



Universidade do Porto
Faculdade de Engenharia

FEUP



Rui Jorge de Almeida Soares Gonçalves

Sistemas naturais complexos: que tendências? - O estocástico e o determinístico.

Rui Jorge de Almeida Soares Gonçalves

Sistemas naturais complexos: que tendências? - O estocástico e o determinístico.



Departamento de Engenharia Civil

Faculdade de Engenharia da Universidade do Porto

2005

Agradecimentos

Rui Jorge de Almeida Soares Gonçalves

Sistemas naturais complexos: que tendências? - O estocástico e o determinístico.



*Tese submetida à Faculdade de Engenharia da
Universidade do Porto para obtenção do grau de Doutor
em Ciências de Engenharia*

Departamento de Engenharia Civil
Faculdade de Engenharia da Universidade do Porto

2005

UNIVERSIDADE DO PORTO

Faculdade de Engenharia

BIBLIOTECA M

N.º 72322

CDU

Data 23 / 6 / 20 05

Agradecimentos

Gostaria de agradecer penhoradamente ao meu orientador, Senhor Professor Francisco Campelo Calheiros, por nos ter suscitado a paixão pela compreensão dos fenómenos naturais, assim como, ao meu co-orientador, Senhor Professor Alberto Adrego Pinto, pelo apoio e disponibilidade incondicional.

Uma palavra especial de reconhecimento ao Senhor Professor José Tentúgal Valente pela cedência dos primeiros dados usados na dissertação. Somos do mesmo modo gratos ao Senhor Engenheiro Aquino Paredes pelo tempo e preciosas informações dispensadas na visita às estações de medição do caudal do Paiva em Castro Daire e Fragas da Torre.

Os agradecimentos não ficariam completos sem uma menção ao Senhor Professor Luis Almeida Vieira pela sua amizade e companheirismo manifestado no apoio às tarefas docentes que nos competiam. Finalmente, resta a minha gratidão à minha família por todos os sacrifícios que imerecidamente a fiz suportar, muito em especial à minha esposa, pelo seu afecto e paciência infindáveis, sem a qual os nossos esforços não teriam chegado a bom porto.

Resumo

Estuda-se o comportamento dos caudais do rio Paiva a partir de uma série temporal formada pelos caudais médios diários do rio durante um período de 53 anos. O objectivo é obter previsões pontuais do valor do caudal diário num futuro próximo. Utilizamos vários métodos provenientes dos processos estocásticos, sistemas dinâmicos e ainda técnicas de *temporal data-mining* de uma forma inovadora. Assim, apresentamos uma curta síntese de cada uma destas matérias nos capítulos 2,3 e 4, respectivamente. No capítulo 5 procedemos a uma caracterização estatística dos caudais médios diários do rio Paiva e da precipitação na sua bacia. Utilizamos também toda a metodologia apresentada nos capítulos anteriores numa tentativa de compreender a série dos caudais médios diários e prever o valor do caudal no futuro. Só na ausência de chuva conseguimos obter resultados satisfatórios uma vez que a complexidade do sistema diminui. Verificamos que a intermitência dos caudais do Paiva não parece ser uma intermitência dinâmica, no sentido em que, com a informação dada pela série temporal, a dinâmica do rio não parece permitir prever quando o caudal vai aumentar. A intermitência dos caudais está ligada às chuvas, que parecem ser imprevisíveis quando apenas se tem informação sobre os caudais diários, pelo que não estamos na presença de um sistema dinâmico determinístico caótico. Na imprevisibilidade das subidas dos caudais, concentramos a nossa análise de previsão no regime das descidas dos caudais. Assim, utilizando a informação do caudal no dia anterior, conseguimos descrever a distribuição do diferencial de caudal para o dia seguinte. Fazemos previsão pontual utilizando o método dos vizinhos mais próximos, a técnica das janelas de correlação e o modelo linear gaussiano estimado. Notamos

que os resultados preditivos obtidos pelo método dos vizinhos mais próximos são os melhores.

Abstract

The behaviour of the river Paiva runoff is analysed. For that purpose 53 years of a daily runoff time series is used. The aim is to predict a future value of the daily discharge value of the river. Several approaches to the problem are considered. Prediction methods from the following areas, stochastic processes, dynamical systems and temporal data-mining are used. The referred methods are presented in chapters 2,3 and 4 respectively. In chapter 5 we characterize the Paiva river daily runoff as long as the 16 years of available data of daily precipitation in its basin. We also made an attempt to understand the past discharge and predict the future runoff using the referred methods. The complexity of the river discharge unable any reasonable prediction under the influence of precipitation. On the other hand good predictability is found for runoff in the absence of rain. The Paiva river daily runoff intermittency seems not to be of a dynamical type in the sense that using the daily runoff series the methods were unable to predict the runoff increase due to the precipitation occurrence. The Paiva intermittency is linked to the occurrence of rain which seems to be unpredictable and therefore the Paiva river daily series is seems not to be chaotic. As a consequence we focus our analysis on the behaviour of the runoff in the absence of rain. We were able to find the probability distribution of the daily runoff decrease given the present runoff. We also use the nearest neighbours method, correlation windows and an ARIMA model to predict the runoff series. The nearest neighbours was found to be the best in prediction.

Résumé

Sont connus les écoulements du fleuve Paiva pendant 53 ans. L'objectif est l'obtention de prévisions ponctuelles pour les écoulements du futur proche. Sont utilisées des méthodes de processus stochastiques, systèmes dynamiques et de temporal data-mining avec nouvelles formes d'abordage. Une courte synthèse est présentée aux chapitres 2,3 et 4. Au chapitre 5 est présenté une étude statistique des écoulements et des précipitations au bassin du Paiva. Les méthodes des chapitres antérieures sont appliquées pour la interprétation des écoulements journaliers moyens et pour les prévisions. Dans l'absence de pluie la complexité est réduite. L'intermittence des écoulements ne semble pas permettre de prévoir les augmentations de débit. L'intermittence semble être reliée à la pluie qui semble imprévisible devant les écoulement journaliers. Donc, nous concentrons l'étude des prévisions dans le cas de diminution de l'écoulement. Avec l'information de chaque jour, nous calculons la différence d'écoulement le jour suivant. Les prévisions ponctuelles avec les proches voisins, les fenêtres de corrélation et le modèle linéaire sont obtenus. La méthode des voisins plus proches est la meilleure.

Conteúdo

Resumo	4
Abstract	6
Résumé	7
Índice de Tabelas	12
Índice de Figuras	16
1 Introdução	17
2 Modelos lineares gaussianos	20
2.1 Conceitos básicos de sucessões cronológicas	20
2.2 Processo linear	25
2.3 Processos lineares estacionários.	28
2.4 Processos lineares não estacionários	35
2.5 A correlação e o espectro	42
2.6 Limitações dos processos lineares	48

2.7	Testes de linearidade e gaussianidade	51
3	Métodos de sistemas dinâmicos	62
3.1	Sistemas com estabilidade. Atractores convencionais.	63
3.2	Sistemas dinâmicos intermitentes	66
3.3	Entropia de um sistema dinâmico.	71
3.4	Reconstrução do espaço de fase	80
3.4.1	Escolha do desfasamento temporal	87
3.4.2	Escolha da dimensão do espaço de reconstrução	90
3.4.3	Falsos vizinhos próximos	91
3.4.4	Filtros lineares sobre vectores de reconstrução	93
3.5	Medidas <i>SRB</i> e dimensões de medidas invariantes	97
3.6	Estimação de invariantes do sistema	105
3.6.1	Estimação dos Expoentes de Lyapunov	109
3.6.2	Estimação da dimensão-correlação	120
3.6.2.1	Integral de correlação amostral	121
3.6.2.2	Estimador de Grassberger - Procaccia	124
3.6.3	Estimação da Entropia	125
3.7	Medidas de reconstrução	126
3.8	Previsão no espaço de fase	134
4	Janelas de correlação	139
4.1	Temporal data-mining	139

4.2	Aplicações de <i>Temporal data-mining</i>	141
4.3	Previsão usando Janelas de correlação	144
5	Análise dos caudais diários do rio Paiva.	148
5.1	Algumas notas sobre o rio Paiva e sua bacia hidrográfica.	149
5.2	Caracterização dos caudais médios diários e da precipitação na bacia. .	152
5.3	Detecção de não linearidade e não gaussianidade	164
5.4	Reconstrução do espaço de fase	166
5.5	Previsão por métodos locais	196
5.6	Previsão usando janelas de correlação	207
5.7	Estimação de um modelo linear	214
6	Conclusões	219
7	Anexo	221
	Referências	223

Lista de Tabelas

5.1	Várias estatísticas para os caudais diários do rio Paiva.	153
5.2	Algumas estatísticas para a precipitação anual nas 7 estações.	158
5.3	Correlação entre as estações de precipitação.	160
5.4	Correlação entre a precipitação diária nas várias estações e os caudais médios diários em Fragas e Castro.	160
5.5	Parâmetros estimados e respectivo rácio-t.	165
5.6	Resultados do teste de linearidade para a série de Fragas.	165
5.7	Resultados do teste de gaussianidade de Rao-Gabr para vários valores dos parâmetros. Série dos caudais médios em Fragas.	165
5.8	Resultados do teste de reversibilidade para vários valores dos parâmetros. Série dos caudais médios em Fragas.	166
5.9	Dois maiores valores próprios da matriz de covariância para a série de Fragas. . . .	174
5.10	Correlações entre a primeira componente e as variáveis iniciais.	175
5.11	1º valor próprio da matriz de covariância relativa às 3 séries <i>SP1</i> , <i>SP2</i> e <i>SP3</i> para várias dimensões.	179
5.12	Estatísticas para o decaimento. Valores em m^3/s , quando aplicável.	181
5.13	Estatísticas para o decaimento. Valores em m^3/s , quando aplicável.	182

5.14	Estatísticas para o decaimento condicionado.	186
5.15	Resíduo quadrático médio para o ano hidrológico 1997/98 para diferentes dimensões do espaço de fase e espaçamentos temporais em dias. Vizinhança de raio variável. .	198
5.16	Resíduo quadrático médio para o ano hidrológico 1997/98 para diferentes dimensões do espaço de fase e espaçamentos temporais em dias. Vizinhança de raio fixo. . . .	198
5.17	<i>RMQ</i> da previsão para o ano hidrológico 1997/98 e diferentes dimensões e espaçamentos temporais.	200
5.18	Resíduo médio quadrático e resíduo relativo médio para os anos hidrológicos 1995-99 para o regime de descidas de caudais e filtragem de vizinhos.	207
5.19	Resíduo quadrático médio para o ano hidrológico 1997/98 para diferentes tipos de regressão.	208
5.20	Resíduo quadrático médio (coluna (A)) e resíduo relativo médio (coluna (B)) para os anos hidrológicos 1995/99 e apenas para as descidas de caudais usando diferentes preditores.	213
5.21	Correlação entre as estações de precipitação.	215
7.1	Caudais médios diários em m^3/s medidos em Fragas da Torre relativos ao ano hidrológico de 1998/99.	222

Lista de Figuras

1.1	O rio Paiva.	17
2.1	Curva de regressão amostral de X_t sobre X_{t-1} para uma série $ARMA(2, 1)$	49
2.2	Curva de regressão amostral de X_t sobre X_{t-5} para a série dos caudais do Paiva.	49
2.3	Série revertida dos caudais diários do rio Paiva.	49
2.4	Série dos caudais do rio Paiva.	49
4.1	Previsão a um passo para o atractor de Hénon usando janelas de correlação.	146
5.1	Cronograma dos caudais do rio Paiva em Fragas da Torre de 1/10/1946 a 30/09/1999.	152
5.2	Cronograma dos caudais do rio Paiva em Fragas da Torre de 1/10/1976 a 30/09/1979.	153
5.3	Cronograma dos caudais do rio Paiva em Fragas da Torre de 1/10/1977 a 30/09/1978.	153
5.4	Logaritmo do periodograma dos caudais do rio Paiva em Fragas da Torre.	154
5.5	FAC dos caudais médios diários do rio Paiva em Fragas da Torre.	155
5.6	Curva de regressão amostral de X_t sobre X_{t-5}	156
5.7	Curva de regressão amostral de X_t sobre X_{t-20}	156
5.8	Evolução do coeficiente de correlação linear com janelas de tamanho 68.	157
5.9	Diagrama de fase dos caudais em Fragas com desfasamento temporal de 1 dia.	158

5.10	Diagrama de fase dos caudais em Fragas com desfasamento temporal de 3 dias. . .	158
5.11	Alinhamento de sequências de tamanho 45 (dias).	158
5.12	Alinhamento de sequências de tamanho 21 (dias).	158
5.13	Médias da precipitação dos totais mensais 1983/96.	159
5.14	Precipitação mensal em Castro Daire 1916/98.	161
5.15	Precipitação mensal em Mosteiro 1944/00.	161
5.16	Estatísticas dos caudais médios mensais em Fragas.	162
5.17	<i>FAC</i> das 7 estações de precipitação.	162
5.18	Diferencial de caudal a um dia na ausência de chuva.	163
5.19	Diferencial de caudal a 5 dias na ausência de chuva.	164
5.20	Informação mútua média para a série dos caudais médios diários em Fragas. . . .	168
5.21	Evolução da percentagem de falsos vizinhos em função do espaço para espaçamentos temporais de 1 a 7 dias.	169
5.22	Evolução da percentagem de falsos vizinhos em para a série de Fragas e a série de precipitação.	170
5.23	$\log(C_m(r)) - \log(r)$ para várias dimensões de 2 a 19	171
5.24	Evolução dos declives dos segmentos de recta do gráfico (5.23).	171
5.25	Declives da curva correlação para valores inferiores a $20 \text{ m}^3/\text{s}$	173
5.26	Série original e a obtida por projecção no 1º componente principal.	176
5.27	Representação dos indivíduos no espaço dos 3 eixos principais.	176
5.28	Curva-correlação para a série obtida por projecção sobre a primeira componente principal	177

5.29 Declives da curva-correlação para a série obtida por projecção sobre a primeira componente principal	178
5.30 Diagrama de fase para valores entre 0 e $8m^3/s$	180
5.31 Diagrama de fase para valores entre 1 e $2m^3/s$	181
5.32 Histograma dos decaimentos como valor do caudal condicionado à classe C2 e densidade meia-normal ajustada.	182
5.33 Histograma dos decaimentos como valor do caudal condicionado à classe C8 e densidade meia-normal ajustada.	183
5.34 Média do decaimento vs caudal instantâneo.	183
5.35 Desvio padrão do decaimento vs caudal instantâneo.	184
5.36 Densidades <i>meias-normais</i> obtidas por ajustamento às 10 primeiras classes.	185
5.37 Histograma dos caudais médios para o período 1946/99.	190
5.38 Histograma dos caudais médios para o período 1946/72.	191
5.39 Funções de distribuição empíricas das amostras obtidas usando <i>bootstrap</i>	193
5.40 Curva $S(\tau)$, para várias dimensões de 2 a 19 e vários raios de vizinhança	194
5.41 $\ln(C_2(m, \varepsilon/s)/C_2(m+1, \varepsilon/s))$ vs $\ln(\varepsilon)$ para várias dimensões de 3 a 20.	195
5.42 Resíduo médio quadrático para a várias dimensões e desfasamentos temporais usando a opção de número de vizinhos fixo.	199
5.43 Resíduo médio quadrático para a várias dimensões e desfasamentos temporais usando a opção de número de vizinhos variável.	199
5.44 Previsão a um passo usando um número de vizinhos variável.	201
5.45 Previsão a um passo usando um número de vizinhos fixo.	202

