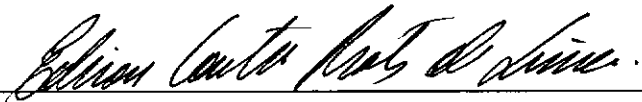


ANÁLISE ESTATÍSTICA DE VALORES EXTREMOS DE SÉRIES TEMPORAIS  
GAUSSIANAS E NÃO GAUSSIANAS

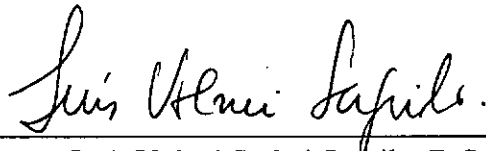
**Byron Iván Guerrero Zurita**

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA CIVIL.

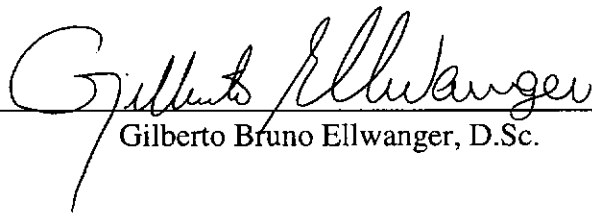
Aprovada por:



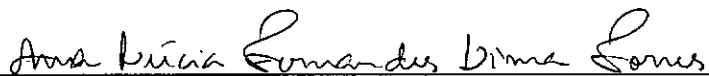
Edison Castro Prates de Lima, D.Sc.



Luís Volnei Sudati Sagrilo, D.Sc.



Gilberto Bruno Ellwanger, D.Sc.



Ana Lúcia Fernandes Lima Torres, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

ABRIL DE 1999

BYRON IVAN GUERRERO ZURITA

Análise Estatística de Valores Extremos de  
Séries Temporais Gaussianas e Não Gaussianas  
[Rio de Janeiro] 1999

IX, 90 p. 29,7 cm (COPPE/UFRJ, M.Sc.,  
Engenharia Civil, 1999)

Tese - Universidade Federal do Rio de  
Janeiro, COPPE

1. Estatística de extremos de sinais não  
gaussianos

I. COPPE/UFRJ II. Título ( série )

**A la memoria de mi Padre,  
A esa grande Mujer que es mi  
Madre,  
A mi Hermana del alma Grace,  
A mi Hermano Mario e Flia,  
A la Mujer que quiero Sonia.**

## AGRADECIMENTOS

Aos professores Luís Volnei Sudati Sagrilo e Edison Castro Prates de Lima pela orientação ao longo do desenvolvimento deste trabalho.

À minha Família.

Ao Corpo Docente e aos funcionários da Coppe, especialmente para Beth e a todos os amigos e colegas que contribuíram para que este trabalho fosse realizado.

À Capes.

A meus colegas Rosangela, Roberto, Ebenhezer, Adriano e Luis Alfredo

Aos Profesores Ellwanger, Landau, Batista, Edison e Sagrilo por sua transmissão de conhecimentos.

RESUMO DA TESE APRESENTADA À COPPE/UFRJ COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (M.Sc.).

ANÁLISE ESTATÍSTICA DE VALORES EXTREMOS DE SÉRIES TEMPORAIS  
GAUSSIANAS E NÃO GAUSSIANAS

Byron Iván Guerrero Zurita

Abril 1999

Orientadores: Edison Castro Prates de Lima  
Luís Volnei Sudati Sagrilo

Programa : Engenharia Civil

O objetivo deste trabalho é uma análise crítica de vários métodos, disponíveis na literatura, para análise estatística de valores extremos de séries temporais oriundas de processos gaussianos e não gaussianos.

Foram investigados quatro métodos, um método de extrapolação direta e três métodos baseados no ajuste da cauda da distribuição de máximos da série temporal. Os métodos investigados foram os seguintes:

- A. Método de extrapolação direta baseado no postulado da estabilidade da distribuição de Gumbel;
- B. Método de ajuste da cauda baseado na distribuição de Weibull ajustada a partir de uma Gumbel (*Weibull Tail*);
- C. Método de ajuste da cauda baseado na distribuição de Weibull ajustada por regressão linear a partir de um treshold de probabilidades da amostra de máximos;
- D. Método de ajuste da cauda baseado na distribuição de Pareto ajustada para valores da amostra de máximos acima de um treshold;

Na avaliação dos métodos acima citados foram utilizadas séries temporais gaussianas e não gaussianas obtidas a partir da simulação numérica da superfície das elevações do mar.

De uma maneira geral, todos os métodos avaliam bem a média do valor extremo. Com exceção do método C, o desvio padrão do valor extremo calculado pelos demais métodos apresenta uma maior incerteza.

ABSTRACT OF THESIS PRESENTED TO COPPE/UFRJ AS PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (M.Sc.).

## EXTREME VALUE ANALYSIS OF GAUSSIAN AND NON GAUSSIAN TIME SERIES

Byron Iván Guerrero Zurita

April 1999

Advisors: Edison Castro Prates de Lima

Luis Volnei Sudati Sagrilo

Department: Civil Engineering

The aim of this work is a critical evaluation of some methods available in the literature for extreme value statistics of gaussian and non gaussian time series.

Four methods have been investigated, one based on direct extrapolation and three based on the tail fitting of maxima sample of the time series. The methods are:

- A. Direct extrapolation method based on stability postulate of Gumbel distribution;
- B. Weibull tail fitting method for maxima sample based on a Gumbel distribution for a extreme value of a short period;
- C. Weibull tail fitting method for maxima sample based on linear regression analysis of maxima values exceeding a treshold level of probability;
- D. Tail fitting method for maxima sample based on Pareto distribution which is fitted to maxima values exceeding a threshold level;

Gaussian and non gaussian time series were derived from numerical simulations of the sea surface elevation.

In general, all of the methods give good results for the mean of the extreme value. In all methods but method C, the calculated standard deviation of the extreme value presents more uncertainty.

# ÍNDICE

## CAPITULO 1 INTRODUÇÃO E MOTIVAÇÃO DO TRABALHO

## CAPITULO 2 ESTATÍSTICA DOS EXTREMOS

2.1	Introdução.....	4
2.2	Definição de Amostra de Valores Máximos Extremos.....	4
2.3	Distribuição Probabilidades de Extremos Exata .....	7
2.4	Distribuições Assintóticas.....	8
2.5	Forma Assintótica Tipo I.....	10
2.6	Forma Assintótica Tipo II.....	13
2.7	Forma Assintótica Tipo III.....	14
2.8	Algumas Distribuições de Probabilidade.....	15
2.8.1	Distribuição de Gauss .....	15
2.8.2	Distribuição de Weibull.....	15

## CAPITULO 3 ANALISES ESTATÍSTICA DE SÉRIES TEMPORAIS.

3.1	Introdução.....	18
3.2	Funções Densidade de Probabilidade de uma Série Temporal e dos Seus máximos.....	20
3.3	Distribuição de Probabilidades de Valores Extremos.....	24
3.4	Distribuição de Probabilidades de Cartwright e Lonquet Higgins.....	25

## CAPÍTULO 4 ESTATÍSTICA DE VALORES MÁXIMOS EXTREMOS DE SÉRIES TEMPORAIS

4.1	Introdução.....	27
4.2	Método da Extrapolação Direta Baseado na Distribuição de Gumbel.....	28
4.3	Método de Ajuste da Cauda da Distribuição Parente Baseado na Distribuição de Weibull.....	30
4.4	Método de Ajuste da Cauda da Distribuição Parente pelo Método de Regressão Linear.....	35

4.5	Método Baseado na Distribuição Generalizada de Pareto ( Peak Over Threshold - POT) .....	37
<b>CAPITULO 5 GERAÇÃO DOS SINAIS</b>		
5.1	Introdução.....	41
5.2	Parâmetros Estatísticos de Um Estado de Mar.....	41
5.3	Geração de um Sinal Gaussiano de Elevações do Mar.....	45
5.4	Geração de um Sinal Não Gaussiano.....	51
<b>CAPITULO 6 APLICAÇÃO DOS MÉTODOS</b>		
6.1	Introdução.....	55
6.2	Sinal Gaussiano.....	55
6.2.1	Método da Extrapolação Direta Baseado na Distribuição de Gumbel.....	55
6.2.2	Método de Ajuste da Cauda da Distribuição Parente Baseado na Distribuição de Weibull.....	61
6.2.3	Método de Ajuste da Cauda da Distribuição Parente pelo Método de Regressão Linear.....	64
6.2.4	Método Baseado na Distribuição Generalizada de Pareto ( Peak Over Threshold - POT) .....	69
6.2.5	Resumo dos Métodos Aplicados na Análise de um Sinal Gaussiano.....	74
6.3	Sinal Não Gaussiano.....	75
6.3.1	Método da Extrapolação Direta Baseado na Distribuição de Gumbel.....	75
6.3.2	Método de Ajuste da Cauda da Distribuição Parente Baseado na Distribuição de Weibull.....	78
6.3.3	Método de Ajuste da Cauda da Distribuição Parente pelo Método de Regressão Linear.....	80
6.3.4	Método Baseado na Distribuição Generalizada de Pareto ( Peak Over Threshold – POT) .....	82

6.3.5	Resumo dos Métodos Aplicados na Análise de um Sinal Não Gaussiano.....	84
CAPITULO 7 CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS		85
REFERÊNCIAS BIBLIOGRÁFICAS.....		89

# CAPÍTULO 1

## INTRODUÇÃO E MOTIVAÇÃO DO TRABALHO

À medida que a prospecção de petróleo na plataforma continental foi ocorrendo em lâminas d'água cada vez mais profundas, as plataformas fixas tornaram-se inviáveis e foram substituídas por estruturas complacentes do tipo plataformas semi-submersíveis, plataformas de pernas atirantadas (TLP), navios estacionários de produção (FPSO) e etc.

No projeto dessas unidades estacionárias de produção, alguns componentes estruturais merecem especial atenção, tais como os risers de perfuração, completação e produção, as linhas de ancoragem das plataformas semi-submersíveis e de FPSOs e os tendões das TLPs.

Entre os principais aspectos que devem ser analisados nestes componentes estruturais destacam-se:

- Esforços e tensões máximas decorrentes das cargas estáticas e dinâmicas atuantes diretamente sobre os elementos estruturais e das induzidas pelos movimentos impostos da unidade flutuante;
- Fadiga estrutural devida às cargas anteriormente mencionadas e devida às vibrações induzidas pelo fenômeno de desprendimento de vórtices.

No caso da análise de risers flexíveis, observa-se que a resposta dos mesmos é nitidamente não-Gaussiana [3]. Deste modo torna-se importante o desenvolvimento de métodos que permitam o estabelecimento de distribuições de probabilidade adequadas para a descrição dos valores máximos e dos valores extremos da resposta.

Existem na literatura vários métodos propostos para estimar os valores extremos da resposta [3,12,14], a partir do ajuste de funções teóricas conhecidas de

distribuição de probabilidade para representar as correspondentes distribuições de valores máximos e de valores extremos das amostras.

A motivação principal deste trabalho foi a realização de uma análise crítica comparativa entre os vários métodos propostos para a determinação da estatística dos valores máximos extremos de séries temporais, provenientes de sinais de respostas estacionárias não Gaussianas.

É importante observar que, no ajuste de funções teóricas de distribuição de probabilidades, deve-se ter especial atenção para a correta representação dos valores máximos da amostra, associados ao extremo superior da função de distribuição (tramo superior ou *tail*).

O ajuste de funções de distribuição de probabilidade, baseadas em técnicas convencionais, são direcionadas para uma boa representação nas vizinhanças do valor médio da amostra, podendo introduzir erros significativos na estimativa dos valores extremos, caso o modelo de distribuição adotado seja incapaz de descrever corretamente a estatística dos maiores valores da amostra.

Como consequência, foi observado que os métodos, baseados no ajuste da cauda (*tail-fitting*) das distribuições de probabilidade, tendem a fornecer uma melhor estimativa para os valores extremos da resposta.

O **Capítulo 2** apresenta uma revisão da teoria clássica da estatística de valores extremos.

No **Capítulo 3** é feita a descrição estatística de uma série temporal.

O **Capítulo 4** apresenta a descrição dos diferentes métodos propostos na literatura para a estimativa de valores máximos extremos de um sinal não Gaussiano.

No **Capítulo 5** é descrita a geração de um sinal Gaussiano, correspondente a uma realização de um processo aleatório para a simulação das elevações da superfície do mar, e é também apresentada a geração de um processo não Gaussiano.

No **Capítulo 6** é feita a análise comparativa entre os diversos métodos, descritos no Capítulo 4, para a estimativa de valores máximos extremos de séries

temporais, provenientes de sinais Gaussianos e não Gaussianas.

Finalmente, no **Capítulo 7** são apresentados as principais conclusões e sugestões para desenvolvimentos futuros.

# CAPÍTULO 2

## ESTATÍSTICA DE VALORES EXTREMOS

### 2.1 – INTRODUÇÃO

Neste capítulo serão apresentados os principais resultados da estatística de valores extremos de variáveis aleatórias. Por valores extremos entende-se o *maior* e/ou o *menor* valor atingido por uma variável aleatória dentro de um determinado tempo.

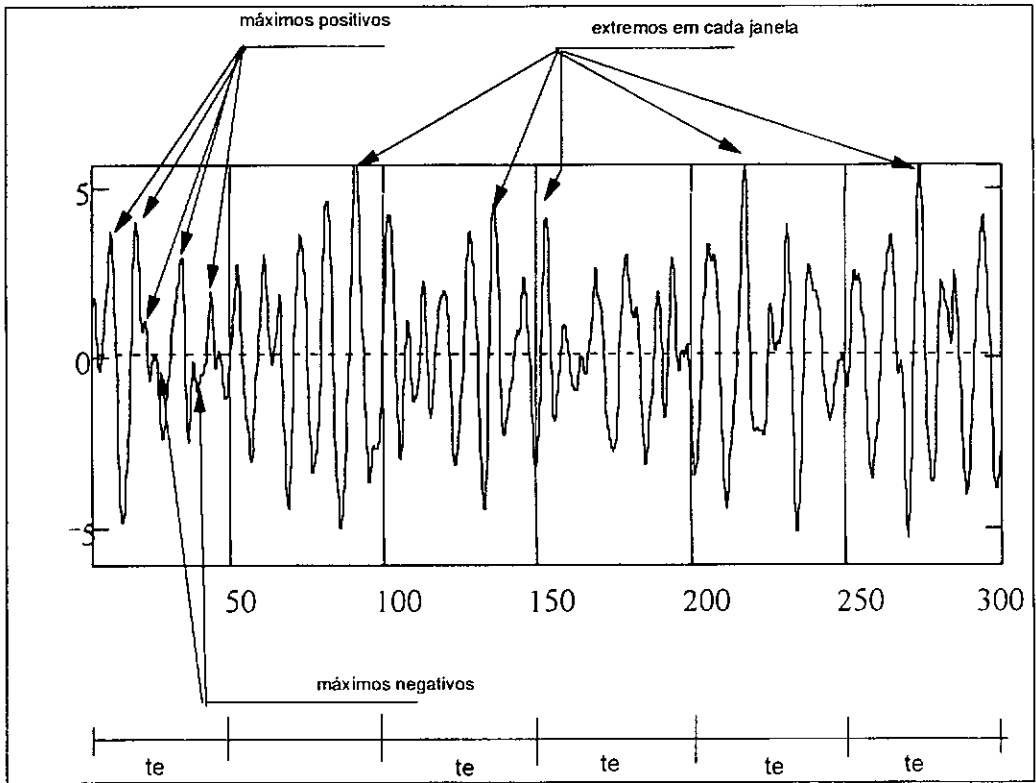
Observa-se que, para cada novo conjunto de valores da variável aleatória (realização), ocorrerá um valor extremo que será geralmente distinto dos valores extremos observados em realizações anteriores. Deste modo, o valor extremo de uma variável aleatória é também uma variável aleatória associada a uma distribuição de probabilidades. A distribuição de probabilidades de valores extremos está relacionada com a distribuição de probabilidades associada aos valores máximos da variável aleatória, sendo considerada neste caso como a distribuição parente da referida distribuição de probabilidades de valores extremos.

### 2.2 – DEFINIÇÃO DE AMOSTRA DE VALORES MÁXIMOS E EXTREMOS

A resposta de uma análise aleatória no domínio do tempo é obtida sob a forma de uma série temporal de valores igualmente espaçados. Para efeito de estudo da estatística de valores extremos, esta série é denominada de **amostra parente**  $X_p$ .

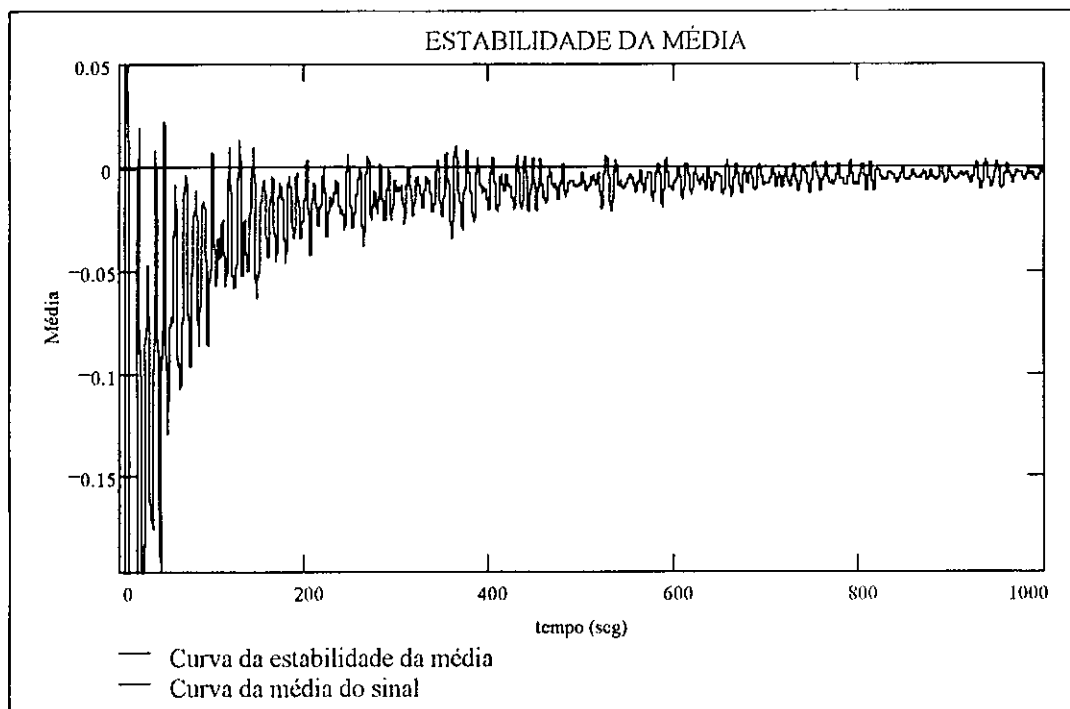
A **amostra de máximos**  $X_m$ , é constituída de todos os valores máximos individuais da resposta, tal como está indicado na Figura 2.1. A **amostra de extremos**  $X_e$ , é formada pelo conjunto dos maiores valores máximos ocorridos em cada um dos  $N_e$  intervalos de duração  $t_e$  em que foi subdividida a série original (amostra parente).

Cada intervalo contém em média  $m = v_m t_e$  valores máximos, onde  $v_m$  é a frequência média de valores máximos.

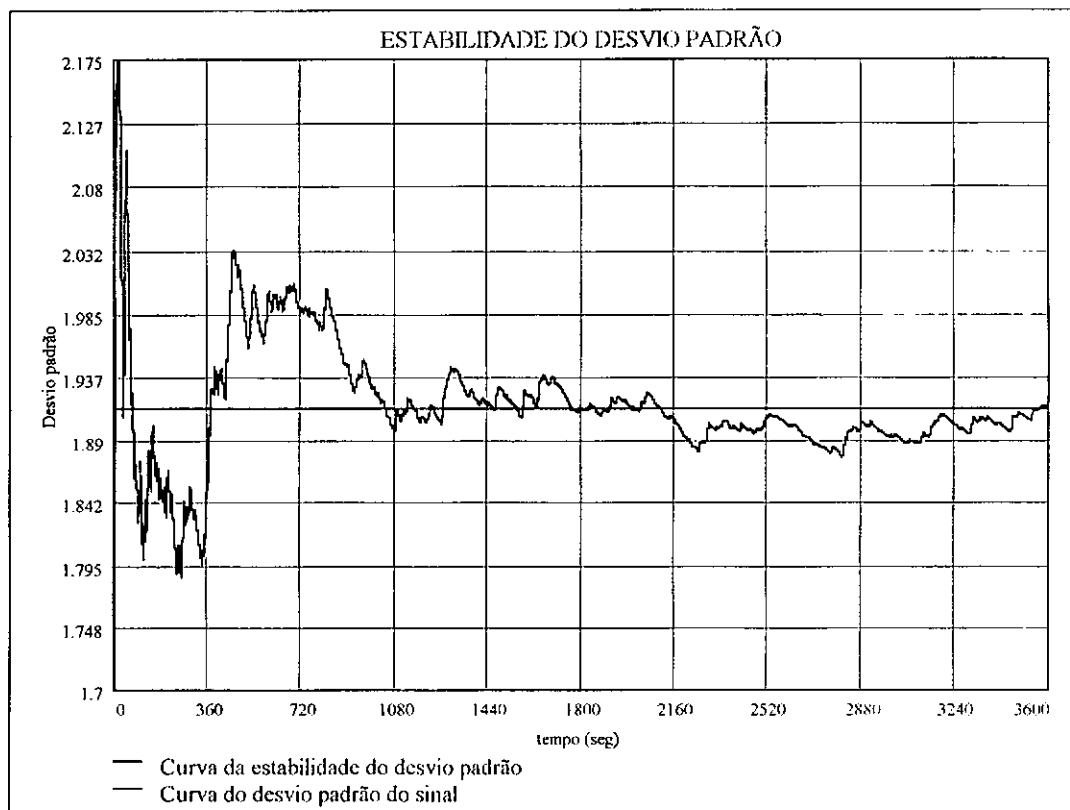


*Figura 2.1 – Definição das amostras parente, de valores máximos e de valores extremos.*

Neste trabalho assume-se que as séries temporais são amostras de processos aleatórios estacionários e ergódicos [5], i.e., uma única série temporal é capaz de representar estatisticamente o processo. Desta forma, é importante observar que o intervalo ou janela  $t_e$  deve ter uma duração suficiente de forma a garantir a estabilidade estatística do sinal, ou seja a estabilidade da média e a estabilidade do desvio padrão do mesmo, como está indicado, respectivamente, nas Figuras (2.2) e (2.3).



*Figura 2.2 – Estabilidade da média*



*Figura 2.3 – Estabilidade do desvio padrão*

### 2.3 - DISTRIBUIÇÃO PROBABILIDADES DE EXTREMOS EXATA

Seja  $X$  uma variável aleatória inicial, possuindo uma função cumulativa de probabilidades  $F_X(x)$  conhecida e assumindo-se a existência de várias amostras de tamanho  $n$  da variável aleatória  $X$ , i.e.  $(x_1, x_2, \dots, x_n)$ , onde os índices representam o primeiro, o segundo, ..., e o  $n$ -ésimo valor observado em cada uma das amostras. Uma vez que o valor da variável aleatória é imprevisível antes da ocorrência da respectiva observação, pode-se assumir que cada observação corresponde ao valor de uma variável aleatória, sendo que o conjunto de observações  $(x_1, x_2, \dots, x_n)$  pode ser considerado como uma realização de um conjunto de  $n$  variáveis aleatórias  $(X_1, X_2, \dots, X_n)$ . Portanto, o valor máximo extremo de uma amostra de tamanho  $n$  pode ser considerado uma variável aleatória definida por [1]:

$$Y_n = \max(X_1, X_2, \dots, X_n) \quad (2.1)$$

O valor  $Y_n$  é o maior valor ocorrido entre os valores de  $(X_1, X_2, \dots, X_n)$ . Neste caso se o valor de  $Y_n$  for menor que um dado valor  $y$ , então necessariamente todas as variáveis  $(X_1, X_2, \dots, X_n)$ , devem ser menores que  $y$ . Assumindo-se que os valores de  $X_1, X_2, \dots, X_n$  sejam estatisticamente independentes e identicamente distribuídos à variável  $X$ , ou seja:

$$F_{X_1}(x) = F_{X_2}(x) = F_{X_3}(x) \dots = F_{X_n}(x) \quad (2.2)$$

A função cumulativa do valor máximo extremo  $Y_n$  é então definida como:

$$F_{Y_n}(y) = P(Y_n \leq y) \quad (2.3)$$

$$F_{Y_n}(y) = P(X_1 \leq y; X_2 \leq y, \dots, X_n \leq y) \quad (2.4)$$

$$F_{Y_n}(y) = [F_X(y)]^n \quad (2.5)$$

e a correspondente função densidade de probabilidades por:

$$f_{Y_n}(y) = \frac{dF_{Y_n}(y)}{dy} = n[F_X(y)]^{n-1} f_X(y) \quad (2.6)$$

onde  $f_X(y)$  é função densidade de probabilidades da variável inicial X.

O valor mínimo de uma amostra de tamanho n, como pode ser visto em [1], é definido de forma análoga, porém, não é de interesse no presente trabalho.

## 2.4 - DISTRIBUIÇÕES ASSINTÓTICAS

Observa-se que as distribuições de extremos tendem para distribuições assintóticas quando n tende para o infinito. Isto pode ser observado graficamente na Figura 2.4.

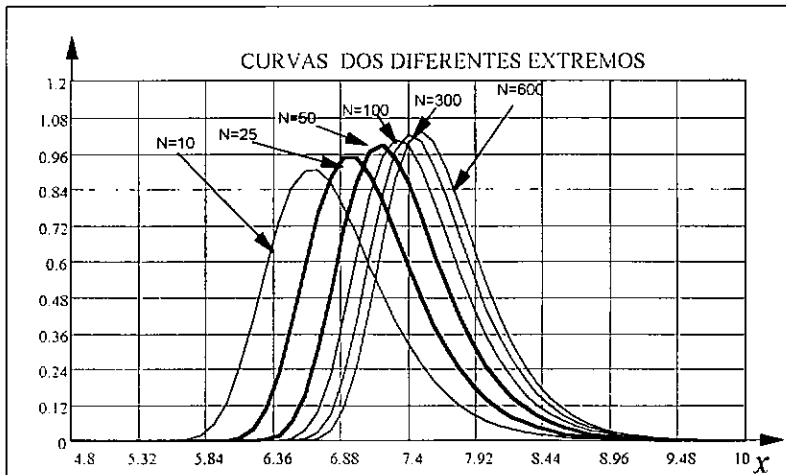


Figura 2.4 – Distribuições de extremos para diferentes valores de n

É importante salientar que a forma das distribuições assintóticas não depende da forma exata da distribuição de probabilidades inicial da variável aleatória, mas somente do comportamento do tramo (cauda) desta distribuição na direção da extremidade de interesse (máximo ou mínimo) da variável inicial.

