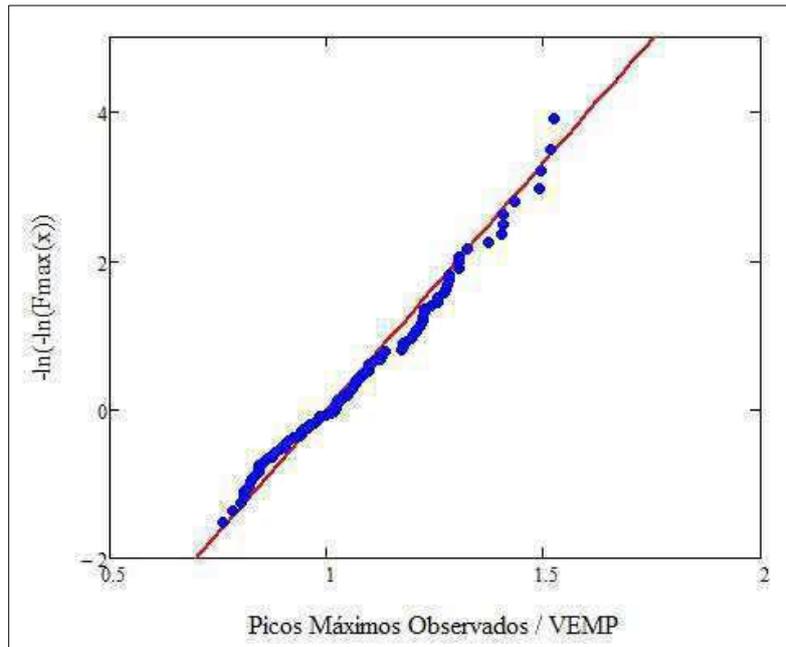


Figura VI.17 – Estudo de Caso 1: Ajuste dos picos máximos observados (normalizados) das 50 séries do FU a uma distribuição Tipo I

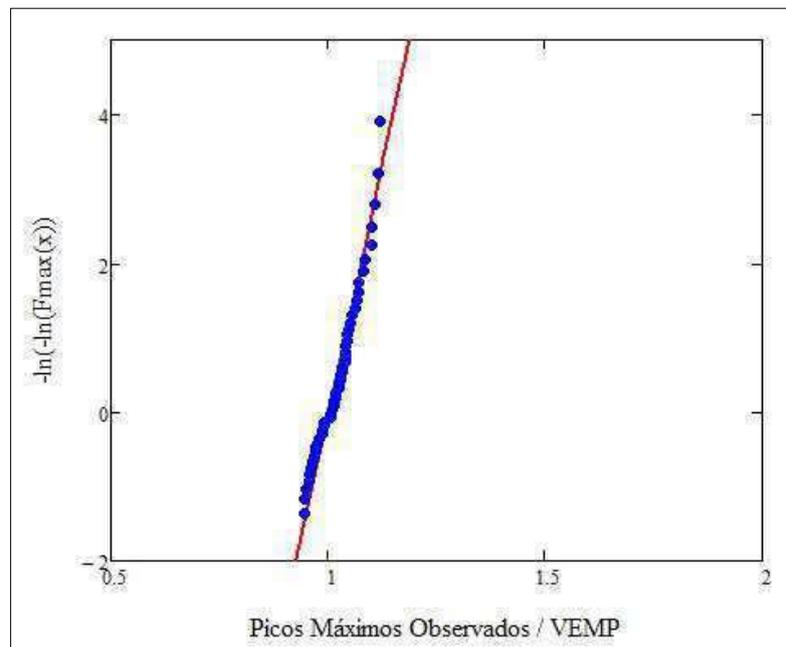


Figura VI.18 – Estudo de Caso 2: Ajuste dos picos máximos observados (normalizados) das 50 séries do FU a uma distribuição Tipo I

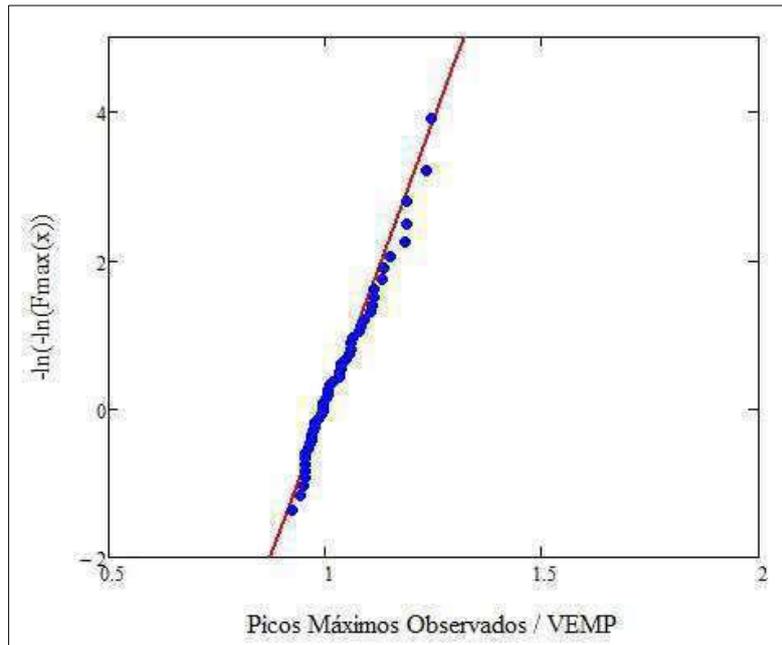


Figura VI.19 – Estudo de Caso 3: Ajuste dos picos máximos observados (normalizados) das 50 séries do FU a uma distribuição Tipo I

#### VI.4. Análise de Valores Extremos das Séries Temporais dos Fatores de Utilização – Resultados dos Métodos Estudados

Cada um dos métodos descritos no Capítulo V foi utilizado para estimar o valor mais provável do fator de utilização em cada um dos três casos de estudo. Lembrando que estes métodos se baseiam em apenas uma realização do processo investigado, os mesmos foram utilizados individualmente para cada uma das 50 séries temporais variando-se os seus respectivos comprimentos (tempo de simulação). Assim, conforme descrito no item V.3. , a acurácia e variabilidade de cada método foi investigada através da tendência e do coeficiente de variação dos 50 resultados para cada tamanho de simulação (Eqs. (V-22) e (V-23), respectivamente).

As Figuras VI.20, VI.22 e VI.24 apresentam os resultados de tendência do valor extremo mais provável normalizado para um período de  $T = 3h$  (média do valores estimados dividida pelo valor de referência), respectivamente, para os três casos de estudo. As Figuras VI.21, VI.23 e VI.25 apresentam os correspondentes coeficientes de variação dos resultados estimados.

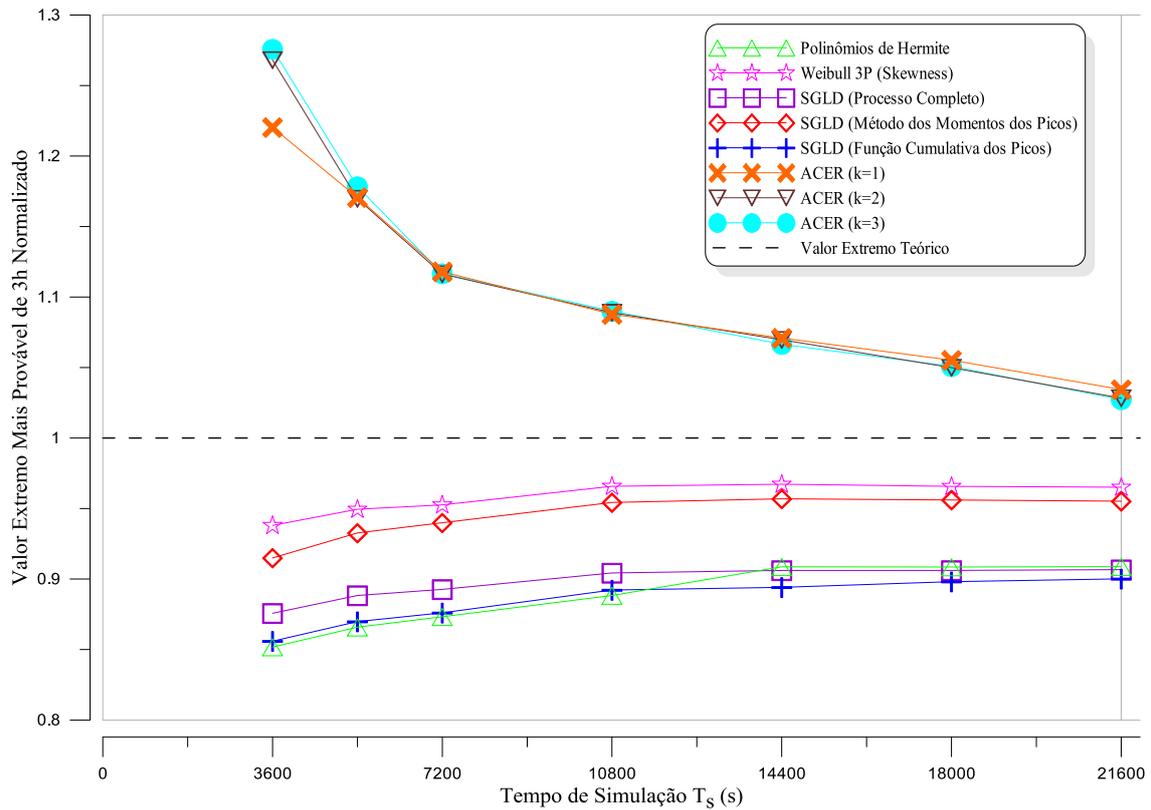


Figura VI.20 – Estudo de Caso 1: Estimativas do valor extremo mais provável de 3 horas (Viés/tendência do estimador)