5.46	Série residual para a previsão a um passo do ano 1997/98 usando os vizinhos mais próximos	203
5.47	FAC da série residual para a previsão a um passo do ano 1997/98 usando os vizinhos mais próximos	203
5.48	<i>DVS plot</i> para o rio Paiva em Fragas. Erro médio absoluto em função da percentagem de vizinhos usada.	206
5.49	Vizinhos no espaço de fase em função do caudal do rio. Ano hidrológico 1997/8. . .	206
5.50	Vizinhos encontrados a uma distância ϵ no espaço de fase em função do valor em cada instante. Atractor de Hénon.	206
5.51	Previsão a um passo usando janelas de correlação e <i>LSR</i> . Ano hidrológico 1997/8. .	209
5.52	Previsão a um passo usando janelas de correlação e <i>RSR</i> . Ano hidrológico 1997/8. .	209
5.53	Alguns segmentos correlacionados.	210
5.54	Alguns segmentos correlacionados com valores previstos e valor real.	210
5.55	Histograma de todos os valores extrapolados com a regressão e o valor observado. .	211
5.56	Série residual usando a regressão <i>LSR</i>	212
5.57	<i>FAC</i> da série residual.	212
5.58	Comportamento do resíduo médio quadrático da previsão obtida usando como predictor a média e o primeiro quartil.	213
5.59	Número de segmentos correlacionados em função do valor do caudal.	214
5.60	Série residual 1995/99.	216
5.61	<i>FAC</i> da série residual 1995/99.	217
5.62	Previsão a um passo usando o modelo <i>AR(15)</i> . Ano hidrológico 1998/99.	218

Capítulo 1

Introdução

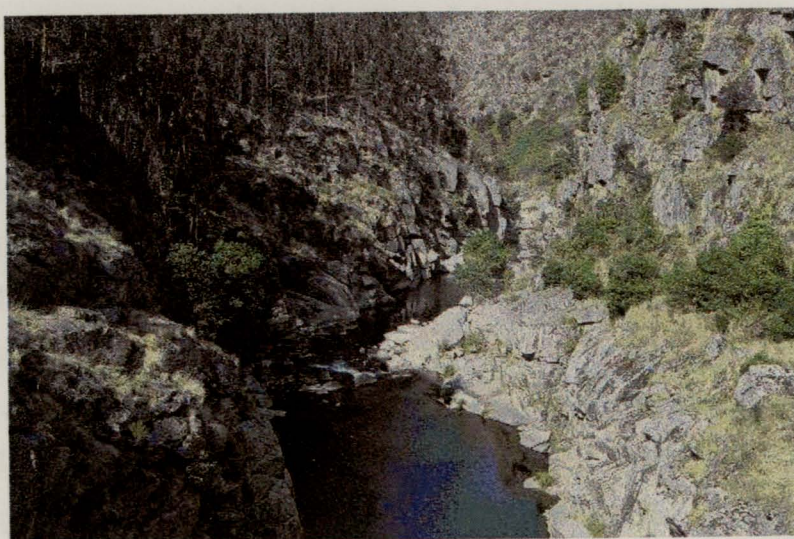


Figura 1.1: O rio Paiva.

Qual a natureza da complexidade de um sistema natural como o rio? O caudal de um rio é o resultado de uma medição de um sistema complexo em que intervêm um grande número de variáveis. Nestas se incluem as que mais directamente se relacionam com o estado do tempo: precipitação, temperatura e humidade relativa do ar. Outras, como a impermeabilidade dos solos e o estado da vegetação, têm uma variação mais regular, isto é, mais lenta. Algumas destas variáveis, como por exemplo a precipitação, apresentam uma significativa variação espacial e sofrem, também, drásticas alterações

num curto período de tempo, inferior a um dia, que é usualmente o período mínimo de observação. Esta complexidade torna o sistema quase imprevisível, excepto se, numa situação ideal, se conseguisse observar tantas variáveis quantas as que se quisesse com a frequência que se desejasse. Vamos estudar o comportamento dos caudais do rio Paiva a partir de uma série temporal formada pelos caudais médios diários do rio medidos durante um período de 53 anos. O objectivo mais importante da compreensão dos fenómenos naturais é a previsão, e esta será encarada neste trabalho na forma mais clássica, a previsão pontual. Entenda-se por previsão pontual a determinação do caudal do rio num dia futuro utilizando o conhecimento de caudais anteriores. Só conseguimos obter resultados satisfatórios na ausência de chuva uma vez que a complexidade do sistema diminui (tabelas 5.16, 5.18).

Vamos utilizar métodos de previsão provenientes dos processos estocásticos, sistemas dinâmicos e técnicas de *temporal data-mining*. Assim, dedicamos o Capítulo 2 ao estudo de modelos lineares gaussianos [Box76b]; o Capítulo 3 aos sistemas dinâmicos, após Poincaré, introduzindo a entropia de Kolmogorov-Sinai [Kol58], o modelo de Lorenz [Lor63], os sistemas intermitentes de Pomeau e Manneville [Pom80], os métodos de reconstrução de Takens [Tak80], assim como a estimação do integral-correlação de Grassberger e Procaccia [GP]; o Capítulo 4 explica a utilização das janelas de correlação para fazer previsões em séries temporais [Mei98]; o Capítulo 5 contém as aplicações de todos os métodos aos dados dos caudais diários do rio Paiva.

Notamos a existência de uma analogia comportamental entre a intermitência no comportamento dos caudais do Paiva e a intermitência em sistemas dinâmicos, dado que o rio revela um comportamento sob influência da chuva muito diferente daquele na ausência de chuva. Assim, pode fazer-se uma comparação entre a fase laminar do sistema dinâmico intermitente e o comportamento dos caudais do Paiva na ausência de chuva, e entre a fase caótica do sistema dinâmico intermitente e o comportamento dos caudais do Paiva sob influência da chuva. No entanto, verificamos que a intermitência dos caudais do Paiva não parece ser uma intermitência dinâmica, no sentido em que, com a informação dada pela série temporal, a dinâmica do rio não permite prever quando o caudal vai aumentar, enquanto que nos sistemas dinâmicos intermitentes

o conhecimento do estado do sistema permite ter essa informação. A intermitência dos caudais provém das chuvas, que parecem ser imprevisíveis quando apenas se tem informação sobre os caudais diários. Pelo que não estamos na presença de um sistema dinâmico determinístico caótico. Esta conclusão deve-se ao facto de, após aplicado o método de reconstrução de Takens à série temporal, o erro de previsão obtido a partir do método dos vizinhos mais próximos ser elevado (5.16).

Dada a imprevisibilidade das subidas dos caudais, concentramos a nossa análise de previsão no regime das descidas dos caudais. Assim, utilizando a informação do caudal no dia anterior, conseguimos descrever a distribuição do diferencial de caudal para o dia seguinte para valores pequenos de caudais uma vez que apenas para estes existem observações suficientes (5.33). Para além de descrever a distribuição dos diferenciais dos caudais, conseguimos fazer previsão pontual utilizando o método dos vizinhos mais próximos, a técnica das janelas de correlação e o modelo linear gaussiano estimado. Os resultados preditivos obtidos pelo método dos vizinhos mais próximos são melhores (5.18). Notamos a existência de uma curva que aproxima as médias dos caudais, dado o caudal no dia anterior, que apresenta dois declives bem distintos um para valores pequenos e outro para valores grandes de caudais (5.6).

Um outro aspecto da auto-semelhança encontrada na série temporal é revelado pelos três primeiros harmónicos da função de autocorrelação (5.5). O primeiro harmónico revela a sazonalidade anual do sistema, com grande precisão, enquanto que os segundos e terceiros harmónicos mostram uma variação aproximada de 45 dias em torno de dois anos. Esta bifurcação do segundo terceiro harmónico resulta, eventualmente, da variabilidade dos períodos secos e de chuva ao longo dos anos.

Em anexo apresentamos os dados relativos aos caudais médios diários do rio Paiva medidos em Fragas da Torre relativos ao ano hidrológico de 1998/99.

Capítulo 2

Modelos lineares gaussianos

Neste capítulo introduzem-se alguns conceitos fundamentais de análise de séries temporais e processos estocásticos. Apresentam-se os modelos lineares gaussianos e suas principais características que dominaram o panorama da modelação de séries temporais durante 60 anos. Enuncia-se o teorema de Wold que destaca o processo linear geral, na classe dos processos estacionários. A classe de processos não estacionários *ARIMA* e as suas variantes são introduzidas assim como as técnicas de estacionarização fundamentais para a modelação linear. Como não poderia deixar de ser, faz-se referência às mais relevantes limitações dos modelos lineares gaussianos e para isso introduzem-se, também, alguns testes de linearidade e gaussianidade.

2.1 Conceitos básicos de sucessões cronológicas

Dado um espaço de probabilidade $(\Omega, \mathcal{A}, \mathcal{P})$ e um qualquer conjunto T , um processo estocástico é a função $X(t, \omega)$, definida em $T \times \Omega$, que para cada $t \in T$ é uma variável aleatória. A natureza de T e do conjunto onde $X(t, \omega)$ toma valores, seja, S , determinam a natureza do processo estocástico. Um processo estocástico diz-se discreto se T for um conjunto numerável e diz-se contínuo se T for um intervalo da recta real. Para cada ω' fixo, $\omega' \in \Omega$, $X(t, \omega')$ é uma função determinística de t que toma

valores em T . Cada uma destas funções é uma realização do processo e o conjunto das realizações do processo denomina-se, o *ensemble* ou espaço de trajectórias do processo. Uma sucessão cronológica é apenas uma das infinitas realizações do *ensemble* associado ao processo, o que corresponde a um registo de observações de algum fenómeno, feito sequencialmente no tempo.

Na análise subsequente, salvo menção em contrário, designa-se por $\{X_t\}$, $t \in T \equiv \{0, \pm 1, \pm 2, \dots\}$, um processo estocástico discreto ¹. A estrutura probabilística de X_t fica completamente determinada se se conhecer a função de distribuição conjunta de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ onde $t_1, t_2, \dots, t_n$ é um conjunto finito de valores admissíveis de t , e n um inteiro positivo arbitrário.

O processo $\{X_t\}$ diz-se estritamente estacionário se as suas propriedades estatísticas não se alterarem com o tempo, isto é, se para um conjunto admissível, $\{t_1, t_2, \dots, t_n\}$ e quaisquer inteiros h, n a relação,

$$F_{t_1+h, \dots, t_n+h}(X_1, \dots, X_n) = F_{t_1, \dots, t_n}(X_1, \dots, X_n) \quad (2.1)$$

se verificar. No entanto a relação (2.1) é de difícil, por vezes impossível, verificação. Em alternativa usa-se a estacionaridade fraca ou no sentido lato. O processo $\{X_t\}$ diz-se fracamente estacionário até ordem m se, os vectores aleatórios n -dimensionais,

$$\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\} \text{ e } \{X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}\}$$

possuírem momentos até à m -ésima ordem e se esses momentos forem iguais. Desta forma, um processo $\{X_t\}$ é fracamente estacionário (dir-se-à apenas estacionário) até ordem 1 se, $E(X_t) = \mu$ para todo o t e $\mu < \infty$. O processo diz-se estacionário até 2ª ordem se for estacionário até 1ª ordem e sendo μ a média,

$$E(X_t^2) = \mu_2, \text{ constante, } \forall t$$

¹A definição vai fazer-se para processos com parâmetro discreto. A generalização para o caso contínuo faz-se sem dificuldade.

deduzindo-se destas expressões que a variância, σ^2 é constante e igual a $\mu_2 - \mu^2$. O valor esperado, $E(X_t X_s)$, depende apenas da diferença, $t - s$, isto é, $E(X_t X_s) = E(X_{t-s} X_0)$ para todo o s, t . Seja $\{X_t : t = 0, \pm 1, \pm 2, \dots\}$ um processo estacionário até 2ª ordem com média, μ e variância, σ^2 , o valor esperado,

$$\gamma_k = Cov(X_t X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)], k = 0, \pm 1, \pm 2, \dots \quad (2.2)$$

que existe por hipótese, define a função de autocovariância (*FACOV*) do processo. Para cada k , γ_k mede a intensidade com que covariam pares de valores do processo separados por um desfasamento (*lag*) de amplitude k . Escrevendo,

$$\rho_k = \gamma_k / \gamma_0, \quad (2.3)$$

obtém-se, para cada k , a função de autocorrelação (*FAC*) do processo considerado. Estas funções têm algumas propriedades,

- 1) $\gamma_0 = \sigma^2$ $\rho_0 = 1$
- 2) $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq \rho_0$
- 3) $\gamma_k = \gamma_{-k}$, $\rho_k = \rho_{-k}$
- 4) $\forall t_1, \dots, t_n \forall \beta_1, \dots, \beta_n \in \mathbb{R} \ 0 \leq V \{ \sum_{k=i}^n \beta_k X_{t_k} \} = \sum_{k=i}^n \sum_{m=1}^n \gamma_{|t_r - t_s|} \beta_k \beta_m$

A propriedade 3) mostra que ρ_k e γ_k são funções pares. A propriedade 4), [Pri81], mostra que a *FAC* e a *FACOV* são funções semidefinidas positivas, à representação gráfica da *FAC* chamamos um correlograma.

A correlação parcial que existe entre X_t e X_{t+k} , isto é, a correlação entre estas duas variáveis, quando se fixam as variáveis aleatórias intermédias $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$

dá, principalmente no caso dos processos lineares, uma importante contribuição à caracterização do processo. Suponha-se, sem perda de generalidade, que $E(X_t) = 0$ e $V(X_t) = 1$ e considere-se a regressão múltipla de X_{t+k} sobre os valores passados do processo até X_t , seja, [Mur93],

$$X_{t+k} = \phi_{k1}X_{t+k-1} + \phi_{k2}X_{t+k-2} + \cdots + \phi_{kk}X_t + \xi_{t+k} \quad (2.4)$$

onde ϕ_{kj} , $j = 1, 2, \dots, k$ são os coeficientes da regressão e ξ_{t+k} o erro não correlacionado com X_{t+k-j} , $j \geq 1$. O coeficiente ϕ_{kk} representa a função autocorrelação parcial (*FACP*).

Um processo estacionário até 2ª ordem de grande interesse na análise de sucessões cronológicas é o chamado ruído branco. Um processo $\{\varepsilon_t\}$, $t = 0, \pm 1, \pm 2, \dots$ diz-se ruído branco se for formado por uma sucessão de variáveis aleatórias não correlacionadas, igualmente distribuídas e com média e variância constantes, $E(\varepsilon_t) = \mu$, $V(\varepsilon_t) = \sigma_\varepsilon^2$ tal como obriga a estacionaridade fraca até 2ª ordem. Dessa forma a *FACOV* e *FAC* deste processo são dadas pelas seguintes expressões,

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 & , \quad k = 0 \\ 0 & , \quad k \neq 0 \end{cases}, \quad \rho_k = \begin{cases} 1 & , \quad k = 0 \\ 0 & , \quad k \neq 0 \end{cases}$$

A importância do processo não advém de si próprio mas do papel que desempenha na construção de outros processos estacionários importantes na modelação de sucessões cronológicas.