Gumbel [2] dependendo da distribuição inicial as distribuições assintóticas de extremos convergem para um dos três tipos definidos a seguir:

- **Distribuição Assintótica TIPO I (ou de Gumbel):** distribuição assintótica na forma exponencial dupla:

$$F_{Y_n}(y) = \exp(-\exp(-A(n)y)) \quad (2.7)$$

- **Distribuição Assintótica TIPO II (ou de Frechet):** distribuição assintótica na forma exponencial:

$$F_{Y_n}(y) = \exp\left(-\frac{A(n)}{y^k}\right) \quad (2.8)$$

- **Distribuição Assintótica TIPO III:** distribuição na forma exponencial, porém, truncadas, num limite superior  $w$ :

$$F_{Y_n}(y) = \exp(-A(n)(w-y)^k) \quad (2.9)$$

Nas Equações (2.7) a (2.9),  $A(n)$  e  $k$  são os parâmetros das distribuições.

Pode-se demonstrar [1] que a distribuição de valores extremos, associada a uma distribuição parente com decaimento exponencial na direção dos mesmos, converge para a distribuição assintótica de Tipo I. Caso a distribuição inicial apresente um decaimento polinomial na direção dos extremos, haverá convergência para a distribuição assintótica de valores extremos de Tipo II. Finalmente, se os valores extremos possuírem um limite finito superior ou inferior, a correspondente distribuição de extremos converge para a forma assintótica de Tipo III.

## 2.5 FORMA ASSINTÓTICA DE TIPO I

A função cumulativa de probabilidades da distribuição de valores máximos extremos Tipo I ou Gumbel é definida pela expressão:

$$F_{X_n}(x) = \exp\left[-\exp\left[-\alpha_n(x - u_n)\right]\right] \quad (2.10)$$

onde  $\alpha_n$  e  $u_n$  são, respectivamente, os parâmetros de locação e de escala.

A correspondente função de densidade de probabilidades é:

$$f_{X_n}(x) = \alpha_n \exp\left[-\alpha_n(x - u_n)\right] \exp\left[-\exp\left[\alpha_n(x - u_n)\right]\right] \quad (2.11)$$

A média e o desvio padrão da distribuição de extremos do Tipo I são dados por:

$$\mu_{Y_n} = u_n + \frac{0.57722}{\alpha_n} \quad (2.12)$$

$$\sigma_{Y_n} = \frac{\pi}{\sqrt{6}\alpha_n} \quad (2.13)$$

sendo  $n$  o número de valores da amostra considerada.

Caso seja conhecida a distribuição de probabilidades parente da variável aleatória, os parâmetros  $\alpha_n$  e  $u_n$  podem ser determinados pelas equações:

$$F_X(u_n) = 1 - \frac{1}{n} \quad (2.14)$$

$$\alpha_n = n f_X(u_n) \quad (2.15)$$

onde  $F_X()$  e  $f_X()$  são, respectivamente, a função cumulativa e a função de densidade de probabilidade da distribuição de probabilidades parente da variável aleatória e  $n$  é o número de ocorrências da variável aleatória num período de tempo pré-estabelecido.

A distribuição Tipo I é a distribuição assintótica de extremos para várias distribuições de probabilidade parentes de interesse em engenharia [1] tais como: normal, Rayleigh, exponencial, Weibull, etc.

A distribuição Tipo I pode ser usada diretamente no ajuste de uma amostra de valores extremos. Por exemplo, os parâmetros  $\alpha_n$  e  $u_n$  podem ser obtidos diretamente da amostra de valores extremos  $X_e$ , determinada a partir da divisão da série temporal original em  $N_e$  intervalos de duração  $t_e$ , como indicado na Figura 2.1. Observa-se que cada intervalo  $t_e$  contém um número  $n=vt_e$  de valores máximos. A distribuição Tipo I assim obtida é a distribuição do valor extremo associado ao período  $t_e$ .

Para uma amostra de valores extremos, os parâmetros  $\alpha_n$  e  $u_n$  podem ser obtidos com o auxílio de regressão linear [ 2 , 3 ] e são dados pelas seguintes expressões:

$$\alpha_n = \frac{C_1}{S_e} \quad u_n = M_e - \frac{C_2}{C_1} S_e \quad (2.16)$$

onde  $M_e$  e  $S_e$  são, respectivamente, a média e o desvio padrão da amostra de valores extremos  $X_e$ , calculados pelas expressões:

$$M_e = \frac{1}{N_e} \sum_{k=1}^{N_e} X_{e_k} \quad , \quad S_e^2 = \frac{1}{N_e - 1} \left( \sum_{k=1}^{N_e} X_{e_k}^2 - N_e M_e^2 \right) \quad (2.17)$$

Os coeficientes de regressão linear  $C_1$  e  $C_2$  são dados por:

$$C_1 = \sqrt{C_3 - C_2^2} \quad , \quad C_2 = \frac{1}{N_e} \sum_{k=1}^{N_e} \left[ -\ln \left[ -\ln \left( \frac{k}{N_e + 1} \right) \right] \right] \quad (2.18)$$

$$C_3 = \frac{1}{N_e} \sum_{k=1}^{N_e} \left[ \ln \left[ -\ln \left( \frac{k}{N_e + 1} \right) \right] \right]^2 \quad (2.19)$$

O *postulado de estabilidade* [1] da distribuição Tipo I ou de Gumbel demonstra que se  $F_{X_e}(X_e)$  é a função cumulativa de probabilidades da distribuição de Gumbel, então a  $[F_{X_e}(X_e)]^k$  também é uma função cumulativa de probabilidade de Gumbel cujos parâmetros  $\alpha_n$  e  $u_n$  são dados por:

$$\alpha_k = \alpha_n \quad , \quad u_k = u_n + \frac{\ln(k)}{\alpha_n} \quad (2.20)$$

O postulado de estabilidade estabelece, portanto, que o desvio padrão  $\sigma_k$  (ver Eq. 2.17) da distribuição  $[F_{X_e}(X_e)]^k$  é idêntico ao desvio padrão  $\sigma_e$  da distribuição original  $F_{X_e}(X_e)$  e que a média  $\mu_k$  da distribuição  $[F_{X_e}(X_e)]^k$  está relacionada à média  $\mu_e$  da distribuição original através da relação:

$$\mu_k = \mu_e + \beta_k \sigma_e \quad (2.21)$$

onde o parâmetro  $\beta_k$  é dado por [3]:

$$\beta_k = \frac{0.57722 - C_2}{C_1} + \frac{\ln(k)}{C_1} \quad (2.22)$$

onde os coeficientes  $C_1$  e  $C_2$  são definidos pelas equações (2.18) e (2.19).

## 2.6 FORMA ASSINTÓTICA TIPO II

Os valores extremos de uma distribuição inicial com uma extremidade polinomial em direção ao apropriado extremo, converge para uma distribuição assintótica Tipo II (ou de Frechet), cuja a função cumulativa de probabilidades é dada pela expressão:

$$F_{Y_n}(y) = \exp\left(-\frac{v_n}{y}\right)^k \quad (2.23)$$

A correspondente função de densidade de probabilidades é:

$$f_{Y_n}(y) = \frac{k}{v_n} \left(\frac{v_n}{y}\right)^{k+1} \exp\left[-\left(\frac{v_n}{y}\right)^k\right] \quad (2.24)$$

onde:

$v_n$  = parâmetro de escala

$k$  = parâmetro de forma

Os momentos (média e variância) da curva assintótica Tipo II podem ser obtidos com base nos parâmetros:

$$\mu_{Y_n} = v_n \Gamma\left[1 - \frac{1}{k}\right] \quad (2.25)$$

$$\sigma_{Y_n}^2 = v_n^2 \left[ \Gamma\left(1 - \frac{2}{k}\right) - \Gamma^2\left(1 - \frac{1}{k}\right) \right] \quad (2.26)$$

## 2.7 DISTRIBUIÇÃO ASSINTÓTICA TIPO III

A função cumulativa de probabilidades da distribuição assintótica de valores extremos do Tipo III é dada por:

$$F_{Y_n}(y) = 1 - \exp\left(-\left[\frac{y - U_n}{\alpha_n}\right]^{k_n}\right) \quad (2.27)$$

Onde  $U_n$ ,  $\alpha_n$  e  $k_n$  são, respectivamente, os parâmetros de locação, escala e forma da distribuição.

A distribuição assintótica do Tipo III é também conhecida na literatura como distribuição de Weibull [1].

A distribuição assintótica do Tipo III serve para distribuições iniciais que tenham um limite superior (truncamento). Devido a isto ela é de pouco interesse como distribuição de extremos. Porém, como distribuição inicial, como veremos mais adiante, esta tem um largo emprego.

## 2.8 ALGUMAS DISTRIBUIÇÕES DE PROBABILIDADE

A seguir serão apresentadas algumas distribuições de probabilidades que apresentam interesse neste trabalho. Juntamente com estas distribuições serão apresentadas as distribuições de extremos assintóticas associadas às mesmas.

### 2.8.1 – Distribuição de Gauss

Uma variável aleatória  $X$  tem uma distribuição de probabilidades Gaussiana se sua função densidade de probabilidades pode ser representada por:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right] \quad (2.28)$$

onde  $\mu$  e  $\sigma$  são, respectivamente, a média e o desvio padrão da variável aleatória  $X$ . Uma variável Gaussiana possui coeficiente de assimetria igual a zero e coeficiente de kurtosis igual a três.

O valor extremo de uma amostra com  $n$  valores de uma variável aleatória Gaussiana tem uma distribuição de extremos do Tipo I com os parâmetros definidos por:

$$\alpha_n = \frac{\sqrt{2 \ln(n)}}{\sigma} \quad (2.29)$$

$$u_n = \left[ \sqrt{2 \ln(n)} - \frac{\ln(\ln(n)) + \ln(4\pi)}{2\sqrt{2 \ln(n)}} \right] \sigma + \mu \quad (2.30)$$

## 2.8.2 – Distribuição de Weibull

Uma variável aleatória  $X$  tem uma distribuição de probabilidades do tipo Weibull se sua função densidade de probabilidades puder ser representada por:

$$f(x) = \frac{(x-u)^{k-1}}{\alpha^k} k \exp \left[ - \left( \frac{x-u}{\alpha} \right)^k \right] \quad x > u \quad (2.31)$$

onde  $u$ ,  $\alpha$  e  $k$  são respectivamente os parâmetros de posição, escala e forma da distribuição. A função cumulativa de probabilidades da distribuição de Weibull é dada pela expressão (2.27).

A média  $\mu$  e o desvio padrão  $\sigma$  da distribuição de Weibull são, respectivamente, dados por [1]:

$$\mu = u + \alpha \Gamma \left[ 1 + \frac{1}{k} \right] \quad (2.32)$$

$$\sigma = \alpha \left[ \Gamma \left( 1 + \frac{2}{k} \right) - \Gamma^2 \left( 1 + \frac{1}{k} \right) \right]^{\frac{1}{2}} \quad (2.33)$$

O coeficiente de assimetria  $\gamma_1$  e o de kurtosis  $\gamma_2$  da distribuição de Weibull são calculados, respectivamente, por [1]:

$$\gamma_1 = \frac{\Gamma \left( 1 + \frac{3}{k} \right) - 3\Gamma \left( 1 + \frac{1}{k} \right) \Gamma \left( 1 + \frac{2}{k} \right) + 2\Gamma^3 \left( 1 + \frac{1}{k} \right)}{\left[ \Gamma \left( 1 + \frac{2}{k} \right) - \Gamma^2 \left( 1 + \frac{1}{k} \right) \right]^{\frac{3}{2}}} \quad (2.34)$$

$$\gamma_2 = \frac{\Gamma\left(1 + \frac{4}{k}\right) - 4\Gamma\left(1 + \frac{1}{k}\right)\Gamma\left(1 + \frac{3}{k}\right) + 6\Gamma^2\left(1 + \frac{1}{k}\right)\Gamma\left(1 + \frac{2}{k}\right) - 3\Gamma^4\left(1 + \frac{1}{k}\right)}{\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right]^2} \quad (2.35)$$

Pode ser demonstrado que o valor máximo extremo de uma amostra com  $n$  valores de uma variável aleatória cuja distribuição de probabilidades é a distribuição de Weibull, tem uma distribuição de extremos do Tipo I com os parâmetros  $U_n$  e  $\alpha_n$  definidos por [3]:

$$U_n = u + \alpha(\ln n)^{\frac{1}{k}} \quad (2.36)$$

$$\alpha_n = \frac{k}{\alpha} (\ln n)^{\frac{k-1}{k}} \quad (2.37)$$

A distribuição de Rayleigh e a distribuição exponencial são importantes casos particulares da distribuição de Weibull, i.e.:

$$\text{Rayleigh } F(x) = 1 - \exp\left(-\frac{1}{2}\left(\frac{x-u}{\alpha}\right)^2\right) \quad x > u \quad (2.38)$$

$$\text{Exponencial } F(x) = 1 - \exp\left(-\left(\frac{x-u}{\alpha}\right)\right) \quad x > u \quad (2.39)$$

Segundo a Eq. (2.27), a distribuição de Rayleigh é obtida fazendo-se  $k=2$  e  $\alpha = \sqrt{2}\alpha$  e a distribuição exponencial é o caso particular da distribuição de Weibull para  $k=1$ .

Através das Eq. (2.27) o desvio padrão dos valores extremos de variável aleatória que tem uma distribuição parente exponencial é constante e independente do número  $n$ .

## CAPÍTULO 3

### ANÁLISE ESTATÍSTICA DE SÉRIES TEMPORAIS

#### 3.1 - INTRODUÇÃO

Uma dada resposta  $X$  de uma análise aleatória no domínio do tempo é obtida sob a forma de uma série temporal  $X(t)$  com  $N$  valores  $X_i(t_i)$  igualmente espaçados de um intervalo de tempo  $\Delta t$ . Os principais parâmetros estatísticos da série são definidos na Tabela 3.1 [8,16].

Parâmetro	Definição	Equação
Número de cruzamentos ascendentes	$N_c$	(3.1)
Número de picos positivos	$N_p^+$	(3.2)
Número total de picos	$N_{TP}$	(3.3)
Período de cruzamento zero	$T_z = \frac{T}{N_c} = \frac{N\Delta t}{N_c}$	(3.4)
Período de pico	$T_p = \frac{T}{N_p^+}$	(3.5)
Largura de banda	$\varepsilon = \sqrt{1 - \left[\frac{T_p}{T_z}\right]^2}$	(3.6)
Média	$\bar{x} = \sum_{i=1}^N \frac{X_i(t)}{N}$	(3.7)
Valor médio quadrático	$VMQ = \sum_{i=1}^N \frac{X_i(t)^2}{N}$	(3.8)
Variância	$Var = \sum_{i=1}^N \frac{\left(X_i(t) - \bar{x}\right)^2}{N-1}$	(3.9)

*Tabela 3.1 – Parâmetros estatísticos de uma série temporal*

Desvio padrão	$\sigma = \sqrt{Var}$	( 3.10 )
Coefficiente de assimetria	$\gamma_1 = \frac{\sum_{i=1}^N \frac{[X_i(t) - \bar{x}]^3}{N-1}}{\sigma^3}$	( 3.11 )
Coefficiente de Kurtosis	$\gamma_2 = \frac{\sum_{i=1}^N \frac{[X_i(t) - \bar{x}]^4}{N-1}}{\sigma^4}$	( 3.12 )

*Tabela 3.1 (cont.) – Parâmetros estatísticos de uma série temporal*

A densidade espectral  $S(\omega)$  de uma série temporal pode ser obtida através de uma transformada de Fourier [5]. Os parâmetros estatísticos tais como períodos, momentos, largura de banda do sinal, etc. podem ser obtidos, também, a partir da densidade espectral através das relações apresentadas na Tabela 3.2.

Parâmetro	Definição	Equação
n-ésimo momento	$M_n = \int_{\omega=0}^{\alpha} \omega^n S(\omega) d\omega$	( 3.13 )
Período de cruzamento zero	$T_z = 2\pi \sqrt{\frac{M_0}{M_2}}$	( 3.14 )
Período de pico	$T_p = 2\pi \sqrt{\frac{M_2}{M_4}}$	( 3.15 )
Largura de banda	$\varepsilon = \sqrt{1 - \left[ \frac{T_p}{T_z} \right]^2} = \sqrt{1 - \left[ \frac{M_2^2}{M_0 M_4} \right]}$	( 3.16 )
Variância	$Var = \sigma^2 = M_0$	( 3.17 )

*Tabela 3.2 – Parâmetros estatísticos de uma série temporal a partir do espectro*

Além dos parâmetros definidos anteriormente é possível obter as distribuições de densidade de probabilidade da série temporal e dos picos (máximos) positivos e negativos (picos totais). Estas distribuições são as distribuições teóricas para identificar qual a distribuição teórica que se ajusta ao sinal.

### 3.2 – FUNÇÕES DENSIDADE DE PROBABILIDADE DE UMA SÉRIE TEMPORAL E DOS SEUS MÁXIMOS

Apresenta-se a seguir a metodologia utilizada neste trabalho para efetuar a análise estatística de uma série temporal.

Seja a série temporal  $X(t)$  com  $N$  valores  $X_i(t_i)$  igualmente espaçados de um intervalo de tempo  $\Delta t$ , como está indicado na Figura 3.1. São determinados, inicialmente, os seus valores máximo  $X_{\max}$  e mínimo  $X_{\min}$  sendo este intervalo dividido em  $m$  intervalos iguais  $\Delta X = (X_{\max} - X_{\min}) / m$ .

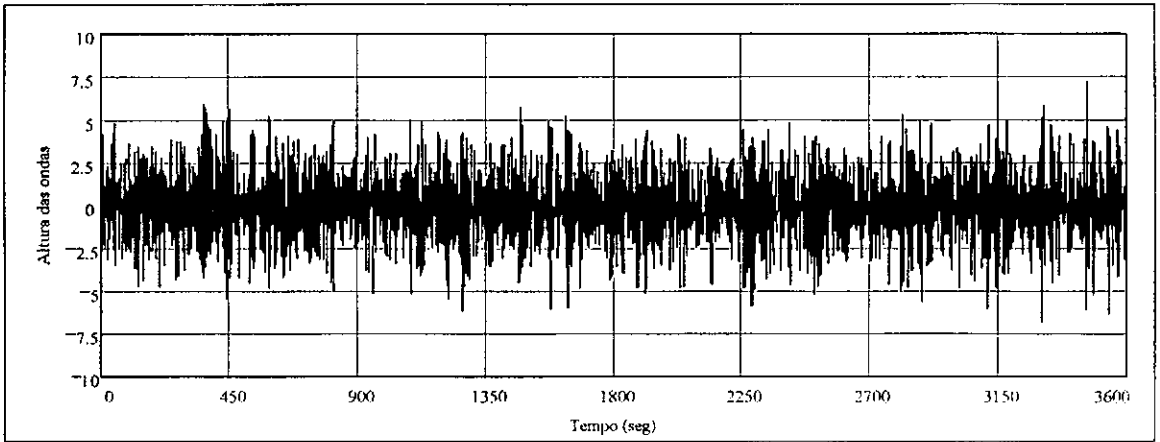


Figura 3.1 - Série temporal  $X(t)$  com  $N$  valores igualmente espaçados de  $\Delta t$

Para cada intervalo  $i$  ( $i = 1, 2, \dots, m$ ), cujos limites são definidos por:

$$X_{\min} + (i - 1) \Delta X \leq X_i < X_{\min} + i \Delta X \quad (3.18)$$

podem ser identificados:

- o número de pontos  $n_i$  da função  $X(t)$  que estão localizados neste intervalo.
- número de picos (máximos) positivos  $N_{p_i}^+$ . Um pico positivo satisfaz as seguintes relações:

$$X_i(t) > 0 \quad (3.19)$$

$$X_i(t) > X_{i-1}(t) \quad \text{e} \quad X_i(t) > X_{i+1}(t) \quad (3.20)$$

- número total de picos (máximos) positivos e negativos  $N_{TP}$ . Um máximo satisfaz as seguintes relações :

$$X_i(t) > X_{i-1}(t) \quad \text{e} \quad X_i(t) > X_{i+1}(t) \quad (3.21)$$

Na definição da função densidade de probabilidade da série temporal parte-se de uma função  $X(t)$  contínua, como indicado na Figura 3.2. A função densidade de probabilidade pode ser calculada pela expressão [ 5 ]:

$$f_x(x)\Delta x = \frac{dt_1 + dt_2 + dt_3 + \dots + dt_n}{T} = \frac{1}{T} \sum_{i=1}^n dt_i \quad (3.22)$$

Onde  $n$  é o número de vezes que a série temporal permanece no intervalo  $[x, x+\Delta x]$ .

Fazendo-se  $T$  tender ao infinito, resulta:

$$f_x(x)dx = \lim_{T \rightarrow \infty} \frac{\sum_{i=1}^{n_x} dt_i}{T} \quad (3.23)$$

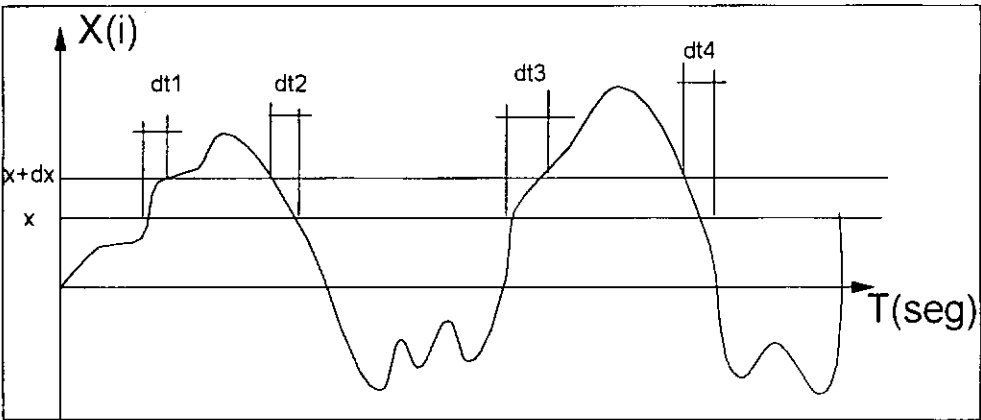


Figura 3.2 - Função  $X(t)$  contínua

Observa-se na Eq. (3.22) que, no caso de uma série discreta, o somatório  $\sum dt_i$  é equivalente ao número de pontos  $n_i$  da função  $X_i(t)$  no intervalo definido em (3.18) multiplicado por  $\Delta t$ . Assim o valor na função densidade de probabilidade da série temporal pode ser obtida por:

$$f_x(x_i) = \frac{n_i \Delta t}{T \Delta x} = \frac{n_i \Delta t}{N \Delta t \Delta x} = \frac{n_i}{N \Delta x} \quad (3.24)$$

A função densidade de probabilidade calculada em (3.24) define uma distribuição de probabilidade teórica que melhor se ajuste à série temporal. Técnicas de ajuste de distribuições podem ser vistas na Ref. [17]. A Figura 3.3 apresenta um exemplo onde uma distribuição normal representa muito bem a função densidade de probabilidades de uma série temporal

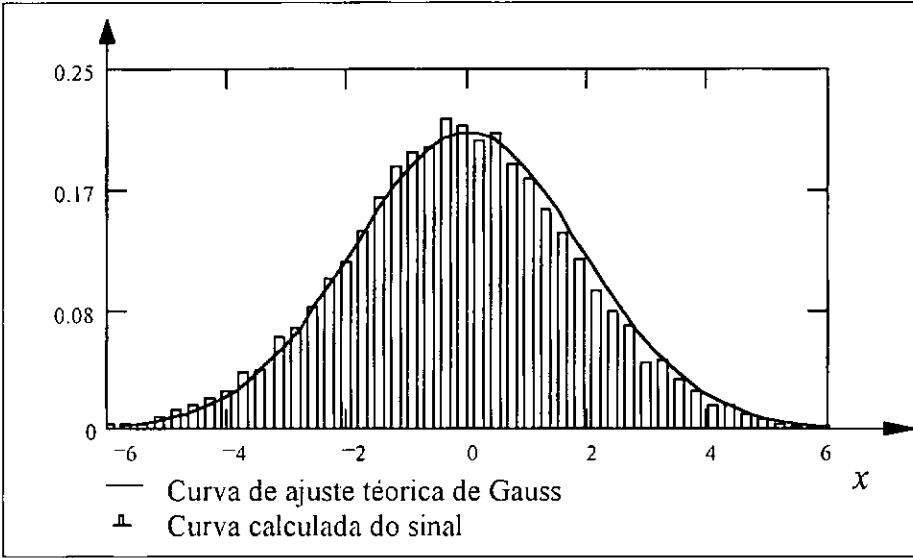


Figura 3.3 – Comparação entre a densidade de probabilidade calculada e a de Gauss

A função densidade de probabilidade dos picos positivos é determinada de forma análoga:

$$f_{x_{pi}^+}(x_{pi}^+) = \frac{n_{pi}^+}{Np^+ \Delta x} \quad (3.25)$$

onde  $Np^+ = \sum n_{pi}^+$ .

A densidade de probabilidade de picos (máximos) positivos e negativos é dada por:

$$f_{x_{tpi}}(x_{tpi}) = \frac{n_{tpi}}{N_{TP} \Delta x} \quad (3.26)$$

onde  $N_{TP} = \sum n_{tpi}$ .

A função cumulativa dos máximos (totais ou só positivos) pode ser obtida ordenando, dos menores para os maiores, todos os valores observados e empregando a seguinte [ 17 ]:

$$F_{x_m}(x_{m_i}) = \frac{i}{N+1} \quad (3.27)$$

onde  $X_m$  é aqui usado para identificar os máximos considerados (totais ou só positivo),  $x_{m_i}$  é o  $i$ -ésimo valor ordenado.

Caso o sinal for Gaussiano e de banda estreita ( $\varepsilon \rightarrow 0.0$ ), a distribuição dos picos positivos segue a distribuição de Rayleigh [5]:

$$f_{x_p}(x) = \frac{x}{\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \quad (3.28)$$

onde  $\sigma$  é o desvio padrão da série temporal obtido pela Eq. 3.10 ou Eq. 3.17

Caso o sinal for Gaussiano e de banda larga, a distribuição dos picos (máximos) positivos e negativos segue a distribuição de Rice [ 6 ]:

$$f_{x_p}(x) = \frac{1}{\sigma\sqrt{2\pi}} \left[ \varepsilon \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma\varepsilon}\right)^2\right) + \sqrt{1-\varepsilon^2} \frac{x}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \int_{-\alpha}^{\left(\frac{x}{\sigma}\right)\sqrt{1-\varepsilon^2}} \exp\left(-\frac{x^2}{2}\right) dx \right] \quad (3.29)$$

onde  $\varepsilon$  é o fator de largura de banda e  $\sigma$  é o desvio padrão da série temporal .