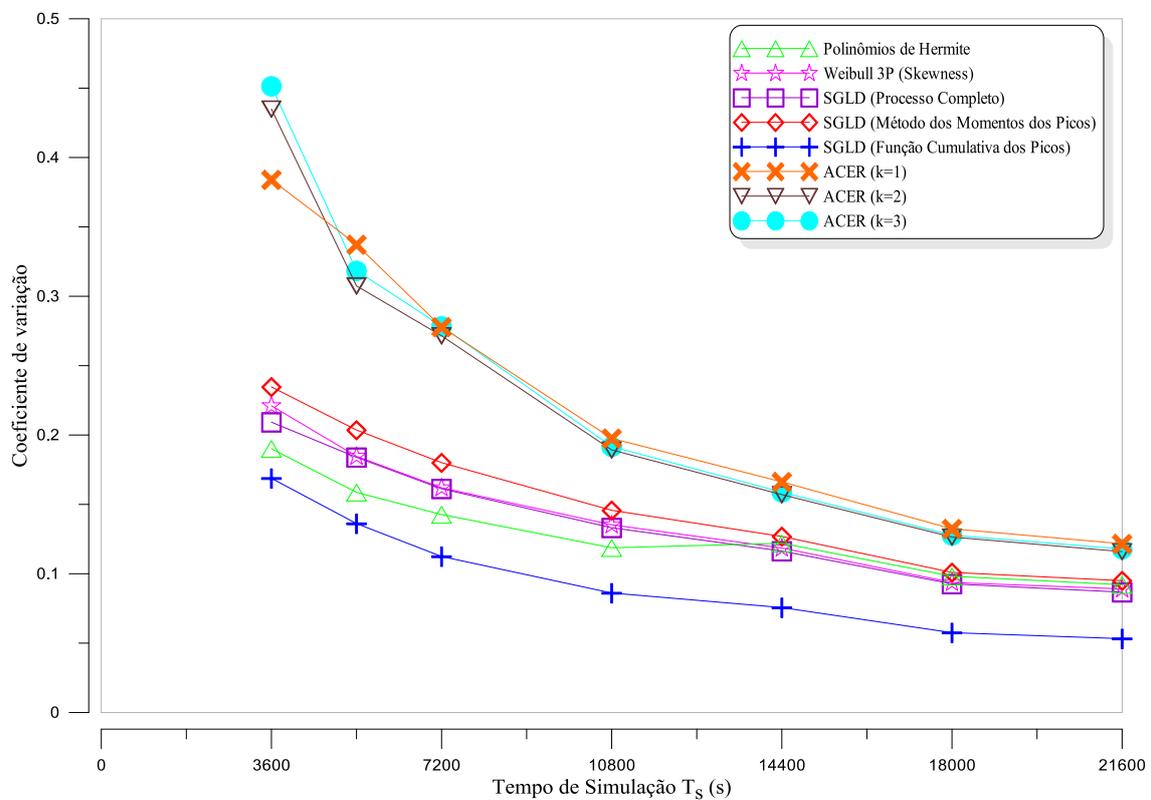


Figura VI.21 – Estudo de Caso 1: Coeficientes de variação do estimador do valor extremo mais provável de 3 horas

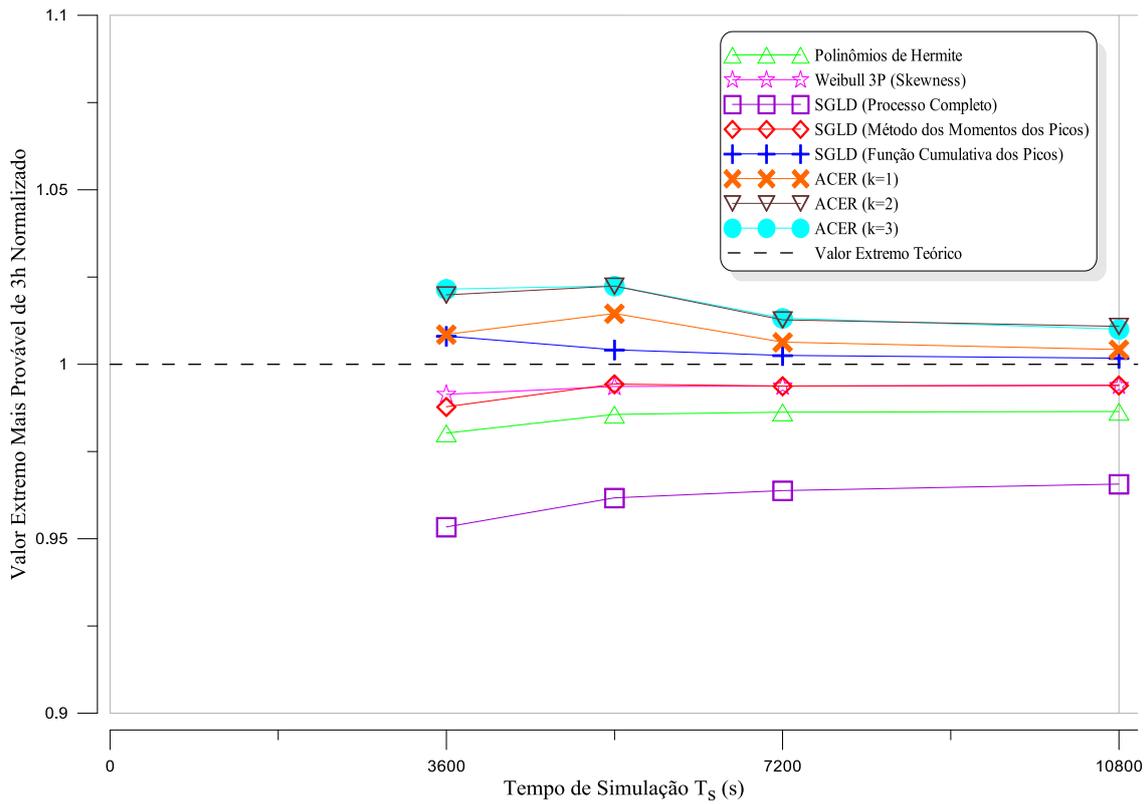


Figura VI.22 – Estudo de Caso 2: Estimativas do valor extremo mais provável de 3 horas (Viés/tendência do estimador)

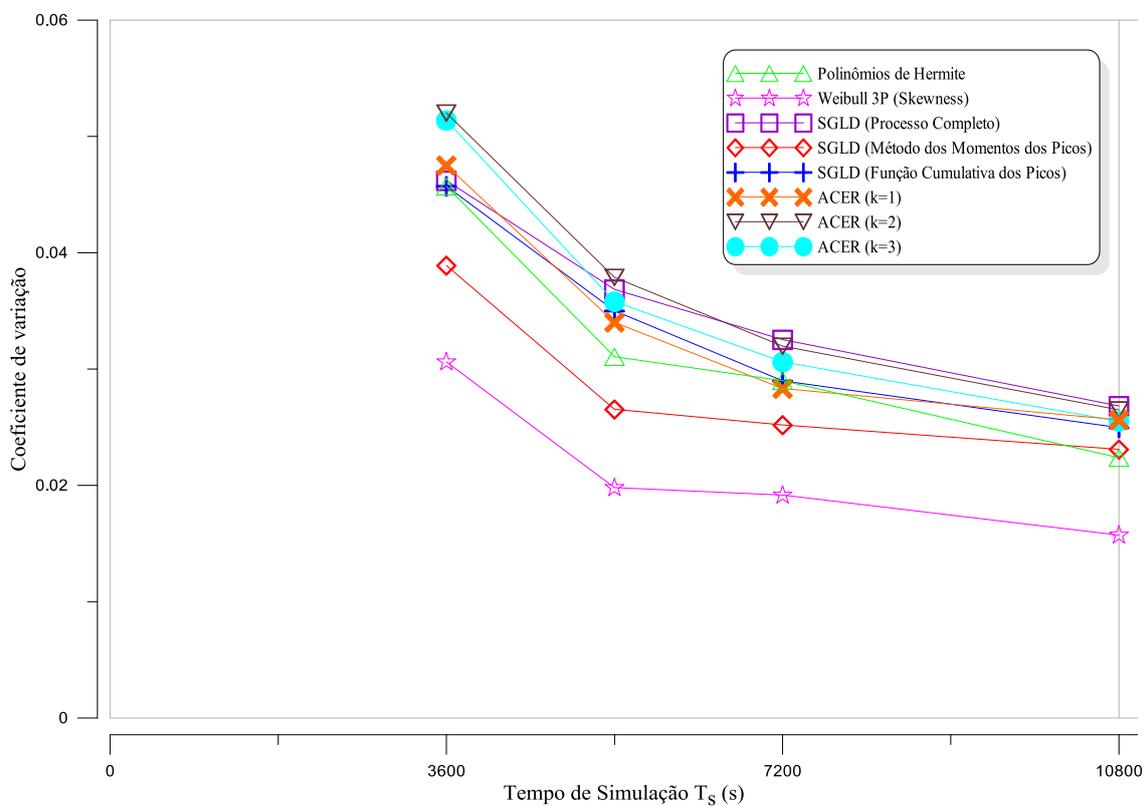


Figura VI.23 – Estudo de Caso2: Coeficientes de variação do estimador do valor extremo mais provável de 3 horas