Na prática de análise de sucessões, tem grande importância para a caracterização e modelação de uma sucessão, estimar a média, variância, *FACOV*, *FAC* e a *FACP*. Seja $X_1, X_2, \dots, X_N$, uma sucessão cronológica observada. O estimador que se impõe para a média é,

$$\hat{\mu} = \bar{X} = \frac{1}{N} \sum_{t=1}^N X_t \quad (2.5)$$

Um estimador para a variância é, $\hat{\sigma}_X^2 = \frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2$. Um estimador muito usado para estimar a *FACOV* é dado pela seguinte expressão,

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X}) (X_{t+k} - \bar{X}) \quad (2.6)$$

onde $0 \leq k \leq N - 1$ e $\hat{\gamma}_k = \hat{\gamma}_{-k}$. Apesar de não ser consistente, $\hat{\gamma}_k$ de (2.6) é o estimador mais eficiente da *FACOV*, [Mur93]. A preferência por $\hat{\gamma}_k$ como estimador da *FACOV*, leva a que se opte como estimador da *FAC*, a função,

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X}) (X_{t+k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} \quad (2.7)$$

A representação gráfica desta função designa-se por correlograma estimado. Para processos estacionários Gaussianos, [Bar46] mostrou que quando N , tamanho da amostra, é suficientemente grande, $\hat{\rho}(k)$ tem distribuição aproximadamente normal. No caso do ruído branco a distribuição tem média nula e variância $1/N$. Se se representar num correlograma estimado de ruído branco as bandas $\pm 2/\sqrt{N}$ poderá controlar-se estatisticamente a evolução da *FAC* estimada uma vez que essas bandas representam os limites inferiores e superiores do intervalo de confiança a 95% para aquela função. Se 95% dos valores estimados da *FAC* ficarem dentro desse intervalo, então o comportamento está de acordo com aquilo que se espera de uma realização de ruído branco. Os autores Levinson, e Durbin construíram um algoritmo recursivo para o cálculo da *FACP* teórica e estimada [Mur93].

Deve-se a Wold, [Wol38] a demonstração de um teorema fundamental para processos

estocásticos estacionários. Qualquer processo estacionário pode decompor-se na soma de um processo *regular* com um processo *singular*².

Teorema 2.1.1 *Qualquer processo estacionário $\{X_t\}$ pode exprimir-se na forma*

$$X_t = U_t + V_t \quad (2.8)$$

onde,

- U_t e V_t são processos não correlacionados.
- U_t é regular com representação linear parcial, $U_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u}$, com $g_0 = 1$, $\sum_{u=0}^{\infty} g_u^2 < \infty$ e $E[\varepsilon_t V_s] = 0 \forall s, t$. As sucessões $\{g_u\}$ e $\{\varepsilon_t\}$ são unicamente determinadas.
- V_t é singular, isto é, é um processo determinístico com variância de previsão zero.

para mais pormenores ver [Pri81]. O teorema de Wold está relacionado com o teorema da decomposição do espectro integrado de Lebesgue.

2.2 Processo linear

O processo estocástico linear deve a sua origem à suposição de que uma sucessão é uma agregação linear de choques aleatórios (também designados por inovações), [Yul27]. Esta suposição é natural uma vez que existe dependência entre as variáveis aleatórias referentes a sucessivos pontos no tempo. Em tal hipótese, escreve-se,

²um processo diz-se regular se for puramente aleatório e singular se for puramente determinístico.

$$X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (\psi_0 = 1) \quad (2.9)$$

$\{X_t\}$ diz-se um processo linear geral. A equação (2.9) também se designa por uma média móvel, simbolicamente $MA(\infty)$. O processo linear geral, (2.9), permite-nos representar X_t como uma soma ponderada de valores passados e presente do processo $\{\varepsilon_t\}$. Como as variáveis aleatórias $\{\varepsilon_t\}$ não são correlacionadas, a sua função de autocovariância é dada pela seguinte expressão,

$$\gamma_k = E[\varepsilon_t \varepsilon_{t+k}] = \begin{cases} \sigma_\varepsilon^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad (2.10)$$

onde $V(X_t) = (\sum_{j=0}^{\infty} \psi_j^2 \varepsilon_{t-j})^2 \sigma_\varepsilon^2$. Para que o processo $\{X_t\}$ tenha variância finita e seja estacionário, é condição necessária é que a série $\sum_{j=0}^{\infty} \psi_j^2$ seja convergente. Como (2.9) envolve uma soma infinita de parcelas deve ser questionado o sentido probabilístico da definição, [Gon96], em geral admite-se que se tem convergência em média quadrática, isto é,

$$\lim_{n \rightarrow \infty} E \left\{ \left(X_t - \sum_{j=0}^n \psi_j \varepsilon_{t-j} \right)^2 \right\} = 0 \quad (2.11)$$

O operador atraso, B , definido pelas relações $BX_t = X_{t-1}$ e $B^k X_t = X_{t-k}$, permite escrever o processo $\{X_t\}$ na seguinte forma mais condensada,

$$X_t = \psi(B) \varepsilon_t, \quad \psi(B) = \sum_{j=0}^{\infty} \psi_j B^j. \quad (2.12)$$

Usando propriedades do ruído branco, pode obter-se a função de autocovariância do processo linear geral,

$$\begin{aligned}\gamma_k = E(X_t X_{t+k}) &= E\left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t+k-j}\right) \\ &= \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+k}\right) \sigma_\varepsilon^2\end{aligned}\quad (2.13)$$

Facilmente se obtém desta a expressão da *FAC*, $\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}$. Uma maneira mais conveniente de obter a autocovariância de um processo linear é através da função geradora de autocovariância, [Mur93],

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k, \quad k = 0, \pm 1, \pm 2, \pm 3, \dots \quad (2.14)$$

Donde para a *FAC* se obtém, $\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0}$, $k = 0, \pm 1, \pm 2, \pm 3, \dots$.

A convergência da série do 2º membro de (2.13) assegura que o processo têm variância finita. Para o processo linear, $\{X_t\}$, definido em (2.9) a condição de estacionaridade é que a série $\psi(B)$, função geradora dos ψ' s, tem que convergir para $|B| \leq 1$. Sob certas condições, o processo $\{X_t\}$ também se pode escrever como uma soma ponderada dos valores passados $\{X_s\}$, $s < t$, adicionados de um choque aleatório, ε_t , isto é,

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \varepsilon_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t \quad (2.15)$$

ou $X_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \dots = \varepsilon_t$. Ou ainda, $\pi(B)X_t = \varepsilon_t$ com,

$$\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \dots) = 1 - \sum_{j=1}^{\infty} \pi_j B^j, \text{ onde } 1 + \sum_{j=1}^{\infty} |\pi_j| < \infty \quad (2.16)$$

A representação (2.15) designa-se autoregressiva, simbolicamente $AR(\infty)$, e facilita a compreensão dos mecanismos de previsão, uma vez que estas se baseiam em regressões. A relação entre os dois tipos de representação obtém-se da comparação directa entre (2.12) e (2.16), dando origem à equação,

$$\pi(B) = [\psi(B)]^{-1} \Rightarrow \pi(B) \psi(B) = 1 \quad (2.17)$$

A relação permite o cálculo dos π 's conhecendo os ψ 's e vice-versa. A condição de invertibilidade é independente da de estacionaridade, um processo linear diz-se invertível se for passível de representação autoregressiva, matematicamente, é invertível se o polinómio $\pi(B)$ convergir para $|B| \leq 1$. Na prática, as representações $MA(\infty)$ e $AR(\infty)$ não são de grande utilidade, ofendem o princípio da parcimónia e por isso apenas se exploram versões finitas destes processos.

2.3 Processos lineares estacionários.

Um processo $\{X_t\}$ diz-se $AR(p)$ quando satisfaz a seguinte equação,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t \quad (2.18)$$

ou $\phi_p(B)X_t = \varepsilon_t$ onde, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$. A solução formal de (2.18), X_t , é a soma da solução complementar, X_t^* , com uma solução particular, X_t^{**} , ou seja, $X_t = X_t^* + X_t^{**}$. Onde X_t^* é a solução de $\phi_p(B)X_t = 0$, e é da forma,

$$A_1 G_1^t + A_2 G_2^t + \cdots + A_p G_p^t, \quad G_i = 1/B_i, \quad i = 1 \cdots p \text{ onde } \forall_{i \neq j} B_i \neq B_j \quad (2.19)$$

os B 'is são as raízes do polinómio autoregressivo ³. Para que $\{X_t\}$ represente um processo assintoticamente estacionário é necessário que $|G_i| < 1$ para $i = 1 \cdots p$, isto é, todas as raízes do polinómio autoregressivo têm que ficar fora círculo unitário, só assim $\lim_{n \rightarrow \infty} X_t^* = 0$. Também se chama solução complementar, parte transitória, por causa do seu decaimento para zero. Esse facto permite a aproximação do sistema em estado de estacionário (*steady state*), isto é, reduzido à solução estacionária,

$$X_t = [\phi(B)]^{-1} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (2.20)$$

que é uma solução particular da equação (2.18). Verifica-se a convergência em média quadrática, $\sum_{j=0}^{\infty} |\psi_j|$ converge e por consequência também $\sum_{j=0}^{\infty} \psi_j^2$. Quando o processo se admite estacionário obtêm-se a seguinte relação, [Box76b],

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} - \cdots - \phi_p \rho_{k-p}, \quad k \geq 1, \quad (2.21)$$

onde $\sigma^2 = \sigma_\varepsilon^2 / (1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \cdots - \phi_p \rho_p)$. A relação (2.21) mostra que a *FAC* dos processos autoregressivos de ordem p satisfaz a equação às diferenças homogénea associada a (2.18), que se pode escrever como, $\phi(B)\rho_k = 0$, $B\rho_k = \rho_{k-1}$ e tem solução geral,

$$\rho_k = C_1 G_1^k + C_2 G_2^k + \cdots + C_p G_p^k, \quad k \geq 0, \quad i = 1, \cdots, p \quad (2.22)$$

³Admitindo que o polinómio autoregressivo tem raízes (B_i) , $i = 1, 2, \cdots, p$ distintas

onde $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ são as raízes da equação característica, $\pi(B) = 0$. Já se viu que para haver estacionaridade é necessário que $|G_i| < 1$. Nesse caso duas situações podem ocorrer,

1. Uma raiz G_i é real e $C_i G_i^k$ em (2.19) decai geometricamente para zero à medida que k aumenta. O comportamento da *FAC* nestas situações é o de uma exponencial amortecida.
2. Um par de raízes G_i, G_j é complexo, nesse caso contribuem com um termo do tipo $d^k \sin(2\pi f k + F)$ para a *FAC* em (2.22), o que provoca uma curva sinusoidal amortecida.

Na generalidade das situações de um processo autoregressivo estacionário consistirá numa mistura de exponenciais amortecidas com sinusoidais do mesmo tipo. Se fizer $k = 1, 2, \dots, p$ em (2.21), obtém-se um sistema de equações lineares para $\phi_1, \phi_2, \dots, \phi_p$ em termos de $\rho_1, \rho_2, \dots, \rho_p$, isto é,

$$\begin{cases} \phi_1 & + \phi_2 \rho_1 & + \dots + \phi_p \rho_{p-1} & = \rho_1 \\ \phi_1 \rho_1 & + \phi_2 & + \dots + \phi_p \rho_{p-2} & = \rho_2 \\ \dots & \dots & \dots & \\ \phi_1 \rho_{p-1} & + \phi_2 \rho_{p-2} & + \dots + \phi_p & = \rho_p \end{cases} \quad (2.23)$$

O sistema é formado pelas chamadas equações de Yule-Walker que são de extrema importância na estimação. O sistema (2.23) pode escrever-se de forma equivalente,

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_2 & & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & & \rho_{p-3} \\ & & & \dots & 1 \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix} \quad (2.24)$$

A *FACP* obtém-se das equações (2.23), [Box76b], verificando-se que no caso desta função existe um decaimento brusco para zero, isto é $\phi_{kk} = 0$ para $k \geq p + 1$. A análise multivariada permite que se escreva,

$$X_{t+h} = \phi_{h1}X_{t+h-1} + \phi_{h2}X_{t+h-2} + \cdots + \phi_{hh}X_t + \xi_{t+h}, \quad (2.25)$$

é a própria natureza do processo $AR(p)$ que faz com que $\phi_{hh} = 0$ para $h \geq p + 1$, o valor do processo no instante t depende apenas dos $p - 1$ valores anteriores e do choque no mesmo instante.

Um processo $\{X_t\}$ diz-se $MA(q)$ quando é da forma,

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \quad (2.26)$$

ou $X_t = \theta_q(B)\varepsilon_t$ com, $\theta_q(B) = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q$. Onde $\theta_q(B)$ é o polinómio médias móveis de ordem q .

A representação de processos deste tipo é uma versão finita da representação do processo linear geral. Consiste, como se referiu anteriormente, em exprimir $\{X_t\}$ como função de um processo mais simples, $\{\varepsilon_t\}$.

A natureza da representação torna óbvia a estacionaridade do processo dado que em vez de uma sucessão infinita em $\{\varepsilon_t\}$ se tem uma soma dos valores do mesmo processo.

A *FACOV* é facilmente determinada, ver [Box76b],

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q) \sigma_\varepsilon^2 & k = 1, \dots, q \\ 0, & k > q \end{cases} \quad (2.27)$$

a variância é facilmente deduzida desta expressão, $\gamma_0 = \sigma^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_\varepsilon^2$.

Para a FAC tem-se,

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_q\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, & 0 < k \leq q \\ 0, & k > q, \end{cases} \quad (2.28)$$

Analogamente à *FACP* dos processos autoregressivos a *FAC* dos processos $MA(q)$ tem um decaimento brusco para zero. A invertibilidade, característica dos processos autoregressivos, nem sempre se verifica para processos de médias móveis. Consiste na possibilidade de transformação $MA(q) \rightarrow AR(\infty)$ mais concretamente,

$$[\theta_q(B)]^{-1} X_t = \varepsilon_t \quad (2.29)$$

O processo $\{X_t\}$ é invertível se $\theta_q(B) = 0$ tiver todas as raízes fora do círculo unitário. Não existe nestes processos uma expressão geral para a *FACP*, apenas se dirá que tem um comportamento igual ao da *FAC* dos processos $AR(p)$, ou seja, exponencial amortecida ou sinusoidal amortecida. Desde que um processo seja estacionário ou invertível ele admite uma representação médias móveis assim como uma representação autoregressiva.

Suponha que o processo que gera a sucessão cronológica é combinação das duas componentes, nesse caso e de acordo com o que foi referido, qualquer tentativa de representação na forma médias móveis ou autoregressiva resulta num processo com número de parâmetros excessivo. Um processo $\{X_t\}$ diz-se $ARMA(p, q)$ quando satisfaz a equação às diferenças estocástica,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_p \varepsilon_{t-p} \quad (2.30)$$

ou $\phi_p(B)X_t = \theta_q(B)\varepsilon_t$ onde $\phi_p(B)$ e $\theta_q(B)$ são os polinómios autoregressivo e médias

móveis de ordem p e q , respectivamente.