Observa-se que a distribuição de Rayleigh é uma caso particular da distribuição de Rice quando  $\varepsilon = 0.0$  .

A comparação entre as funções densidade de probabilidade de máximos calculada com as teóricas de Rayleigh e de Rice (caso em que  $\varepsilon \neq 0.0$  ) é mostrado na Figura 3.4.

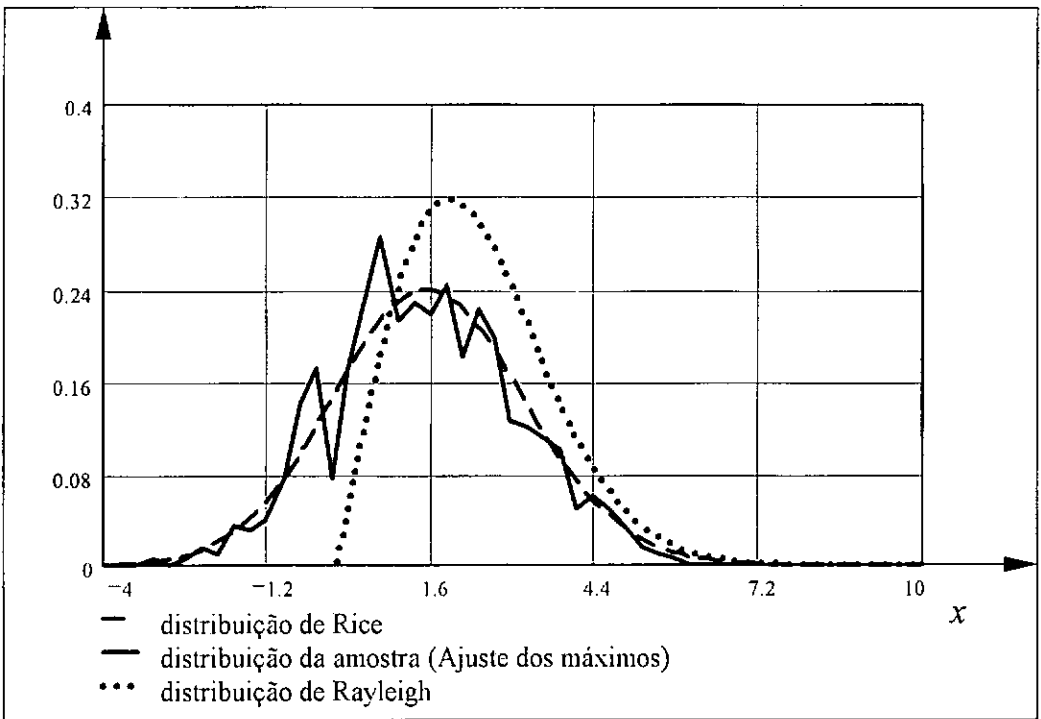


Figura 3.4 – Comparação entre as distribuições calculada, de Rayleigh e de Rice

### 3.3 - DISTRIBUIÇÃO DE PROBABILIDADES DE VALORES EXTREMOS

A distribuição exata de valores máximos extremos de uma série temporal pode ser calculada a partir da função densidade de probabilidades dos valores máximos totais e a correspondente cumulativa, obtidas no item anterior, através da metodologia apresentada no Capítulo 2. No caso de uma série gaussiana emprega-se a distribuição de probabilidades de Rice e do número de máximos  $N_{TP}$  (picos positivos e negativos) correspondente à duração do sinal.

Considerando-se a duração do sinal  $T$ , o período de pico  $T_p$  e a frequência de máximos  $\nu_m = \frac{1}{T_p}$ , a distribuição de probabilidades de extremos exata pode ser escrita através das seguintes expressões:

$$N_m = \nu_m T \quad (3.30)$$

$$F_{Y_{N_m}}(y) = [F_{X_{tp}}(y)]^{N_m} \quad (3.31)$$

$$f_{Y_{N_m}}(y) = N_m F_{X_{tp}}(y)^{N_m-1} f_{X_{tp}}(y) \quad (3.32)$$

onde

$F_{Y_{N_m}}(\cdot)$  é a função cumulativa de probabilidades de valores extremos;

$f_{Y_{N_m}}(\cdot)$  é função de densidade de probabilidades de valores extremos.

A distribuição de extremos de um caso genérico e a correspondente distribuição inicial (parente) de Rice são apresentados na Figura 3.5.

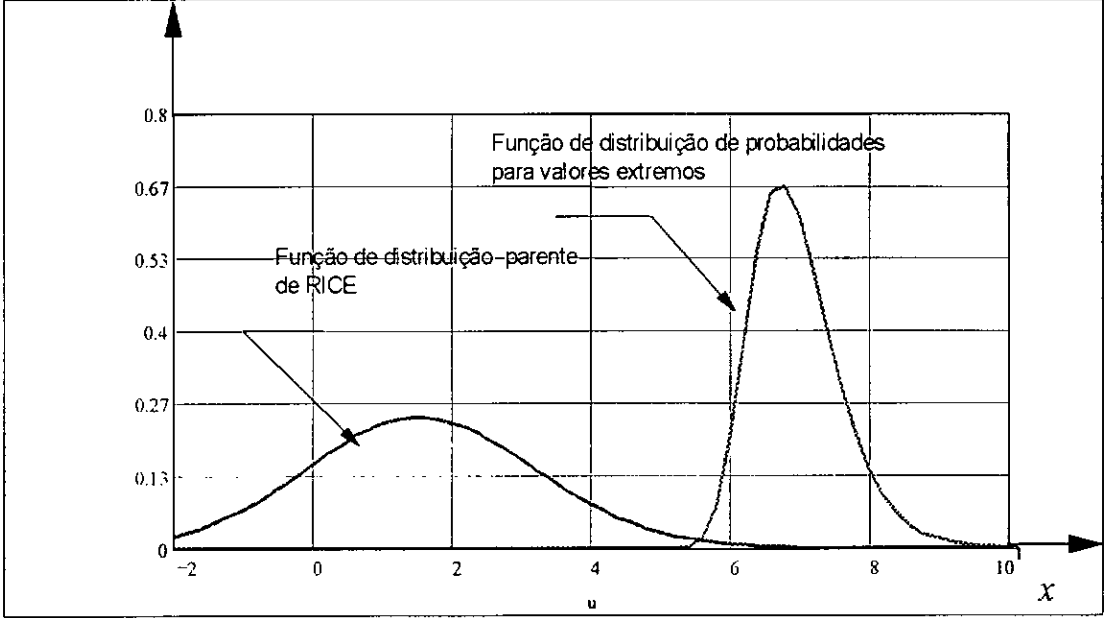


Figura 3.5 – Distribuição de Probabilidades Parente e de Extremos

### 3.4 – DISTRIBUIÇÃO DE PROBABILIDADES DE CARTWRIGHT E LONQUET -HIGGINS

Como pode ser visto na Ref. [6], assintoticamente a distribuição exata de valores extremos de uma série temporal gaussiana converge para a distribuição de Cartwright e Lonquet-Higgins dada pelas seguintes expressões:

$$F_{Y_{N_m}}(y) = \exp\left[-v_0 T \exp\left(-\frac{y^2}{2\sigma^2}\right)\right] \quad (3.33)$$

$$f_{Y_{N_m}}(y) = \frac{dF_{Y_{N_m}}(y)}{dy} \quad (3.34)$$

onde

$$v_0 \text{ é a frequência de cruzamento zero } v_0 = \frac{1}{Tz};$$

$T$  é a duração considerada do sinal;  
 $\sigma$  é o desvio padrão da série temporal;  
 $N_m$  é o número de máximos ( $= v_0 T$ );  
 $F_{Y_{N_m}}(y)$  é a função cumulativa de probabilidades de valores extremos e;  
 $f_{Y_{N_m}}(y)$  é função de densidade de probabilidades de valores extremos.

O gráfico da Figura 3.6 apresenta a concordância entre a função densidade de probabilidades calculada segundo Cartwright e Lonquet-Higgins e a função densidade de probabilidade teórica de valores extremos calculada a partir da distribuição parente de Rice.

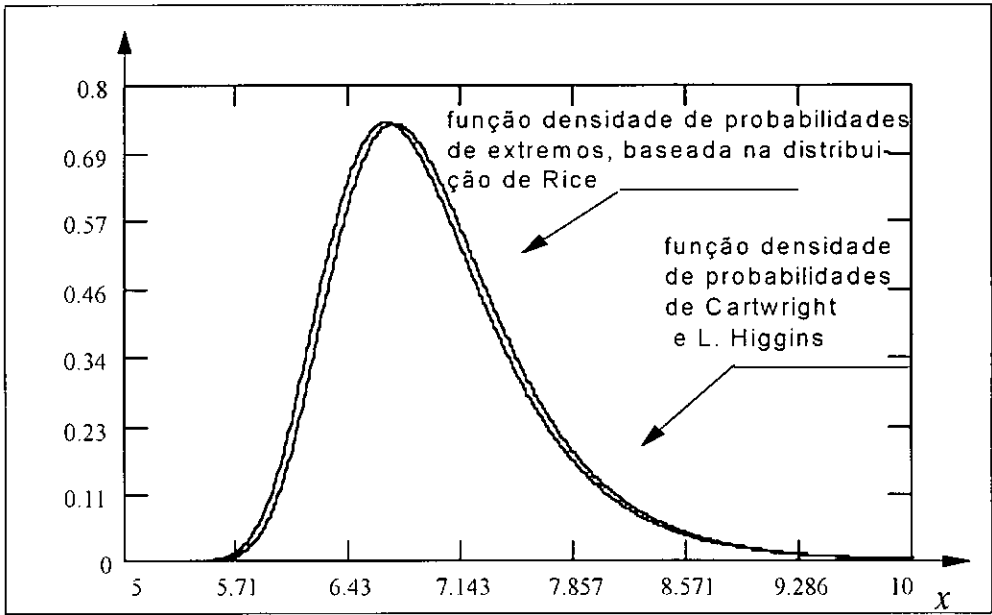


Figura 3.6 – Distribuição de extremos teórica e de Cartwright e Lonquet-Higgins

O valor esperado (média) dos extremos, de acordo com a função de distribuição de probabilidades para valores extremos dada por Cartwright e Lonquet-Higgins é dado por:

$$\mu_{Y_{N_m}} = \sqrt{2 \ln(v_0 T)} \sigma + \frac{0.57722 \sigma}{\sqrt{2 \ln(v_0 T)}} \quad (3.35)$$

e o correspondente desvio padrão é dado por:

$$\sigma_{Y_{N_m}} = \frac{\pi}{\sqrt{6}} \frac{\sigma}{\sqrt{2 \ln(v_0 T)}} \quad (3.36)$$

## CAPÍTULO 4

# ESTATÍSTICA DE VALORES MÁXIMOS EXTREMOS DE SÉRIES TEMPORAIS

### 4.1 – INTRODUÇÃO

A definição de distribuições de extremos para séries temporais gaussianas foi apresentada no capítulo anterior. Nota-se que neste caso, a distribuição de extremos é definida completamente a partir da densidade espectral da série temporal. Porém, para sinais não-gaussianos, na grande maioria das vezes, não é possível obter expressões que definem diretamente a distribuição de extremos. Neste caso é necessário utilizar técnicas baseadas na distribuição parente dos máximos (picos) do sinal para uma estimativa desta distribuição como será visto a seguir.

Existem na literatura vários métodos propostos de estimativa de valores extremos de amostras provenientes de séries temporais, a partir do ajuste de funções teóricas conhecidas de distribuição de probabilidade para representar as distribuições de valores máximos e de valores extremos [1,3].

É importante observar que o ajuste de distribuições de probabilidades (da amostra inicial ou parente), baseado em técnicas convencionais como, por exemplo, o método dos momentos [17], é direcionado para uma boa representação da distribuição nas vizinhanças do valor médio da amostra (parte central da distribuição), podendo, entretanto, introduzir erros na estimativa dos valores situados na cauda da mesma, o que prejudicaria de forma significativa a estimativa dos valores extremos.

Como consequência, foi observado que os métodos baseados no ajuste da cauda (*tail-fitting*) das distribuições de probabilidade, tendem a fornecer uma melhor estimativa para os valores extremos da resposta.

A seguir serão apresentados os métodos para análise de extremos investigados neste trabalho que se baseam no ajuste da cauda da distribuição parente. Nota-se que estes métodos não distinguem se a série temporal é gaussiana ou não.

## 4.2 - MÉTODO DE EXTRAPOLAÇÃO DIRETA BASEADO NA DISTRIBUIÇÃO DE GUMBEL

O primeiro passo no método de extrapolação direta [3] consiste na definição de uma distribuição de Gumbel  $F_{X_e}(X_e)$  diretamente a partir da amostra de valores extremos  $X_e$  obtidas da série temporal original, conforme descrito no Capítulo 2 deste trabalho.

Observa-se que  $X_e$  é obtido pela divisão da série temporal original em  $N_e$  intervalos de duração  $t_e$ , cada um contendo  $m$  máximos. Portanto a distribuição de Gumbel  $F_{X_e}(X_e)$ , é uma aproximação da distribuição de valores extremos para uma amostra de  $m$  valores máximos ou um intervalo de tempo  $t_e$ .

A aproximação da distribuição de valores extremos associada a um número de máximos  $q \neq m$  pode ser obtida, através da utilização do *postulado de estabilidade* da distribuição de Gumbel apresentado no Capítulo 2, pela expressão:

$$F_{X_q}(X_q) = [ F_{X_e}(X_e) ]^k \quad , \quad (4.1)$$

$$k = q / m \quad (4.2)$$

O postulado de estabilidade estabelece que o desvio padrão  $\sigma_q$  da distribuição,  $F_{X_q}(X_q)$ , é idêntico ao desvio padrão  $\sigma_e$  da distribuição original  $F_{X_e}(X_e)$ , e que a média  $\mu_q$  da distribuição  $F_{X_q}(X_q)$  está relacionada à média  $\mu_e$  da distribuição original através da relação:

$$\mu_q = \mu_e + \beta_k \sigma_e \quad (4.3)$$

onde o parâmetro  $\beta_k$  dado por [3]:

$$\beta_k = \frac{0.57722 - C_2}{C_1} + \frac{\ln(k)}{C_1} \quad (4.4)$$

onde os coeficientes  $C_1$  e  $C_2$  foram determinados no Capítulo 2, respectivamente, pelas equações (2.15) e (2.16).

É importante observar neste caso que a distribuição de valores extremos  $F_{X_q}(X_q)$ , determinada com a utilização do postulado de estabilidade, é obtida por simples translação da distribuição de valores extremos original  $F_{X_0}(X_0)$ , como está indicado na Figura 4.1.

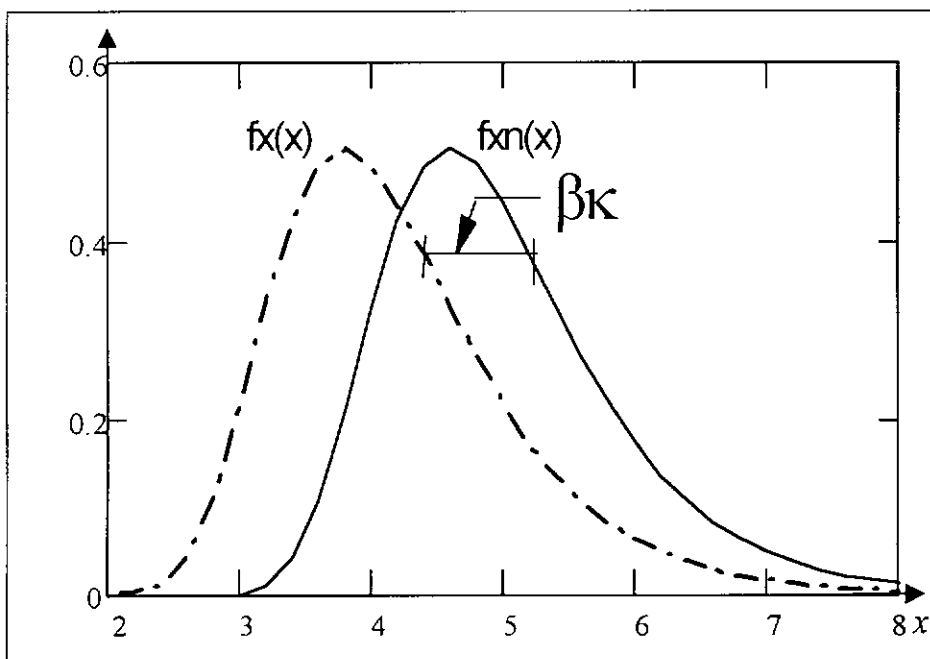


Figura 4.1 Método da extrapolação direta (postulado da estabilidade da Tipo I)

Uma vez obtida a distribuição de extremos para um determinado período (ou um dado número de máximos), através desta metodologia, estima-se a distribuição para um outro período qualquer.

Esta metodologia apresenta a característica de manter o mesmo desvio padrão para a distribuição de extremos para distintos intervalos de tempo (ou número de máximos). Pela teoria apresentada no Capítulo 2, observa-se que para as distribuições lá apresentadas, somente quando a distribuição parente (inicial) for uma exponencial esta característica é observada.

Nesta metodologia a distribuição parente dos máximos da série temporal não é considerada.

### 4.3 - MÉTODO DE AJUSTE DA CAUDA DA DISTRIBUIÇÃO PARENTE, BASEADO NA DISTRIBUIÇÃO DE WEIBULL

Conforme a Ref. [3], nesta metodologia a distribuição de Weibull é utilizada para representar a distribuição parente dos valores máximos  $X_m$ , com o seu tramo superior (cauda) ajustada de forma a reproduzir a correspondente distribuição de valores extremos de Gumbel, calculada diretamente a partir da amostra de valores extremos  $X_e$ , como será descrito a seguir.

Considera-se que a distribuição de Weibull seja uma opção conveniente para a distribuição de valores máximos, uma vez que ela engloba as distribuições Rayleigh (distribuição de máximos de um processo gaussiano  $X(t)$  de banda estreita) e exponencial (distribuição de máximos de um processo não-gaussiano  $X(t)|X(t)|$ , como será visto mais adiante, sendo  $X(t)$  gaussiano de banda estreita) como casos especiais, conforme ilustrado no Capítulo 2.

A função cumulativa da distribuição de Weibull é aqui reescrita da seguinte forma:

$$F_{X_w}(x) = 1 - \exp\left(-\left[\frac{x - u_w}{\alpha_w}\right]^{\lambda_w}\right) \quad (4.5)$$

onde  $u_w$ ,  $\alpha_w$  e  $\lambda_w$  são, respectivamente, os parâmetros de locação, escala e forma da distribuição.

Esta metodologia é dividida em duas etapas como será descrito a seguir.

A *primeira etapa* consiste na determinação dos parâmetros  $u_n$  e  $\alpha_n$  da distribuição de valores extremos Tipo I de Gumbel a partir da amostra de valores extremos  $X_e$ :

$$F_{X_e}(X_e) = \exp[-\exp[-\alpha_n(X_e - u_n)]] \quad (4.6)$$

Como já foi afirmado anteriormente, a amostra  $X_e$  é obtida pela divisão da série temporal original em  $N_e$  intervalos de duração  $t_e$ , contendo cada um, em média,  $m$  máximos. Neste caso, a distribuição de Gumbel se constitui numa aproximação da distribuição exata de valores extremos correspondente a uma amostra de  $m$  valores máximos.

Os parâmetros da distribuição de Gumbel podem ser determinados pelas expressões (2.12), baseadas no método de regressão linear, ou seja:

$$\alpha_n = \frac{C_1}{S_e} \quad u_n = M_e - \frac{C_2}{C_1} S_e \quad (4.7)$$

onde  $M_e$  e  $S_e$  são, respectivamente, a média e o desvio padrão da amostra de valores extremos  $X_e$ , calculados pelas expressões definidas em (2.17). Os coeficientes de regressão  $C_1$  e  $C_2$  são determinados pelas Equações (2.18) e (2.19).

Os parâmetros  $\lambda_w$  e  $\alpha_w$  da distribuição parente de Weibull poderão ser obtidos a partir das Eq. 2.32 e 2.33, em função do parâmetro  $u_w$ , pelas expressões:

$$\lambda_w = \frac{(u_n - u_w) \alpha_n}{\ln(m)} \quad , \quad \alpha_w = \frac{\lambda_w}{\alpha_n} [\ln(m)]^{\frac{\lambda_w - 1}{\lambda_w}} \quad (4.8)$$

Como o principal objetivo deste método é o ajuste da cauda da distribuição de Weibull, observa-se que o parâmetro de locação  $u_w$  não desempenha um papel importante na definição do comportamento da cauda da distribuição. Deste modo é suficiente realizar uma estimativa simples desse parâmetro a partir da média da amostra de valores máximos  $X_m$ , através da equação:

$$u_w + \alpha_w \Gamma\left(1 + \frac{1}{\lambda_w}\right) = \frac{1}{m} \sum_{k=1}^m X_{m_k} \quad (4.9)$$

onde  $\Gamma(\cdot)$  é a Função Gamma e os parâmetros  $\lambda_w$  e  $\alpha_w$  são definidos, em função do parâmetro  $u_w$ , nas equações apresentadas em (4.8). A solução simultânea das Eq. (4.8) e (4.9) permite a avaliação dos parâmetros da distribuição de Weibull.

Outra alternativa para definição dos parâmetros de Weibull é simplesmente fazer  $u_w = 0.0$  e calcular os demais parâmetros através das Eq. (4.8).

Caso os parâmetros  $u_n$  e  $\alpha_n$  da distribuição de valores extremos de Gumbel tenham sido calculados pelas expressões definidas em (4.7), os parâmetros  $\lambda_w$  e  $\alpha_w$  da distribuição parente de Weibull poderão ser obtidos diretamente em função da média  $M_e$  do desvio padrão  $S_e$  da amostra de extremos, substituindo-se os parâmetros  $u_n$  e  $\alpha_n$  em (4.7) nas expressões definidas em (4.8), resultando finalmente:

$$\lambda_w = \frac{1}{\ln(m)} \left[ \frac{C_1}{S_e} (M_e - u_w) - C_2 \right] \quad (4.10)$$

$$\alpha_w = \frac{\lambda_w S_e}{C_1} [\ln(m)]^{\frac{\lambda_w - 1}{\lambda_w}} \quad (4.11)$$

Deste modo, obtém-se a distribuição parente de valores máximos de Weibull, oriunda da correspondente distribuição de valores extremos de Tipo I de Gumbel, como está indicado na Figura 4.2

A *segunda etapa* consiste na determinação da distribuição de valores extremos de Gumbel, associada a um número de máximos  $q > m$ , diretamente a partir da distribuição parente de valores máximos de Weibull, calculada na primeira etapa.

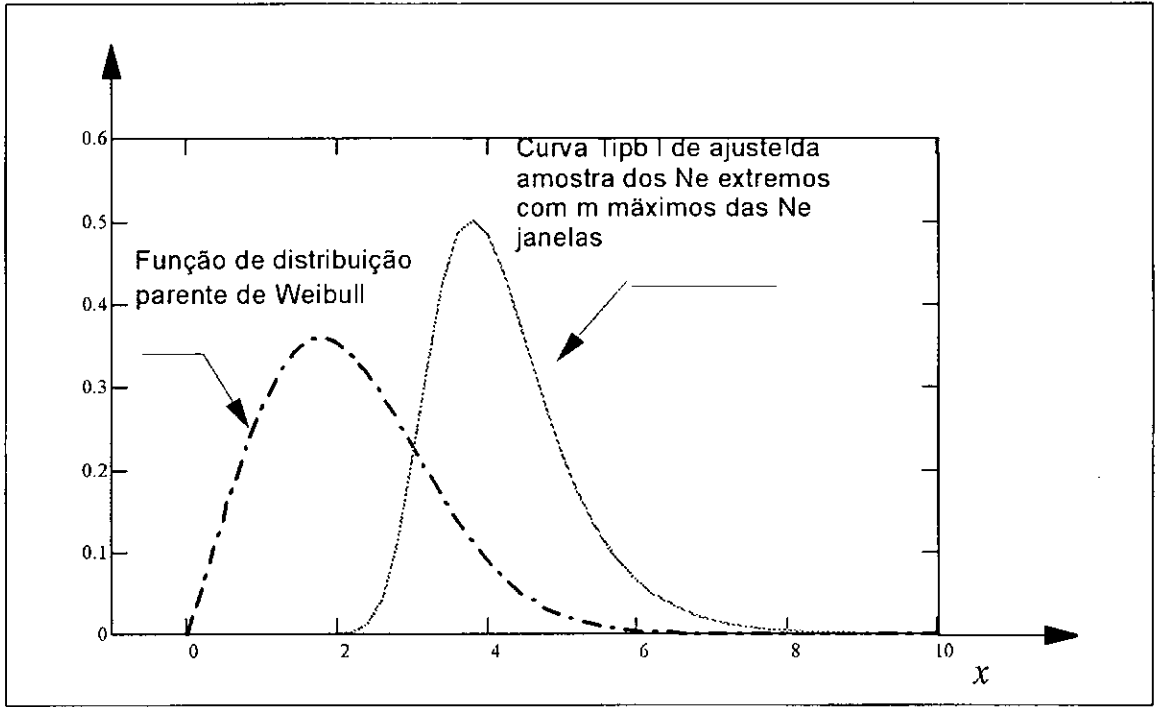


Figura 4.2 – Obtenção da distribuição de Weibull (parente)

Conforme o Capítulo 2, os parâmetros da *distribuição corrigida* de valores extremos de Gumbel, para um número de máximos  $q > m$ , serão dados pelas expressões:

$$\alpha_n^{(q)} = \frac{\lambda_w}{\alpha_w} [\ln(q)]^{\frac{\lambda_w-1}{\lambda_w}} ; \quad u_n^{(q)} = u_w + \alpha_w [\ln(q)]^{\frac{1}{\lambda_w}} \quad (4.12)$$

A média e o desvio padrão da distribuição corrigida de Gumbel serão dados pelas expressões:

$$\mu_{X_e}^{(q)} = U_w + \alpha_w \left[ \ln(q)^{\frac{1}{\lambda_w}} + \frac{0.57722}{\lambda_w} [\ln(q)]^{\frac{1-\lambda_w}{\lambda_w}} \right] \quad (4.13)$$

$$\sigma_{X_e}^{(q)} = \frac{\pi \alpha_w}{\sqrt{6} \lambda_w} [\ln(q)]^{\frac{1-\lambda_w}{\lambda_w}} \quad (4.14)$$

A Figura 4.3 mostra a *distribuição corrigida* de valores extremos de Gumbel, para um número de máximos  $q > m$ , obtida a partir da distribuição parente de valores máximos de Weibull calculada na primeira etapa.