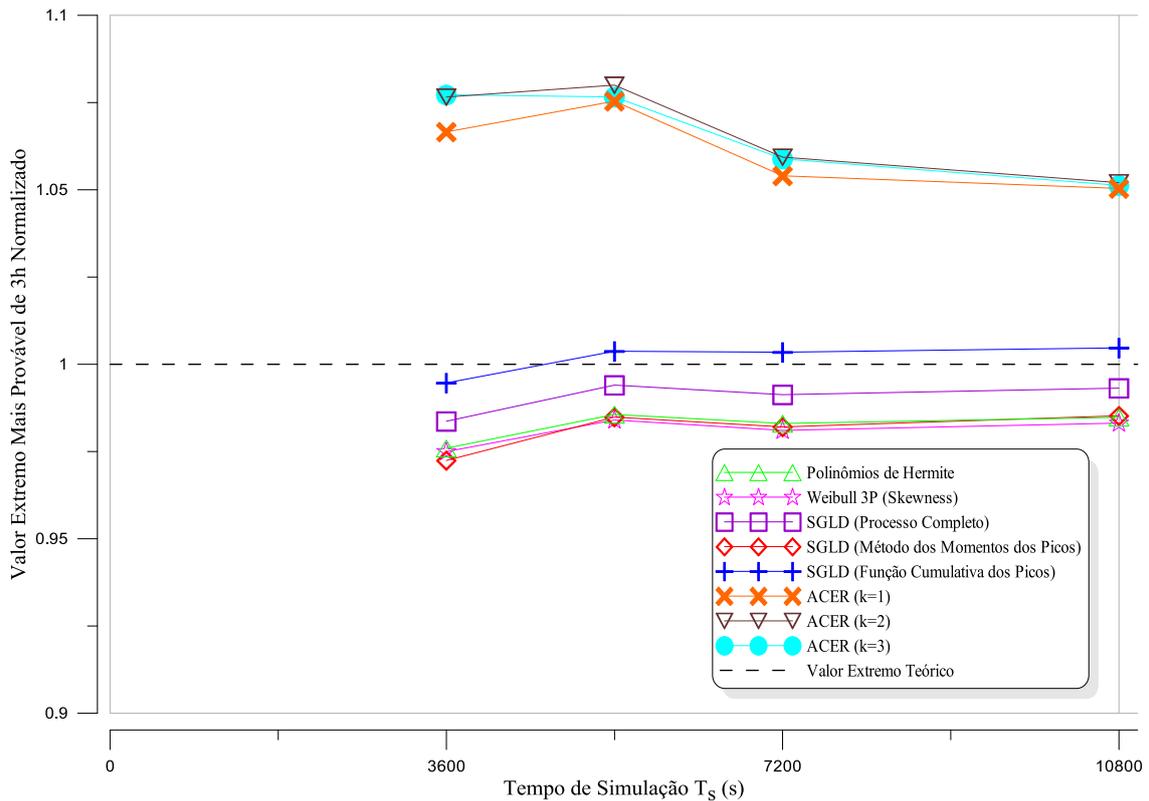


Figura VI.24 – Estudo de Caso 3: Estimativas do valor extremo mais provável de 3 horas (Viés/tendência do estimador)

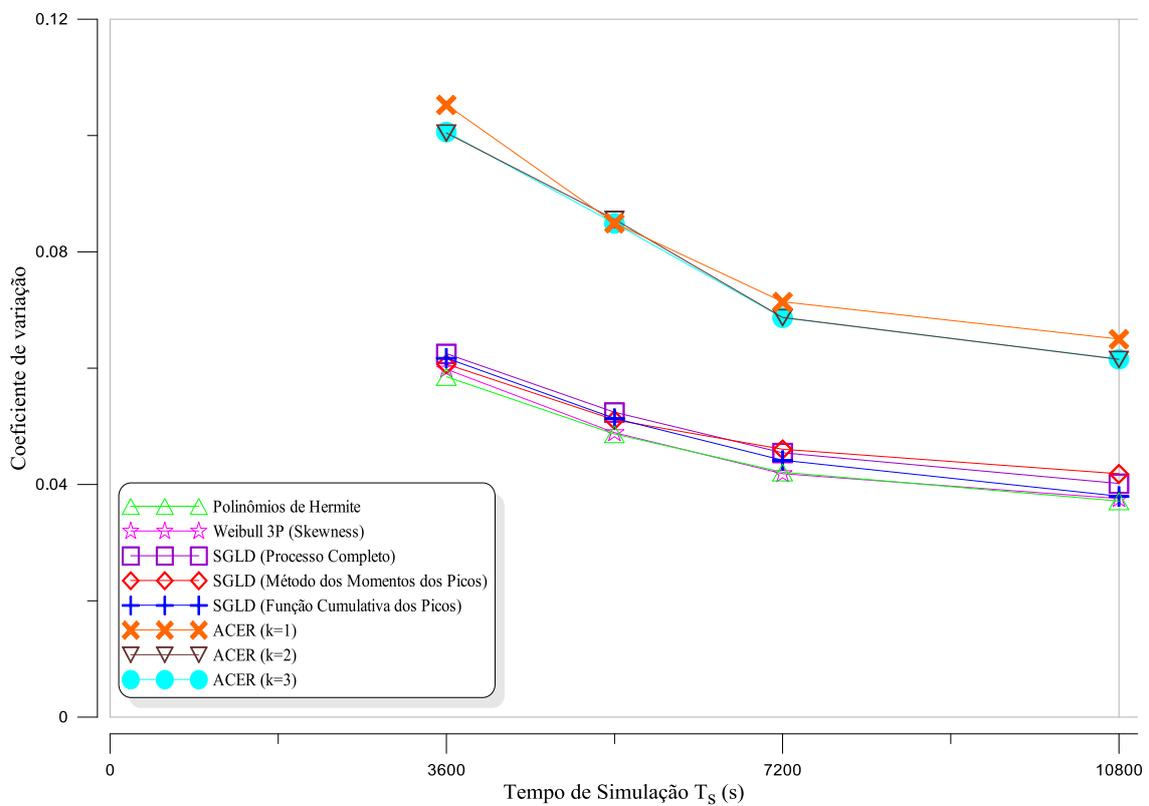


Figura VI.25 – Estudo de Caso 3: Coeficientes de variação do estimador do valor extremo mais provável de 3 horas

Em geral, todos os métodos tendem a estimar valores constantes à medida que o tamanho da simulação aumenta. O coeficiente de variação diminui em função do tempo de simulação.

Como regra geral, o método ACER tende sempre a estimar valores conservadores e também apresenta uma maior variabilidade nos seus resultados. Os resultados do método SGLD tendem a estar mais próximos àqueles obtidos pelo método baseado na distribuição de Weibull e pelo método baseado nos polinômios de Hermite.

Especificamente, em relação à cada um dos estudos de caso faz-se as seguintes observações:

- Estudo de Caso 1

Neste caso o fator de utilização corresponde à um processo aleatório extremamente não-gaussiano e com grande variabilidade nos parâmetros estatísticos de uma realização para outra. Esta variabilidade se refletiu nos estimadores do valor extremo mais provável, principalmente, para o caso de simulações menores que 3-h. Para simulações menores que 3 horas (10800 segundos) todos os métodos foram tendenciosos na estimativa do VEMP, sendo que em simulações de 1 hora (3600 segundos) o ACER superestimou em quase um 30%, o método baseado nos Polinômios de Hermite subestimou em até 15% e os métodos baseados na SGLD subestimaram em até 10%. Contudo, o método baseado na distribuição de Weibull foi o único que forneceu resultados com tendência menor que 5%. Igualmente, todos os métodos apresentaram uma grande variabilidade nos seus resultados, sendo que o ACER apresentou um coeficiente de variação de até 45%, e os demais métodos variaram entre 15% e 20%.

No caso de simulações de 3 horas (10800 segundos) ou mais longas, todos os métodos apresentaram convergência ao valor teórico e pouca variabilidade entre os seus resultados. Sendo que o método ACER superestimou por volta de 5% os resultados e os demais métodos subestimaram em 10% ou menos o VEMP. Igualmente, todos os métodos tiveram pouca variabilidade, com coeficientes de variação entre 20% e 5%.