Para o processo ser estacionário, as raízes de $\phi_p(B) = 0$ têm que estar fora do círculo unitário, para ser invertível as raízes de $\theta_q(B) = 0$ têm que estar fora do círculo unitário. Na análise subsequente admitir-se-á que o processo $\{X_t\}$ é estacionário e invertível e também que as duas equações referidas não têm raízes comuns. A *FAC* de um processo $ARMA(p, q)$ obtém-se facilmente, escrevendo (2.30) na forma,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \quad (2.31)$$

multiplicando por X_{t-k} , $K > 0$, tomando valores esperados e por ser $E(\varepsilon_{t-j} X_{t-k}) = 0$, $k > j$, tem-se, $\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}$, $k \geq q+1$, obteve-se a equação às diferenças para a *FACOV*, dividindo por γ_0 obtém-se a *FAC*,

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}, \quad k \geq q+1. \quad (2.32)$$

A expressão (2.32) mostra que para valores de k maiores que $q+1$ a *FAC* dos processos $ARMA(p, q)$ comporta-se como a de um processo $AR(p)$, isto é, apresenta um decaimento gradual para zero. Como consequência o comportamento da *FAC* não depende dos parâmetros da componente médias móveis, ao contrário dos valores iniciais $\rho_1, \dots, \rho_q$, que dependem dos parâmetros de ambas as componentes e são indispensáveis para o cálculo da função de autoregressão para $k \geq q+1$.

A *FACP* de um processo $ARMA(p, q)$ tem um comportamento semelhante à dos processos médias móveis puros, isto é, decaimento gradual para zero do tipo exponencial e/ou sinusoidal amortecido.

Em [Mur93], reúnem-se as principais características dos processos referidos. Note-se que $AR(p) \equiv ARMA(p, 0)$ e $MA(q) \equiv ARMA(0, q)$.

Os modelos $ARMA(p, q)$ formam uma subclasse de processos estacionários lineares bastante rica. Muitas das sucessões estacionárias podem representar-se aproximada-

mente por um modelo desse tipo. As sucessões cronológicas podem apresentar, por vezes, uma componente sazonal ou componente periódica de período $S > 1$, ou seja, o seu comportamento exibe similaridades em intervalos de amplitude fixa, S . Estatisticamente verifica-se a existência de correlação significativa entre as observações espaçadas por um intervalo de tempo múltiplo de S . Os modelos lineares mais adequados para este tipo de séries são os modelos mistos autoregressivos e médias móveis estritamente sazonais $ARMA(P, Q)_S$. Os processos $ARMA(P, Q)_S$ são da forma ,

$$\Phi_P(B^S)X_t = \Theta_Q(B^S)\varepsilon_t \quad (2.33)$$

onde, $\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}$, e, $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}$ são os polinómios autoregressivos e médias móveis sazonais. As condições de estacionaridade e invertibilidade continuam as mesmas, as raízes de ambos os polinómios têm que estar fora do círculo unitário. Não se vai aqui fazer um estudo detalhado deste processo visto ser semelhante ao $ARMA(p, q)$. Refere-se como características mais importantes que a *FAC* e a *FACP* deste processo têm valor nulo para desfasamentos temporais⁴, não múltiplos de S e o comportamento nos desfasamentos múltiplos é semelhante ao processo misto não sazonal: decaimento sinusoidal e/ou exponencial amortecido para zero. Todas as características mencionadas até aqui podem reunir-se na classe de modelos lineares mais abrangente, a dos processos multiplicativos, $ARMA(p, q) \times (P, Q)_S$. Estes processos reúnem todas as componentes até agora estudadas, sazonais e não sazonais, autoregressiva e médias móveis, são da forma,

$$\phi_p(B)\Phi_P(B^S)X_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t, \quad (2.34)$$

⁴lags

Se todas as raízes dos polinómios envolvidos na equação estiverem fora do círculo unitário ficam garantidas a estacionaridade e a invertibilidade.

Nesta secção focou-se apenas a classe dos processos estacionários lineares, na secção seguinte vai focar-se uma não menos importante classe, a dos processos não estacionários lineares.

2.4 Processos lineares não estacionários

A maior parte das sucessões cronológicas observadas, especialmente as económicas e as associadas com fenómenos físicos são não estacionárias. Tomando como referência a classe mais importante de processos, os estacionários até 2ª ordem, um processo pode ser não estacionário por a média e/ou a variância serem funções do tempo não constantes. A não estacionaridade pode tomar as mais variadas formas. Uma das situações mais comuns é caso em que o processo observado, $\{X_t\}$, toma a forma,

$$X_t = \mu_t + Z_t, \quad (2.35)$$

onde μ_t é uma função determinística de t , não constante e $\{Z_t\}$ é um processo estacionário de média zero. Diz-se nesse caso que a sucessão gerada por este processo tem tendência. O modelo (2.35) é a soma de um resíduo estacionário com uma tendência. Permite que a média do processo, $E\{X_t\}$, varie no tempo arbitrariamente. Se μ_t for uma função linear a tendência é linear se for um polinómio de grau superior, a tendência diz-se quadrática, cúbica, etc. Se o processo for do tipo, $X_t = A \cos(\frac{2\pi t}{12}) + Z_t$ apresenta uma componente sazonal uma vez que a função média é periódica de período 12. A sucessão diz-se com componente sazonal. Estas componentes podem aparecer juntas, na forma aditiva, multiplicativa, mista combinando estas operações com o já mencionado processo estacionário de média nula, $\{Z_t\}$, como é o caso de,

$$X_t = (a + bt + ct^2) \left(A \sin \left(\frac{2\pi t}{6} \right) Z_t \right) \quad (2.36)$$

A sucessão pode ser estacionária em média, mas não em variância,

$$X_t \approx \begin{cases} N(0, \sigma_1^2), \sigma_1^2 = 1, t \leq t_1 \\ N(0, \sigma_2^2), \sigma_2^2 = 2, t > t_1 \end{cases} \quad (2.37)$$

A variância pode aumentar com a tendência, $X_t = \mu_t Z_t$.

Na prática, estes casos de não estacionaridade podem combinar-se e formar situações bastante complexas. Um processo estacionário em média não é necessariamente estacionário em variância, por outro lado, se o processo for não estacionário em média não o pode ser em variância e covariância.

As funções *FAC* e *FACP* exibem por vezes características interessantes e úteis que, por vezes, indicam o tipo de não estacionaridade. A maior parte da teoria de probabilidade de sucessões cronológicas diz respeito a sucessões estacionárias e por esta razão não raramente se tem que recorrer a transformações que estabilizam a média e/ou a variância convertendo uma sucessão cronológica não estacionária numa sucessão estacionária. O tema das transformações é o que vai ser tratado a seguir.

A estimação de uma média não constante é um problema complexo. Pode tratar-se de duas formas, com um modelo paramétrico, que não vai ser abordado nesta tese, ou usando filtragem de modo a conseguir uma estabilização. A estabilização da média é conseguida com operadores que pertencem à classe dos filtros lineares. Um filtro linear é da forma,

$$Y_t = \sum_{n=-\infty}^{\infty} \alpha_n X_{t-n} = \sum_{n=-\infty}^{\infty} \alpha_n B^n X_t = \alpha(B) X_t, \quad (2.38)$$

onde, $\alpha(B) = \sum_{n=-\infty}^{\infty} \alpha_n B^n$, $\sum_{n=-\infty}^{\infty} |\alpha_n| < \infty$.

Chama-se a $\{X_t\}$ é a sucessão *input* ou entrada, $\{Y_t\}$ a sucessão *output* ou sucessão transformada e α_n os pesos. Diz-se estar em presença de um sistema linear⁵. Em todas as sucessões associadas a fenómenos físicos a saída dependerá, como é óbvio, apenas dos valores presente e passados da entrada.

Um tipo especial de filtragem, que é particularmente útil para eliminar tendência de uma sucessão, é a diferenciação. Esta transformação foi introduzida por [Box76b].

Para dados não sazonais ou não periódicos uma diferenciação de 1ª ordem é geralmente suficiente para obter uma aparente estacionaridade. A série diferenciada, $\{Y_t\}$, é formada a partir da original, $\{X_t\}$, por,

$$Y_t = \nabla X_t = X_t - X_{t-1} = (1 - B)X_t, \quad (2.39)$$

Por vezes uma só diferenciação não chega necessitando-se de uma 2ª diferenciação simples resultando no total numa diferenciação de 2ª ordem,

$$Y_t = \nabla^2 X_t = \nabla(\nabla X_t) = (1 - B^2)X_t = X_t - 2X_{t-1} + X_{t-2}, \quad (2.40)$$

O operador pode generalizar-se e para qualquer inteiro $d \geq 0$, $\nabla^d X_t = (1 - B)^d X_t$, onde d é um inteiro a que se chama ordem das diferenças ou da diferenciação.

O uso da diferenciação tem efeitos secundários nefastos, acarreta um aumento da variância e perda de observações, por isso deve usar-se com parcimónia para evitar a sobrediferenciação que acontece quando o grau de diferenciação usado é superior ao necessário para tornar a sucessão estacionária.

Os processos não estacionários em média que se tornam estacionários por meio de diferenciação denominam-se processos não estacionários homogêneos. Quando a sucessão

⁵Existe um suporte teórico a este tema que não vai ser aqui desenvolvido.

observada tem uma componente sazonal, a diferenciação usual, usando termos consecutivos da série a transformar, não elimina esse efeito. O facto mais importante das sucessões sazonais de período S é que as observações separadas no tempo por S intervalos são similares. Pode então esperar-se que o operador de diferenciação sazonal,

$$\nabla_S X_t = X_t - X_{t-S} = (1 - B^S)X_t, \quad (2.41)$$

desempenhe um importante papel nestas sucessões, uma vez que anula o efeito sazonal. Este operador é também um caso particular do filtro (2.40).

Tal como o operador anterior, o operador sazonal é também susceptível de generalização e para qualquer inteiro, $D \geq 0$, $\nabla_S^D X_t = (1 - B^S)^D X_t$.

O inteiro D é a ordem das diferenças ou da diferenciação sazonal. é de salientar que o uso deste operador faz com que além da introdução de variância espúria já referida, se percam $D \times S$ observações na transformação. A diferenciação ocupa-se apenas da não estacionaridade em média. Se um processo for não estacionário em variância continuará nesse estado após diferenciação. Os autores em, [Box76a], introduziram uma transformação paramétrica para estabilizar a variância,

$$TX_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ X_t & \text{se } \lambda = 0 \end{cases}, \lambda \in [-1, 1]$$

Note-se que o aparecimento do logaritmo se deve ao facto de $\lim_{\lambda \rightarrow 0} \frac{X_t^\lambda - 1}{\lambda} = \ln X_t$.

Claro está que a transformação não se pode aplicar se os dados da sucessão em análise forem negativos, nesse caso soma-se aos termos da sucessão um número maior que o módulo do valor mínimo da sucessão, este procedimento não vai alterar o valor dos momentos de 2ª ordem da sucessão.

[Box76a] estimaram λ pelo método da máxima verosimilhança mas também se pode fazer por aproximações sucessivas, usando como critério a variância da sucessão trans-

formada.

Quando necessária esta transformação deve ser feita antes de qualquer diferenciação. Convém não esquecer que o cálculo de diferenças introduz variância nos dados e isso dificultaria ainda mais a escolha da transformação adequada para estabilizar a variância. O terceiro requisito para a estacionaridade de uma sucessão é a estabilização da *FACOV* e *FAC*. Não há processos formais para o fazer, regra geral, quando a sucessão observada é suficientemente grande para justificar este passo vale a pena dividi-la em duas partes e comparar a *FAC* estimada de cada uma de modo a verificar se o processo é estacionário em covariância. Um processo *ARMA*(p, q) é estacionário se as raízes do polinómio autoregressivo estiverem fora do círculo unitário, caso contrário, isto é, se houver raízes dentro do círculo, a sucessão tem um comportamento não estacionário explosivo. O único caso que fica por estudar é aquele em que o processo tem raízes no círculo unitário.

Um processo $\{X_t\}$ diz-se *ARIMA*(p, d, q) se,

$$\phi_p(B)(1 - B)^d X_t = \theta_0 + \theta_q(B)\varepsilon_t, \quad (2.42)$$

onde $\phi_p(B)$ e $\theta_q(B)$ são, respectivamente, os polinómios autoregressivo estacionário e médias móveis invertível. Portanto, $\{X_t\}$ é um processo não estacionário uma vez que o polinómio $\phi_p(B)(1 - B)^d$ tem d raízes no círculo unitário, no entanto se se diferenciar $\{X_t\}$ d vezes, obtém-se um processo estacionário invertível,

$$\nabla^d X_t = \phi_p(B)X_t = \theta_0 + \theta_q(B)\varepsilon_t \quad (2.43)$$

Também se pode dizer que $\{X_t\}$ é um processo não estacionário *ARMA*($p + d, q$) da forma,

$$\varphi_{p+d}(B)X_t = \theta_0 + \theta_q(B)\varepsilon_t \quad (2.44)$$

onde, $\varphi_{p+d}(B)X_t = \phi_p(B)(1-B)^d X_t$, é um polinómio autoregressivo de ordem $p+d$, com d raízes no círculo unitário. Sendo $Y_t = \nabla^d X_t$, o processo $\{X_t\}$ é definido pelas equações,

$$\phi_p(B)Y_t = \theta_0 + \theta_q(B)\varepsilon_t \quad (2.45)$$

e $Y_t = \nabla^d X_t$.

Uma forma alternativa de olhar para o processo quando $d \geq 1$ resulta da inversão de (2.46),

$$X_t = S^d Y_t \quad (2.46)$$

onde S é o operador soma infinito,

$$SX_t = \sum_{h=-\infty}^t X_h = (1 + B + B^2 + \dots)X_t = (1 - B)^{-1}X_t = \nabla^{-1}X_t \quad (2.47)$$

pelo que,

$$S = (1 - B)^{-1} = \nabla^{-1} \quad (2.48)$$

A equação (2.48) implica que o processo $\{X_t\}$ pode ser obtido somando (ou *integrando*) o processo estacionário $\{Y_t\}$ d vezes. Daí a sigla *ARIMA* em contraste com o

estacionário *ARMA*. O parâmetro θ_0 de (2.45) é consequência de uma generalização. Sem ele não se poderia representar sucessões com tendências estocásticas. No entanto se se quiser incluir uma função determinística do tempo, $f(t)$, por exemplo para modelar uma tendência determinística do tipo polinomial de grau d , é necessário incluir no modelo a constante θ_0 .

Quando $d = 0$ o processo $\{X_t\}$ é estacionário e θ_0 está directamente relacionado com a média, $\mu = E(X_t)$,

$$\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p) \quad (2.49)$$

neste caso θ_0 é um parâmetro dependente da média, $\mu = E(X_t)$. Se $\mu = 0$, hipótese geralmente admitida, pode omitir-se θ_0 .