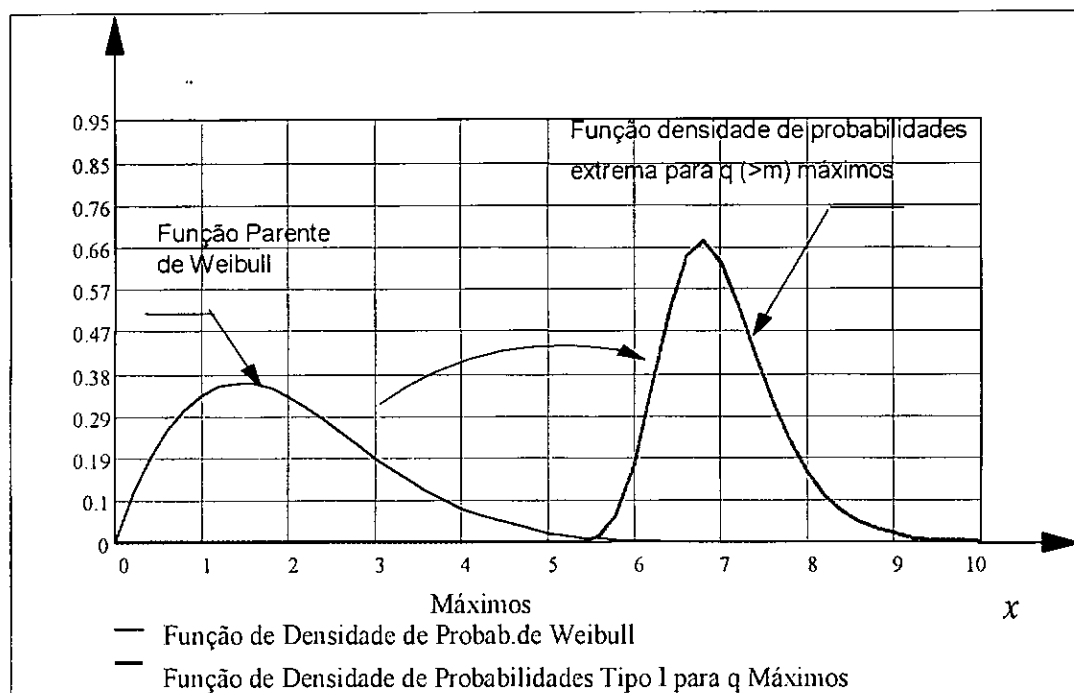


Figura 4.3 – Distribuição de extremos extrapolada para  $q > m$ .

Como será visto no Capítulo 6, no presente trabalho variou-se o número de janelas da série temporal (duração  $t_e$ ) e tomou-se como parâmetros finais da distribuição de Gumbel a média dos parâmetros obtidos para diferentes janelas.

#### 4.4 - MÉTODO DE AJUSTE DA CAUDA DA DISTRIBUIÇÃO PARENTE, PELO MÉTODO DE REGRESSÃO LINEAR

Da mesma forma que no item anterior, a distribuição de Weibull será utilizada para representar a distribuição parente de valores máximos  $X_m$ , que é uma amostra que engloba todos os máximos positivos da série temporal (por facilidade de notação  $X_m$  equivale-se com a amostra  $X_{p_i}^+$  apresentada no Capítulo 3). Neste caso, para uma melhor representação da cauda, os seus parâmetros são determinados por regressão linear, aplicado nos valores da amostra cuja função cumulativa de probabilidades apresente valores situados *acima de um valor limite (threshold) determinado*.

Como o principal objetivo deste método é o ajuste da cauda da distribuição de Weibull, observa-se que o seu parâmetro de locação  $u_w$  não desempenha um papel importante na definição do comportamento da cauda da distribuição. Deste modo será utilizada diretamente a distribuição de Weibull de dois parâmetros:

$$F_{X_w}(x) = 1 - \exp\left[-\left(\frac{x}{\alpha_w}\right)^{\lambda_w}\right] \quad (4.15)$$

Os parâmetros  $\alpha_w$  e  $\lambda_w$  são calculados a partir dos *coeficientes da reta de regressão* de Weibull, que é determinada aplicando-se o logaritmo natural em ambos os lados de (4.15):

$$\ln[-\ln[1 - F_{X_w}(x)]] = \lambda_w [\ln(x) - \ln(\alpha_w)] \quad (4.16)$$

A reta de regressão fica então definida pela equação:

$$y = \lambda_w z + b_w \quad (4.17)$$

onde:

$$y = \ln[-\ln[1 - F_{X_w}(x)]] \quad (4.18)$$

$$b_w = -\lambda_w \ln(\alpha_w) \quad e \quad z = \ln(x) \quad (4.19)$$

Os valores da amostra  $X_m$  são *ordenados de forma incremental*, de tal forma que o valor de  $F_w(X)$  em (4.18) possa ser calculado pela expressão:

$$F_{X_w}(x_n) = \frac{n}{m+1} \quad (4.20)$$

onde  $n$  é o número de elementos da amostra com valores menores ou iguais a  $x_n$ .

Os parâmetros  $\lambda_w$  e  $b_w$  são determinados pela *análise de regressão linear* (Eq. 4.16 a 4.19), a partir da amostra de máximos  $X_m$ , considerando-se apenas os elementos da amostra cujos valores da função cumulativa de probabilidade sejam *maiores que um dado valor de threshold de probabilidade P*, ou seja:

$$F_{X_w}(X_p) = P \quad (4.21)$$

O valor de  $n_p$ , correspondente a  $X_p$ , é determinado a partir de (4.20) e (4.21), resultando:

$$n_p = P(m+1) \quad (4.22)$$

Os parâmetros  $\lambda_w$  e  $\alpha_w$  da distribuição de Weibull em (4.15) dependem do valor limite de probabilidade  $P$ . Neste trabalho foram os seguintes valores para  $P$ : 0.60, 0.65, 0.70, 0.75, 0.80, 0.85 e 0.90. Para cada um destes limites são definidos um conjunto parâmetros  $\lambda_{wi}$  e  $\alpha_{wi}$  da distribuição de Weibull. Para diminuir a incerteza estatística, os parâmetros finais adotados  $\lambda_w$  e  $\alpha_w$  foram os valores médios dos parâmetros  $\lambda_{wi}$  e  $\alpha_{wi}$ .

Finalmente, a distribuição de valores extremos é uma Tipo I e obtida a partir da distribuição ajustada de Weibull considerada como distribuição parente. Os parâmetros da Tipo I são calculados pelas Equações 2.32 e 2.33.

#### 4.5.- MÉTODO BASEADO NA DISTRIBUIÇÃO GENERALIZADA DE PARETO ( PEAK OVER THRESHOLD - POT )

A distribuição de Pareto é uma distribuição de dois parâmetros que tem uma aplicação na análise estatística de valores extremos [10, 11].

Este método, que também consiste no ajuste da cauda da distribuição parente, está baseado na hipótese que a distribuição de valores da amostra dos valores máximos da série temporal excedentes a um dado valor de *threshold*, segue, aproximadamente, a distribuição de Pareto [12].

A função cumulativa de probabilidades da distribuição de Pareto, para valores que excedem um dado *threshold*, é dada pela expressão [13]:

$$F(x/x \geq thold) = \begin{cases} 1 - \left[ 1 - k \left( \frac{x - thold}{\alpha} \right)^k \right]^{\frac{1}{k}}, & \text{para } k \neq 0 \\ 1 - \exp \left[ - \frac{(x - thold)}{\alpha} \right], & \text{para } k = 0 \end{cases} \quad (4.23)$$

A correspondente função de densidade de probabilidades é dada por:

$$f(x/x \geq thold) = \begin{cases} \frac{1}{\alpha} \left[ 1 - k \left( \frac{x - thold}{\alpha} \right)^k \right]^{\frac{1}{k} - 1}, & \text{para } k \neq 0 \\ \frac{1}{\alpha} \exp \left[ - \frac{(x - thold)}{\alpha} \right], & \text{para } k = 0 \end{cases} \quad (4.24)$$

onde:

*thold* é o valor de *threshold*;

$\alpha$  é o parâmetro de escala da distribuição;

*k* é o parâmetro de forma da distribuição.

O intervalo de variação de *x* é:

$$\text{thold} \leq x \leq \infty, \quad \text{para } k \leq 0.0 \quad (4.25)$$

$$\text{thold} \leq x \leq \frac{\alpha}{k} + \text{thold} \quad \text{para } k > 0.0 \quad (4.26)$$

Verifica-se que esta distribuição é truncada num limite superior  $\left(\frac{\alpha}{k} + \text{thold}\right)$  quando  $k > 0.0$ .

Observa-se que para os casos especiais de  $k = 0$  e  $k = 1$  a distribuição de Pareto fica reduzida, respectivamente, na distribuição exponencial com média igual a  $\alpha$  e na distribuição uniforme definida no intervalo  $[\text{thold}, \alpha + \text{thold}]$ .

Os parâmetros  $\alpha$  e  $k$  podem ser estimados, por exemplo, pelo método dos momentos [10,11], resultando:

$$\alpha = \frac{1}{2} \bar{X} \left[ \left( \frac{\bar{X}}{S} \right)^2 + 1 \right] \quad (4.27)$$

$$k = \frac{1}{2} \left[ \left( \frac{\bar{X}}{S} \right)^2 - 1 \right] \quad (4.28)$$

onde  $\bar{X}$  e  $S$  são, respectivamente, a média e a variância dos valores excedentes ao treshold

É importante ressaltar que, do ponto de vista de aplicações práticas, o parâmetro  $k$  deve ficar dentro do intervalo  $(-0.5 < k < 0.5)$  [10].

Quando a distribuição de Pareto é utilizada como uma alternativa da distribuição exponencial, os valores de  $k$  próximos de zero ( $k=0$ ) apresentam maior interesse, uma vez que a distribuição de Pareto fica reduzida na distribuição exponencial quando  $k = 0$ , como está indicado na Figura 4.4.

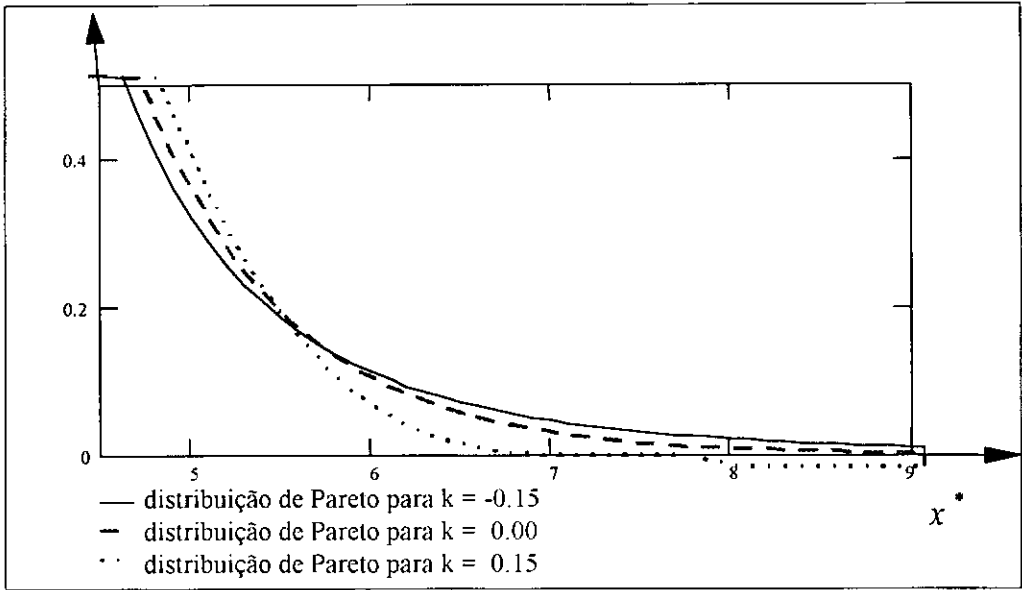


Figura 4.4 – Forma da distribuição de Pareto para vários valores de  $k$  ( $x^* = x\text{-thold}$ )

Ferreira e Guedes Soares [14] chegaram a conclusão de que a distribuição de Pareto, com parâmetro de forma  $k = 0$  (distribuição exponencial), é adequada para modelar os picos que excedem um *threshold* alto.

Gross et Al. [12] sugerem que os melhores resultados são obtidos escolhendo-se um valor de *threshold* correspondente a um número de excedências médio da ordem de 10 por ano.

Finalmente, observa-se que a distribuição de valores extremos será obtida a partir da distribuição de Pareto (distribuição parente), como está indicado na Figura 4.5, utilizando a teoria clássica de extremos conforme o Capítulo 2 deste trabalho.

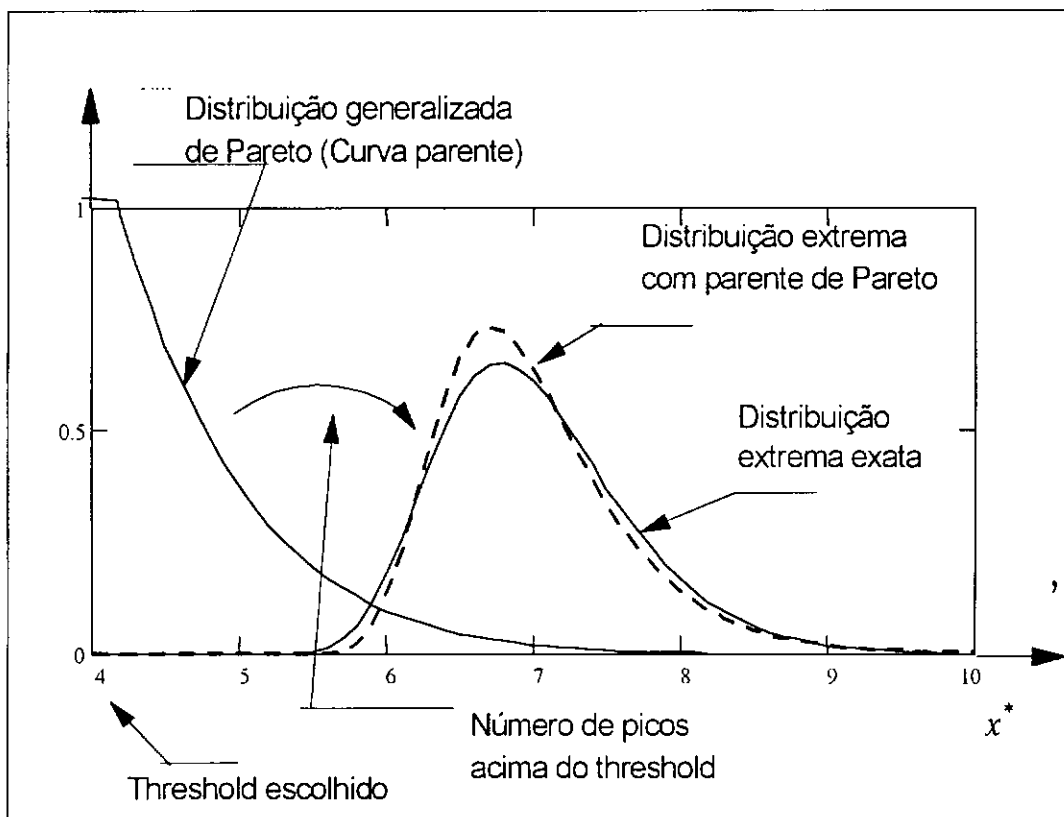


Figura 4.5 – Distribuição de extremos baseado na distribuição de Pareto ( $x^* = x\text{-thold}$ )

# CAPÍTULO 5

## GERAÇÃO DOS SINAIS

### 5.1 - INTRODUÇÃO

Neste capítulo serão apresentadas as técnicas de geração de dois sinais que serão utilizados neste trabalho. O primeiro será um sinal Gaussiano e o segundo não Gaussiano, ambos com durações médias de 6 horas.

Uma forma aproximada para geração de uma realização de um processo Gaussiano consiste na soma de um número finito,  $N$ , de sinais harmônicos, com ângulos de fases aleatórias, distribuídos uniformemente no intervalo  $(0, 2\pi)$  radianos, e com amplitudes determinadas diretamente a partir da densidade espectral.

Como exemplo de densidade espectral, foi utilizada a densidade espectral das elevações de um estado de mar.

### 5.2 – PARÂMETROS ESTATÍSTICOS DE UM ESTADO DE MAR

Os parâmetros estatísticos que caracterizam um estado de mar são usualmente dados pelo período de cruzamento zero  $T_z$  e pela altura significativa de onda  $H_s$ , calculada como sendo a média das alturas do 1/3 superior das maiores ondas [8]. Empregando os dados da Figura (5.1), estes parâmetros são definidos por:

$$T_z = \frac{T_r}{N} \quad (5.1)$$

onde:  $N$  = número de cruzamentos positivos.

$$H_s = \frac{\sum_{i=1}^N H_i}{\left[ \frac{N}{3} \right]}, \dots, M = N - 2 / 3N + 1 \quad (5.2)$$

onde:

$N$  = número de cruzamentos positivos;

$H_i$  = maior ordenada em cada período de cruzamento;

$T_i$  = período de cruzamento, como indicado na Figura 5.1.

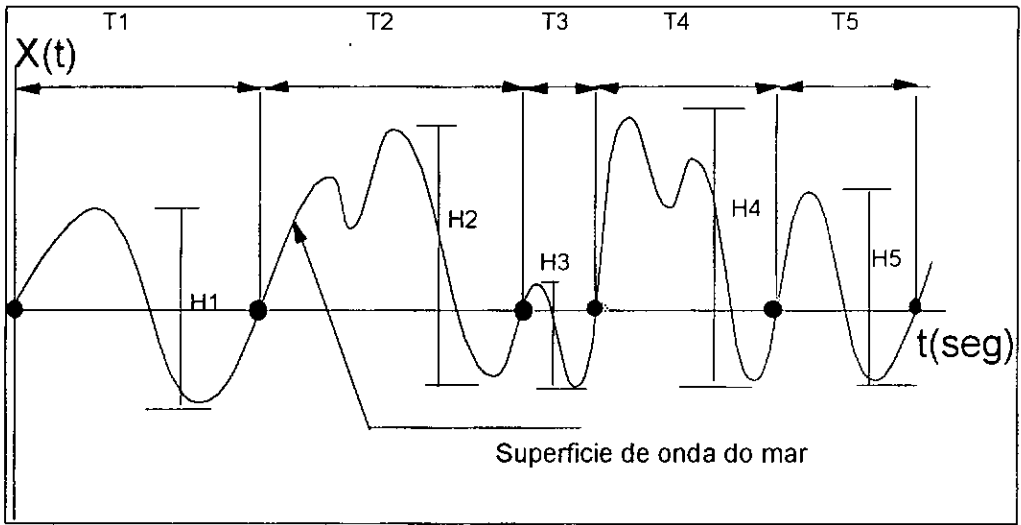


Figura 5.1 – Período de cruzamento zero e altura das ondas

Considerando-se o espectro de elevações do estado de mar como de banda estreita, as amplitudes das ondas (valores máximos das elevações) seguirão a distribuição de Rayleigh. Desta forma,  $H_s$  e  $T_z$  poderão ser determinados, a partir dos momentos do espectro, através das expressões:

$$H_s = 4\sqrt{m_0} \quad (5.3)$$

$$T_z = 2\pi \sqrt{\frac{m_0}{m_2}} \quad (5.4)$$

Onde (ver Tabela 3.1):

$m_0$  = momento de ordem zero do espectro associado;

$m_2$  = momento de ordem dois do espectro associado.

Dentre os diversos modelos matemáticos de espectros propostos para representar a densidade de energia das elevações de um estado de mar, um dos mais usados é dado por:

$$S(\omega) = \frac{A}{\omega^p} \exp\left[-\frac{B}{\omega^q}\right] \quad (5.5)$$

onde:

$S(\omega)$  = ordenada da função de densidade espectral;

$A, B, p, q$  = parâmetros do espectro.

O espectro adotado neste trabalho foi o espectro ISSC - Pierson Moskowitz de dois parâmetros [8]:

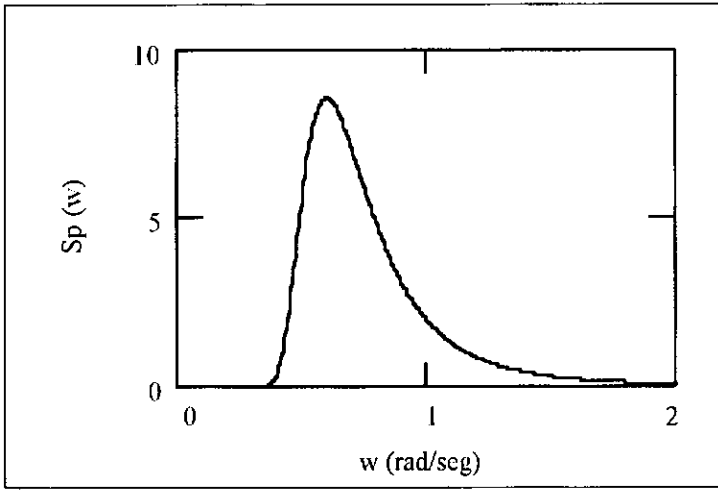
$$S(\omega) = \frac{A}{\omega^5} \exp\left[-\frac{B}{\omega^4}\right] \quad (5.6)$$

onde:

$$A = \frac{[4\pi^3 H_s^2]}{T_z^4} \quad (5.7)$$

$$B = \frac{[4A]}{H_s^2} \quad (5.8)$$

Este espectro, representado na Figura 5.2, é o mais utilizado para representar os estados de mar da Bacia de Campos [9].



*Figura 5.2 – Espectro de Pierson–Moskowitz de elevações de um estado de mar  
( $H_s=7.60m$ ,  $T_z=8.70seg$ )*

Cada uma das realizações das elevações da superfície do mar no domínio do tempo  $X(t)$  é obtida pela soma de um número finito,  $N$ , de sinais harmônicos, cujas amplitudes são determinadas pela densidade espectral do estado de mar em questão e os ângulos de fase são gerados aleatoriamente a partir de uma distribuição uniforme de probabilidades no intervalo  $(0, 2\pi)$  radianos [8]. Neste trabalho a faixa de frequências do espectro é dividida em  $N\omega$  intervalos iguais  $\Delta\omega$ , porém, a frequência associada ao intervalo é definida aleatoriamente dentro de cada intervalo.

Desta forma, o sinal  $X(t)$  será calculado de acordo com as seguintes expressões:

$$X(t) = \sum_{j=1}^{N\omega} An_j \cos[-\omega_j \cdot t + Fase_j] \quad (5.9)$$

sendo:

$$An_j = \sqrt{2S(w_j)\Delta\omega} \quad (5.10)$$

onde:

$An_j$  = amplitude do  $j$ -ésimo harmônico;

$\omega_j$  = frequência do  $j$ -ésimo harmônico;

$Fase_j$  = fase aleatória, gerada a partir de uma distribuição uniforme no intervalo  $(0, 2\pi)$ ;

$\Delta\omega$  = tamanho do intervalo de frequência;

$N\omega$  = número finito de harmônicos.

Neste trabalho limitou-se a representação da função densidade espectral no intervalo significativo de frequências, i.e.,  $[(\omega_i)=0.05]$ rad/s a  $[(\omega_f)=4.00]$  rad/s. Desta forma, para um valor definido de  $\Delta\omega$  o número de harmônicos é dado por:

$$N\omega = \frac{\omega_f - \omega_i}{\Delta\omega} \quad (5.11)$$

O número total de pontos de uma série temporal gerada é dada por

$$N = \frac{T}{\Delta t} + 1 \quad (5.12)$$

onde T é o tempo total de simulação e  $\Delta t$  é respectivo intervalo de amostragem.

### 5.3 GERAÇÃO DE UM SINAL GAUSSIANO DE ELEVAÇÕES DO MAR

Conforme observações, a elevação do mar constitui-se de um processo aleatório gaussiano [8] e, portanto, a distribuição exata de valores extremos pode ser obtida diretamente da densidade espectral da mesma.

A seguir será investigado numericamente qual é o número de frequências  $N\omega$  e o intervalo de amostragem  $\Delta t$  necessários para se gerar, através do somatório de harmônicos, uma série temporal de elevações do mar que represente corretamente a distribuição de extremos associada ao correspondente espectro de mar. Nesta investigação empregou-se um estado de mar com  $H_s = 7.60$  m e  $T_z = 8.70$  s.

Conforme o Capítulo III, a distribuição dos máximos de um sinal gaussiano é dada pela distribuição de Rice. A distribuição de extremos exata pode ser obtida a partir desta. Alternativamente, a distribuições de extremos pode ser obtida pela fórmula de Cartwright e Lonquet-Higgins. Na Tabela (5.1) são apresentados os parâmetros, média e

desvio padrão, da distribuição de extremos da amplitude máxima do estado de mar correspondente ao período de uma hora.

<b>Parâmetro (m)</b>	<b>Exata (parente Rice)</b>	<b>Cartwright e L. Higgins</b>
<b><math>\mu</math> extrema</b>	7.005	7.042
<b><math>\sigma</math> extremo</b>	0.646	0.645

*Tabela 5.1 - Valores exatos da média e do desvio padrão do valor extremo para 1 hora*

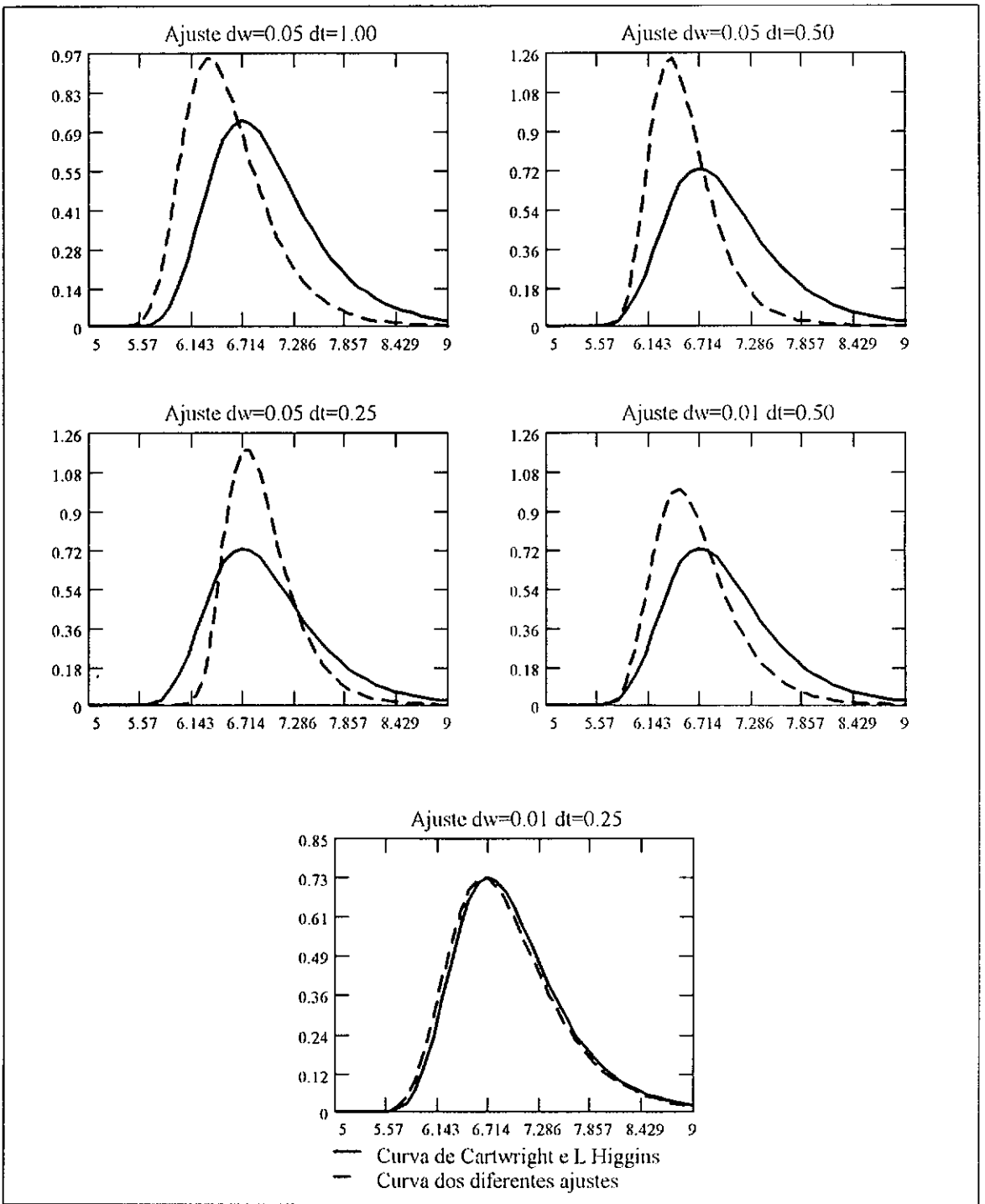
A seguir descreve-se a estratégia adotada para investigar numericamente o número de harmônicos e o intervalo de amostragem mínimos para que a série temporal gerada de acordo com a Eq. (5.9) represente adequadamente os extremos do estado de mar aqui analisado.