- Estudo de Caso 2

Neste caso o processo que representa o fator de utilização da seção possui características bem gaussianas. Neste caso, todos os métodos forneceram resultados não muito tendenciosos (diferenças menores que 5%) até mesmo para tempos de simulação curtos (3600 segundos) e atingindo o valor teórico de referência para séries com durações maiores. Igualmente todos os métodos apresentaram resultados com coeficientes de variação menores que 5%.

- Estudo de Caso 3

Neste caso a série temporal do fator de utilização caracteriza-se também por um processo não-gaussiano. Neste caso todos os métodos, exceto o ACER, também forneceram resultados não tendenciosos (viés menor que 5% para todos os tamanhos de simulação considerados) para o valor extremo mais provável de 3-h. O método ACER foi único que superestimou ligeiramente (acima de 5%) o VEMP para todos os tamanhos de simulação investigados. Igualmente todos os métodos apresentaram resultados com coeficientes de variação não muito altos sendo que o método ACER apresentou valores próximos a 8% e os demais no entorno de 5%.

## CAPÍTULO VII

### CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

#### VII.1 Conclusões

Este trabalho teve como objetivo principal a avaliação de distintas metodologias para a estimativa de valores extremos de séries temporais não gaussianas com aplicações específicas para o Fator de Utilização de seções transversais de risers de aço segundo a norma DNV (DnV-OS-F201. 2010). Os quatro métodos estudados foram: modelo baseado nos Polinômios de Hermite, modelo baseado na distribuição Lognormal generalizada deslocada (SGLD), método baseado nas taxas médias de excedências condicionadas (ACER) e Modelo baseado na distribuição de Weibull. No caso da SGLD, várias técnicas de aplicação foram implementadas. Todos estes métodos estão detalhadamente descritos no CAPÍTULO V. Estes métodos utilizam apenas uma série temporal de duração  $T_{sim}$  para fazer a estimativa da distribuição de extremos para um período de curto prazo de duração  $T$ . Na prática,  $T_{sim}$  pode ser menor, igual ou maior que  $T$ .

No trabalho foram investigados três casos distintos de séries de fatores de utilização de risers na configuração SLWR. Um caso representa um processo praticamente gaussiano (Caso 2) e os demais processos não gaussianos, sendo um deles com características não-gaussianas extremamente acentuadas (Caso 1).

Para investigar a acurácia dos valores estimados por cada um dos métodos foi obtida uma amostra de valores extremos, obtida através de várias (50) realizações distintas de duração  $T$ , para se ter um valor de referência para comparação. Os valores extremos foram estimados ajustando-se uma distribuição Tipo I a esta amostra.

As incertezas nos valores estimados por cada um dos métodos foi avaliada através da tendência dos valores estimados e também pela variabilidade dos mesmos (representada pelo coeficiente de variação), em função do tamanho da série temporal  $T_{sim}$ . Observou-se que tanto a variabilidade estatística quanto a tendenciosidade das estimativas é grande para séries temporais curtas, o que significa que simulações longas são necessárias para diminuir as incertezas destas estimativas. Os maiores vieses nas estimativas ocorrem para simulações curtas (menores que 10800s) e o mesmo ocorrendo

para o coeficiente de variação do valor extremo mais provável estimado, independentemente do método utilizado.

No caso particular da SGLD, a sua performance foi avaliada através de três técnicas diferentes, sempre demonstrando sua extensa capacidade de modelar satisfatoriamente qualquer situação. Assim, estamos em condições de concluir que os métodos baseados nesta nova distribuição, fornecem resultados confiáveis (praticamente não tendenciosos) inclusive para tempos de simulação pequenos (menores que 10800s). Além disso, em geral, os resultados obtidos pelo método através das simulações individuais convergem para o valor teórico, independentemente das propriedades estatísticas do processo. Dentre os três procedimentos investigados, o que apresentou menor variabilidade foi o procedimento baseado na distribuição cumulativa empírica dos picos.

O modelo de ACER se mostra como uma alternativa singular pois, ao contrário de outras metodologias, ele leva em consideração a dependência estatística entre picos próximos. Em todos os casos de estudo ele estimou valores extremos convergentes ao valor teórico, porém, foi o método que apresentou maior variabilidade nos resultados (maiores coeficientes de variação). Em comparação com o resto das metodologias, ele foi o que mais superestimou os resultados, porém, sempre dentro dos intervalos admissíveis (menores que 5%). Nos casos estudados não se mostrou grande diferença entre os valores estimados considerando diferentes valores para o índice  $k$ , tendo casos que o valor estimado foi praticamente o mesmo. Isto indica que a dependência estatística entre os picos do processo não existe (ou desprezível).

O modelo baseado na distribuição de Hermite mostrou uma incerteza estatística relativamente grande para séries curtas (com durações menores que 10800s) e igualmente possui uma tendência notória de superestimar os valores extremos. Já para séries maiores (com durações superiores a 10800s) o método forneceu resultados pouco tendenciosos (com vieses menores a 5%), mas conservando a particularidade de superestimar os valores extremos.

O modelo baseado na distribuição Weibull, também se mostrou não tendencioso e altamente confiável, mesmo para tempos de simulação curtos. Em termos gerais, confirma-se que este método venha sendo estudado como uma solução tradicional.

A principal contribuição, obtida através de todas as análises, é que ambos os modelos ACER e SGLD, publicados recentemente (2015 e 2013, respectivamente), funcionam satisfatoriamente para estimativa de valores extremos de processos aleatórios, seja qual for a distribuição que descreve o processo: gaussiana ou não-gaussiana. Observa-se, entanto, que este trabalho é pioneiro no uso da SGLD para estimativa de valores extremos.

Finalmente, em função dos resultados obtidos e baseado na experiência do autor, dentre as alternativas estudadas, a escolhida como a mais promissora é aquela baseada na Distribuição Lognormal Deslocada Generalizada (SGLD) com ajuste da função cumulativa empírica dos picos do processo. Esta observação se deve ao fato de que esta metodologia foi a que forneceu estimativas não desviadas do valor teoricamente exato e com pouca variabilidade (dispersão) para simulações distintas, inclusive para tempos de simulação curtos.

## VII.2 Sugestões para trabalhos futuros

Como sugestões para trabalhos futuros podem-se listar as seguintes:

- Ampliar os testes sobre a aplicabilidade da distribuição SGLD em outras áreas de interesse, tais como modelagem de dados ambientais, etc.
- Desenvolver um método de calibração para o índice “k” no modelo ACER.
- Expandir a aplicação dos métodos ACER e SGLD para a estatística de longo prazo da resposta.
- Implementar métodos para determinar Intervalos de Confiança dos valores estimados utilizando a técnica Bootstrap.
- Desenvolver um método para o cálculo direto dos parâmetros  $r$  e  $\sigma$  da SGLD em função da combinação de skewness-kurtosis, sem necessidade de ter que resolver utilizando métodos numéricos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABS 124. (2004). “Dynamics Analysis Procedure for Self-Elevating Drilling Units”. Houston, U.S.A.: ABS Guidance Notes.
- ANG, A.H.S., & TANG, W.H., (1984). “Probability Concepts in Engineering, Planning and Design”, Vol. II, John Willey and Sons, New York, USA.
- API RP-2RD (1998), “Design of risers for Floating Production Systems (FPSs) and Tension-Leg Platforms (TLPs)” 1<sup>st</sup> Edition, American Petroleum Institute, Washington DC, USA.
- CALDERÓN-IBARRA, M.A.; SAGRILO, L.V.S.; SOUSA, F.J.M.; LOW, Y.M. (2017); “Extreme response prediction of steel risers using a four parameter distribution”, 36th International Conference on Offshore Mechanics and Arctic Engineering, OMAE 2017, June 25-30, 2017, Trondheim, Norway.
- DnV-OS-F201. (2010). “Dynamic Risers”, Det Norske Veritas, Hovik, Norway.
- HUIJSMANS, R. H. M., and ADEGEEST, L. J. M. (1998). “Fatigue assessment in FPSO mooring design using moment based Hermite Approximation”. Proceedings of the 17th International Conference on Offshore Mechanics and Arctic Engineering OMAE. July 5-9, 1998, Lisbon, Portugal.
- ISO. (2006). “Site-specific Assessment of Mobile Offshore Units - Part 1. Jack-ups” ISO Petroleum and Natural Gas Industries. Geneva, Switzerland:
- JUN HE, JINGHAI GONG, (2016). “Estimate of small first passage probabilities of nonlinear random vibration systems by using approximation of extreme distributions”, Structural Safety 60, pp. 28-36.
- KARPA, O. (2012). “Development of Bivariate Extreme Value Distributions for Applications in Marine Technology”, Doctoral Theses, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