Quando $d \geq 1$, a presença de θ_0 só é justificada se houver uma tendência determinística polinomial de ordem superior a d , sendo θ_0 o coeficiente do termo de maior grau do polinómio de tendência determinística. Em muitas aplicações não existe uma razão física para a existência de uma tendência determinística, a média de $\{X_t\}$ pode assumir-se nula a menos que tal suposição se mostre contrária ao indicado pelos dados. Se uma componente cíclica com período S se encontra presente nos dados é de se esperar que exista uma relação forte entre $\cdots, X_{t-2S}, X_{t-S}, X_t$ e $\cdots, \varepsilon_{t-2S}, \varepsilon_{t-S}, \varepsilon_t$ e, logicamente, esta relação deve ser tida em conta no modelo. Há, então, que generalizar o modelo, (2.42), tendo em consideração o facto de que aqueles modelos se poderem alargar de modo a cobrir processos estacionarizáveis mediante diferenciação. Assim, $\{X_t\}$, diz-se um processo multiplicativo integrado sazonal, quando [Mur93],

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B)^D X_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (2.50)$$

onde todos os polinómios intervenientes têm as propriedades de estacionaridade e invertibilidade. O modelo (2.50) é designado abreviadamente por $SARIMA(p, d, q)X(P, D, Q)_S$.

Por exemplo, o processo,

$$\phi_1(B)(1-B)^2(1-B^{12})X_t = \theta_2(B)(1-\Theta B^{12})\varepsilon_t \quad (2.51)$$

é um $SARIMA(1, 2, 2)X(0, 1, 1)_{12}$. O modelo (2.51) é ainda susceptível de generalização numa forma com m polinómios autoregressivos sazonais e n polinómios médias móveis sazonais, que é capaz de descrever uma situação bastante complexa em que existam componentes sazonais ou periódicas com diferentes períodos [Mur93],

$$\begin{aligned} \phi_p(B) \left\{ \prod_{i=1}^m \Phi_{P_i}(B)^{S_i} \right\} (1-B)^d \left\{ \prod_{i=1}^{m^*} (1-B^{S_i})^{D_i} \right\} X_t = \\ = \theta_0 + \theta_q(B) \left\{ \prod_{i=1}^n \Theta_{Q_i}(B^{S_i}) \right\} \varepsilon_t \end{aligned} \quad (2.52)$$

Como se pode calcular a identificação, a partir da *FAC* e *FACP* de tais modelos, não é tarefa fácil, no modelo (2.51) são 7 parâmetros, no (2.52) poderão ser muitos mais. Felizmente, na prática tem-se quase sempre, $0 \leq d, D \leq 2$ e os parâmetros de ordem dos polinómios sazonais e não sazonais raramente ultrapassam 2.

2.5 A correlação e o espectro

Nesta secção irá fazer-se referência às ferramentas disponíveis para caracterizar processos e séries do ponto de vista estocástico que para os processos lineares são absolutas, isto é, fazem uso de toda a informação presente nos dados.

Uma dessas ferramentas é a função de densidade espectral, a transformada de Fourier da função de autocovariância ou autocorrelação (caso padronizado). A análise no domínio da frequência, embora relacionada com a análise no domínio do tempo tem uma perspectiva diferente, em vez de se concentrar na estrutura de dependência

temporal investiga em que medida diferentes frequências/componentes periódicas contribuem para a variabilidade, embora, do ponto de vista da teoria da informação sejam equivalentes.

Em que condições se fala de uma densidade espectral? Seguindo [Gon03], o teorema de Herglotz-Bochner está na base do estudo do espectro,

Teorema 2.5.1 *Uma função $f : \mathbb{Z} \rightarrow \mathbb{C}$ é semidefinida positiva, ou seja verifica,*

$$\forall n \in \mathbb{N}, \forall u_1, \dots, u_n \in \mathbb{C}^n, \sum_{i=1}^n \sum_{j=1}^n u_i \bar{u}_j f(i-j) \geq 0 \quad (2.53)$$

se e só se puder escrever na forma, $f(h) = \int_{-\pi}^{\pi} e^{i\omega h} d\mu(\omega)$. Onde μ é uma medida positiva limitada sobre $[-\pi, \pi]$. A medida μ é única e denomina-se medida espectral associada a f .

Como se referiu anteriormente, as funções de autocorrelação e autocovariância estão nas condições do teorema de Herglotz-Bochner e assim pode associar-se, de forma única, uma medida positiva e limitada em $[-\pi, \pi]$. Sendo $\{X_t\}_{t \in \mathbb{Z}}$ um processo real estacionário, chama-se medida espectral de $\{X_t\}$ à medida μ associada à *FACOV*. Atendendo às propriedades de simetria, e ao facto de $\gamma(\cdot)$ ser real, pode escrever-se, $\gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} d\mu(\omega)$, $h \in \mathbb{Z}$.

Estabelece-se aqui uma equivalência entre uma função de autocovariância de um processo real e uma medida positiva simétrica e limitada em $[-\pi, \pi]$.

Se a medida for absolutamente contínua relativamente à medida de Lebesgue em $[-\pi, \pi]$ pode enunciar-se o seguinte teorema,

Teorema 2.5.2 *Uma condição necessária e suficiente para que a função definida por, $\{\rho(r) = 0; r = 0 \pm 1, \pm 2, \dots\}$ seja a função de autocorrelação de um processo estacionário de parâmetro discreto, $\{X_t; t = 0, \pm 1, \pm 2, \dots\}$, é que exista uma função $F(\omega)$ com as propriedades de uma função de distribuição no intervalo $(-\pi, \pi)$ (ou seja $F(-\pi) = 0$, $F(\pi) = 1$, e $F(\omega)$ não-decrescente) tal que,*

$$\rho(r) = \int_{-\pi}^{\pi} e^{(i\omega r)} dF(\omega), \quad r = 0 \pm 1, \pm 2, \dots \quad (2.54)$$

Diz-se, neste caso, que $\{X_t\}$ tem um espectro absolutamente contínuo e [Pri81] mostra que a sucessão admite a seguinte representação linear,

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k} \quad (2.55)$$

Impondo a condição de Paley-Wiener, [Rao84], $\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty$, estão verificadas as condições do teorema da representação espectral. Desta forma existe uma única sucessão $\{h_t\}_{t=0}^{\infty}$ com h_0 real e positivo tal que, [Pri81],

1. $\sum_{u=0}^{\infty} |h_u|^2 < \infty$
2. $H(z) = \sum_{u=0}^{\infty} h_u z^u \neq 0$
3. $f(\omega) \propto |H(e^{-i\omega})|^2$

A expressão (2.55) pode simplificar-se podendo considerar-se apenas os valores passados e presente,

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \quad (2.56)$$

A transformada de Fourier, F é também chamada de espectro integrado. Se a série correspondente convergir absolutamente então F tem derivada contínua, f dada por,

$$f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \rho(r) e^{-i\omega r} \quad -\pi \leq \omega \leq \pi \quad (2.57)$$

A função f é chamada função de densidade espectral (normalizada).

Usando as propriedades da transformada de Fourier e as propriedades da $FAC/FACOV$, pode escrever-se (2.57) de forma alternativa,

$$f(\omega) = \frac{1}{2\pi} \left[1 + 2 \sum_{r=-\infty}^{\infty} \rho(r) \cos(\omega r) \right] \quad -\pi \leq \omega \leq \pi \quad (2.58)$$

A função de densidade espectral goza de importantes propriedades,

- $f(\omega)$ é real, contínua, não negativa.
- $f(\omega)$ é periódica - com período 2π .
- $f(\omega)$ é par ou simétrica pelo que a representação gráfica apenas se faz para um dos eixos.

Sob certas condições, [Gon03], o conhecimento da função de autocovariância ou autocorrelação é equivalente ao conhecimento do espectro absoluto ou normalizado e vice-versa. Usando (2.54) obtém-se,

$$\gamma_0 = V(X_t) = \int_{-\pi}^{\pi} f(\omega) d\omega \quad (2.59)$$

esta última expressão ilustra bem a interpretação da variância que se faz no chamado espaço das frequências. A quantidade, $f(\omega)d\omega$ é a parte da variância causada por frequências no intervalo $(\omega, \omega+d\omega)$, assim, as frequências correspondentes aos máximos da função de densidade espectral tem contribuição preponderante para a variância do processo.

A expressão da densidade espectral para os processos lineares é mais facilmente obtida usando uma representação baseada na função geradora das autocovariâncias, usando as expressões (2.57) e (2.58) obtém-se,

$$f(\omega) = \frac{1}{2\pi} \gamma(e^{-i\omega}) \quad (2.60)$$

para um exemplo de um processo estocástico simples, o ruído branco vem,

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 & , \quad k = 0 \\ 0 & , \quad k \neq 0 \end{cases} \quad (2.61)$$

o que leva a, $f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}$, o que justifica o adjetivo branco usado para este processo em analogia com o que passa com o espectro da luz branca.

A obtenção da função de densidade espectral para as três classes de processos lineares passa pela utilização de filtros lineares já abordados anteriormente. A importância dos filtros lineares invariantes para a obtenção de expressões para a densidade espectral reside no facto de estes serem a generalização da dependência linear entre dois processos. Assim para a situação descrita em (2.56) e não descurando os problemas de existência, tem-se para a função de autocovariância,

$$\begin{aligned} \gamma_Y(k) = E\{Y_t Y_{t+k}\} &= E\left\{\left(\sum_{m=-\infty}^{\infty} \alpha_m X_{t-m}\right) \left(\sum_{n=-\infty}^{\infty} \alpha_n X_{t+k-n}\right)\right\} \\ &= \sum_{m=-\infty}^{\infty} \alpha_m \sum_{n=-\infty}^{\infty} \alpha_n \gamma_X(k+m-n) \end{aligned} \quad (2.62)$$

verificando-se,

$$\sum_{k=-\infty}^{\infty} |\gamma_X(k)| < \infty \Rightarrow \sum_{k=-\infty}^{\infty} |\gamma_Y(k)| < \infty \quad (2.63)$$

[Mur93] então existe a densidade espectral de $\{Y_t\}$.

$$\begin{aligned}
f_Y(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Y(k) e^{-i\omega k} \\
&= \sum_{m=-\infty}^{\infty} \alpha_m e^{i\omega m} \sum_{n=-\infty}^{\infty} e^{-i\omega n} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k+m-n) e^{-i\omega(k+m-n)} \quad (2.64)
\end{aligned}$$

ou seja,

$$f_Y(\omega) = \alpha(e^{i\omega}) \overline{\alpha(e^{i\omega})} f_X(\omega) = |\alpha(e^{i\omega})|^2 f_X(\omega), \quad -\pi \leq \omega \leq \pi \quad (2.65)$$

onde $\alpha(e^{i\omega}) = \sum_{m=-\infty}^{\infty} \alpha_m e^{i\omega m}$. A função,

$$|\alpha(e^{i\omega})|^2 = \alpha(e^{i\omega}) \overline{\alpha(e^{i\omega})} \quad (2.66)$$

chama-se função de transferência do processo $\{Y_t\}$. A expressão mostra que o valor da densidade espectral da sucessão $\{Y_t\}$ para uma determinada frequência depende apenas do valor da densidade espectral da sucessão filtrada na mesma frequência.

Os processos $MA(q)$ abordados na secção anterior podem considerar-se uma filtragem,

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} |\theta_q(e^{i\omega})|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \theta_q(e^{i\omega}) \overline{\theta_q(e^{i\omega})} \quad (2.67)$$

da mesma forma para a densidade espectral de um processo $AR(p)$,

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{1}{\phi_p(e^{i\omega})} \right|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \phi_p(e^{i\omega}) \overline{\phi_p(e^{i\omega})} \quad (2.68)$$

e para um processo $ARMA(p, q)$,

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{\theta_q(e^{i\omega})}{\phi_p(e^{i\omega})} \right|^2 = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\theta_q(e^{i\omega}) \overline{\theta_q(e^{i\omega})}}{\phi_p(e^{i\omega}) \overline{\phi_p(e^{i\omega})}} \quad (2.69)$$

Como se pode verificar, para o caso do processo $ARMA(p, q)$, a função de densidade espectral é função racional da variável $e^{i\omega}$. Esta é uma propriedade caracterizadora daqueles processos. A função de densidade espectral dos processos com componentes sazonais obtém-se de modo análogo.

2.6 Limitações dos processos lineares

Nesta secção vão evidenciar-se algumas características que os processos lineares pela sua natureza linear e gaussiana não podem exibir.

Tong, [Ton90], refere várias características que se podem procurar numa série observada e que são um indício de falta de linearidade. Essas ferramentas/técnicas podem constituir uma etapa de identificação de características embora não se pretenda que funcione como uma espécie de algoritmo a seguir. A construção da classe de modelos lineares tem como consequência um conjunto de propriedades. Seja $\{\varepsilon_t\}$ uma sucessão i.i.d. com distribuição $N(0, \sigma_\varepsilon^2)$ e seja $\{X_t\}$ o processo $ARMA(p, q)$ construído a partir daquele de expressão (2.30) com média, $\mu_X = 0$ e variância, $\sigma_X^2 = \int_{-\pi}^{\pi} \frac{\sigma_\varepsilon^2}{2\pi} \phi_p(e^{i\omega}) \overline{\phi_p(e^{i\omega})}$. O processo $\{X_t\}$ é estacionário e tem lei temporal normal, isto é, forma uma sucessão gaussiana, admite uma representação linear (2.9) invertível. Um processo ou modelo linear gaussiano fica completamente especificado conhecida a sua média, μ_X e função de autocovariância, γ_k ou a função de densidade espectral, $f(\omega)$. O autor em [Pri81] refere que o grau de generalidade conferido pelo modelo $ARMA$ na classe dos processos estacionários é, idêntico ao grau de generalidade atingido pelas funções racionais na classe das funções continuas.

- A classe de processos $ARMA$ é normal e por conseguinte simétrica, não são o modelo mais indicado para dados com fortemente assimétricos. Como são exemplo os dados do tipo hidrológico.
- Os diagramas do tipo (X_t, X_{t-j}) , $j = \pm 1, \pm 2, \dots, \pm p$, ligando por segmentos de

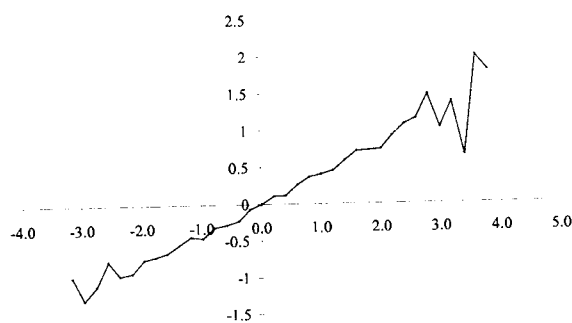


Figura 2.1: Curva de regressão amostral de X_t sobre X_{t-1} para uma série $ARMA(2, 1)$.

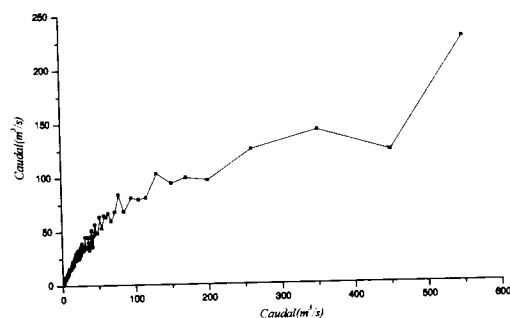


Figura 2.2: Curva de regressão amostral de X_t sobre X_{t-5} para a série dos caudais do Paiva.

recta os pontos adjacentes, (X_{t+1}, X_{t+1-j}) são os análogos em tempo discreto dos diagramas de fase. Permitem visualizar estruturas típicas de sistemas não lineares como órbitas periódicas entre outras que não podem surgir, por exemplo, em séries lineares gaussianas.