Para cada conjunto  $\Delta\omega$  e  $\Delta t$  escolhido, geraram-se 20 séries temporais com uma hora de duração diferentes, variando-se o conjunto de fases aleatórias. Tomaram-se os máximos valores obtidos de cada uma destas séries e calcularam-se os parâmetros estatísticos, bem como os parâmetros da distribuição Tipo I associados, desta amostra de 20 valores extremos. Os parâmetros estatísticos (média e desvio padrão) obtidos estão na Tabela (5.2) e a comparação da distribuição Tipo I de cada conjunto  $\Delta\omega$  e  $\Delta t$  com a distribuição de extremos exata é apresentada na Figura (5.3)

De acordo com os resultados apresentados verifica-se que o melhor resultado corresponde à simulação número 5, realizada com  $\Delta\omega = 0.01$  rad/seg e  $\Delta t = 0.25$  seg, conseguindo reproduzir a função de distribuição de valores extremos exata (1 hora)

Simulação	1	2	3	4	5
Realização	$\Delta\omega=0.05$	$\Delta\omega=0.05$	$\Delta\omega=0.05$	$\Delta\omega=0.01$	$\Delta\omega=0.01$
	$\Delta t=1.00$	$\Delta t=0.50$	$\Delta t=0.25$	$\Delta t=0.50$	$\Delta t=0.25$
1	6.530	6.618	6.756	6.457	6.347
2	6.964	6.186	6.952	6.098	6.077
3	6.630	7.223	6.492	6.313	6.665
4	7.355	6.982	6.437	6.780	8.181
5	7.421	5.978	7.636	7.010	7.926
6	6.962	6.714	6.798	6.276	7.233
7	6.064	6.943	6.042	8.017	6.823
8	5.991	6.426	6.457	6.177	6.867
9	6.337	6.382	6.617	7.060	7.394
10	7.450	6.977	6.972	7.131	6.352
11	6.148	6.913	6.218	6.466	7.107
12	6.819	6.167	7.300	6.316	6.343
13	5.868	6.480	7.239	6.163	7.723
14	6.875	6.639	6.833	6.359	6.218
15	6.233	7.201	6.225	6.896	6.603
16	6.677	6.440	6.785	6.514	6.218
17	5.928	6.196	6.037	7.304	6.514
18	6.200	5.960	6.313	6.512	6.654
19	6.033	6.137	7.430	6.656	8.266
20	6.747	6.444	7.582	7.115	6.425
média	6.561	6.550	6.925	6.681	6.947
desvio padrão	0.495	0.380	0.392	0.470	0.648

Tabela 5.2 Valores máximos extremos para 5 simulações de 3600s



*Figura 5.3 - Distribuições de valores extremos de Gumbel para cada simulação comparadas com a distribuição exata.*

Pode-se melhorar os resultados anteriores, aumentando o número de realizações para 50. Na Tabela 5.3 estão reunidos os resultados dos valores máximos extremos das 50 realizações, correspondentes à simulação de número 5.

Valores Máximos Extremos									
6.347	6.077	6.665	8.181	7.926	7.233	6.823	6.867	7.394	6.352
7.107	6.343	7.723	6.218	6.603	6.218	6.514	6.654	8.266	6.425
6.435	6.343	6.586	9.415	7.632	6.942	6.218	7.157	6.970	6.122
6.774	6.794	6.944	6.618	7.020	6.812	6.955	6.688	6.443	7.055
7.039	6.487	7.338	7.530	7.641	6.218	7.867	8.780	7.218	6.898

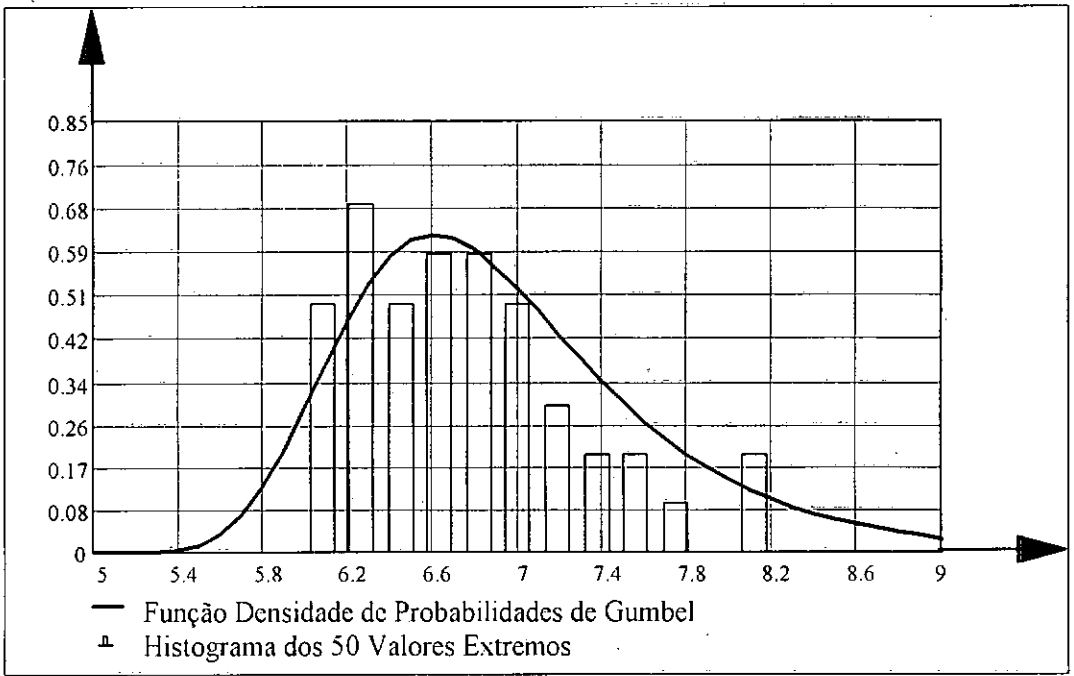
*Tabela 5.3 - Valores máximos extremos das 50 realizações da simulação nº 5*

Os valores exatos da média e do desvio padrão do sinal Gaussiano, calculados, respectivamente, a partir da distribuição parente de Rice e da distribuição assintótica de Cartwright e L. Highings são apresentados na Tabela 5.4.

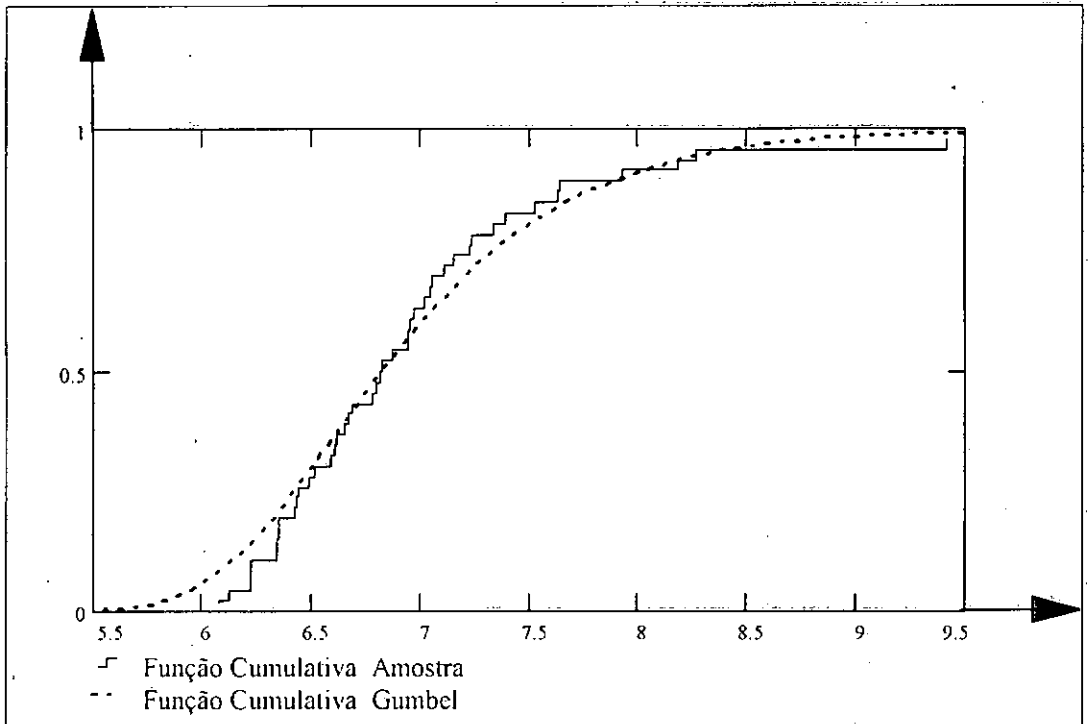
	Simulação Nº 5	Rice	Cartwright e L. Higgins
$\mu$ extrema	6.920	7.005	7.042
$\sigma$ extremo	0.642	0.646	0.645

*Tabela 5.4 - Valores exatos e calculados da média e do desvio padrão do valor extremo de um sinal Gaussiano com 1 hora de duração.*

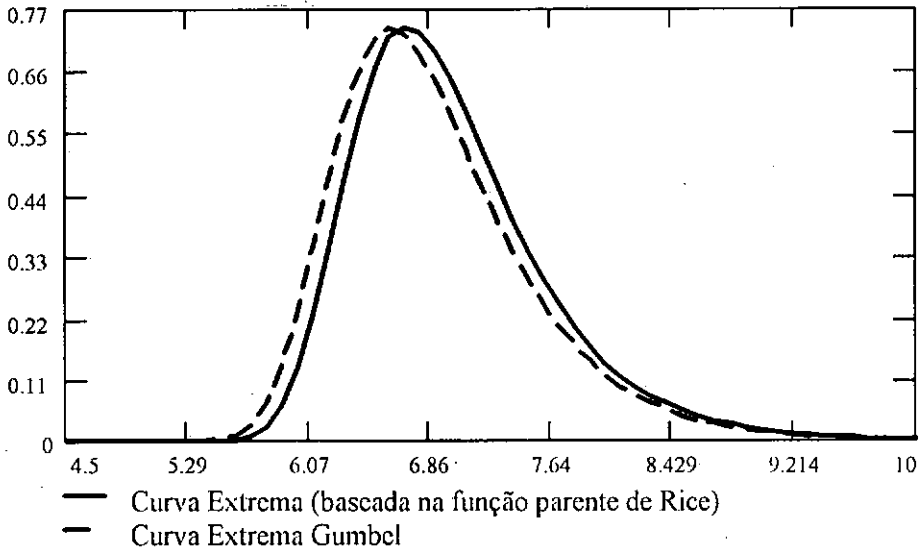
As Figuras 5.4 a 5.6 apresentam o ajuste dos 50 valores da amostra de extremos da simulação 5 à distribuição de Gumbel (Tipo I), bem como a comparação desta com a distribuição teórica de um sinal gaussiano.



*Figura 5.4 – Histograma e a distribuição de Gumbel da simulação n° 5*



*Figura 5.5 – Função cumulativa: amostra X Gumbel*



*Figura 5.6 - Comparação entre a distribuição de Gumbel (simulação de nº 5) e a distribuição exata de valores extremos (a partir parente de Rice)*

De acordo com o que foi apresentado anteriormente, é necessário um grande número de harmônicos (>400) e um intervalo de amostragem pequeno para que a simulação da superfície através da soma de séries harmônicas represente adequadamente o comportamento dos extremos de um estado de mar.

## 5.5 - GERAÇÃO DE UM SINAL NÃO GAUSSIANO

Neste item será gerado artificialmente uma série não-gaussiana a partir das elevações da superfície do mar (processo gaussiano) para investigação do comportamento dos extremos da mesma.

A geração de um processo não gaussiano pode ser diretamente a partir de uma transformação não-linear num processo aleatório gaussiano. Neste trabalho gera-se um processo não gaussiano a partir da série temporal  $X(t)$  das elevações da superfície do mar a partir da seguinte expressão:

$$Y(t) = X(t)X(t) \quad (5.13)$$

Observar que  $Y(t)$  é semelhante a parcela de arraste da equação de Morison, comumente empregada para avaliação das forças hidrodinâmicas em estruturas esbeltas [8].

Se  $X(t)$  for de banda estreita a distribuição dos máximos é dada pela distribuição de Rayleigh, conforme o Capítulo 2. Neste caso, a distribuição exata dos máximos de  $Y(t)$  pode ser obtida através da transformação de variáveis aleatórias [1]:

$$f_y(y) = f_x(y) \left| \frac{dx}{dy} \right| \quad (5.14)$$

que para o presente caso resulta em

$$f_y(y) = \frac{1}{2\sigma_x^2} \exp\left(-\frac{1}{2\sigma_x^2} y\right) \quad (5.15)$$

onde  $\sigma_x$  é o desvio padrão do processo aleatório gaussiano  $X(t)$ .

A distribuição dos máximos do processo  $Y(t) = X(t)|X(t)|$  nada mais é do que uma distribuição exponencial com parâmetro  $\alpha = 2\sigma_x^2$ .

Para um processo gaussiano  $X(t)$  de banda larga a distribuição dos máximos do processo  $Y(t) = X(t)|X(t)|$  pode ser igualmente obtida a partir da distribuição de Rice. Como a distribuição dos máximos, a distribuição de extremos pode também ser obtida.

A seguir investiga-se o comportamento dos extremos do sinal não gaussiano, associados a um período de uma hora, a partir da simulação deste sinal através da geração de 50 realizações independentes (conjunto de fases diferentes) de  $X(t)$  com  $\Delta\omega = 0.01$  rad/s e  $\Delta t = 0.25$  s, usando o mesmo estado de mar do item anterior.

Os valores máximos de cada uma das 50 realizações estão indicados na Tabela 5.5. A Figura 5.7 mostra a série temporal de destas realizações.

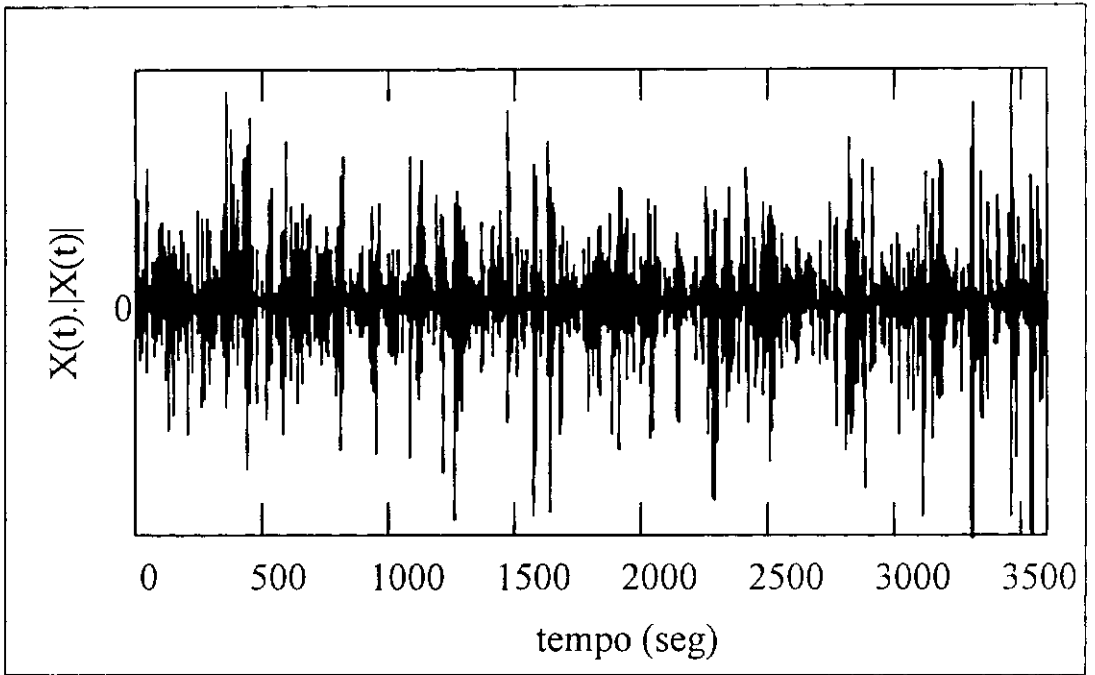
Valores Máximos Extremos									
37.798	44.223	57.927	36.556	48.753	37.431	37.447	51.257	47.555	48.108
3.266	53.071	52.718	38.165	52.155	56.993	49.357	44.525	61.201	40.051
42.663	38.970	48.229	40.444	45.517	51.926	49.261	47.995	40.126	85.859
41.281	41.409	40.234	43.375	88.642	58.247	66.929	62.821	51.223	48.581
37.479	45.887	46.158	48.219	43.798	49.280	46.403	48.372	44.729	41.512

Tabela 5.5 - Valores máximos extremos das 50 realizações de  $Y(t)$

Na Tabela 5.6 são apresentados os valores da média e do desvio padrão do valor extremo de  $Y(t) = X(t)X(t)$  obtidos pelas simulações e pela distribuição exata.

Parâmetro (m.m)	Valor Exato	Valor Simulado
$\mu$ extrema	49.00	49.15
$\sigma$ extremo	9.22	10.60

Tabela 5.6 - Valores exatos da média e do desvio padrão do valor extremo para 1 hora do processo não-gaussiano  $Y(t)$ .



*Figura 5.7 - Sinal não gaussiano  $Y(t)$*

A Figura 5.8 apresenta um histograma de uma série temporal de  $Y(t)$  comparada a distribuição de Gauss. A Figura 5.9 compara os máximos do sinal com as curvas das distribuições de Rice e Rayleigh. Através desta duas figuras, observa-se que o sinal é não gaussiano. Corroborando com esta observação estão os valores calculados dos coeficientes de assimetria e de kurtosis de  $Y(t)$ , que foram, respectivamente, 1.14 e 5.4, são bastante diferentes de um processo gaussiano, que são respectivamente 0.0 e 3.0.

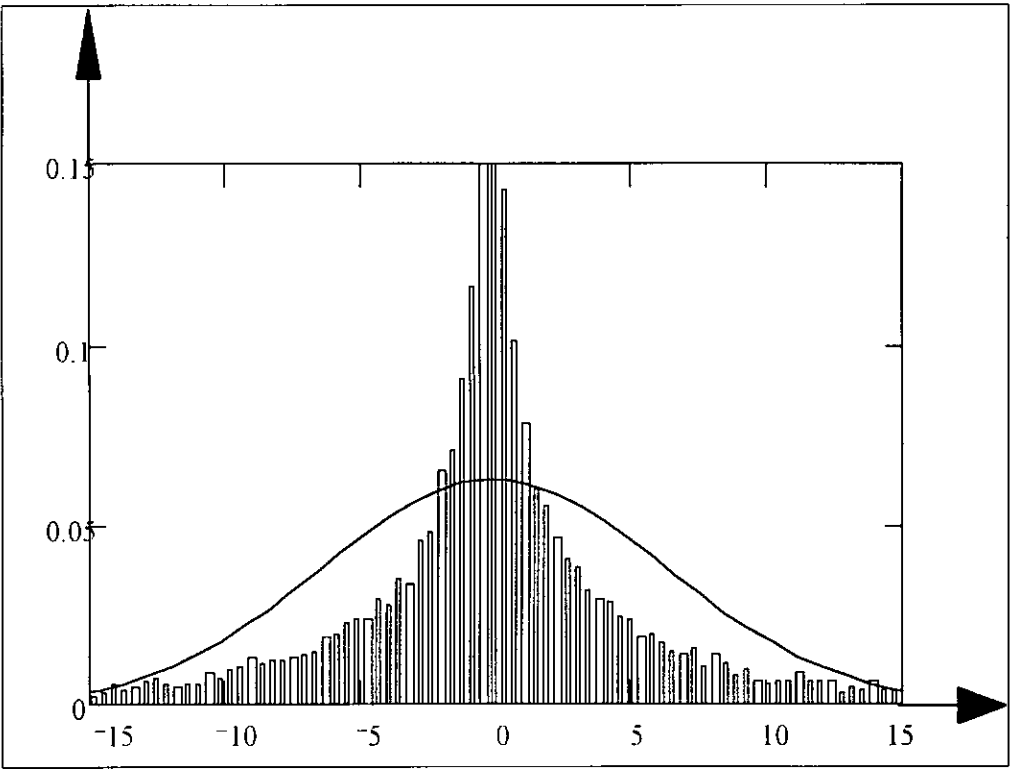


Figura 5.8 – Histograma da série temporal e a distribuição de gauss

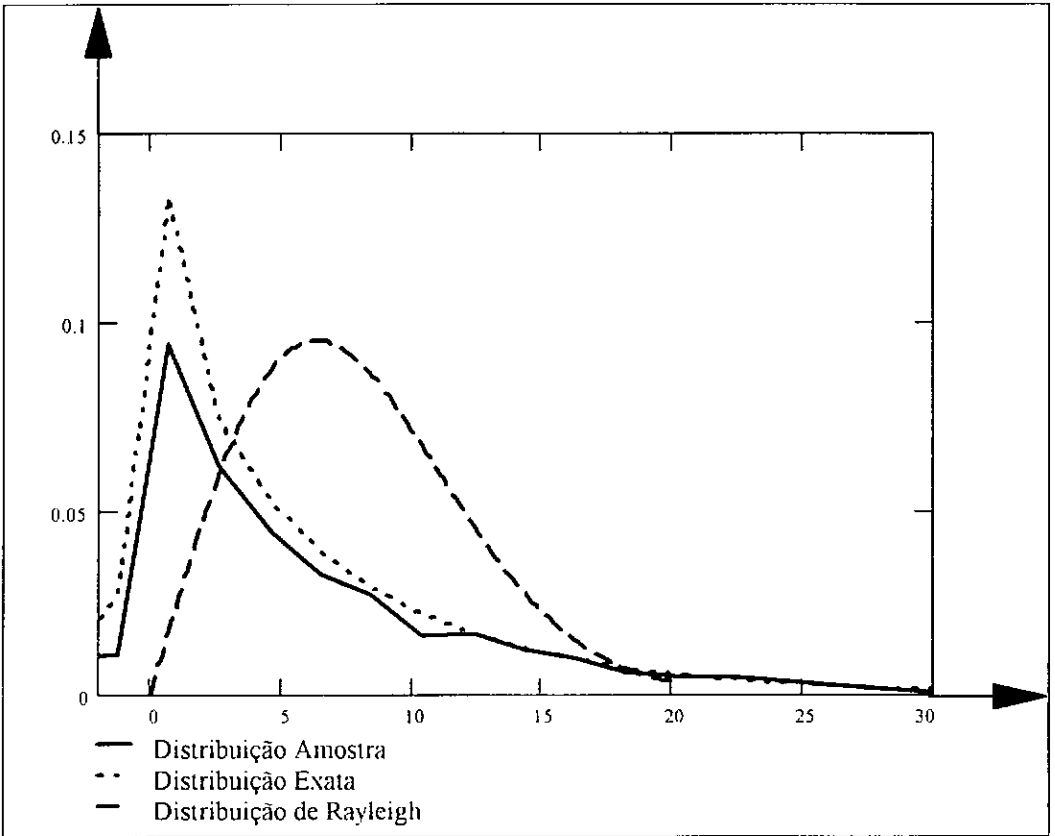
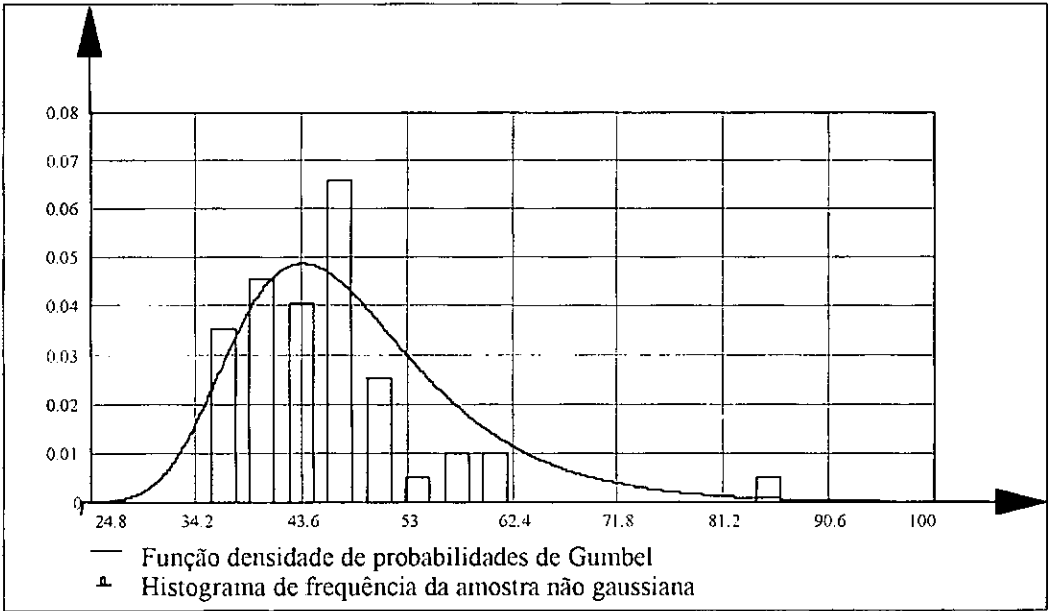
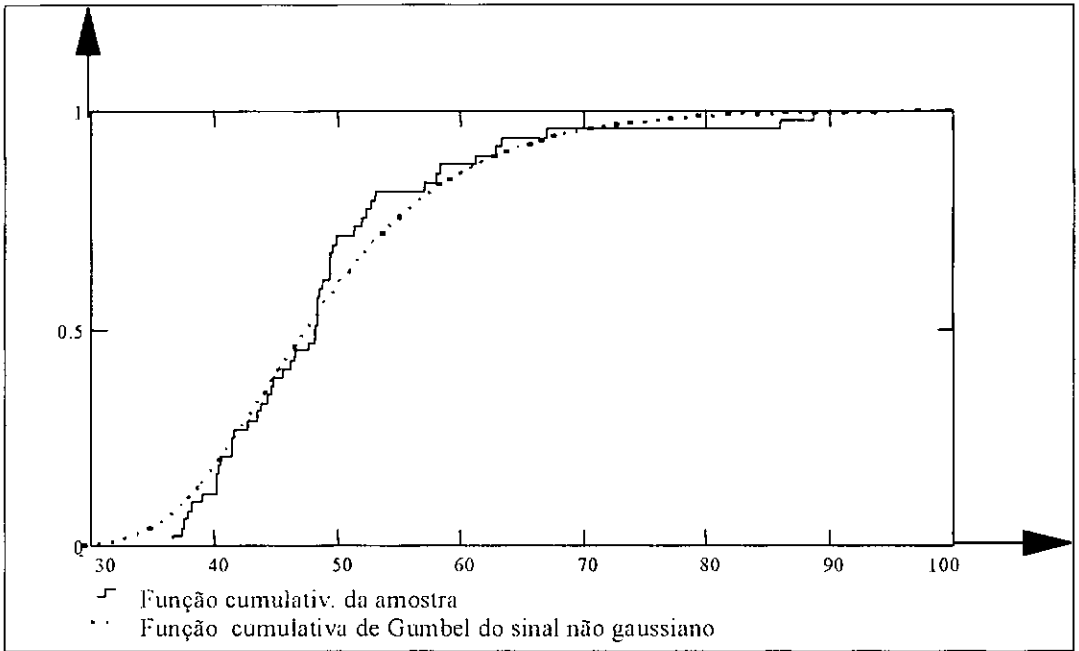


Figura 5.9 – Máximos do sinal comparados com a distribuição Rayleigh, e com a distribuição exata



*Figura 5.10 – Histograma de 50 extremos X Gumbel*



*Figura 5.11 – Função cumulativa dos extremos: amostra X Gumbel*

Pelos resultados apresentados, observa-se que a discretização do espectro, com  $\Delta\omega = 0.01$  rad/s e  $\Delta t = 0.25$  s, definida no item anterior, também consegue representar adequadamente os extremos do processo aleatório não gaussiano analisado.