- KARPA, O. (2015). “ACER User Guide”, Centre for Ships and Ocean Structures (CeSOS), Norwegian University of Science and Technology (NTNU), Trondheim, Norway. <https://folk.ntnu.no/arvidn/ACER/>
- LEHN, E & MOE, V. (2008). “Statistics for You and Me: With emphasis on peak and extreme value distributions”, Norwegian Marine Technology Research Institute AS (MARINTEK), Trondheim, Norway.
- LIMA, E. C., & SAGRILO, L. V. (2008). “Notas de Aula: Variáveis Aleatórias; Distribuições de Probabilidades; Estatística de Extremos; Inferência Estatística.” COPPE/UFRJ, Rio de Janeiro, Brasil.
- LOW Y. M. (2013). “A New Distribution Model for Fitting Four Moments and its Application to Reliability Analysis”, Structural Safety, Vol. 42, pp. 12-25.
- MADSEN H. O., KRENK S., LIND N. C. (1986) “Methods of Structural Safety” Englewood Cliffs, N. J., Prentice-Hall, New Jersey, USA.
- MANSOUR, A. E., and JUNCHER JENSEN, J. (1995). “Slightly Non-Linear Extreme Loads and Load Combinations” Journal of Ship Research, Vol. 39, No. 2, June 1995, pp. 139-149.
- NAESS, A., (2001). “Crossing rate statistics of quadratic transformations of Gaussian processes”, Probabilistic Engineering Mechanics, No. 16, pp. 209-217.
- NAESS A., GAIDAI O. (2009). “Estimation of Extreme Values From Sampled Time Series”, Structural Safety, Vol. 31, pp. 325-334.
- NAESS, A., O. GAIDAI, AND O. BATSEVYCH, (2009). “Extreme value statistics of combined load effect processes.” Structural Safety, Vol. 31, 298–305.
- NAESS A., MOAN T. (2012). “Stochastic Dynamics of Marine Structures”, Cambridge University Press; 1st Edition, England, United Kingdom.
- NAESS, A., O. GAIDAI, AND O. KARPA, (2013), “Estimation of Extreme Values by the Average Conditional Exceedance Rate Method” Journal of Probability and Statistics, Volume 2013, Article ID 797014, 15 pages.

- NASCIMENTO, L. S. (2009). “Análise de Valores Extremos de Parâmetros de Resposta Dinâmica de Plataformas Auto-Elevatórias.” Dissertação de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- NASCIMENTO, L.S., SAGRILO L.V.S., ELLWANGER G.B. (2012). “On the Extreme Value Analysis of Jack-Up Dynamic Response Parameters”, *Marine Systems and Ocean Technology*, Vol. 6, No. 2, pp. 75-86.
- NASCIMENTO, L. S. (2016). “Análise Estatística de Respostas Dinâmicas de Estruturas Tubulares Sob Carregamento de Onda Irregular Não Linear” Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- NEWLAND, D.E., (1993). “An Introduction to Random Vibrations, Spectral & Wavelet Analysis”, 3rd Edition, Longman Scientific and Technical, Harlow, England.
- PAIVA JÚNIOR, J. M. (2010) “Análise de extremos utilizando a hipótese de Poisson”, Dissertação de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- RICE, S. O. (1945), “Mathematical Analysis of Random Noise”. *Bell System Technical Journal* 24 46–156. American Telephone and Telegraph Company (AT&T), New York, USA.
- SAGRILO L.V.S., NAESS A., GAO Z. (2012). “On the Extreme Value Analysis of the Response of a Turret-Moored FPSO”, *Journal of Offshore Mechanics and Arctic Engineering*, Vol. 134, No. 2. American Society of Mechanical Engineers.
- SOUSA, F. J. M. (2011), “Calibração de condições de projeto baseada na tração de longo-prazo para risers em catenária”, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Brasil.
- TORHAUG, R. (1996). “Extreme Response of Nonlinear Ocean Structures: Identification of Minimal Stochastic Wave Input for Time-Domain Simulation” Doctoral Thesis, Reliability of Marine Structures (RMS) Program, Stanford University, USA.
- WINTERSTEIN, S. R. (1985). “Non-normal responses and fatigue damage”, *Journal of Engineering Mechanics*, ASCE, Vol. 111, pp. 1291-1295.

- WINTERSTEIN, S. R. (1987). "Moment-based Hermite Models of Random Vibration.", Department of Structural Engineering, Report No. 219, Technical University of Denmark, Lyngby, Denmark:
- WINTERSTEIN, S. R. (1988). "Nonlinear Vibration Models for Extremes and Fatigue", Journal of Engineering Mechanics, ASCE, Vol. 114, pp. 1772-1790.
- WINTERSTEIN, S. R.; UDE, T. C.; & KLEIVEN, G. (1994). "Springing and Slow-drift Responses: Predicted Extremes and Fatigue vs. Simulation", BOSS-1994, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
- WINTERSTEIN, S. R. & LANGE, C.H. (1995). "Moment-Based Probability Models for Wind Engineering Applications" Proceedings of the Tenth Engineering Mechanics Speciality Conference, ASCE, Vol. 1, pp. 159-162.
- ZURITA, B.I., (1999). "Análise Estatística de Valores Extremos de Séries Temporais Gaussianas e Não Gaussianas". Dissertação de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

## ANEXO A

### CRITÉRIO DE PROJETO SEGUNDO A NORMA DNV LRFD

Em 2010, a Det Norske Veritas (DNV) apresentou critério LRFD (*Load and Resistance Factor Design*, o que significa: Fator de Projeto baseado em Cargas e Resistências) para projetos de risers metálicos, definido pela norma DnV-OS-F201 (DnV-OS-F201. 2010). Este critério combina parâmetros de carregamentos e de resistências para determinar o denominado Fator de Utilização (neste trabalho se faz referência a ele como FU), a partir do qual é possível definir a viabilidade de um projeto de um riser metálico.

Em particular, o FU avalia todas as tensões atuantes sobre uma seção transversal do riser e as pondera através de fatores parciais de segurança para calcular um valor característico para cada seção analisada do riser. Ele é definido em forma genérica pela seguinte série temporal  $FU(t)$ :

$$FU(t) = f(T(t), M(t), P, E, G) \quad (A-1)$$

onde:

$T(t)$	Esforço axial atuante na seção;
$M(t)$	Momentos fletores atuantes na seção;
$P$	Pressões interna e externa na seção;
$E$	Conjunto de propriedades referentes ao material e a geometria do riser;
$G$	Conjunto de fatores de segurança associados a diversas incertezas de cargas e resistências (definidos na DnV-OS-F201. , 2010);
$t$	Tempo em segundos.

Intrinsecamente, o Fator de Utilização representa uma probabilidade de falha pré-estabelecida e, em razão disso, ele deve-se manter dentro de certos limites. Desta maneira, o parâmetro de aceitação do projeto do riser será:

$$FU(t) \leq 1 \quad (\text{A-2})$$

para todas e cada uma das seções ao longo do riser.

Os esforços  $T(t)$  e  $M(t)$  são geralmente obtidos através de análises dinâmicas globais aleatórias (no domínio do tempo) obtendo como resultado séries temporais representativas de processos estocásticos. Em consequência, as séries do Fator de Utilização  $FU(t)$  serão também representativas de processos estocásticos. Assim, para a obtenção dos seus valores extremos de curto-prazo poderão ser aplicados os conceitos e procedimentos descritos no CAPÍTULO III deste trabalho.

## ANEXO B

### SOLUÇÕES APROXIMADAS PARA O SISTEMA DE EQUAÇÕES DOS POLINÔMIOS DE HERMITE

Lembrando que para a aproximação dos Polinômios de Hermite, adotando-se  $N = 4$ , a série infinita dada pela Eq. (II-60), adquire a seguinte forma:

$$g(u) = c_1 + c_2 u + c_3(u^2 - 1) + c_4(u^3 - 3u) \quad (\text{B-1})$$

A maneira ideal de se estimar os coeficientes  $c_1$ ,  $c_2$ ,  $c_3$  e  $c_4$  é através da solução do sistema de equações não lineares descrito na Eq. (II-71). Entretanto, para evitar a solução deste sistema várias soluções aproximadas foram propostas na literatura. Considerando um processo  $X(t)$  de média zero, tem-se as seguintes aproximações:

a) Modelo de Winterstein (WINTERSTEIN, 1987):

$$\begin{aligned} c_1 &= 0 & c_2 &= 1 \\ c_3 &= \frac{\gamma_x}{4 + 2(\sqrt{1 + 1.5(\kappa_x - 3)} - 1)} & c_4 &= \frac{\sqrt{1 + 1.5(\kappa_x - 3)} - 1}{18} \end{aligned} \quad (\text{B-2})$$

b) Modelo de Mansour e Jensen (MANSOUR & JENSEN, 1995):

$$\begin{aligned} c_1 &= 0 & c_2 &= 1 \\ c_3 &= \frac{\gamma_x}{5.8 + 2(\sqrt{1 + 1.5(\kappa_x - 3)} - 1)} & c_4 &= \frac{\sqrt{1 + 1.5(\kappa_x - 3)} - 1}{30} \end{aligned} \quad (\text{B-3})$$

c) Modelo de Torhaug (TORHAUG, 1996):

$$\begin{aligned}
 c_1 &= 0 & c_3 &= \frac{\gamma_x}{6} \left[ \frac{1 - 0.015|\gamma_x| + 0.3(\gamma_x)^2}{1 + 0.2(\kappa_x - 3)} \right] \\
 c_2 &= 1 & c_4 &= \frac{\sqrt[3]{1 + 1.25(\kappa_x - 3)} - 1}{10} \left[ 1 - \frac{1.43(\gamma_x)^2}{\kappa_x - 3} \right]^{1 - 0.1(\kappa_x)^{0.8}}
 \end{aligned} \tag{B-4}$$

d) Ajuste de Primeira Ordem (HUIJSMANS, 1998):

$$c_1 = 0 \quad c_2 = 1 \quad c_3 = \frac{\gamma_x}{6} \quad c_4 = \frac{\kappa_x - 3}{24} \tag{B-5}$$

Nos modelos acima  $\gamma_x$  e  $\kappa_x$  correspondem aos coeficientes de assimetria e de curtose de  $X(t)$ , respectivamente.