- Características nos dados como a existência de valores muito afastados do centro, os *picos* súbitos ou variações de grande amplitude, torna-os pouco interessantes para a modelação linear uma vez que nestes modelos a probabilidade de ocorrência é extremamente baixa.

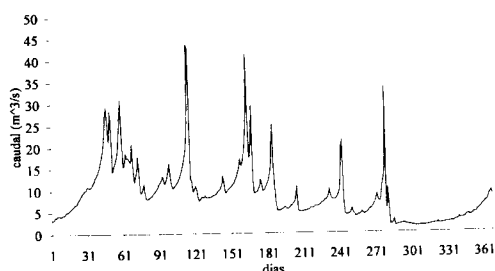


Figura 2.3: Série revertida dos caudais diários do rio Paiva.

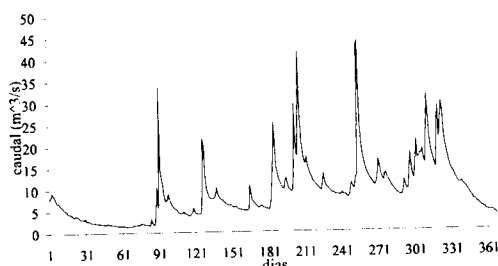


Figura 2.4: Série dos caudais do rio Paiva.

- Devido à normalidade das distribuições conjuntas nos modelos $ARMA$ as curvas de regressão do tipo, $E(X_t/X_{t-j})$, $j \in \mathbb{Z}$ são rectas. A ligação funcional entre os dois valores está afastada da linearidade se um valor de $\gamma_0 \rho_j^2 / V(E(X_t|X_{t-j}))$ estiver afastado de 1. Por exemplo, se os dados revelarem uma forte periodicidade e o valor da FAC se mantiver baixo então existe uma forte associação não

linear. As figuras (2.1) e (2.2) ilustram esse facto para um processo linear e um não linear.

- Em muitas situações, especialmente para fenómenos físicos, a alteração de valor de um parâmetro pode alterar a estrutura de dependência das variáveis do sistema. Nos modelos lineares a estrutura de dependência temporal está fixa.
- Nos modelos lineares a estrutura probabilística é simétrica como se referiu, $\gamma_k = \gamma_{-k}$. Dessa forma se a série em análise for revertida no tempo, $\{X_{-t}\}$ deve-se obter uma série de estrutura semelhante. Os gráficos (2.3) e (2.4) mostram a falta de reversibilidade da série dos caudais diários do rio Paiva. A normalidade implica que os cumulantes de ordem igual ou superior à segunda sejam nulos. Rao e Gabr, [Rao84] desenvolveram um teste estatístico baseado no bispectro, a transformada de Fourier do cumulante de 3ª ordem [Pri81]. A classe de modelos *ARMA* gaussiana tem bispectro nulo.
- A detecção de uma componente não estocástica ou determinística pode ser levada a cabo usando uma versão discreta de uma ferramenta introduzida por Poincaré, os diagramas de fase.

A este conjunto de ferramentas podem adicionar-se outras mais que surgiram no contexto da análise de séries não lineares observadas. A estas far-se-ão referência mais à frente. A existência nos dados de características fora do alcance de trajectórias de modelos lineares gaussianos levaram tanto os estatísticos como os físicos experimentais a procurar alternativas. Como referem [Cha01], foi Henri Poincaré que, no virar do século XX, reconheceu que a dependência sensível de condições iniciais era uma fonte fundamental de aleatoriedade. Foi Lorenz que descobriu o primeiro sistema em tempo contínuo exibindo aquela propriedade [Lor63].

2.7 Testes de linearidade e gaussianidade

Uma característica que distingue as sucessões cronológicas de outros tipos de dados é a existência de correlações temporais. O objectivo central em análise de séries temporais é conseguir um modelo probabilístico ou semi-probabilístico que explique o comportamento da correlação entre os seus valores em diferentes instantes. Um modelo nessas condições deve ter como resíduos de um ajuste apenas ruído branco. A metodologia de Box-Jenkins baseia-se em duas hipóteses. A primeira é a de a sucessão ser ou poder tornar-se estacionária por transformação e a segunda, a de que a sucessão se adapta a um modelo linear. A estacionaridade, que atrás se descreveu, significa que as propriedades estatísticas da sucessão mantêm-se constantes ao longo do tempo, a linearidade significa que os valores da série podem ser representados como combinação linear dos valores presente e passados de um ruído branco. Como referem [Mur93], ambas as situações correspondem a idealizações matemáticas, na prática servem como aproximações à realidade.

Se existir uma dependência de todos os valores passados do processo, o problema da modelação linear implica igualmente que perante uma sucessão cronológica de tamanho finito seja impossível escolher entre o conjunto de todas as funções, aquela que corresponde à forma funcional exacta da dependência,

$$h(\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots) = \varepsilon_t \quad (2.70)$$

A modelação linear consiste em considerar, apenas, o conjunto de funções da forma,

$$\sum_{u=-\infty}^{\infty} h_u X_{t-u} = e_t \quad (2.71)$$

sendo $\{h_u\}$ uma sucessão e $\{e_t\}$ um ruído branco independente. Estas funções denominam-

se correntemente modelos lineares. Em modelação, apenas se considera a situação de dependência do passado, $h_u = 0$, $u < 0$, diz-se então que o modelo é não antecipativo. A tradução deste modelo linear de representação finita em termos do número de parâmetros foi levada a cabo na secção anterior.

A gaussianidade e a linearidade são conceitos que se intersectam na teoria das sucessões cronológicas. É um resultado bem conhecido que se a gaussianidade estiver presente a inexistência de correlação é sinónimo de independência. Um processo com espectro absolutamente contínuo admite uma representação linear,

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \quad (2.72)$$

onde $\{\varepsilon_t\}$ é um processo não correlacionado. Se impusermos à densidade espectral, $f(\omega)$, a condição de Paley-Wiener, $\int_{-\pi}^{\pi} \ln(f(\omega)) d\omega > -\infty$, o processo $\{X_t\}$ pode agora ser representado como uma combinação linear apenas do presente e do passado de um processo não correlacionado, $\{\varepsilon_t\}$,

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \quad (2.73)$$

o que não quer dizer que o processo $\{X_t\}$ possa ser descrito pelo modelo linear,

$$X_t = \sum_{u=0}^{\infty} h_u e_{t-u} \quad (2.74)$$

De facto, o processo $\{\varepsilon_t\}$ é não correlacionado e o $\{e_t\}$ é independente. Se $\{\varepsilon_t\}$ for gaussiano então não existe diferença entre as duas representações. Assim, se um processo for gaussiano a autocorrelação e o espectro descrevem totalmente o processo, se tal não se verificar, terá que se olhar para momentos e respectivas transformadas de Fourier, os espectros de ordem superior ou poliespectro.

Suponha-se que o processo $\{X_t\}$ é estacionário até à ordem m e seja $C(s_1, s_2, \dots, s_{m-1})$ o coeficiente de $(z_1, z_2, \dots, z_m)$ no desenvolvimento da função geradora de cumulantes,

$$K(z_1, z_2, \dots, z_m) = \ln E\{e^{z_1 v X_t + z_2 X_{t+s_1} + \dots + z_m X_{t+s_{m-1}}}\} \quad (2.75)$$

tal como acontece para segunda ordem, a estacionaridade até ordem m faz com que $C(s_1, s_2, \dots, s_{m-1})$ não dependa de t .

O poliespectro de ordem m do processo estocástico $\{X_t\}$ é definido por,

$$h(\omega_1, \dots, \omega_{m-1}) = \sum_{s_1=-\infty}^{\infty} \dots \sum_{s_{m-1}=-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{m-1} C(s_1, \dots, s_{m-1}) e^{-i \sum_{i=1}^{m-1} \omega_i} \quad (2.76)$$

O cumulante $C(s_1, s_2, \dots, s_{m-1})$ é um polinómio de ordem m em momentos de ordem não superiores a m . As relações entre momentos e cumulantes (ou semi-invariantes) foram deduzidas por Leonov e Shiraev.

O cumulante de segunda ordem, $C(s_1)$ é a covariância entre X_t e X_{t+s_1} pelo que o poliespectro de 2ª ordem é o espectro convencional. O cumulante de 3ª ordem pode escrever-se como,

$$\begin{aligned} C(s_1, s_2) &= \mu_3(s_1, s_2) - \{\mu_2(s_1) + \mu_2(s_2) + \mu_2(s_2 - s_1)\}\mu + 2\mu^3 \\ &= E(X_t - \mu)(X_{t+s_1} - \mu)(X_{t+s_2} - \mu) \end{aligned} \quad (2.77)$$

Os segundos e terceiros cumulantes são iguais aos momentos centrados correspondentes. O poliespectro de 3ª ordem chama-se correntemente bispectro.

O facto de a distribuição normal multivariada ter todos os cumulantes de ordem superior à segunda iguais a zero vem que, por consequência, todo o poliespectro correspondente é identicamente nulo, constituindo assim uma base para a construção de testes de gaussianidade em processos estocásticos. Um afastamento da normalidade revela-se, geralmente, nos cumulantes de ordem superior e no bispectro, ou seja, um

processo com cumulante de 3ª ordem e, por consequência bispectro, não identicamente nulos não é, certamente, gaussiano. O recíproco não é verdadeiro, basta considerar um processo que admita a representação (2.74) onde $\{e_t\}$, que é i.i.d., tenha uma distribuição simétrica vem que o bispectro de $\{X_t\}$ é nulo.

Mostra-se, [Rao84], que o bispectro do modelo linear (ruído branco independente) se pode escrever como,

$$f(\omega_1, \omega_2) = \frac{\mu_3}{2\pi^2} H(\omega_1) H(\omega_2) H(\omega_1 + \omega_2) \quad (2.78)$$

onde $H(\omega) = \sum_{u=0}^{\infty} h_u e^{-i\omega u}$. Sendo que o espectro se pode representar como,

$$f(\omega) = \frac{\sigma_e^2}{2\pi} H(-\omega) H(\omega) \quad (2.79)$$

definindo a função,

$$X(\omega_1, \omega_2) = \frac{|f(\omega_1, \omega_2)|^2}{f(\omega_1) f(\omega_2) f(\omega_1 + \omega_2)} \quad (2.80)$$

obtém-se para os modelos lineares,

$$X(\omega_1, \omega_2) = \frac{\mu_3^2}{2\pi\sigma_e^6}, \quad \forall \omega_1, \omega_2 \quad (2.81)$$

Qualquer processo que admita uma representação em modelo linear apresenta uma função *rácio* constante. Se a sucessão $\{e_t\}$ tiver distribuição normal então o rácio tem valor nulo, uma vez que $\mu_3 = 0$, caso contrário, o rácio tem valor constante. O rácio pode servir como estatística para testar a normalidade ou linearidade de uma sucessão, para isso é necessário obter estimadores consistentes para o espectro e bispectro. A estimação centrada e consistente do espectro e do bispectro (ou poliespectros de ordem superior), a partir de uma série numérica, implica o uso de alisamento sobre as frequências [Pri81], [Rao84]. O alisamento ou combinação linear ponderada é aplicado ao estimador suficiente do espectro obtido usando a expressão (2.79) substituindo a

autocovariância teórica pela amostral,

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{N-1} \lambda(k) \hat{\gamma}_k e^{-i\omega k} \quad (2.82)$$

onde $\lambda(k)$ que se denomina janela retardada, [Mur93], é uma função real e par com várias propriedades [Pri81].

No caso bidimensional existem janelas bidimensionais que tornam o estimador do bispectro consistente. Assim a expressão para o estimador consistente do bispectro é,

$$\hat{f}(\omega_1, \omega_2) = \left(\frac{1}{2\pi} \right)^2 \sum_{s_1=-(N-1)}^{N-1} \sum_{s_2=-(N-1)}^{N-1} \lambda(s_1, s_2) \hat{\mu}_k e^{-i(\omega_1 s_1 + \omega_2 s_2)} \quad (2.83)$$

onde $\lambda(s_1, s_2)$ é uma janela retardada bidimensional que, tal como a janela unidimensional, tem que satisfazer um conjunto de requisitos. Em [Pri81] e [Rao84] encontra-se uma excelente exposição de várias janelas propostas na literatura, estes últimos autores propõem ainda uma janela óptima no sentido dos mínimos quadrados, a janela de Rao-Gabr.

O teste de gaussianidade de Rao-Gabr tem como hipótese nula que o bispectro é igual a zero para todas as frequências ,

$$H_0 : h(\omega_1, \omega_2) = 0 ; H_1 : h(\omega_1, \omega_2) \neq 0$$

A rejeição de H_0 implica que a sucessão é não gaussiana embora isso não signifique que o processo seja gaussiano basta considerar, por exemplo, um processo com representação linear para o qual o 3º cumulante seja nulo. O ensaio do teste implica a estimação do bispectro numa região específica do plano uma vez que a existem relações de simetria que limitam a área em estudo. Procuram-se então estimativas do bispectro que estejam não correlacionadas o que faz com que o teste, apesar de não paramétrico, envolva diversos parâmetros relativos à ponderação necessária à estimação consistente do bispectro e às distâncias entre as estimativas, [Rao84],[Pri88].

$$D = \begin{bmatrix} \xi_{11} & \cdots & \xi_{1n} \\ \vdots & \ddots & \vdots \\ \xi_{P1} & \cdots & \xi_{Pn} \end{bmatrix} = [\xi_{(1)}, \dots, \xi_{(n)}] \quad (2.84)$$

Constrói-se, então, uma matriz (2.84) cujos elementos são as estimativas do bispectro calculadas sobre a grelha com os valores adequados dos parâmetros. Trata-se portanto de testar a média de uma distribuição de matriz de covariância desconhecida pelo que faz todo o sentido usar a estatística T^2 de Hotteling⁶ uma vez que para um número N de observações suficientemente grande, cada $\xi_{(i)}$ tem aproximadamente distribuição normal complexa com média η e a matriz covariância, Σ_ξ . Nestas condições toma-se como estatística do teste, $F_1 = \{2(n - P)/2P\}T^2$ onde, $T^2 = n\hat{\eta}^* A^{-1} \hat{\eta}$ e $A = n\hat{\Sigma}_\xi$. Os estimadores da média, $\hat{\eta}$ e variância, $\hat{\Sigma}$ são os da máxima verosimilhança. A hipótese nula do teste é $H_0 : \eta = 0$ $T^2 = 0$ e a alternativa, $H_1 : T^2 > 0$. A estatística F_1 tem distribuição *F-Snedcor* com $(2P, 2(n - P))$ graus de liberdade, a hipótese H_0 é rejeitada se F_1 for superior a um determinado nível. Rao e Gabr dividiram o teste em duas etapas, a primeira usando o interior da região e a segunda usando a origem e a fronteira. A segunda etapa apenas tem lugar se, na primeira, a hipótese de nulidade não for rejeitada. Na região de fronteira os valores do bispectro são reais pelo que se recorre à versão real (original) do teste T^2 de Hotelling.