# CAPITULO 6

## APLICAÇÃO DOS MÉTODOS

### 6.1 - INTRODUÇÃO

Os diferentes procedimentos para a determinação da estatística de valores extremos de séries temporais, apresentados no Capítulo 4, serão aplicados, sucessivamente, na análise de dois sinais. O primeiro será constituído por um sinal Gaussiano, representativo da resposta de um sistema linear, e o segundo por um sinal não Gaussiano, análogo a resposta de um sistema não-linear. Estes sinais são baseados no sinal de elevações da superfície do mar apresentado no capítulo anterior. O interessante é que para ambos os sinais é possível comparar a precisão dos diferentes métodos de avaliação de extremos com os resultados teóricos exatos.

A duração dos dois sinais considerados é de aproximadamente 6 horas. Porém, a estatística do extremos é sempre feita para o valor máximo extremo correspondente a uma hora de duração. Todas as análises realizadas neste trabalho foram executadas através do programa MATHCAD 7.0 [18]

### 6.2 – SINAL GAUSSIANO

Apresenta-se a seguir as aplicações dos diferentes procedimentos para a determinação da estatística de valores extremos de um sinal Gaussiano. O sinal escolhido para esta análise é a série temporal da elevação do mar apresentada no Capítulo 5.

#### 6.2.1 - MÉTODO DE EXTRAPOLAÇÃO DIRETA BASEADO NA DISTRIBUIÇÃO DE GUMBEL

O primeiro passo do método de extrapolação direta consiste na determinação dos parâmetros da distribuição de Gumbel, a partir da amostra de extremos  $X_e$ . Esta amostra é obtida pela divisão do sinal em  $N_e$  intervalos de duração  $t_e$  (tamanho da janela),

contendo em média  $m$  máximos cada um. Consequentemente, é necessário definir o valor a ser adotado para a duração  $t_e$  dos  $N_e$  intervalos em que será dividido o sinal.

Para esta finalidade, foram considerados diferentes valores de  $t_e$  na subdivisão da série temporal cujo tempo total de simulação  $T$  foi de 6 horas (21800s). Estes valores foram 500s, 1000s e 1500s. A distribuição de valores extremos de Gumbel, calculada com os máximos das  $N_e=T/t_e$  janelas, i.e., a distribuição do valor extremo para um período  $t_e$ , foi diretamente extrapolada para obter a distribuição do valor máximo correspondente ao período de uma hora.

Observa-se, porém que a duração ótima do intervalo  $t_e$  dependerá das características próprias de cada sinal gerado. Deste modo, optou-se por calcular os parâmetros da distribuição de Gumbel pela *média* dos valores obtidos para os valores de  $t_e$  iguais a 500s, 1000s e 1500s. Numa janela menor que 500s não é obtida a estacionaridade do sinal ( observação visual da estabilidade da média e o desvio padrão), e para janelas maiores que 1500s o número de máximos  $N_e$  fica muito baixo (<14).

Como para cada valor de  $t_e$  gera-se um conjunto de parâmetros para a distribuição de Gumbel, os parâmetros da distribuição final utilizada na comparação com a distribuição teórica foram calculados como os valores médios obtidos para diferentes valores de  $t_e$ .

Observa-se que na extrapolação direta usando o postulado da estabilidade, a distância que separa as duas curvas é determinada pela expressão:

$$d = \frac{1}{\alpha} \ln \left[ \frac{T_e}{t_e} \right] \quad (4.1)$$

onde  $T_e$  = duração do extremo considerado e  $\alpha$  = parâmetro da distribuição de Gumbel.

Esta metodologia foi aplicada a sete diferentes sinais (realizações) da superfície de elevações obtidos através de diferentes conjuntos de fases aleatórias. Cada uma destas realizações foi obtida a partir de sementes (seeds) constituídas por números primos.

A Figura (6.1) ilustra, para uma destas realizações, o comportamento da distribuição do valor extremo para um período de uma hora para os diferentes intervalos de tempo  $t_e$ .

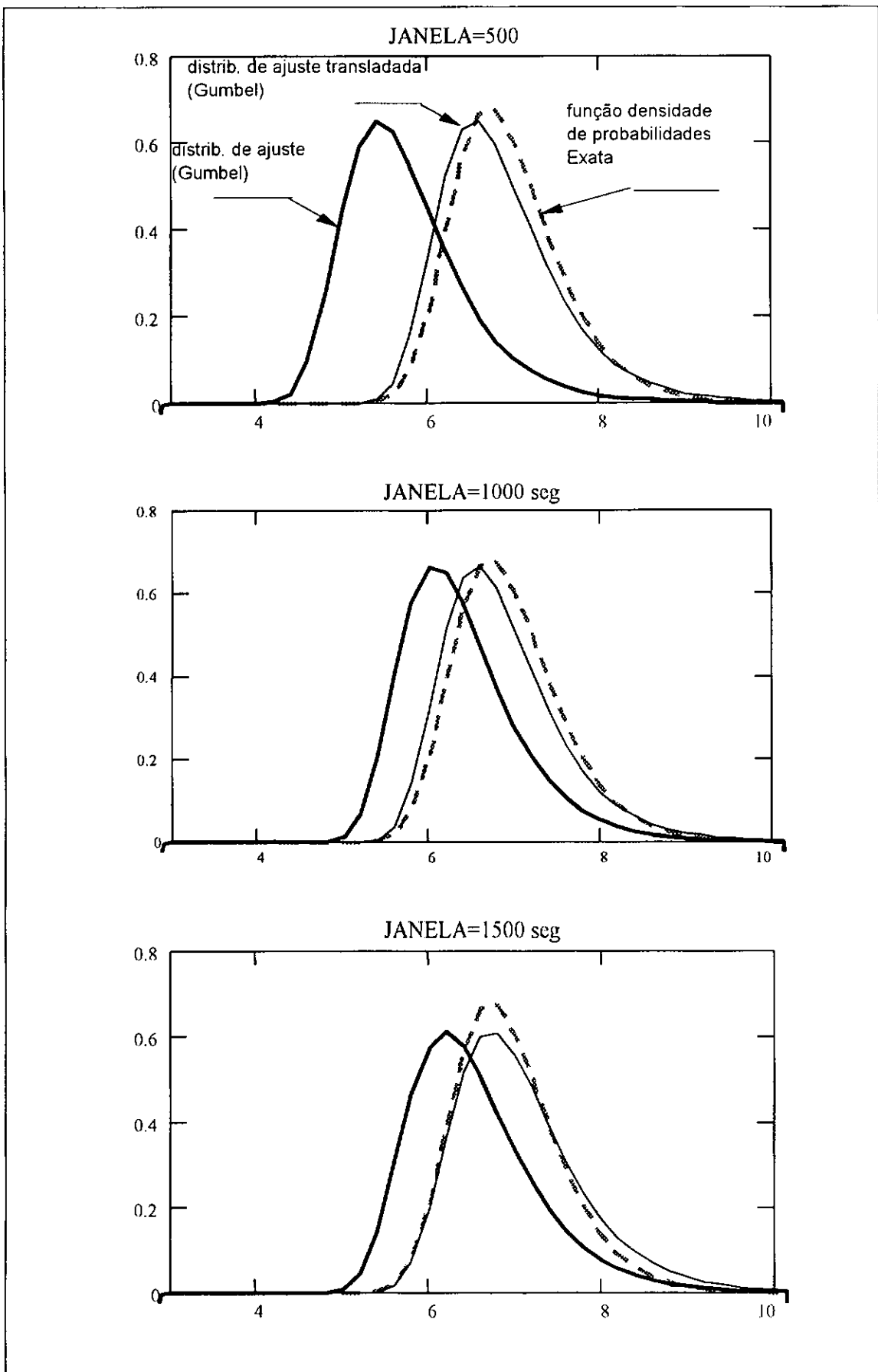


Figura 6.1 - Comparações entre as distribuições exata (linha pontilhada) da amplitude máxima em uma hora e da Gumbel obtida por extrapolação direta (linha fina).

Observa-se também na Figura 6.1, que a estabilidade estatística é alcançada a partir de um intervalo de duração  $t_e = 500s$ .

Na Tabela 6.1 são apresentados os valores da média e do desvio padrão da distribuição de valores extremos de Gumbel, calculados com a metodologia proposta, para as seis realizações de um sinal Gaussiano de 6 horas de duração.

SINAL	MÉDIA	DESVIO PADRÃO
S1NP5	7.026	0.654
S2NP7	7.297	0.910
S3NP11	7.156	0.912
S4NP13	6.980	0.815
S5NP17	6.834	0.611
S6NP23	7.213	0.891
S7NP29	6.871	0.786
<b>Exato</b>	7.005	0.646

Tabela 6.1 Média e Desvio Padrão da Distribuição de Valores Extremos de Gumbel

Observa-se na Tabela 6.1 que os valores médios da distribuição de Gumbel apresentaram um erro máximo de 4% em relação ao valor exato (Rice) igual a 7.00, e que os valores do desvio padrão apresentaram um erro da ordem de 39%, em relação ao valor exato igual a 0.646. O método da extrapolação direta, neste caso, estima bem o valor médio da distribuição do valor extremo e estima de maneira não muito precisa o desvio padrão da mesma.

Os gráficos da Figura 6.2 mostram as comparações entre a distribuição exata de valores extremos e a correspondente distribuição de Gumbel para seis dos sete sinais constantes da Tabela 6.1.

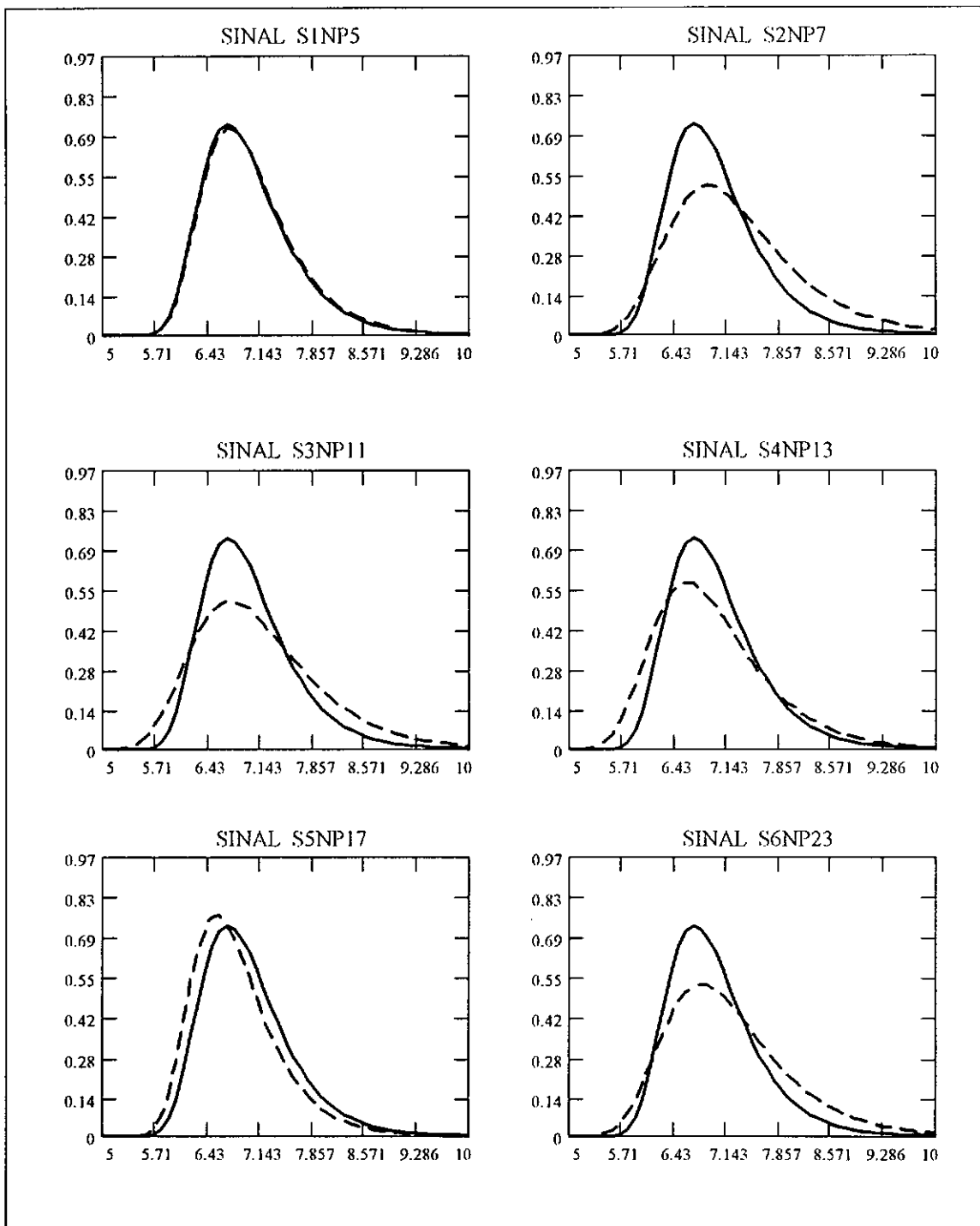


Figura 6.2 - Comparação entre a distribuição exata de valores extremos (linha cheia) e a correspondente distribuição de Gumbel (linha pontilhada), para seis sinais da Tabela 6.1.

## 6.2.2 - MÉTODO DE AJUSTE DA CAUDA DA DISTRIBUIÇÃO PARENTE, BASEADO NA DISTRIBUIÇÃO DE WEIBULL

A primeira etapa do método consiste na determinação dos parâmetros da distribuição de valores extremos Tipo I de Gumbel, a partir da amostra de valores extremos  $X_e$ , de forma análoga ao método anterior. Para esta finalidade foram adotados também os valores de  $t_e$  iguais a 500s, 1000s e 1500 seg.

Uma vez definida a distribuição de extremos de Gumbel, obtém-se a distribuição parente de Weibull equivalente a esta e então, a partir desta última, avalia-se a distribuição de extremos para o período desejado, no caso 1 hora. Os parâmetros da distribuição de Gumbel são calculados pela *média* dos valores obtidos para cada duração adotada de  $t_e$ . Também foram utilizados as sete realizações apresentadas no item anterior.

Na Tabela 6.2 são apresentados os valores da média e do desvio padrão da distribuição corrigida de valores extremos de Gumbel, calculados com esta metodologia, para as sete realizações de um sinal Gaussiano de 6 horas de duração.

SINAL	MÉDIA	DESVIO PADRÃO
S1NP5	6.940	0.575
S2NP7	7.210	0.833
S3NP11	7.072	0.838
S4NP13	6.892	0.736
S5NP17	6.754	0.532
S6NP23	7.126	0.813
S7NP29	6.824	0.690
Exato	7.005	0.646

*Tabela 6.2 - Média e Desvio Padrão da Distribuição Corrigida de Valores Extremos de Gumbel (Weibull Tail)*

Observa-se na Tabela 6.2 que os valores médios da distribuição *corrigida* de Gumbel apresentaram um erro máximo de até 3% em relação ao valor exato ( a partir de

Rice) e que os valores do desvio padrão apresentaram um erro da ordem de até 29% em relação ao valor exato.

Os gráficos da Figura 6.3 mostram as comparações entre a distribuição exata de valores extremos e a correspondente distribuição de Gumbel obtida por esta metodologia para seis dos sete sinais constantes da Tabela 6.2.

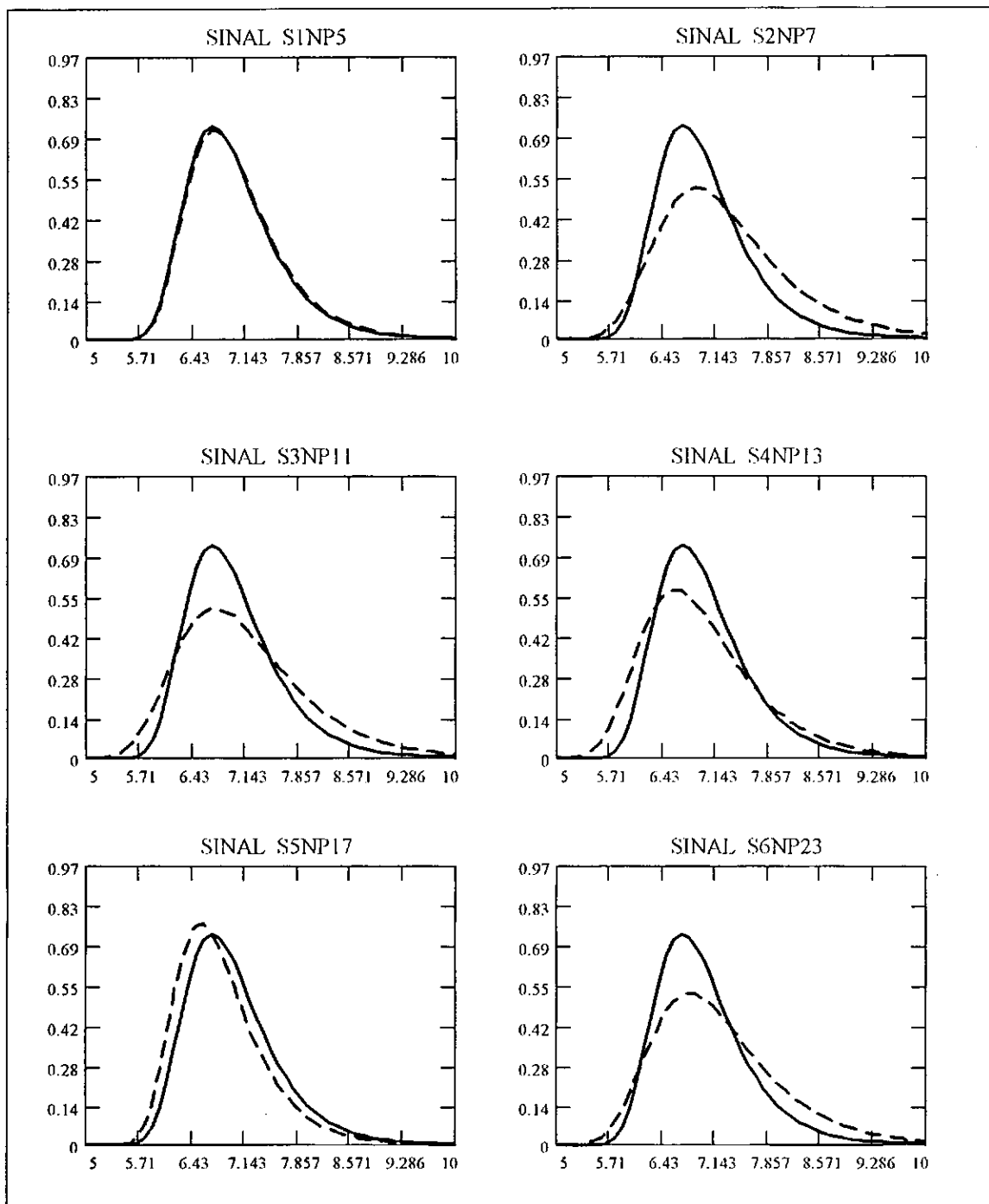


Figura 6.3 - Comparação entre a distribuição exata de valores extremos (linha cheia) e a correspondente distribuição corrigida de Gumbel (linha pontilhada).

De forma análoga ao método da extrapolação direta, o método de ajuste da cauda da distribuição parente baseado na distribuição de Weibull, neste caso, estima bem o valor médio da distribuição do valor extremo e estima de maneira não muito precisa o desvio padrão da mesma.

### 6.2.3 - MÉTODO DE AJUSTE DA CAUDA DA DISTRIBUIÇÃO PARENTE, PELO MÉTODO DE REGRESSÃO LINEAR

Da mesma forma que no item anterior, a distribuição de Weibull é utilizada para representar a distribuição parente de valores máximos  $X_m$ , sendo, neste caso, determinada pelo método de regressão linear, aplicado nos valores da amostra  $X_m$ , correspondente a duração total da série temporal, cuja função cumulativa de probabilidade apresente valores situados *acima de um valor de threshold determinado*.

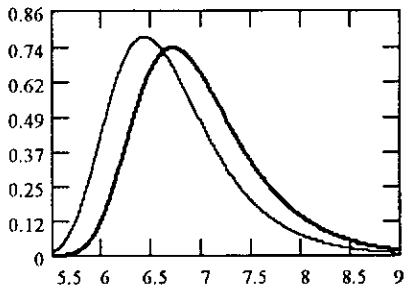
Esta idéia é análoga para o método apresentado no item anterior, que nada mais é do que uma melhor representação da distribuição dos máximos na região da cauda. Porém, a maneira usada para definição da distribuição de Weibull é que é diferente.

Neste trabalho foram adotados, sucessivamente, os seguintes valores para *threshold de probabilidade P*: 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85 e 0.90. Foram investigadas as mesmas sete séries temporais da superfície do mar usadas nos itens anteriores.

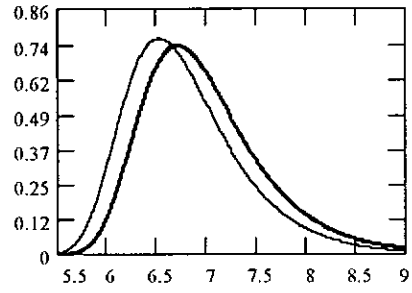
Para a determinação da distribuição de Weibull, definida em (4.15), foram adotados os valores médios dos parâmetros  $\lambda_w$  e  $\alpha_w$ , calculados para cada valor de *threshold*.

Os gráficos da Figura 6.4 mostram as comparações entre as distribuições de valores extremos exata e a calculada a partir da distribuição *parente* de Weibull, ajustada para diversos valores de *threshold* para a série temporal S7NP29 (6 horas de duração).

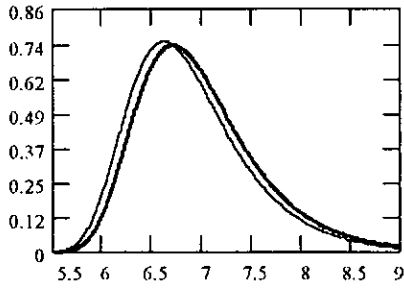
Na Tabela 6.3 são apresentados os valores da média e do desvio padrão da distribuição de valores extremos, calculada a partir da distribuição *parente* de Weibull para cada valor de *threshold* para a série temporal S7NP29.



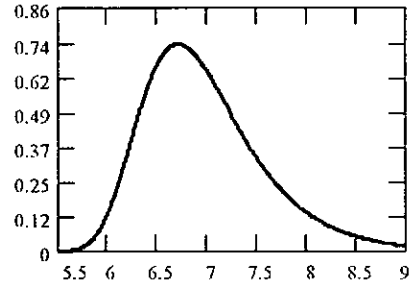
THRESHOLD=0.55



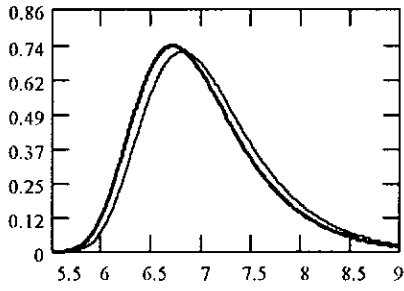
THRESHOLD=0.60



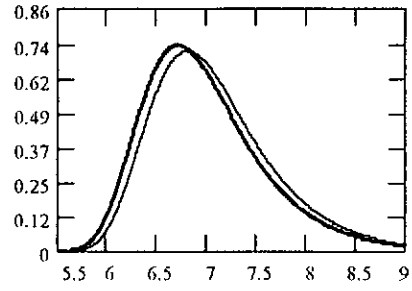
THRESHOLD=0.65



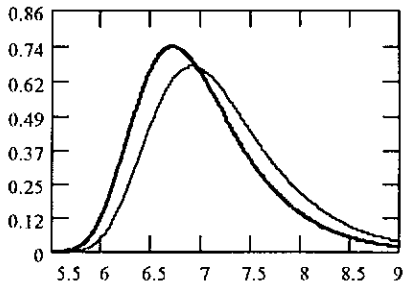
THRESHOLD=0.70



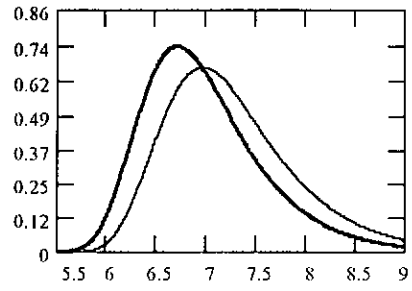
THRESHOLD=0.75



THRESHOLD=0.80



THRESHOLD=0.85



THRESHOLD=0.90

Figura 6.4 – Comparação entre as Distribuições de Valores Extremos Exata (linha fina) e a calculada a partir da Distribuição Parente de Weibull (linha grossa)

Os cálculos foram feitos a partir de um sinal Gaussiano de 6 horas de duração.