Como dito anteriormente, estas soluções são aproximadas. Porém, elas podem ser utilizadas como valores iniciais em métodos iterativos para acelerar a convergência da solução do sistema de equações não lineares (vide Eq. (II-71))

## ANEXO C

### MÉTODO DE NEWTON-RAPHSON PARA CÁLCULO DOS PARÂMETROS ( $r, \sigma$ ) DA SHIFTED GENERALIZED LOGNORMAL DISTRIBUTION (SGLD)

Uma vez identificadas as equações (II-42) e (II-43), definimos o vetor  $u_n$  como sendo:

$$u_n = \begin{bmatrix} r_n \\ \sigma_n \end{bmatrix} \quad n = 1, 2, 3... \quad (C-1)$$

Por outro lado, o Jacobiano da função é dado pela seguinte matriz de derivadas:

$$J(u_n) = \begin{bmatrix} \frac{d}{dr} \gamma_y(u_n) & \frac{d}{d\sigma} \gamma_y(u_n) \\ \frac{d}{dr} \kappa_y(u_n) & \frac{d}{d\sigma} \kappa_y(u_n) \end{bmatrix} \quad (C-2)$$

Pela definição das funções, o objetivo é satisfazer o seguinte sistema de equações:

$$G(u_n) = \begin{bmatrix} \gamma_y(u_n) - \gamma_x \\ \kappa_y(u_n) - \kappa_x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (C-3)$$

Para satisfazer o sistema anterior aplica-se sucessivamente  $n + 1$  número de vezes o método de Newton-Raphson resolvendo-se a seguinte equação:

$$u_{n+1} = u_n - (J(u_n))^{-1}(G(u_n)) \quad (C-4)$$

Para acelerar a convergência do método, é recomendável estabelecer os valores iniciais do vetor  $u_1$  como sendo  $r_1 = 2$  e  $\sigma_1$  satisfazendo à seguinte equação (LOW, 2012):

$$\sigma_1 = \sqrt{\ln \left[ 1 + \left( 2 \sinh \left( \frac{1}{3} \sinh^{-1} \left( -\frac{\gamma_y}{2} \right) \right) \right)^2 \right]} \quad (C-5)$$

## ANEXO D

### PARÂMETROS ESTATÍSTICOS DAS SÉRIES TEMPORAIS DE 3 HORAS

Para cada Estudo de Caso são apresentados os parâmetros estatísticos de cada uma das 50 realizações individuais de 3 horas.

*Tabela 12 Estudo de Caso 1: Parâmetros estatísticos das 50 realizações de 3 horas*

Realização	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Máx. Observ.
1	0.12003	0.041306	4.3642	43.199	0.9634
2	0.12055	0.041627	3.5138	24.289	0.6514
3	0.12065	0.042604	3.6949	25.607	0.6303
4	0.12037	0.042873	4.7199	49.97	1.0248
5	0.12092	0.046091	5.7804	72.077	1.1742
6	0.12061	0.044109	4.9906	50.41	0.8901
7	0.12011	0.039615	3.3828	24.602	0.6707
8	0.12089	0.044536	4.4976	39.902	0.8289
9	0.1204	0.044449	4.9086	44.599	0.8363
10	0.12086	0.044068	4.4923	41.618	0.9305
11	0.12015	0.040674	4.0088	36.763	0.8623
12	0.12021	0.041935	4.3748	41.361	0.9466
13	0.12022	0.042019	4.448	42.713	0.8374
14	0.12043	0.04251	4.692	49.727	0.9583
15	0.12045	0.043067	4.4527	39.137	0.7404
16	0.12051	0.043583	4.4659	38.521	0.7652
17	0.12009	0.041518	4.08	33.196	0.8026
18	0.12018	0.044158	5.7033	66.034	0.9967
19	0.12058	0.043338	4.5329	42.888	0.9469
20	0.12022	0.040541	3.4664	24.697	0.686
21	0.12041	0.043922	5.2829	61.951	1.1051
22	0.12008	0.041752	5.1136	63.115	1.0072
23	0.1203	0.040911	3.8436	33.128	0.8623
24	0.12083	0.04344	4.461	45.009	0.9234
25	0.11995	0.039976	3.6241	27.328	0.7259
26	0.11989	0.042445	5.2173	57.772	1.0002
27	0.12026	0.039754	3.0433	18.795	0.6349
28	0.12021	0.040231	3.3124	22.187	0.6931
29	0.12062	0.043989	4.6526	42.86	0.8825
30	0.11984	0.038862	3.4518	27.061	0.7552
31	0.12007	0.039376	3.1806	21.362	0.6614
32	0.12028	0.041463	3.9504	32.036	0.7413
33	0.1209	0.048274	6.0809	66.084	0.9867
34	0.12019	0.04254	5.3868	68.486	1.1048

35	0.12038	0.04233	3.8343	27.339	0.6623
36	0.1205	0.044482	5.0227	50.837	1.0249
37	0.12062	0.042704	4.2077	38.736	0.9427
38	0.12041	0.04163	3.8637	31.077	0.8128
39	0.12018	0.045401	7.0508	105.26	1.1918
40	0.12026	0.041183	4.1276	39.659	0.938
41	0.12081	0.043771	4.2076	35.801	0.8854
42	0.12028	0.041675	4.1131	35.854	0.8471
43	0.12089	0.0432	3.8412	29.414	0.7386
44	0.12109	0.047305	5.5779	60.384	1.0068
45	0.11998	0.039653	3.6584	31.235	0.8405
46	0.12093	0.048087	6.7563	92.272	1.1691
47	0.11986	0.039668	3.7756	31.724	0.8057
48	0.12004	0.040524	3.5822	25.103	0.6455
49	0.12112	0.044698	4.6419	47.323	1.0407
50	0.12047	0.045461	6.1624	80.995	1.1982
<b>Média</b>	<b>0.120400</b>	<b>0.042666</b>	<b>4.471900</b>	<b>43.630000</b>	<b>0.853500</b>
<b>Desvio Padrão</b>	<b>0.000333</b>	<b>0.002209</b>	<b>0.906400</b>	<b>18.523000</b>	<b>0.151800</b>
<b>Coef. Var.</b>	<b>0.002765</b>	<b>0.051772</b>	<b>0.202690</b>	<b>0.424540</b>	<b>0.173570</b>

*Tabela 13 Estudo de Caso 2: Parâmetros estatísticos das 50 realizações de 3 horas*

Realização	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Máx. Observ.
1	0.27227	0.06748	0.05691	2.78720	0.56550
2	0.27199	0.06759	0.03117	2.89110	0.51230
3	0.27218	0.06750	0.02136	2.90290	0.51680
4	0.27213	0.06801	0.03478	2.98260	0.52470
5	0.27092	0.06739	0.09866	3.17520	0.54430
6	0.27276	0.06760	0.07459	2.87250	0.55240
7	0.27189	0.06712	0.04371	2.87140	0.51540
8	0.27119	0.06685	0.08770	3.05170	0.54390
9	0.27114	0.06723	0.07094	2.95530	0.50270
10	0.27234	0.06766	0.11352	3.05010	0.55540
11	0.27242	0.06715	0.01352	2.92270	0.55370
12	0.27222	0.06791	0.03122	2.95740	0.55800
13	0.27203	0.06756	0.01316	2.95230	0.55700
14	0.27094	0.06685	0.04757	2.92010	0.51120
15	0.27171	0.06726	0.05409	2.95340	0.50910
16	0.27072	0.06669	0.12566	3.08520	0.59290
17	0.27161	0.06740	0.10638	2.96870	0.53960
18	0.27129	0.06749	0.07618	3.03170	0.55340
19	0.27249	0.06783	0.08400	2.95520	0.59570
20	0.27251	0.06735	-0.01969	2.80080	0.50310
21	0.27190	0.06788	0.05788	2.97150	0.53500
22	0.27117	0.06705	0.04815	2.96390	0.52610
23	0.27205	0.06774	0.07990	3.05130	0.55470
24	0.27179	0.06707	0.02155	2.88870	0.53750
25	0.27075	0.06712	0.09781	3.00870	0.54840
26	0.27190	0.06704	0.06539	2.94230	0.53890
27	0.27250	0.06757	0.04801	2.90610	0.51710
28	0.27131	0.06749	0.08567	3.01760	0.54880
29	0.27124	0.06763	0.03424	2.98620	0.54770
30	0.27127	0.06726	0.05067	2.97460	0.56640
31	0.27036	0.06701	0.07427	2.98540	0.56840
32	0.27186	0.06807	0.10741	2.90640	0.52450
33	0.27182	0.06778	0.08580	3.13060	0.58440
34	0.27103	0.06705	0.06689	2.90680	0.50790
35	0.27131	0.06714	0.06185	2.89650	0.56810
36	0.27118	0.06694	0.05413	2.94640	0.58990
37	0.27191	0.06733	0.08658	3.04890	0.56030
38	0.27163	0.06685	0.00841	2.91420	0.52020
39	0.27295	0.06783	0.03974	2.90370	0.53520
40	0.27140	0.06765	0.04787	3.00330	0.57660
41	0.27148	0.06735	0.07487	3.05320	0.57410
42	0.27195	0.06804	0.06770	3.04970	0.55270
43	0.27095	0.06678	0.03948	2.91830	0.51950