O teste de linearidade usa como estatística o rácio (2.80) estimado na mesma grelha de frequências. Com os valores estimados de $X_{ij} = X(\omega_i, \omega_j)$, constrói-se, n vectores de dimensão P , $Y_i = (y_1, \dots, y_P)$. A hipótese de nulidade aqui é $H_0 : X_t$ é linear que caso se verifique, implica que os valores médios de $\{Y_i\}$ contenham elementos idênticos, ou seja, o equivalente a testar a igualdade das médias, $\mu_1 = \mu_2 = \dots, \mu_P$, o que coincide com o teste de simetria em análise multivariada. Sendo,

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, S_Y = \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})^T, \bar{\beta} = B\bar{Y} \quad (2.85)$$

⁶numa versão complexa

onde B é matriz $Q \times P$ com forma,

$$B = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

A nova variável aleatória tem média nula e matriz de covariância $B\Sigma_Y B'$. Sob a hipótese nula pode demonstrar-se que a estatística do teste,

$$F_2 = \left[\frac{(n-Q)}{Q} \right] t^2, T^2 = n\bar{\beta}S^{-1}\bar{\beta} \quad (2.86)$$

tem distribuição *F-Snedcor* com $(Q, n-Q)$ graus de liberdade, sendo $Q = P - 1$ e $S = BS_Y B'$.

[Cha86] denotaram grandes dificuldades na *afinação* dos parâmetros de alisamento nos testes de Rao e Gabr, pequenas alterações nos parâmetros podem resultar em estimativas muito diferentes para o rácio (2.80).

Nos testes de gaussianidade e linearidade de Rao e Gabr não se faz uso da variância assintótica do bispectro amostral, este procedimento é uma opção razoável se as amostras forem pequenas. Assim, [Hin82] propôs uma modificação aos testes de Rao e Gabr fazendo uso das propriedades assintóticas do bispectro amostral. Em vez de usar funções de alisamento tendo em vista obter consistência no estimador, [Hin82] propôs fazerem-se médias do bispectro amostral numa grelha apropriada de frequências.

Seguindo [Hin82], supondo que o processo $\{X_t\}$ admite uma representação linear, (2.55), verifica-se, (2.78), definindo-se a sucessão, $\omega_n = 2\pi/N$ para $n = 0, 1, \dots, N-1$. Para cada par de inteiros j, k defina-se,

$$F(j, k) = \frac{1}{N} X(\omega_j) X(\omega_k) X^*(\omega_{j+k}) \quad (2.87)$$

onde $X(\omega_k) = \sum_{t=0}^{N-1} X(t) \exp(-i\omega_k t)$. Devido às relações de simetria do bispectro, [Rao84], o domínio de interesse para o estudo de $F(j, k)$ é uma grelha triangular,

$D = 0 < j \leq N/2$ onde $0 < k \leq j$ e $2j + k \leq N$.

Fazendo $X(0) = 0$, ou alternativamente, supondo média nula, vem que $F(j, 0) = F(0, k) \equiv 0$. Usando um resultado de [Bri75] vem que,

$$\begin{aligned} E[F(n-3j)j, k] &= f(\omega_j, \omega_k) + O(N^{-1})E|F(j, k) - f(\omega_1, \omega_2)|^2 \\ &= Nf(\omega_j)f(\omega_k) + [1 + \delta(j-k) + \\ &\quad + \delta(N-2j-k) + 4\delta(N-3j)\delta(N-3k)] + O(1) \end{aligned} \quad (2.88)$$

onde $\delta(k) = 0$, $k > 0$ e $\delta(0) = 1$. Usando todos os pontos da grelha no triângulo D . Seja $M = N^c$ para $1/2 < c < 1$. Para a grelha $L = \{((2m-1)M/2, (2n-1)M/2)\}_{m=1}^n$ e $m \leq N/2M - n/2 + 3/4$ em D obtém-se consistência ao tomar médias de $F(j, k)$ em quadrados de M^2 pontos centrados em $((2m-1)M/2, (2n-1)M/2)$. [Hin82] propõe como estimador do bispectro,

$$\hat{f}(m, n) = \frac{1}{M^2} \sum_{j=(m-1)M}^{mM-1} \sum_{k=(n-1)M}^{nM-1} F(j, k) \quad (2.89)$$

A variância de $\hat{f}(m, n)$ depende do tamanho e número de quadrados usados no cálculo da média.

[Hin82] mostra que os estimadores $\hat{f}(m, n)$ são assintoticamente independentes e a distribuição de,

$$\begin{aligned} X_{m,n} &= \frac{Q_{m,n}}{N^{4c-1}}^{-1/2} \left[f\left(2\pi(2m-1)\frac{M}{2N}\right) \right. \\ &\quad \left. \left[f\left(\frac{2\pi(2n-1)M}{2N}\right) f\left(\frac{2\pi(m+n-1)M}{N}\right) \right]^{-1/2} \hat{f}(m, n) \right] \end{aligned} \quad (2.90)$$

é normal complexa com variância unitária. Aqui $Q_{m,n}$ corresponde ao número de pontos no quadrado em D (mas não na fronteira onde $j = k$ ou $2j + k = N$) mais 2 vezes o número dessas fronteiras. Por exemplo, se o quadrado estiver contido em D , então $Q_{m,n} = M^2$.

Consequentemente, $2|X_{m,n}|^2$ tem distribuição χ^2 com 2 graus de liberdade e média, $\lambda_{m,n} = 2(N^{1-4c}Q_{m,n})^{-1}\gamma_x(m,n)$ onde $\gamma_x(m,n)$ é a função assimetria.

Mostra-se igualmente que $S = 2 \sum_{(m,n) \in L} |X_{m,n}|^2$ é aproximadamente $\chi^2_{2P}(\lambda)$ onde $\lambda = \sum_{(m,n) \in L} \lambda_{m,n}$ e P o número de (m,n) em L . Esta estatística é basicamente a de Rao e Gabr com a matriz de covariâncias assintótica em vez da sua congénere amostral. Sob a hipótese de nulidade, $f(\omega_1, \omega_2) \equiv 0$, S é aproximadamente $\chi^2_{2P}(0)$ para um grande N , isto sugere a estatística teste, $\hat{S} = 2 \sum_{(m,n)} |\hat{X}_{(m,n)}|^2$ onde $\hat{X}_{(m,n)}$ é dada por (2.90) substituindo o espectro teórico pelo amostral na mesma grelha de frequências. Devido à inconsistência do estimador do espectro, Hinich, sugere o uso do alisamento proposto por Fuller(1976) onde $\hat{f}(\omega) = f(\omega)[1 + (M/N)Y]$ onde Y é normal *standard*.

A hipótese de nulidade do teste de gaussianidade é rejeitada se $\hat{S} > t_\alpha$ onde $\alpha = P(\chi^2_{2P} > t_\alpha)$.

Hinich mostra que o teste é consistente quando $N \rightarrow 0$. Segundo aquele autor a obtenção de grande potência no teste depende muito do tamanho da amostra e da relação com o valor da função de assimetria.

O teste de linearidade proposto por Hinich é não paramétrico e baseado numa medida robusta de dispersão. O teste proposto por Rao e Gabr que usa a distribuição *F-Snedcor* é sensível a *outliers* daí que Hinich proponha a distância interquartil da estatística $\{2|X_{m,n}|^2\}$ que é robusta a grandes *outliers*.

Sendo R a distância interquartil de $\{2|X_{m,n}|^2\}$ para quadrados em D , deve rejeitar-se a linearidade se R for significativamente maior ou menor que a distância interquartil de $\chi^2_2(\hat{\lambda}_0)$ onde $\hat{\lambda}_0$ é um estimador consistente de λ_0 . O estimador deste parâmetro tem a forma,

$$\hat{\lambda}_0 = 2(PM^2)^{-1} \sum_{(m,n) \in L} Q_{m,n}(|X_{m,n}|^2 - 1) = \lambda_0 + O_P(P^{-1/2}) \quad (2.91)$$

Hinich mostra que quando P_I , número de quadrados no interior de D é grande então R é aproximadamente normal com média $\zeta_3 - \zeta_1$ e variância $\sigma = (16P_I)^{-1}[3f^{-2}(\zeta_1) - 2f^{-1}(\zeta_1)f^{-1}(\zeta_3) + 3f^{-2}(\zeta_3)]$ onde ζ_1, ζ_3 são, respectivamente o 1 e 3 quartis de $f(\xi^2)$. As propriedades assintóticas do intervalo interquartil amostral de uma distribuição são dadas por um teorema de David, [Dav80]. Estas propriedades, para o caso de uma distribuição $\chi^2_2(\lambda_0)$, podem ser usadas para obter uma região crítica para um teste de nível α usando a estatística R .

A designação *teste de gaussianidade* aos testes acima descritos designados por [Rao84] e também por [Hin82] não devem induzir o utilizador em erro. De facto, ao testar a nulidade do bispectro ou do cumulante de terceira ordem está a testar-se a não assimetria do processo. Um bispectro nulo não é prova de gaussianidade. O bispectro de um processo reversível é identicamente nulo, por exemplo, um processo com a representação (2.74) onde $\{e_t\}$ é i.i.d. com distribuição simétrica tem bispectro nulo. Por exemplo, se $\{e_t\}$ tiver distribuição de *Laplace* então todo o poliespectro de ordem impar, bispectro incluído, é identicamente nulo. Por outro lado, existem processos com representação linear simétricos cujo bispectro é não nulo, basta considerar um processo que admita uma representação do tipo (2.74) onde $\mu_3 \neq 0$ e onde a sucessão $\{h_t\}$ seja tal que $\sum_j h_j^3 = 0$. Nestas condições, facilmente se verifica que o processo $\{X_t\}$ tem assimetria zero, o que não significa que o cumulante de terceira ordem ou o bispectro sejam identicamente nulos. Os resultados descritos traduzem as limitações destes testes, no entanto, na prática, quando os dados apresentam um grande desvio da linearidade e consequentemente da gaussianidade nenhum dos cenários atrás descritos se pode verificar pelo que faz todo o sentido usar os testes.

[Cha86] indicam que a modificação de Hinich estabiliza os resultados e torna-os menos sensíveis a *outliers*. No entanto, a experiência daqueles autores mostra que os testes são exigentes em termos do tamanho da amostra e revelam baixa potência. Verificam ainda

a incapacidade dos testes em detectar um afastamento da gaussianidade/linearidade de um processo autoregressivo não-linear com estrutura simétrica perturbado por ruído, $\{\varepsilon_t\}$ com distribuição simétrica.

Capítulo 3

Métodos de sistemas dinâmicos

Neste capítulo introduzem-se alguns conceitos fundamentais da teoria dos sistemas dinâmicos diferenciáveis. Indicam-se as formas conhecidas de perda de estabilidade e a transição para um regime caótico. Apresentam-se os sistemas dinâmicos intermitentes de Pomeau e Manneville [Pom80]. Enuncia-se o teorema de Birkoff(ergódico) e o teorema de Oseledec(multiplicativo ergódico) que garante a existência dos expoentes de Lyapunov. Os expoentes Lyapunov estão associados a subespaços nos quais ocorre expansão/contração exponencial e caracterizam a evolução e o comportamento global de um sistema dinâmico. Introduce-se a definição de entropia de Kolmogorov-Sinai de uma medida [Kol58]. Enunciamos o teorema de reconstrução de Takens [Tak80] e o teorema de Sauer [Sau91]. Explicamos também, como determinar experimentalmente algumas dessas quantidades a partir de um sistema reconstruído. Os testes de reversibilidade e diferenças em medidas de reconstrução de Diks, [Dik99] são relevantes para o estudo de estacionaridade de uma série e por isso são apresentados assim como a definição de medida de reconstrução. Por último, abordamos a previsão no espaço de fase descrevendo o funcionamento dos métodos locais no espaço de fase em particular de Pikovski [Pik86] e May e Sugihara [Sug].

3.1 Sistemas com estabilidade. Atractores convencionais.

Um sistema dinâmico é a evolução no tempo de uma ou mais variáveis, é dado pela solução de um sistema de equações diferenciais com tempo contínuo,

$$\frac{dX(t)}{dt} = \mathbf{F}(X(t)), X \in \mathcal{M} \quad (3.1)$$

onde $X = (x, \dot{x}, x^{(2)}, \dots, x^{(N-1)})$ representa um estado do sistema no espaço de fase, $\mathcal{M}$ que pode ser um espaço euclidiano, $\mathbb{R}^N$, uma variedade diferenciável ou um espaço de dimensão infinita. Sendo X_0 o valor inicial a curva gerada pela evolução do sistema, $\{X_t\}_{t \geq 0}$ designa órbita ou trajectória.

No caso do tempo ser uma variável discreta a descrição passa por um sistema de equações com diferenças representada por uma função N -dimensional.

$$X_{n+1} = f(X_n), X_n \in \mathcal{M} \quad (3.2)$$

A órbita ou trajectória é, nesse caso, um conjunto de pontos, $X_0, X_1, \dots, X_n$.

As funções, F e f supõem-se diferenciáveis ou até mesmo de classe C^∞ a suposição tem o seu suporte no facto de que as quantidades físicas serem geralmente funções *bem comportadas*. Introduza-se o operador não-linear, f^t , t real ou inteiro com as propriedades,

$$\begin{cases} f^0 &= I \\ f^s f^t &= f^{s+t} \end{cases} \quad (3.3)$$

desta forma f^t é grupo ou pelo menos semigrupo ($t \geq 0$). Supondo que o espaço de fase é um espaço linear então define-se o operador linear Df_x^n , matriz de derivadas

parciais, usando a regra da cadeia,

$$Df_x^n = Df_{f^{n-1}x} \cdots Df_{f^2x} Df_{fx} Df_x \quad (3.4)$$

Um sistema contínuo pode ser reduzido a um sistema em tempo discreto usando uma técnica denominada secção de Poincaré. Esta técnica consiste na redução da dimensão do fluxo a um sistema discreto com menos uma dimensão usando as intersecções das órbitas com a superfície no espaço de fase.

Outra via possível para discretizar o sistema consiste em amostrar o fluxo em instantes $t_n = t_0 + nT$ ($n = 0, 1, 2, \dots$) onde T é escolhido convenientemente. Desta forma uma órbita contínua, X_t dá origem a uma órbita discreta, $X_n \equiv X(t_n)$.

A teoria dos sistemas dinâmicos não-lineares em tempo discreto, em geral, e as equações com diferenças não lineares, em particular, está incompleta. Quase sempre, para efeitos de análise, introduzem-se definições de *trabalho* que decorrem das análogas em tempo contínuo.

Os sistemas em tempo discreto exibem comportamentos complexos com apenas uma dimensão.

Os sistemas hamiltonianos são uma classe de sistemas dinâmicos que ocorrem em muitas circunstâncias, por exemplo, os sistemas associados às equações do movimento de Newton são hamiltonianas. Os sistemas hamiltonianos caracterizam-se por serem conservativos, ou seja, a função que os rege preserva, a menos de uma mudança de variáveis, volumes no espaço de fase. Os sistemas não conservativos dizem-se dissipativos cuja principal característica é o facto de possuírem atractores ou conjuntos atractores.