<i>Threshold de Prob.</i>	<b>N° de máximos</b>	$\gamma$	$\sigma$	<b>Média</b>	<b>Desvio Padrão</b>
0.60	1400	1.778	2.404	7.132	0.720
0.65	1237	1.787	2.413	7.115	0.714
0.70	1094	1.814	2.429	7.049	0.697
0.75	883	1.854	2.456	6.965	0.673
0.80	688	1.872	2.473	6.942	0.664
0.85	529	1.882	2.484	6.937	0.660
0.90	352	1.970	2.557	6.895	0.650
<b>Valor Médio</b>				7.000	0.683

*Tabela 6.3 - Média e Desvio Padrão da Distribuição de Valores Extremos (S7NP29)*

Na Tabela 6.4 são apresentados os valores da média e do desvio padrão da distribuição de valores extremos, calculada a partir da distribuição *parente* de Weibull, para as sete realizações da série temporal da superfície de elevações do mar.

<b>SINAL</b>	<b>MÉDIA</b>	<b>DESVIO PADRÃO</b>
<b>S1NP5</b>	7.097	0.697
<b>S2NP7</b>	7.089	0.705
<b>S3NP11</b>	7.030	0.700
<b>S4NP13</b>	7.084	0.703
<b>S5NP17</b>	7.079	0.692
<b>S6NP23</b>	7.188	0.732
<b>S7NP29</b>	7.009	0.683
<b>Exato</b>	7.005	0.646

*Tabela 6.4 - Média e desvio padrão da distribuição de valores extremos*

Os gráficos da Figura 6.5 mostram as comparações entre as distribuições de valores extremos exata e a calculada a partir da distribuição *parente* de Weibull, para seis dos sete sinais constantes da Tabela 6.4.

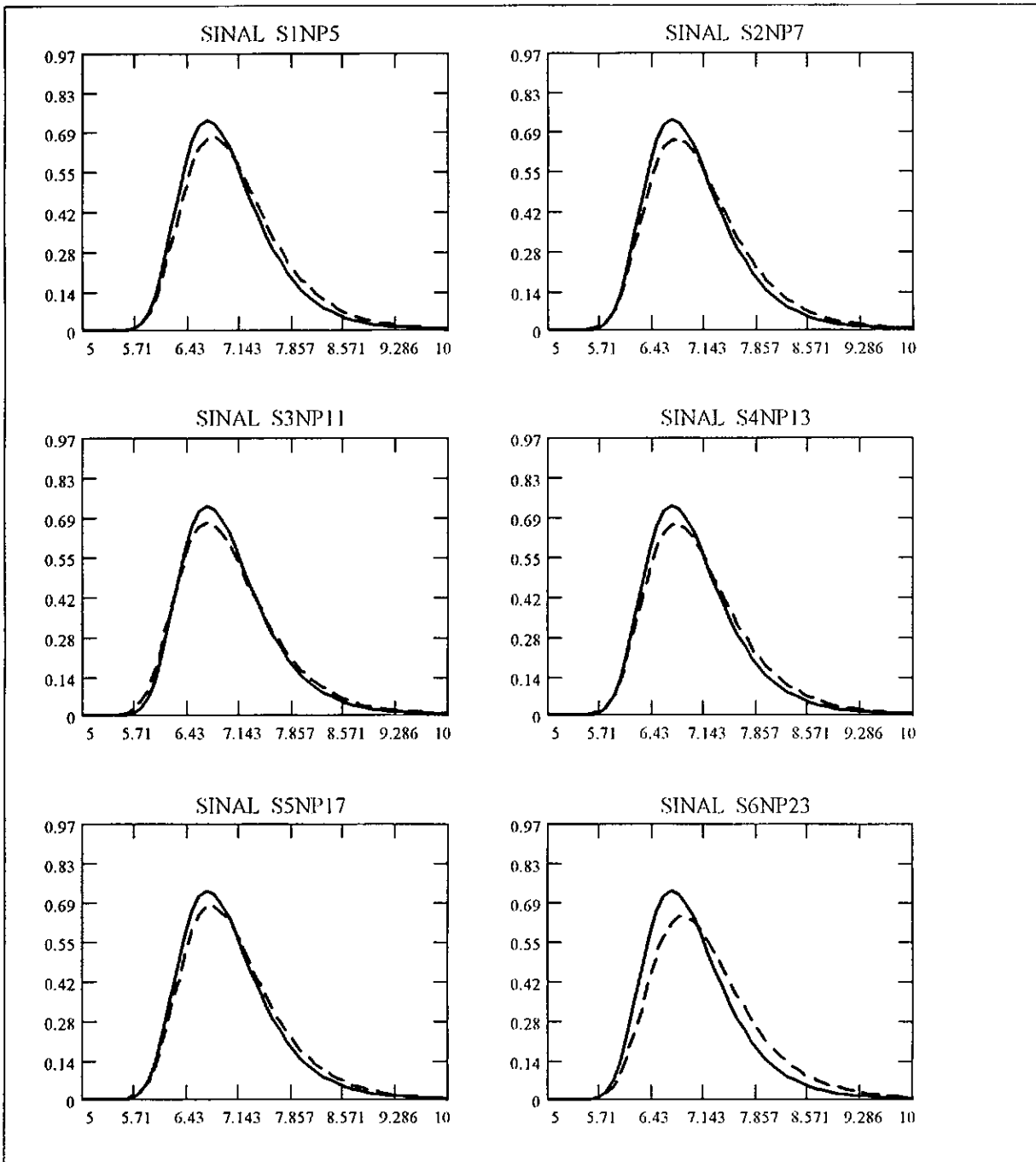


Figura 6.5 – Comparações entre as Distribuições de Valores Extremos Exata (linha pontilhada) e a Calculada a partir da Distribuição Parente de Weibull (linha cheia).

Observa-se na Tabela 6.4 que os valores médios da distribuição de extremos apresentaram um erro máximo de até 3% em relação ao valor exato ( a partir de Rice) e que os valores do desvio padrão apresentaram um erro da ordem de até 12% em relação ao valor exato. O método de ajuste da cauda baseado na distribuição de Weibull com regressão para valores de probabilidade acima de um threshold, no presente exemplo, estima bem o valor médio da distribuição do valor extremo e estima de maneira bem razoável o desvio padrão da mesma.

#### 6.2.4 – MÉTODO BASEADO NA DISTRIBUIÇÃO GENERALIZADA DE PARETO ( PEAK OVER THRESHOLD - POT )

Este método está baseado na constatação de que a distribuição dos valores da amostra de todos os máximos da série temporal  $X_m$ , *excedentes* em relação a um dado valor de *threshold*, pode ser representada pela distribuição de Pareto, conforme foi comentado no Capítulo 4.

Para esta metodologia foram também analisadas as mesmas séries temporais investigadas nos itens anteriores.

A primeira etapa do método consiste, portanto, na determinação dos parâmetros de escala  $\alpha$ , e do parâmetro de forma  $k$  da distribuição parente de Pareto em função do *threshold* escolhido.

Deve-se ter especial atenção na determinação do parâmetro de forma  $k$  que deve ficar com *valores próximos de zero*, para garantir um bom ajuste da distribuição de Pareto, de acordo com o que foi apresentado no Capítulo 4. Para atingir esta finalidade deve-se, geralmente, utilizar um valor relativamente alto de *threshold*.

Na análise da realização denominada SINP5 da série temporal gaussiana da elevação da superfície do mar, foram utilizados valores de *threshold* a partir de 3m com incrementos de 0.1m até atingir o valor final de 5.4m, como está indicado na Tabela 6.5.

Pelos resultados, apresentados na Tabela 6.5, pode-se observar que, à medida que o valor de  $k$  se aproxima de zero, os correspondentes valores da média e do desvio padrão da distribuição de valores extremos se aproximam dos valores exatos. As melhores aproximações foram obtidas para os valores do parâmetro  $k$  situados no intervalo (0.00 a 0.15).

Deste modo, foi verificado que os sinais, para os quais o parâmetro  $k$  apresenta valores no intervalo (0.00 a 0.15), fornecem as melhores aproximações para a média e o desvio padrão da distribuição de valores extremos.

Nos sinais S1NP5, S2NP7, S4NP13 e S6NP23, ocorreram muitos valores de *threshold* (desde 3.7, 3.8, . . . até 5.0 ), com parâmetros *k* dentro do intervalo (0.00 a 0.15).

Thold	$\mu$ Pareto	$\sigma$ Pareto	<i>k</i> Pareto	$\alpha$ Pareto	$\mu$ D.E.	$\sigma$ D.E.	Nº Picos
3.0	0.935	0.797	0.188	1.110	6.759	0.463	822
3.1	0.923	0.782	0.195	1.103	6.733	0.449	780
3.2	0.901	0.770	0.183	1.066	6.773	0.471	756
3.3	0.867	0.761	0.149	0.996	6.882	0.540	711
3.4	0.867	0.747	0.172	1.016	6.815	0.493	650
3.5	0.845	0.730	0.155	0.974	6.866	0.526	511
3.6	0.828	0.723	0.156	0.956	6.854	0.522	460
3.7	0.815	0.712	0.154	0.941	6.862	0.525	410
3.8	0.772	0.704	0.101	0.851	6.985	0.629	380
3.9	0.784	0.692	0.140	0.892	6.901	0.552	356
4.0	0.745	0.688	0.086	0.809	7.014	0.658	318
4.1	0.721	0.683	0.058	0.763	7.074	0.722	287
4.2	0.729	0.676	0.081	0.787	7.035	0.673	247
4.3	0.718	0.669	0.075	0.772	7.045	0.684	216
4.4	0.685	0.666	0.028	0.703	7.114	0.782	197
4.5	0.750	0.650	0.014	0.682	7.070	0.765	166
4.6	0.745	0.640	0.093	0.782	7.043	0.654	152
4.7	0.761	0.638	0.210	0.924	6.946	0.496	128
4.8	0.735	0.620	0.202	0.884	6.946	0.504	115
4.9	0.723	0.604	0.218	0.881	6.936	0.484	103
5.0	0.781	0.562	0.464	1.445	6.813	0.305	85
5.1	0.767	0.540	0.508	1.156	6.811	0.285	73
5.2	0.711	0.525	0.419	1.009	6.843	0.326	64
5.3	0.678	0.505	0.401	0.955	6.851	0.333	54
5.4	0.764	0.502	0.400	1.000	6.811	0.300	44
<b>Exato</b>					7.005	0.646	

Tabela 6.5 – Parâmetros da distribuição de Pareto para vários Tresholds (Sinal S1NP5)

No sinal S3NP11, os valores de *threshold* que resultaram parâmetros **k** dentro do intervalo (0.00 a 0.15), ficaram situados dentro da faixa ( desde 3.1, 3.2, . . . até 4.5 ).

Finalmente, o sinal S5NP17, só teve três valores do parâmetro **k** compreendidos entre (0.10 e 0.15), correspondendo aos níveis de *threshold* de 4.50, 4.60 e 4.70.

Como os valores finais dos parâmetros **k** e  $\alpha$  de cada sinal são calculados pelo valor médio desses parâmetros obtidos em cada sinal com tresholds que resultam em **k** no intervalo(0.10 e 0.15),, observa-se que, as melhores aproximações da média e do desvio padrão da distribuição de valores extremos, estão associadas aos sinais que produziram o maior número de parâmetros **k** dentro deste intervalo, como pode ser verificado na Tabela 6.5.

SINAL	MEDIA	DESVIO PADRÃO
S1NP5	7.000	0.666
S2NP7	7.009	0.667
S3NP11	6.940	0.582
S4NP13	6.863	0.606
S5NP17	6.757	0.520
S6NP23	7.010	0.615
S7NP29	6.711	0.522
Exato	7.005	0.646

Tabela 6.6 - Média e Desvio Padrão da Distribuição de Valores Extremos pelo Método de Pareto

Observa-se, também, na Tabela 6.6 que o maior erro ocorrido no valor médio da distribuição de valores extremos foi de 3.5%. e no desvio padrão foi de 19%.

Na Figura 6.6 são apresentados, para efeito de comparação, os gráficos das funções de densidade de probabilidade de valores extremos exata e a obtida pelo método da distribuição generalizada de Pareto.

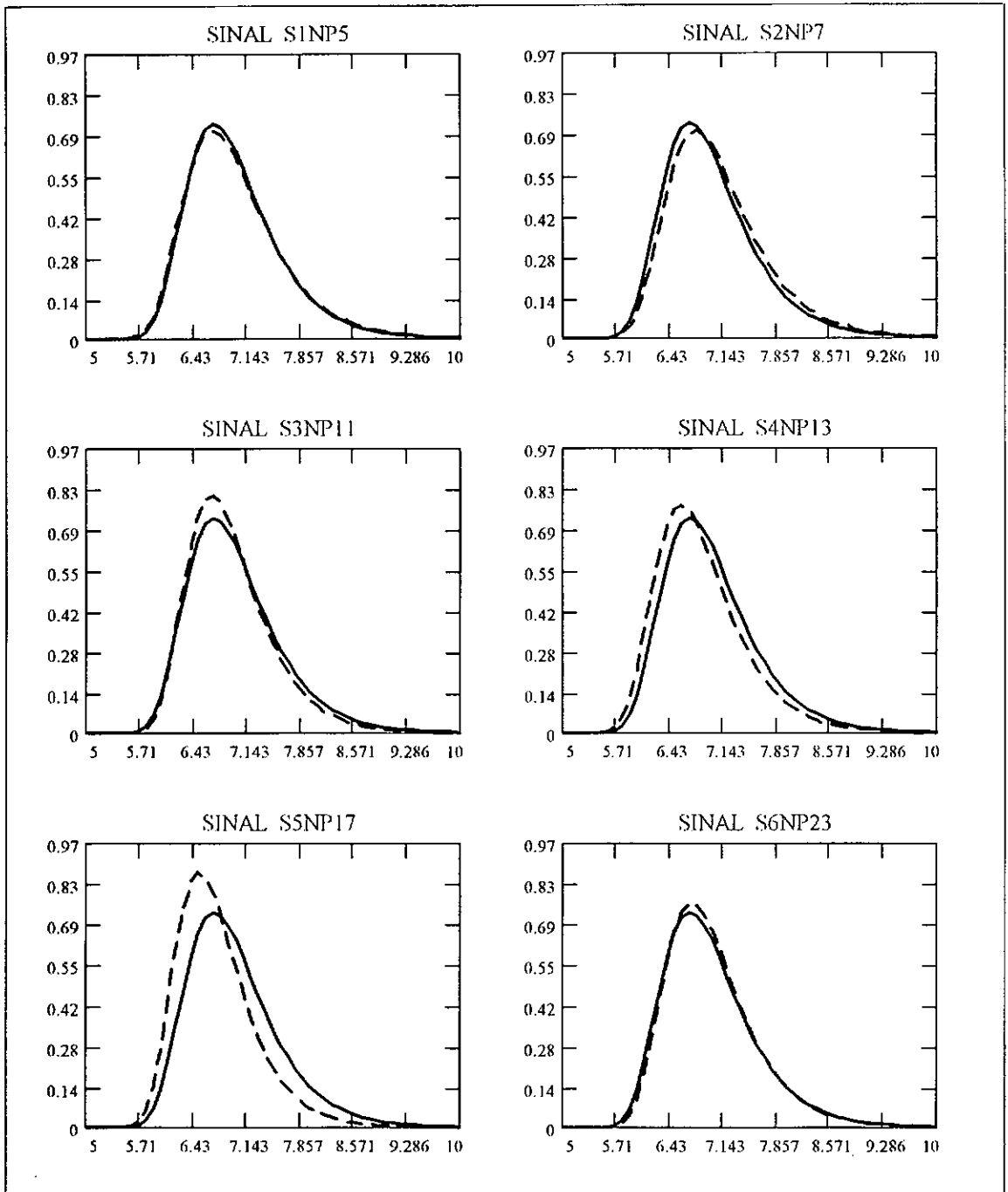


Figura 6.6 – Funções de densidade de probabilidade de valores extremos exata (linha cheia) e a obtida pelo método da distribuição generalizada de Pareto (linha pontilhada).

O método de ajuste da cauda baseado na distribuição de Pareto, no presente exemplo, estima bem o valor médio da distribuição do valor extremo e estima de maneira pouco razoável o desvio padrão da mesma.

## 6.2.5 – RESUMO DOS MÉTODOS APLICADOS NA ANÁLISE DE UM SINAL GAUSSIANO

Na Tabela 6.7 é apresentado um resumo dos valores da média e do desvio padrão, calculados pelos diversos métodos apresentados anteriormente.

SINAIS	Weibull Tail (Gumbel)		Extr. Direta (Gumbel)		Regr. Linear com Threshold		P. O. T. Dist. de Pareto	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
S1NP5	6.940	0.575	7.026	0.654	7.097	0.697	7.000	0.666
S2NP7	7.210	0.833	7.297	0.910	7.089	0.705	7.090	0.667
S3NP11	7.071	0.838	7.156	0.912	7.030	0.700	6.940	0.582
S4NP13	6.892	0.736	6.980	0.815	7.084	0.703	6.863	0.606
S5NP17	6.754	0.532	6.834	0.611	7.079	0.692	6.757	0.540
S6NP23	7.126	0.813	7.213	0.891	7.188	0.732	7.010	0.615
S7NP29	6.824	0.690	6.871	0.786	7.009	0.683	6.711	0.522
<b>Média</b>	6.974	0.717	7.054	0.797	7.082	0.702	6.910	0.600
<b>Erro (%)</b>	-0.4	11.0	0.8	20.3	1.2	8.7	-1.3	-7.1
<b>Desvio Padrão</b>	0.154	0.115	0.162	0.114	0.05	0.014	0.130	0.052
<b>C.O.V. (%)</b>	2.2	16.1	2.3	14.3	0.75	2.15	1.87	8.71

Tabela 6.7 – Média e Desvio Padrão dos Valores Máximos Extremos. (Os valores exatos são: média 7.002 e desvio padrão 0.646)

Pelo resumo apresentado na Tabela (6.7) observa-se que para o sinal gaussiano investigado, de uma maneira em geral, todos os métodos avaliam a média do valor extremo com uma boa precisão. Além disto, a incerteza estatística, medida pelo coeficiente de variação desta grandeza entre todas as simulações, também é pequena. Entretanto, a mesma observação não é válida para o desvio padrão. O desvio padrão é avaliado com menor precisão que a média e, com exceção ao método de regressão linear de Weibull, a incerteza estatística é maior do que a incerteza da média.

### 6.3 – SINAL NÃO GAUSSIANO

A mesma metodologia aplicada ao item anterior para o sinal gaussiano será agora aplicada a um sinal não-gaussiano. O sinal aqui analisado é o  $Y(t) = X(t)X(t)$ , sendo  $X(t)$  o sinal gaussiano de elevações do mar descrito no Capítulo 5 deste trabalho. Observa-se que para o presente caso é possível obter a distribuição exata de valores extremos, como descrito no referido capítulo. Desta forma, é possível comparar a precisão dos diferentes métodos de extremos em relação aos resultados exatos na avaliação de um processo não gaussiano.

Em todas as análises que seguem as mesmas sete realizações das elevações do mar geradas a partir de diferentes conjuntos de fases aleatórias foram utilizados, sendo cada um destes conjuntos gerados com diferentes sementes (seeds) constituídas dos números primos. Para cada realização de  $X(t)$  gerou-se a respectiva realização de  $Y(t) = X(t)X(t)$ .

A seguir não serão especificadas unidades, porém, a unidade de  $Y(t)$ , sua média e seu desvio padrão é  $m^2$ .

#### 6.3.1 - MÉTODO DE EXTRAPOLAÇÃO DIRETA BASEADO NA DISTRIBUIÇÃO DE GUMBEL

Analogamente ao sinal gaussiano, a amostra de valores extremos das séries temporais não-gaussianas foram obtidas pela divisão do sinal em  $N_e$  intervalos de duração (janelas) 500s, 1000s, e 1500s. A distribuição de valores extremos de Gumbel, calculada com os máximos das  $N_e = T/t_e$  janelas, i.e., a distribuição do valor extremo para um período  $t_e$ , foi extrapolada para obter a distribuição do valor máximo correspondente ao período de uma hora. Os parâmetros finais da distribuição são os parâmetros médios das três janelas utilizadas.

A Tabela 6.8 mostra os valores obtidos para a média e o desvio padrão da distribuição do valor extremo do processo não-gaussiano, para o intervalo de uma hora,

a partir de cada uma das sete realizações. Esta tabela também inclui os valores exatos desta distribuição, obtidos no Capítulo 5.

Os gráficos da Figura 6.7 mostram as comparações entre a distribuição exata de valores extremos e a correspondente distribuição de Gumbel, obtida por extrapolação direta, para seis dos sete sinais da Tabela 6.8.

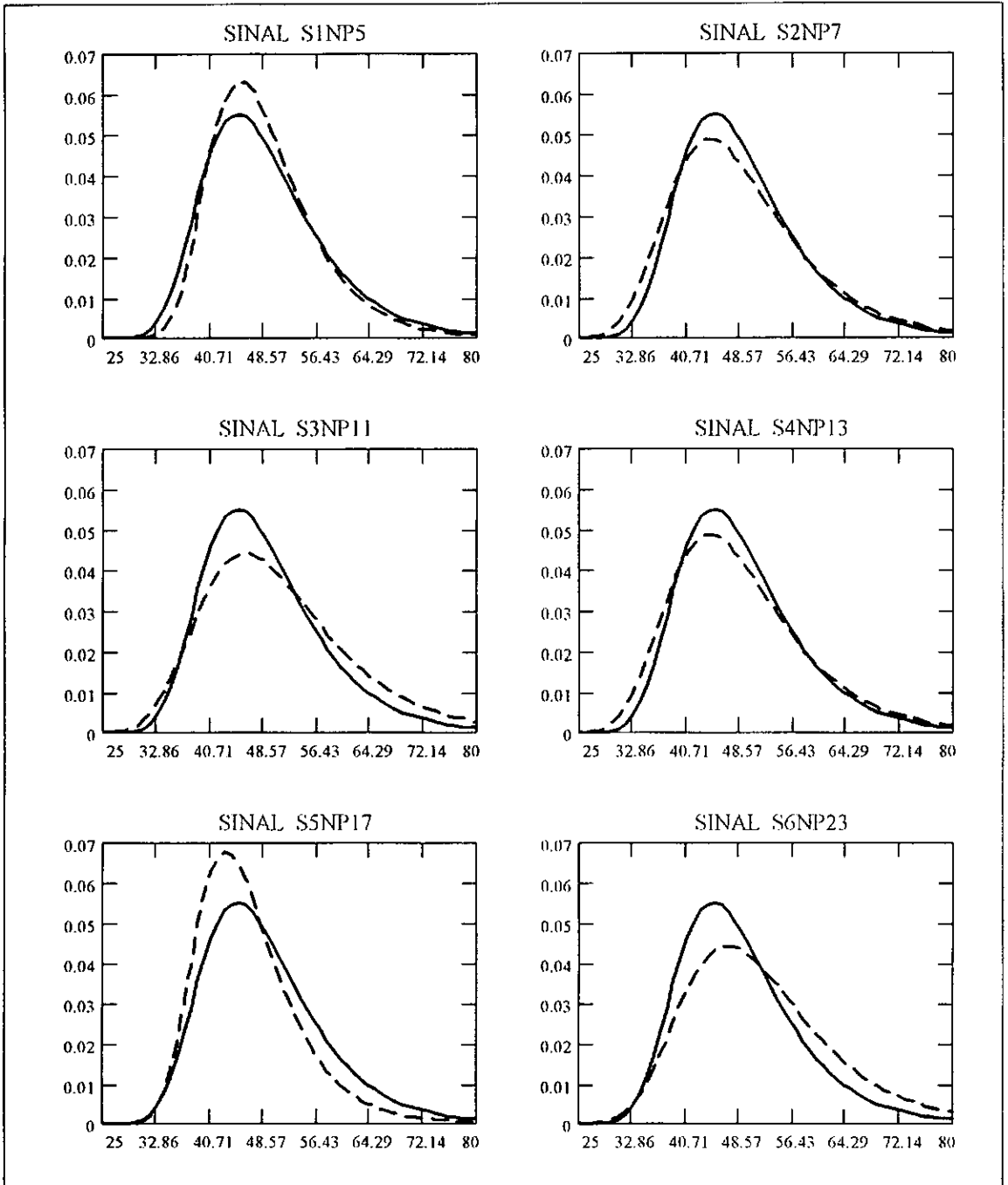


Figura 6.7 - Comparação entre a Distribuição Exata de Valores Extremos (linha cheia) e a correspondente Distribuição de Gumbel extrapolada (linha pontilhada).

SINAL	MÉDIA	DESVIO PADRÃO
S1NP5	48.89	8.04
S2NP7	48.99	10.47
S3NP11	51.19	11.53
S4NP13	48.99	10.47
S5NP17	46.45	7.50
S6NP23	52.15	11.48
S7NP29	47.38	9.73
Exato	48.95	9.22

*Tabela 6.8 – Média e Desvio Padrão da Distribuição de Valores Extremos de Gumbel*

Observa-se na Tabela 6.8 que os valores médios da distribuição de Gumbel apresentaram um erro máximo de 6.5% em relação ao valor exato, e que os valores do desvio padrão apresentaram um erro máximo da ordem de 25% em relação ao valor exato.

### 6.3.2 - MÉTODO DE AJUSTE DA CAUDA DA DISTRIBUIÇÃO PARENTE, BASEADO NA DISTRIBUIÇÃO DE WEIBULL

A mesma metodologia de ajuste da cauda da distribuição parente, baseado na distribuição de Weibull (*Weibull Tail*), aplicada ao processo gaussiano, ver item 6.2.2, foi aplicada para o processo não gaussiano  $Y(t)$  através das sete realizações descritas.

A Tabela 6.9 mostra os valores obtidos através desta metodologia para a média e o desvio padrão da distribuição do valor extremo do processo não-gaussiano, para o intervalo de uma hora, a partir de cada uma das sete realizações. Esta tabela também inclui os valores exatos desta distribuição, obtidos no Capítulo 5.