44	0.27154	0.06751	0.07060	3.01990	0.54760
45	0.27204	0.06770	0.08279	3.02250	0.58410
46	0.27199	0.06774	0.11177	2.97260	0.54140
47	0.27115	0.06700	0.05936	2.94450	0.52630
48	0.27222	0.06725	0.06845	2.96190	0.54050
49	0.27143	0.06723	0.02731	2.77980	0.50450
50	0.27181	0.06711	0.05382	2.90430	0.49850
<b>Média</b>	<b>0.271690</b>	<b>0.067382</b>	<b>0.060876</b>	<b>2.961300</b>	<b>0.543050</b>
<b>Desvio Padrão</b>	<b>0.000571</b>	<b>0.000356</b>	<b>0.030314</b>	<b>0.079027</b>	<b>0.025737</b>
<b>Coef. Var.</b>	<b>0.002103</b>	<b>0.005283</b>	<b>0.497970</b>	<b>0.026686</b>	<b>0.047394</b>

*Tabela 14 Estudo de Caso 3: Parâmetros estatísticos das 50 realizações de 3 horas*

Realização	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Máx. Observ.
1	0.11116	0.01212	0.94206	5.02560	0.19040
2	0.11115	0.01213	0.96241	5.07720	0.19100
3	0.11118	0.01226	1.02560	5.46350	0.19050
4	0.11121	0.01224	0.94976	4.96070	0.19830
5	0.11119	0.01250	1.41150	8.38360	0.21640
6	0.11119	0.01225	0.97535	5.14200	0.19500
7	0.11118	0.01220	0.98879	5.17590	0.19550
8	0.11116	0.01219	1.11670	6.16790	0.20980
9	0.11115	0.01213	1.07030	5.92830	0.19550
10	0.11119	0.01236	1.15050	6.30610	0.21220
11	0.11119	0.01231	1.15750	7.09570	0.22730
12	0.11120	0.01229	1.01240	5.48580	0.20770
13	0.11117	0.01216	0.97383	5.28490	0.20660
14	0.11114	0.01209	1.05520	5.83260	0.21250
15	0.11117	0.01219	1.02320	5.54470	0.19720
16	0.11110	0.01204	1.15260	6.58020	0.22130
17	0.11116	0.01215	1.00350	5.33100	0.19120
18	0.11117	0.01227	1.22920	7.65050	0.24680
19	0.11120	0.01232	1.05900	5.92120	0.21130
20	0.11116	0.01208	0.89648	4.82780	0.19280
21	0.11119	0.01227	1.03030	5.42020	0.18500
22	0.11117	0.01219	1.09870	6.06710	0.20250
23	0.11120	0.01245	1.32920	8.50370	0.24870
24	0.11117	0.01213	0.96677	5.04390	0.19400
25	0.11112	0.01208	1.11460	6.29930	0.20120
26	0.11117	0.01224	1.10960	6.05040	0.20340
27	0.11120	0.01232	1.06030	5.65490	0.19920
28	0.11119	0.01238	1.26990	7.43860	0.22150
29	0.11117	0.01231	1.17220	6.54120	0.20730
30	0.11117	0.01221	1.08090	6.04450	0.21230
31	0.11112	0.01205	1.08090	6.24030	0.21830
32	0.11120	0.01233	1.07200	5.84340	0.19950
33	0.11121	0.01248	1.31520	7.93390	0.22640
34	0.11116	0.01220	1.05710	5.68040	0.20100
35	0.11115	0.01218	1.14640	6.77600	0.22240
36	0.11113	0.01208	1.05210	5.97100	0.22240
37	0.11116	0.01218	1.05190	5.58590	0.19880
38	0.11114	0.01205	0.92348	4.78620	0.19070
39	0.11121	0.01224	0.90388	4.62070	0.18310
40	0.11118	0.01229	1.16510	6.97840	0.23720
41	0.11118	0.01240	1.31960	8.20820	0.23810
42	0.11121	0.01241	1.12100	5.95590	0.19350
43	0.11112	0.01203	1.02510	5.65730	0.20710

44	0.11119	0.01228	1.06800	5.74330	0.20170
45	0.11120	0.01237	1.17200	6.92060	0.23030
46	0.11117	0.01227	1.13310	6.67600	0.23780
47	0.11114	0.01204	0.92327	4.81640	0.18860
48	0.11118	0.01229	1.11460	6.32420	0.21600
49	0.11114	0.01205	0.91045	4.80700	0.19220
50	0.11116	0.01211	0.96810	5.05550	0.19100
<b>Média</b>	<b>0.111170</b>	<b>0.012224</b>	<b>1.078200</b>	<b>6.016600</b>	<b>0.207610</b>
<b>Desvio Padrão</b>	<b>0.000026</b>	<b>0.000123</b>	<b>0.117050</b>	<b>0.977270</b>	<b>0.016536</b>
<b>Coef. Var.</b>	<b>0.000232</b>	<b>0.010039</b>	<b>0.108560</b>	<b>0.162430</b>	<b>0.079649</b>

## ANEXO E

### PARÂMETROS ESTATÍSTICOS DOS PICOS DAS SÉRIES TEMPORAIS DE 3 HORAS

Para cada Estudo de Caso são apresentados os parâmetros estatísticos dos Picos de cada uma das 50 realizações individuais de 3 horas.

*Tabela 15 Estudo de Caso 1: Parâmetros estatísticos dos Picos das 50 realizações de 3h*

Realização	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Correl. Picos Consec.
1	0.17189	0.07161	3.99280	29.59600	0.26602
2	0.17677	0.07115	2.89920	13.95600	0.29150
3	0.17671	0.07467	2.86650	13.23200	0.31560
4	0.17554	0.07721	4.22680	32.51800	0.27031
5	0.18009	0.08774	4.90750	40.14200	0.24867
6	0.17593	0.08139	4.35360	29.19200	0.33826
7	0.17470	0.06491	3.08600	16.89500	0.32512
8	0.17862	0.08145	3.64240	21.32400	0.26933
9	0.17608	0.08348	3.87480	22.08200	0.27983
10	0.17877	0.08012	3.76570	23.82200	0.32687
11	0.17453	0.06975	3.78380	26.19900	0.29355
12	0.17520	0.07538	3.82300	25.48100	0.30838
13	0.17457	0.07342	3.98420	27.70100	0.37348
14	0.17540	0.07552	4.44710	33.90000	0.28240
15	0.17554	0.07639	3.77950	22.39900	0.33335
16	0.17654	0.08028	3.65260	20.67200	0.25543
17	0.17334	0.07231	3.42660	19.30300	0.30417
18	0.17297	0.08253	4.95530	37.57500	0.31661
19	0.17742	0.07855	3.86800	24.95000	0.23213
20	0.17453	0.06829	2.96310	15.64800	0.27999
21	0.17495	0.07998	4.67700	38.41400	0.37226
22	0.17376	0.07433	5.03140	45.47400	0.30713
23	0.17399	0.06890	3.61010	23.75700	0.26222
24	0.17893	0.07729	4.01680	29.51600	0.25337
25	0.17298	0.06710	3.19000	17.53600	0.27688
26	0.17243	0.07739	4.55640	34.16400	0.26935
27	0.17547	0.06326	2.44580	11.16600	0.33488
28	0.17486	0.06663	2.75500	13.36100	0.26096
29	0.17734	0.08062	3.82990	23.34500	0.24428
30	0.17318	0.06353	3.28000	20.11200	0.31647
31	0.17375	0.06450	2.74370	13.97000	0.36315
32	0.17460	0.07190	3.47450	19.65300	0.27325

33	0.17837	0.09527	4.67110	30.34200	0.36730
34	0.17432	0.07671	5.10140	45.74500	0.33659
35	0.17478	0.07420	2.99250	14.13500	0.33687
36	0.17693	0.08331	4.15450	27.90700	0.33827
37	0.17639	0.07526	3.76610	25.43300	0.29728
38	0.17600	0.07125	3.31110	19.00700	0.32638
39	0.17364	0.08614	6.25290	60.63800	0.37004
40	0.17511	0.07079	3.96750	29.53500	0.19197
41	0.17760	0.07841	3.44680	20.53900	0.33439
42	0.17498	0.07253	3.61390	22.55600	0.33817
43	0.17805	0.07514	3.22830	17.52100	0.32489
44	0.18070	0.09114	4.47640	30.72800	0.35594
45	0.17306	0.06571	3.45840	23.10100	0.22970
46	0.17745	0.09416	5.57350	47.69200	0.31394
47	0.17272	0.06697	3.44850	21.75000	0.28099
48	0.17389	0.06902	2.91620	14.20900	0.28271
49	0.17960	0.07866	4.11800	30.50000	0.31338
50	0.17453	0.08552	5.36210	46.56300	0.34579
<b>Média</b>	<b>0.17559</b>	<b>0.07584</b>	<b>3.87540</b>	<b>26.29900</b>	<b>0.30260</b>
<b>Desvio Padrão</b>	<b>0.00210</b>	<b>0.00762</b>	<b>0.79855</b>	<b>10.51100</b>	<b>0.04156</b>
<b>Coef. Var.</b>	<b>0.01196</b>	<b>0.10047</b>	<b>0.20606</b>	<b>0.39967</b>	<b>0.13734</b>

Tabela 16 Estudo de Caso 2: Parâmetros estatísticos dos Picos das 50 realizações de 3h

Realização	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Correl. Picos Consec.
1	0.35068	0.04572	0.57451	3.21900	0.43660
2	0.35027	0.04643	0.58251	2.91440	0.41236
3	0.35036	0.04558	0.62962	3.09110	0.40572
4	0.35036	0.04746	0.62990	3.00410	0.41606
5	0.34880	0.04784	0.86830	3.71630	0.45372
6	0.35190	0.04666	0.58530	3.11080	0.45499
7	0.34984	0.04577	0.52980	2.84940	0.43690
8	0.34858	0.04756	0.69666	3.31850	0.46186
9	0.34961	0.04638	0.63818	3.06070	0.47110
10	0.34955	0.04830	0.78097	3.65500	0.45117
11	0.34989	0.04563	0.61423	3.29620	0.41697
12	0.35122	0.04651	0.67755	3.37190	0.40424
13	0.34938	0.04653	0.57298	3.14770	0.42599
14	0.34752	0.04653	0.66567	3.13190	0.45612
15	0.34983	0.04687	0.59181	2.93520	0.45289
16	0.34819	0.04749	0.81277	3.99340	0.43619
17	0.35026	0.04753	0.72671	3.34790	0.45324
18	0.34936	0.04725	0.73425	3.57670	0.41106
19	0.35110	0.04759	0.71201	3.62200	0.44838
20	0.35161	0.04399	0.52930	2.97370	0.39993
21	0.35079	0.04689	0.67602	3.35550	0.43024
22	0.34875	0.04642	0.62869	3.23900	0.44566
23	0.35039	0.04736	0.76530	3.90190	0.47168
24	0.34940	0.04504	0.66635	3.36150	0.39852
25	0.34715	0.04789	0.73248	3.42870	0.45014
26	0.34920	0.04658	0.67599	3.20320	0.43402
27	0.35079	0.04602	0.64303	3.17000	0.41949
28	0.34851	0.04765	0.69842	3.18470	0.47437
29	0.35061	0.04590	0.70325	3.67110	0.43205
30	0.34913	0.04680	0.60276	3.20780	0.44027
31	0.34725	0.04687	0.73766	3.60090	0.46056
32	0.35197	0.04694	0.65697	3.18470	0.44929
33	0.34819	0.04860	0.86457	3.89710	0.44788
34	0.34916	0.04644	0.57704	2.94590	0.45518
35	0.34994	0.04573	0.61547	3.35720	0.45003
36	0.34951	0.04599	0.68691	3.73190	0.44026
37	0.34931	0.04764	0.74700	3.47010	0.44463
38	0.34888	0.04568	0.56605	3.03910	0.44864
39	0.35194	0.04620	0.63201	3.15260	0.40148
40	0.34966	0.04706	0.69747	3.31840	0.43513
41	0.35008	0.04640	0.83553	4.13630	0.47690
42	0.35018	0.04816	0.67183	3.28350	0.46197
43	0.34820	0.04483	0.69192	3.24930	0.46310

44	0.34950	0.04695	0.79598	3.67010	0.43453
45	0.35174	0.04641	0.72290	3.82750	0.40762
46	0.35123	0.04771	0.62958	3.04230	0.52203
47	0.34887	0.04663	0.68602	3.23530	0.42515
48	0.34997	0.04708	0.65690	3.22440	0.44552
49	0.34958	0.04480	0.53492	3.00630	0.42834
50	0.34915	0.04681	0.56271	2.87570	0.44108
<b>Média</b>	<b>0.34975</b>	<b>0.04666</b>	<b>0.67029</b>	<b>3.32620</b>	<b>0.44083</b>
<b>Desvio Padrão</b>	<b>0.00117</b>	<b>0.00095</b>	<b>0.08354</b>	<b>0.31094</b>	<b>0.02363</b>
<b>Coef. Var.</b>	<b>0.00336</b>	<b>0.02036</b>	<b>0.12463</b>	<b>0.09348</b>	<b>0.05359</b>

Tabela 17 Estudo de Caso 3: Parâmetros estatísticos dos Picos das 50 realizações de 3h

Realização	Média	Desvio Padrão	Coef. Assimetria	Coef. Curtose	Correl. Picos Consec.
1	0.12618	0.01172	1.66490	7.08930	0.49458
2	0.12574	0.01195	1.59940	6.59700	0.48242
3	0.12604	0.01227	1.75750	7.28860	0.49128
4	0.12549	0.01208	1.55350	6.27220	0.51859
5	0.12553	0.01370	2.45190	12.05700	0.54200
6	0.12587	0.01216	1.61080	6.60870	0.50212
7	0.12591	0.01208	1.62540	6.59710	0.50931
8	0.12572	0.01248	1.96900	8.81660	0.52280
9	0.12539	0.01223	1.93170	8.73820	0.48115
10	0.12632	0.01287	1.97440	8.71900	0.51724
11	0.12617	0.01251	2.36340	13.95500	0.53101
12	0.12626	0.01221	1.80110	7.95010	0.46793
13	0.12555	0.01214	1.63660	7.50760	0.48623
14	0.12552	0.01220	1.84920	8.47150	0.49825
15	0.12538	0.01230	1.75050	7.73900	0.48924
16	0.12511	0.01250	2.07170	9.98330	0.51010
17	0.12560	0.01218	1.69010	6.97950	0.48855
18	0.12553	0.01277	2.36040	14.09700	0.50936
19	0.12610	0.01234	1.92050	9.24330	0.48612
20	0.12591	0.01147	1.65090	7.17500	0.42932
21	0.12600	0.01236	1.73740	6.80040	0.46928
22	0.12523	0.01249	1.94780	8.66150	0.51472
23	0.12588	0.01320	2.60930	15.92700	0.55304
24	0.12560	0.01204	1.56830	6.18110	0.50985
25	0.12535	0.01233	2.02860	9.59030	0.48979
26	0.12606	0.01254	1.90820	8.32770	0.50443
27	0.12550	0.01250	1.80310	7.47730	0.52240
28	0.12590	0.01312	2.24430	11.31300	0.53586
29	0.12591	0.01280	2.07280	9.30940	0.53780
30	0.12574	0.01227	1.96130	9.19900	0.48504
31	0.12554	0.01215	1.99640	10.36600	0.52488
32	0.12620	0.01233	1.92720	8.53380	0.52022
33	0.12539	0.01330	2.41610	12.79800	0.51297
34	0.12568	0.01240	1.78830	7.56030	0.52437
35	0.12530	0.01243	2.17040	11.54700	0.51820
36	0.12605	0.01207	1.91540	9.63350	0.42990
37	0.12565	0.01241	1.75650	7.11700	0.50813
38	0.12573	0.01178	1.49670	5.84810	0.53388
39	0.12587	0.01192	1.42010	5.34340	0.47449
40	0.12549	0.01260	2.17260	11.81000	0.53229
41	0.12561	0.01317	2.47390	13.94800	0.51065
42	0.12553	0.01286	1.84620	7.56410	0.54602
43	0.12541	0.01198	1.83000	8.42610	0.46182

44	0.12557	0.01254	1.80820	7.68390	0.51365
45	0.12575	0.01269	2.22560	12.02900	0.47910
46	0.12564	0.01255	2.02630	11.09400	0.55738
47	0.12535	0.01175	1.52860	6.03650	0.49908
48	0.12607	0.01253	2.00520	9.87470	0.47374
49	0.12578	0.01159	1.56260	6.49390	0.49138
50	0.12563	0.01195	1.60820	6.35730	0.47756
<b>Média</b>	<b>0.12572</b>	<b>0.01238</b>	<b>1.90180</b>	<b>8.89470</b>	<b>0.50339</b>
<b>Desvio Padrão</b>	<b>0.00029</b>	<b>0.00044</b>	<b>0.28260</b>	<b>2.47060</b>	<b>0.02765</b>
<b>Coef. Var.</b>	<b>0.00230</b>	<b>0.03589</b>	<b>0.14860</b>	<b>0.27776</b>	<b>0.05493</b>