SINAL	MEDIA	DESVIO PADRÃO
S1NP5	47.01	7.86
S2NP7	51.08	11.04
S3NP11	49.26	10.82
S4NP13	46.96	9.17
S5NP17	44.43	5.70
S6NP23	49.95	10.50
S7NP29	45.33	8.24
Exato	48.95	9.22

*Tabela 6.9 - Média e Desvio Padrão da Distribuição Corrigida de Valores Extremos de Gumbel (Weibull Tail)*

Observa-se na Tabela 6.9 que os valores médios da distribuição *corrigida* de Gumbel apresentaram um erro máximo de 9.2% em relação ao valor exato, e que os valores do desvio padrão apresentaram um erro máximo da ordem de 38% em relação ao valor exato.

Os gráficos da Figura 6.8 mostram as comparações entre a distribuição exata de valores extremos e a correspondente distribuição de Gumbel obtidas através desta metodologia para seis dos sete sinais analisados.

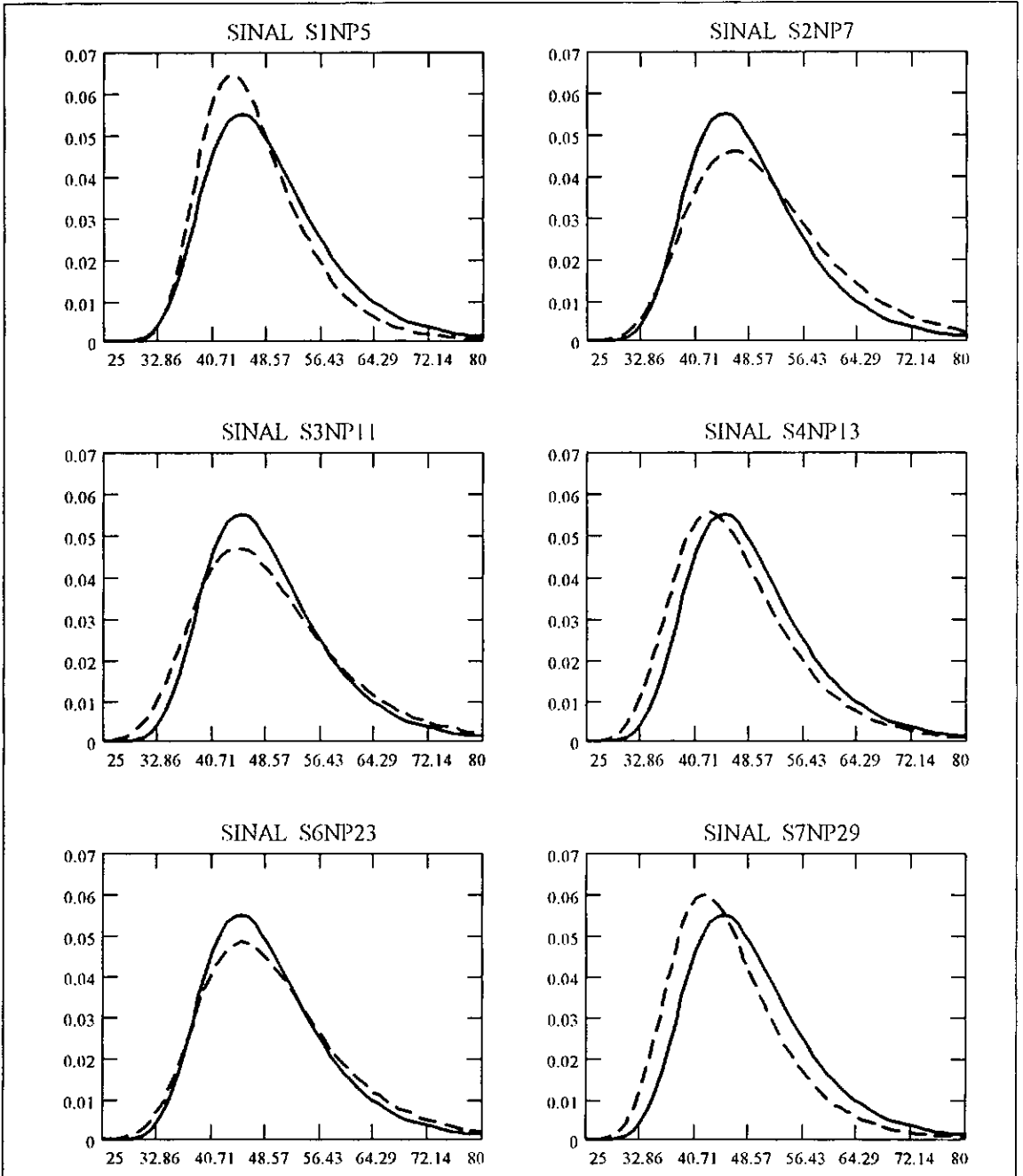


Figura 6.8 - Comparação entre a Distribuição Exata de Valores Extremos (linha cheia) e a correspondente distribuição corrigida de Gumbel (linha pontilhada).

### 6.3.3.- MÉTODO DE AJUSTE DA CAUDA DA DISTRIBUIÇÃO PARENTE, PELO MÉTODO DE REGRESSÃO LINEAR

A mesma metodologia de ajuste da cauda da distribuição parente, baseado na distribuição de Weibull obtida por regressão linear, aplicada ao processo gaussiano, ver item 6.2.3, foi aplicada para o processo não gaussiano  $Y(t)$  através das sete realizações descritas anteriormente.

A Tabela 6.10 mostra os valores obtidos através desta metodologia para a média e o desvio padrão da distribuição do valor extremo do processo não-gaussiano  $Y(t)$ , para o intervalo de uma hora, a partir de cada uma das sete realizações. Esta tabela também inclui os valores exatos desta distribuição, obtidos no Capítulo 5.

SINAL	MÉDIA	DESVIO PADRÃO
S1NP5	51.34	10.37
S2NP7	51.25	10.46
S3NP11	50.20	10.24
S4NP13	51.38	10.43
S5NP17	51.35	10.40
S6NP23	53.70	11.27
S7NP29	51.50	10.47
Exato	48.95	9.22

Tabela 6.10 - Média e Desvio Padrão da Distribuição de Valores

Os gráficos da Figura 6.9 mostram as comparações entre a distribuição exata de valores extremos e a correspondente distribuição de Gumbel obtidas através desta metodologia para seis dos sete sinais analisados.

Observa-se na Tabela 6.10 que os valores médios da distribuição de Gumbel obtidos através da distribuição de Weibull (ajustada por regressão linear) para os máximos do sinal apresentaram um erro máximo de 9.6% em relação ao valor exato, e

que os valores do desvio padrão apresentaram um erro máximo da ordem de 22% em relação ao valor exato.

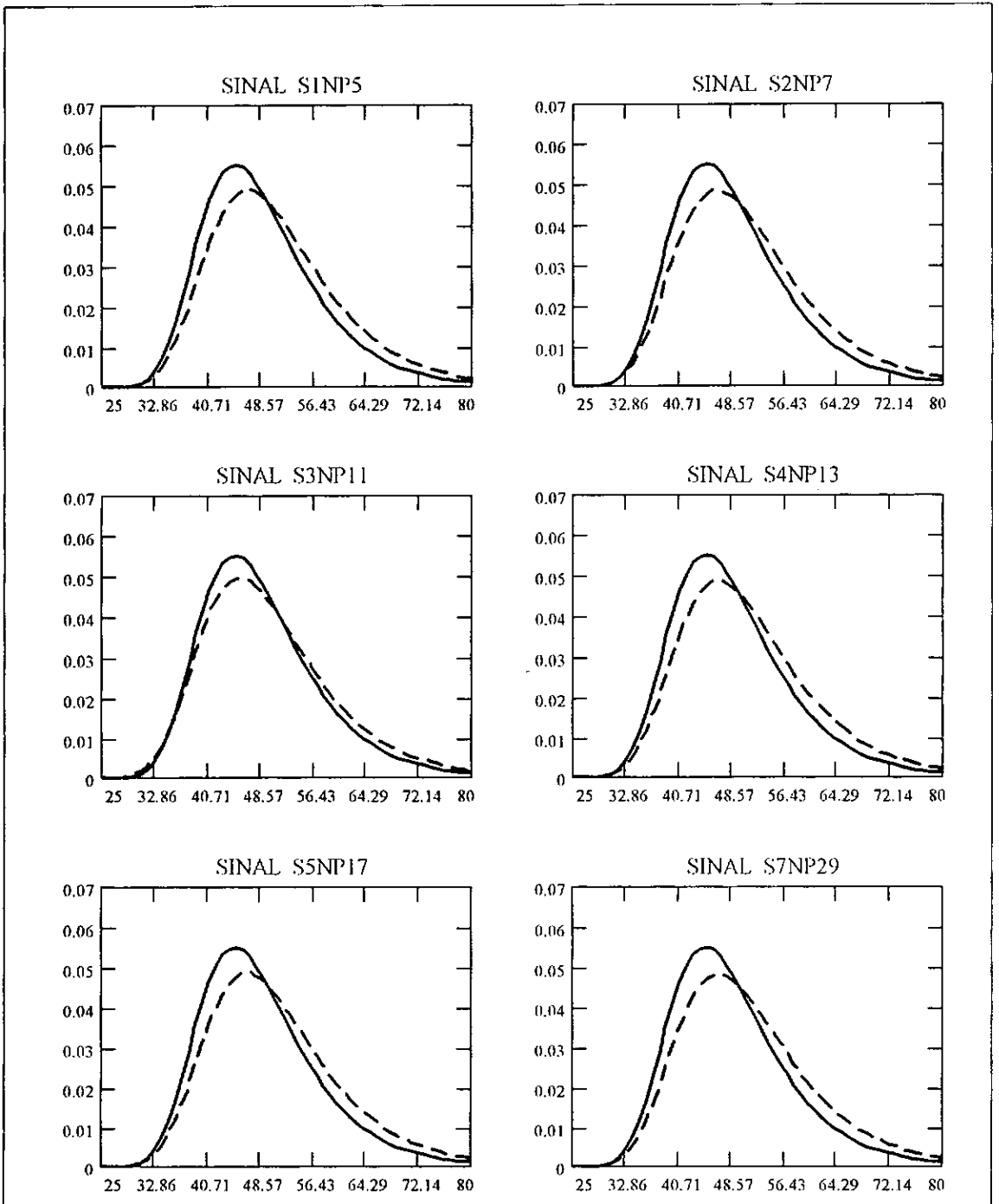


Figura 6.9. Comparações entre as distribuições de Valores extremos Exata (linha cheia) e a calculada a partir da distribuição de Weibul com regressão linear (linha pontilhada).

### 6.3.4 – MÉTODO BASEADO NA DISTRIBUIÇÃO GENERALIZADA DE PARETO ( PEAK OVER THRESHOLD - POT )

A mesma metodologia de ajuste da cauda da distribuição parente, baseado na distribuição na distribuição de Pareto (*Peak Over Threshold*), aplicada ao processo gaussiano, vert item 6.2.4, foi aplicada para o processo não gaussiano  $Y(t)$  através das sete realizações descritas anteriormente.

Observa-se aqui também que, as melhores aproximações da média e do desvio padrão da distribuição de valores extremos, estão associados aos sinais que produziram o maior número de parâmetros  $k$  dentro do intervalo (0.10 e 0.15).

A Tabela 6.11 mostra os valores obtidos através do método baseado na distribuição de Pareto para a média e o desvio padrão da distribuição do valor extremo do processo não-gaussiano  $Y(t)$ , para o intervalo de uma hora, a partir de cada uma das sete realizações.

SINAL	MEDIA	DESVIO PADRÃO
S1NP5	48.99	9.63
S2NP7	50.93	10.42
S3NP11	48.99	9.31
S4NP13	47.60	8.30
S5NP17	45.50	7.23
S6NP23	49.80	9.90
S7NP29	46.00	7.04
Exato	48.95	9.22

Tabela 6. 11 - Média e Desvio Padrão da Distribuição de Valores Extremos

Observa-se na Tabela 6.11 que os valores médios da distribuição de Gumbel obtidos através da distribuição de Weibull (ajustada por regressão linear) para os máximos do sinal apresentaram um erro máximo de 6% em relação ao valor exato, e que

os valores do desvio padrão apresentaram um erro máximo da ordem de 20% em relação ao valor exato.

Na Figura 6.10 são apresentados, para efeito de comparação, os gráficos das funções de densidade de probabilidade de valores extremos exata e a obtida pelo método da distribuição generalizada de Pareto.

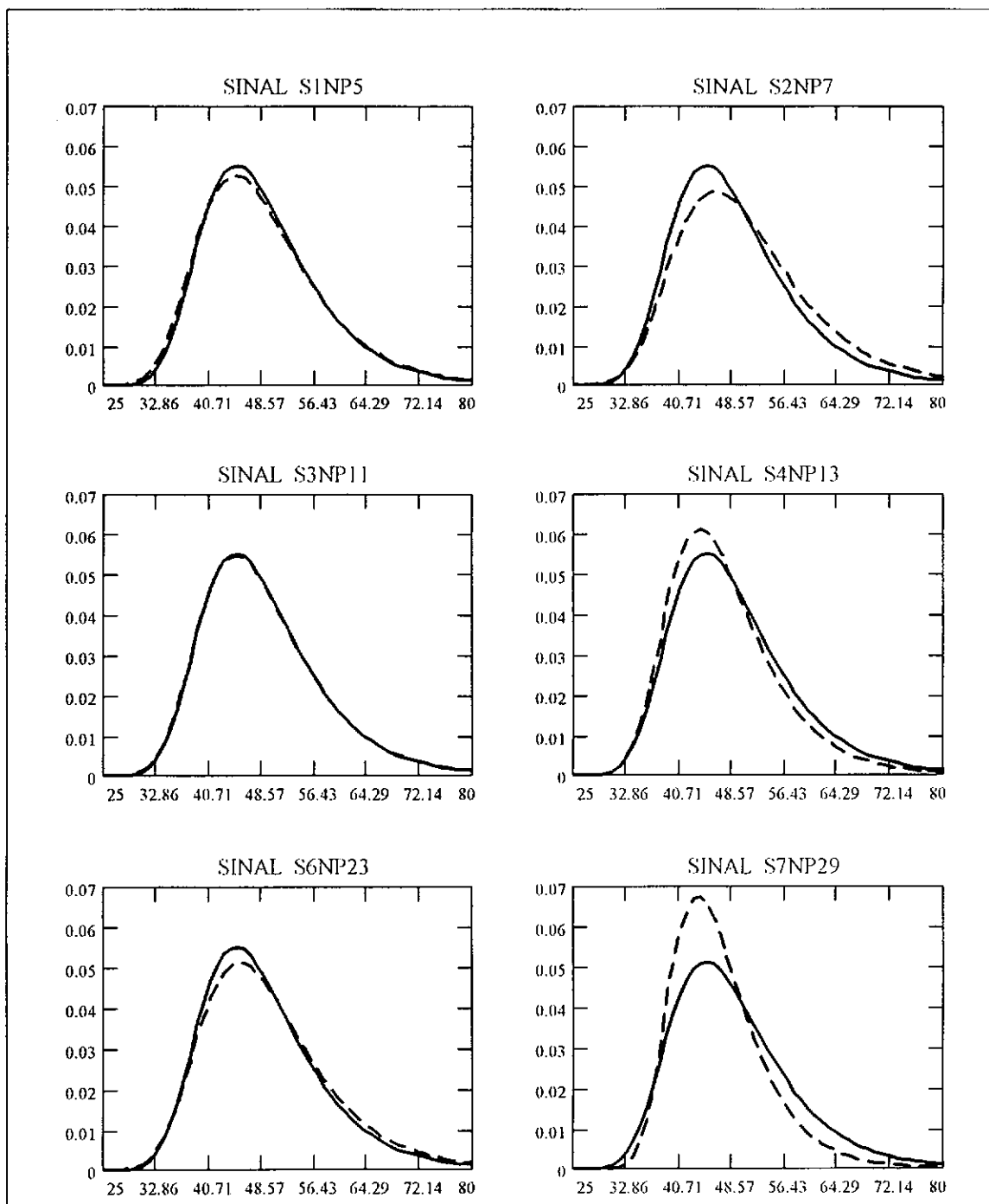


Figura 6.10 Comparações entre as distribuições de Valores extremos Exata (linha cheia) e a obtida pelo método da distribuição generalizada de Pareto (linha pontilhada.)

### 6.3.5.- RESUMO DOS MÉTODOS APLICADOS NA ANÁLISE DE UM SINAL NÃO GAUSSIANO

Na Tabela 6.12 é apresentado um resumo dos valores da média e do desvio padrão, calculados pelos diversos métodos apresentados anteriormente.

SINAIS	Weibull Tail (Gumbel)		Extr. Direta (Gumbel)		Regr. Linear com Threshold		P. O. T. Dist. de Pareto	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
S1NP5	47.01	7.86	48.89	8.04	51.34	10.37	48.99	9.63
S2NP7	51.08	11.04	48.99	10.47	51.25	10.46	50.93	10.42
S3NP11	49.26	10.82	51.19	11.53	50.20	10.24	48.99	9.31
S4NP13	46.96	9.17	48.99	10.47	51.38	10.43	47.60	8.30
S5NP17	44.43	5.70	46.45	7.50	51.35	10.40	45.50	7.23
S6NP23	49.95	10.50	52.15	11.48	53.70	11.27	49.80	9.90
S7NP29	45.33	8.24	47.38	9.73	51.50	10.47	46.00	7.04
<b>Média</b>	47.72	9.05	49.15	9.89	51.53	10.52	48.26	8.83
<b>Erro (%)</b>	-2.6	-9.5	0.3	-1.1	5.2	5.2	-1.5	-11.7
<b>Desvio Padrão</b>	2.28	1.79	1.84	1.47	0.97	0.31	1.84	1.23
<b>C.O.V. (%)</b>	4.77	19.82	3.74	14.83	1.89	2.99	3.92	13.93

*Tabela 6.12 - Média e Desvio Padrão dos Valores Máximos Extremos*

Pelo resumo apresentado na Tabela (6.12) observa-se que para o sinal não gaussiano investigado neste trabalho, de uma maneira em geral, todos os métodos avaliam a média do valor extremo com uma precisão razoável. Além disto, a incerteza estatística, medida pelo coeficiente de variação desta grandeza entre todas as simulações, também é relativamente pequena. Entretanto, a mesma observação não é válida para o desvio padrão. O desvio padrão é avaliado com menor precisão que a média e, com exceção ao método de regressão linear de Weibull, a incerteza estatística é bem maior do que a incerteza da média.

## CAPÍTULO 7

### CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

#### 7.1 CONCLUSÕES

Neste trabalho foi feita uma análise crítica comparativa entre vários métodos propostos para a determinação da estatística dos valores extremos de séries temporais, provenientes de processos estacionários gaussianos e não gaussianos.

Na avaliação dos diferentes métodos foram empregadas séries temporais das elevações do mar para um determinado estado de mar que se constitui de um processo gaussiano. Para se garantir que uma realização do sinal de elevações gerada pelo somatório de harmônicos representasse adequadamente os extremos deste processo, verificou-se que é necessário empregar um grande número de harmônicos ( $>400$ ) e um intervalo de amostragem pequeno ( $<0.25s$ ).

Foram investigados quatro métodos, um método de extrapolação direta e três métodos baseados no ajuste da cauda da distribuição de máximos da série temporal. Os métodos investigados foram os seguintes:

- Método de extrapolação direta baseado no postulado da estabilidade da distribuição de Gumbel;
- Método de ajuste da cauda baseado na distribuição de Weibull ajustada a partir de uma Gumbel (*Weibull Tail*);
- Método de ajuste da cauda baseado na distribuição de Weibull ajustada por regressão linear a partir de um threshold de probabilidades da amostra de máximos;
- Método de ajuste da cauda baseado na distribuição de Pareto ajustada para valores da amostra de máximos acima de um threshold;

Tanto o processo gaussiano bem como o não gaussiano analisados possuem resultados exatos da distribuição de extremos. Desta forma, foi possível comparar os resultados dos diferentes métodos com os resultados teóricos.

Para ambos os processos, verificou-se que os quatro métodos investigados avaliam o valor médio da distribuição de extremos de maneira precisa. Esta precisão foi maior no processo gaussiano. A variabilidade do valor médio extremo para diferentes realizações do mesmo processo também é pequena para todos os métodos analisados.

Verificou-se também que em todos os métodos analisados e para ambos os processos, a precisão do desvio padrão do valor extremo é menor que a precisão da média. Com exceção do método de ajuste da cauda baseado na distribuição de Weibull ajustada por regressão linear a partir de um treshold de probabilidades, a variabilidade do valor médio extremo para diferentes realizações do mesmo processo pode ser grande.

O método de ajuste da cauda baseado na distribuição de Weibull ajustada a partir de uma Gumbel (*Weibull Tail*) foi o método que apresentou a maior variabilidade do desvio padrão em ambos os processos investigados neste trabalho.

O método de extrapolação direta baseado no postulado da estabilidade da distribuição de Gumbel foi o método que apresentou a segunda maior variabilidade do desvio padrão em ambos os processos investigados neste trabalho.

O método de ajuste da cauda baseado na distribuição de Pareto ajustada para valores da amostra de máximos acima de um treshold, usando somente valores de treshold em que o parâmetro  $k$  fique entre 0.0 e 0.15 parece, ser uma alternativa razoável para análise de valores extremos.

O método de ajuste da cauda baseado na distribuição de Weibull ajustada por regressão linear a partir de um treshold de probabilidades da amostra de máximos, apresentou, nos casos investigados, a menor variabilidade tanto da média bem como o desvio padrão. Os valores obtidos para o desvio padrão e para a média em ambos os

casos foram próximos aos exatos, porém, sempre maiores que estes. Isto indica que a distribuição de extremos resultante é próxima à teórica, porém, conservadora.

Deve ser observado que o desvio padrão é quem define a incerteza associada ao valor extremo. Por isto a sua correta avaliação é de fundamental importância para análises em que o mesmo é importante, como por exemplo, a análise de confiabilidade estrutural.

De uma forma em geral e em função dos dois processos analisados neste trabalho, pode-se concluir que se o interesse recai somente na média do valor extremo qualquer um dos métodos pode ser empregado. Quando interesse for a própria distribuição, o desvio padrão é um parâmetro essencial, com isto o método de ajuste da cauda baseado na distribuição de Weibull ajustada por regressão linear a partir de um threshold de probabilidades da amostra de máximos parece ser o mais apropriado.

Um dos fatores importantes a ser observados é o custo computacional envolvido em cada um dos métodos analisados. Os dois primeiros métodos apresentados são baseados em amostras de valores extremos enquanto que os dois últimos são baseados numa amostra de valores máximos de uma realização. Os valores extremos são obtidos de várias realizações (ou janelas) independentes. Desta forma o custo computacional envolvido nos dois últimos métodos pode ser significativamente menor que o custo dos dois primeiros.

## 7.2 SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestões para trabalhos futuros dentro desta linha de estatística de valores extremos pode-se citar:

- avaliação dos métodos apresentados na análise de séries temporais reais advindas da análise aleatória de estruturas oriundas da resposta de sistemas flutuantes de produção de petróleo do tipo FPSO, Spar Buoy, TLP, Semi\_submersível, etc. sob ação de cargas ambientais.

- uso da técnica dos Momentos L para ajuste de distribuições [16]. A análise de extremos esta baseada na cauda superior no comportamento da distribuição dos máximos que é uma função direta dos coeficientes de assimetria skewness e de kurtosis da série temporal. Observa-se que para uma dada realização estes coeficientes são funções cúbica e quádrupla, respectivamente, dos valores amostrais e, portanto, mais suscetíveis a erros. Na técnica dos Momentos L os momentos similares aos coeficientes de assimetria e de kurtosis são funções lineares de valores amostrais e, portanto, mais confiáveis. Assim sendo, o ajuste de distribuições usando a técnica dos Momentos L parece ser uma técnica promissora e robusta para análise da estatística dos extremos.
- uma investigação de estimadores de erro para os resultados obtidos para dar algum indicativo das incertezas associadas aos valores calculados. Nesta linha de investigação podem ser utilizados os intervalos de confiança para cada um dos estimadores calculados [1].

## REFERENCIAS BIBLIOGRÁFICAS

1. ANG H.S. and TANG, A. - *Probability Concepts in Engineering Planning and Design*, Volume II, John Wiley & Sons, 2 ed New York, 1984.
2. GUMBEL, E.J. - "*Statistics of Extremes*", Columbia University Press, 1958.
3. SODAHL N. - *Methods for Design and Analysis of Flexible Risers*, Dsc. Thesis, Division of Marine Structures, The Norwegian Institute of Technology, The University of Trondheim, Norway, 1991.
4. SAGRILO, L.V.S. e LIMA, E.C.P. - *Notas de Aula de Confiabilidade Estrutural (COC 799)*, COPPE/UFRJ, Rio de Janeiro, 1997.
5. NEWLAND, D.E. - *An introduction to random Vibration and espectral Analysis*, Longman, London, 1975.
6. CLOUGH, R. W. and PENZIEN, J. - *Dynamics of Structures*, Mc Graw Hill, Singapore, 1993.
7. KIUREGHIAN, A. - *On Response of Structures to Stationary Excitation*, University of California, Berkeley Report No UCB/EERC\_79/32, Earthquake Engineering Research Center, 1979.
8. CHAKRABARTI, S. K. - *Hidroynamics of Offshore Structures*, Computational Mechanics Publications, Southampton, 1987.
9. MACEDO, P.C. - *Análise Dinâmica de Risers no Dominio do Tempo Utilizando Técnicas de Simulação*, Tese de M.Sc., COPPE/UFRJ, 1993.
10. HOSKING, J.R.M. and WALLIS, J.R. - *Parameter and Quantile Estimation for the Generalized Pareto Distribution.*, Technometrics, Vol 29, No 3, August 1987.
11. DAVIDSON, A. C. and SMITH, R. L. - Models for Exceedances over High Thresholds, *Journal of Royal Statistical. Society, Series B* 52, pp.393-442, 1990.
12. NAESS, A. - *Statistical Extrapolation of Extreme Value Data Based on the Peak over Threshold Method*, OMAE97, Volume II, 1997.
13. STEDINGER, J. R., VOGEL, M.V. and GEORGIU, E.F. - *Frequency Analysis of Extreme Events*, Water Resources, 19 (2), 1991.

14. FERREIRA J.A. and GUEDES SOARES, J. - *An Application of the Peaks over Threshold Method to Significant Wave Height Data*, OMAE97, Vol. II Safety and Reliability, 1997.
15. NESSIM, M. A., HONG H.P., SWAIL, V.R. and HENDERSON, C.A. -*Design Criteria for Offshore Structures Under Combined Wind and Wave Loading*, Journal of Offshores Mechanics and Arctic Engineering. Vol 117/1, 1995.
16. HOSKING J.R.M., *L Moments: Analysis and Estimation of Distribution using Linear Combinations of Order Statistics*, Journal Royal Statistical Soc B 52, No 1, pp. 105-124, 1990.
17. ANG H.S. and TANG, A. - *Probability Concepts in Engineering Planning and Design*, Volume I, John Wiley & Sons, New York, 1975.
18. MANUAL DO USUARIO, MATHCAD 7.0, 1999.