Um conjunto A do espaço de fase é um conjunto atrator com vizinhança fundamental U se satisfizer as propriedades [Rue81],

- Atracção: Todo o aberto $V \supset A$ tem-se $f^t U \subset V$ para t suficientemente grande.
- Invariância: $f^t A = A$ para todo o t

A base de atracção, U é constituída por todos os pontos que se aproximam de A quando $t \rightarrow \infty$, isto é, $\cup_{t>0} f^t U$. [Rue89a] refere que na ausência da definição matemática de um atractor pode fazer-se referência a uma definição operacional. Define-se um atractor como o conjunto de pontos, $f^t x$ que acumulam para um t suficientemente grande. Além das propriedades referidas em cima, um atractor distingue-se do conjunto atractor pelo facto de, não ter necessariamente um aberto como conjunto de atracção e de possuir uma forma de irreducibilidade. Para quaisquer pontos $x \neq x' \in A$ existe um inteiro t tal que $f^t x'$ estará arbitrariamente próximo de x . Esta última condição faz com que um atractor não tenha partes disjuntas. A esta última propriedade também se chama transitividade topológica.

As experiências físicas e computacionais com sistemas dinâmicos geralmente exibem um sistema transitório seguido por um regime assintótico.

Como exemplos de atractores temos,

- O ponto fixo:

Seja P tal que $f^t P = P$, para todo t . Chama-se a P um ponto fixo. A derivada, D_f num ponto fixo é um operador. Se o seu espectro estiver no disco, $\{z : |z| < \alpha\}$ com $\alpha < 1$ e com partes reais negativas então o ponto fixo é atractor. Para o caso discreto, sendo $(P_1, P_2, \dots, P_n)$ uma órbita periódica, $fP_1 = P_2, \dots, fP_n = P_1$ então P_i é ponto fixo de f^n .

- órbita periódica para sistemas contínuos no tempo:

Suponha que existem, a e $T > 0$ tal que, $f^T a = a$ e $f^t a \neq a$ para todo $0 < t < T$. Então $\Gamma = \{f^t a : 0 \leq t \leq T\}$ é uma órbita periódica. Tal como no caso discreto, Df_a^T tem um valor próprio igual a 1. Se os restantes valores próprios estão no interior do círculo unitário então a órbita periódica é atractiva.

- Atractor quasi-periódico:

Uma órbita periódica é um círculo num sistema contínuo e o seu movimento pode ser descrito por,

$$\phi(t) = \phi(0) + \omega t \pmod{2\pi}, \omega = 2\pi/T \quad (3.5)$$

corresponde ao movimento de um oscilador simples. Suponha agora que dispomos de k osciladores com frequências $\omega_1, \omega_2, \dots, \omega_k$ incomensuráveis entre si. Basicamente o movimento quasi-periódico pode ser considerado como uma mistura de movimentos periódicos com diferentes frequências fundamentais. O movimento dos osciladores é descrito por,

$$\phi_i(t) = \phi_i(0) + \omega_i t \pmod{2\pi}, i = 1, \dots, k \quad (3.6)$$

este movimento tem lugar no produto de k círculos, um toro de dimensão k , T^k . Supondo que T^k está mergulhado num espaço euclidiano, $\mathbb{R}^m$ com $m \geq k$ e que é um conjunto atrator então T^k é um atrator quasi-periodico. Assintoticamente, o sistema dinâmico é descrito por

$$X(t) = f^t X(0) = \Phi(\nu_1 t, \dots, \nu_k t) \quad (3.7)$$

onde Φ é periódica de periodo 1 separadamente nos seus k argumentos e $\nu_1, \dots, \nu_k$ são as frequências. O toro, T^k funciona como uma variedade atratora das órbitas. Este tipo de movimento é frequente em sistemas dissipativos.

3.2 Sistemas dinâmicos intermitentes

O atributo que define um atrator com dinâmica caótica é a dependência sensível de condições iniciais (*DSCI*). Assumindo que o sistema tem uma evolução determinística da forma (3.2) com valor inicial, X_0 , seja, $X'_0 = X_0 + \delta X_0$. Devido à continuidade

das soluções espera-se que δX_n seja pequeno, porém no caso de se de a *DSCI* estar presente verifica-se que,

$$\delta X_n = \delta X_0 e^{\lambda n} \quad (3.8)$$

Se órbitas com condições iniciais muito próximas divergem exponencialmente, ou seja, se uma pequena perturbação conduz a resultados muito diferentes então a propriedade *DSCI* gera aleatoriedade ou imprevisibilidade. O crescimento exponencial dos erros torna a evolução temporal independente do seu passado e de certa forma não determinística.

A descoberta de sistemas com dependência sensível de condições iniciais foi feita por Hadamard no final do século XIX. O parâmetro λ de (3.8) corresponde à taxa média de divergência das órbitas, para ocorrer *DSCI* o seu valor terá que ser positivo. Se estiver presente no sistema e se o seu valor não for muito elevado poderá significar a possibilidade de boas previsões a curto prazo.

Os sistemas unidimensionais não invertíveis são os sistemas mais simples que podem exibir caos e por consequência a *DSCI*. Grande parte do comportamento complexo exibido por sistemas contínuos de grandes dimensões é exibido por sistemas discretos com apenas uma dimensão. Daí o seu interesse como ponto de partida para o estudo da teoria do caos que é seguida por vários autores, [Ott93], [All96], [Rue89a].

A relação entre a *DSCI* e um regime dinâmico caótico é estreita. Um regime dinâmico é caótico se o seu espectro de potência for contínuo e de banda larga e tiver um rápido decaimento da autocorrelação/memória do processo, isto é, o conhecimento do estado do sistema num determinado instante não permite, por si só, a previsão a médio/longo prazo do mesmo. Estes são aspectos fundamentais do caos. Sem ser marcada pelo rigor, esta definição encerra e si a incontornável fronteira que separa a previsibilidade da imprevisibilidade.

A descoberta experimental do regime quasi-periódico com transição para caos deve-se a Landau [Lan44]. Nesse célebre trabalho, Landau procurava explicar de que forma um fluido passa de um regime/estado laminar a um regime/estado irregular

quando se aumentava um parâmetro de controlo, o número de Reynolds. Na verdade a experiência levava o sistema a passar de periódico a quasi-periódico e depois a um estado caótico quando se aumentava desmesuradamente o número de Reynolds. Mais tarde Lorenz descobriu um sistema com apenas 3 dimensões, (3.9) que revelava um regime altamente irregular e aparentemente imprevisível [Lor63],

$$\begin{aligned}\frac{dx(t)}{dt} &= -\sigma x(t) + \sigma y(t) \\ \frac{dy(t)}{dt} &= -x(t)z(t) + rx(t) - y(t) \\ \frac{dz(t)}{dt} &= x(t)y(t) - bz(t)\end{aligned}\tag{3.9}$$

onde os valores $\sigma = 10$, $b = 8/3$ e $r = 28$ foram escolhidos depois de um estudo numérico [Lor63]. As equações de Lorenz foram obtidas por truncagem das equações de Navier-Stokes e dão uma descrição aproximada da evolução do estado de um líquido sobre uma superfície aquecida. Este fenómeno é semelhante ao que acontece na superfície terrestre. Para um raio, r suficientemente grande a evolução do sistema exhibe *DSCI* [Lu94]. Alguns autores associam a irregularidade à presença de singularidades na solução da equação de Navier-Stokes mas o problema de turbulência fraca como acontece nesse tipo de experiências, fluxos de fluidos aquecidos, que pode ser descrito por uma equação mais simples não tem a ver com o problema específico da equação de Navier-Stokes. Uma descrição mais simples para esse fenómeno é [Eck85],

$$\frac{dV(t)}{dt} = X_\mu(V)\tag{3.10}$$

onde X_μ é um campo definido numa variedade num espaço euclidiano. A constante μ pode ser um parâmetro experimental (o de Reynolds, por exemplo) que varia com V . O comportamento do fluido torna-se irregular quando μ aumenta. Este facto altera qualitativamente o sistema mas tem também um impacto na geometria do atractor.

Ao fenómeno dá-se o nome de bifurcação [Ott93],[Eck85] e aos conjuntos para V onde μ toma valores a que correspondem regimes qualitativamente diferentes dá-se o nome de conjuntos de bifurcação.

O estudo do aparecimento e evolução de um sistema para o caótico tem as suas raízes na teoria da bifurcação e constitui a teoria geométrica do caos [Eck85]. O objectivo desta teoria é tentar prever, a partir de uma variação de μ e respectiva alteração/mudança qualitativa o que acontecerá se μ continuar a variar. Seguindo [Eck81], [Ott93] indicam 3 possíveis cenários de (transição para)/aparecimento de caos,

- Ruelle e Takens, [RT], através da quasiperiodicidade.
- Feigenbaum(1978), [Fei78], através da duplicação de período.
- Pomeau e Manneville, [Pom80] através da intermitência.

Estando em presença de um sistema dissipativo com um espectro de banda larga, ou que exiba *DSCI*, que tipo de atractor lhe vai estar associado? O atractor de um sistema que exibe *DSCI* diz-se um atractor estranho [RT]. Chama-se atractor estranho ao objecto geométrico descrito pelos pontos da trajectória¹. Como se vai ver mais à frente, o conceito de atractor estranho é demasiado geral tendo em vista o propósito de um estudo qualitativo, por isso vão introduzir ferramentas que permitirão uma caracterização dinâmica e geométrica.

A transição num sistema de um regime regular para um regime caótico está associado a uma perda de estabilidade. Um sistema do tipo (3.1) está num estado estacionário se estiver numa das situações descritas na secção anterior, [Ott93] e [Ber84].

Das 3 formas conhecidas de transição para caos, a intermitência dinâmica é o fenómeno que mais se aproxima dos caudais do rio Paiva. Aqui vamos seguir [Ott93] e [Ber84], vamos supôr que para valores de um parâmetro, p , inferior a um valor crítico o atractor do sistema é uma órbita periódica. Para um valor de p ligeiramente

¹o atractor de Lorenz é estranho [Tuc99]

superior a p_T existem longos períodos de tempo, fases laminares, durante os quais a órbita é regular e aproximadamente periódica e assemelha-se a uma órbita para $p < p_T$, no entanto, este comportamento regular é intermitentemente interrompido por períodos curtos de tempo, *bursts*, onde a órbita do sistema se comporta de uma maneira totalmente diferente. Estes saltos do sistema parecem ocorrer aleatoriamente e pode-se fazer estatísticas sobre as ocorrências dos saltos e as durações das fases laminar e irregular. Se p tomar valores muito maiores do que p_T a duração da fase irregular torna-se bastante maior enquanto que a fase laminar reduz-se na sua duração. Pomeau e Manneville, [Pom80], mostraram esse efeito usando o atrator de Lorenz. Na transição intermitente para caos uma órbita periódica estável torna-se instável ou então é simplesmente destruída quando p ultrapassa p_T . Um outro pormenor importante reside na diferença da dinâmica entre a transição intermitente e transição por duplicação do período, no primeiro caso, a órbita periódica é substituída por outra órbita periódica estável mas que lhe está próxima no espaço de fase, no segundo caso, a órbita ao passar da fase laminar para a irregular afasta-se da vizinhança dos valores da órbita periódica.

Existem 3 tipos de bifurcação que têm as características dos sistemas intermitentes. Os tipos genéricos de bifurcação que se encaixam neste comportamento são a bifurcação de ponto-sela (uma órbita estável e uma instável, correspondentes a valores próprios menores e maiores do que 1 que se designa por tipo I, a de bifurcação (duplicação) inversa de período (uma órbita periódica instável colapsa numa órbita periódica estável de metade do período onde o sistema se encontra a maior parte do tempo, e uma outra turbulenta), tipo III inversa e a de bifurcação subcrítica de Hopf, tipo II [Ott93]. Neste último caso a órbita periódica torna-se instável quando um par de valores próprios conjugados da matriz de Floquet atravessam o raio unitário. O termo transição subcrítica aplicado à bifurcação de Hopf de uma órbita periódica significa que quando o parâmetro p aumenta, uma órbita quasiperiódica instável de 2 frequências colapsa numa órbita periódica estável que depois é substituída por uma instável quando o ponto de bifurcação é ultrapassado. Esta intermitência apresenta as mesmas características qualitativas globais que as do tipo I e III com a exceção

do facto de que as instabilidades que se desenvolvem nas fases laminares têm uma frequência sem relação com a frequência de base do ciclo limite inicial, [Ber84]. Como indica [Ott93], os diferentes tipos de transição intermitente têm associados diferentes escalas de tempo de duração da fase laminar. Designando por $\bar{T}(p)$ a duração média da fase laminar, as escalas de tempo para os vários tipos de bifurcação são,

1. $(p - p_T)^{-1/2}$ para o tipo I
2. $(p - p_T)^{-1}$ para o tipo II e III

Apesar de o tempo médio se revelar o mesmo para o tipo II e III, a distribuição de probabilidade associada pode ser bastante diferente para os dois casos, [Ott93].

3.3 Entropia de um sistema dinâmico.

Nesta secção apresenta-se uma quantidade física que, tal como os expoentes de Lyapunov, são uma forma de quantificar e caracterizar o caos.

A entropia métrica foi introduzida por [Kol58]. No que se segue iremos considerar sistemas em tempo discreto tal como o descrito em, (3.2). Se f é uma transformação que preserva uma medida, ρ , então o invariante K-S (kolmogorov-Sinai) ou entropia, $h(\rho)$ mede a taxa assintótica de criação de informação por iteração de f , [Rue89a]. Kolmogorov procurava um número que fosse invariante sob um isomorfismo. A noção de entropia na qual de baseou Kolmogorov havia sido introduzida por [Sha48].

Shannon, num trabalho que deu origem à teoria da informação, havia definido o conceito de grau de incerteza na previsão do resultado de uma experiência aleatória. Numa experiência aleatória com r resultados possíveis com probabilidades, $\{p_i\}_{i=1}^r$, a entropia de Shanon é, [Ott93], [Eck85],

$$H_S = - \sum_{i=1}^r p_i \ln(p_i) \quad (3.11)$$

por convenção, $p \ln(p) = 0$ se $p = 0$. Por exemplo, a ausência de incerteza, $p_i = 1$ e $p_j = 0$ para todo o $j \neq i$ resulta em $H_S = 0$. A situação mais incerta corresponde a $p_j = p_i$ para todo $j \neq i$ que resulta em $H_S = \ln(r)$.

Para se poder aplicar no âmbito dos sistemas dinâmicos o conceito tem que ser adaptado. Considere-se o exemplo 3.6.1. Se atribuirmos uma probabilidade p_i a cada elemento i de V obtém-se uma medida de probabilidade ρ em Ω . Para o caso em que $V = \{0, 1\}$ e $p = 1 - p = 1/2$ a entropia de Shannon é, $H_S = \ln 2$.

Sendo ρ uma medida invariante para algum sistema dinâmico, a entropia métrica ou K-S, $h(\rho)$, define-se da seguinte forma. Seja W um conjunto compacto, suporte da medida ρ , munido de uma métrica. Se $W = W_1 \cup W_2 \cup \dots \cup W_m$ for uma partição disjunta de W , pode definir-se uma função que corresponde à entropia para a partição,

$$H(\{W\}) = - \sum_{i=1}^m p_i \ln(\rho(W_i)) \quad (3.12)$$

construa-se uma sucessão de partições $\{W_i^{(n)}\}$ cada vez mais refinada usando o seguinte critério, [Rue89a], o diâmetro de $\{W_i\}$ corresponde à precisão com que se trabalha, ou seja, dentro de cada W_i não se distinguem pontos entre si. Pode conseguir-se obter uma partição com essa propriedade usando imagens inversas. Defina-se nova partição, $\{W_i^{(2)}\}$ cujos elementos são as m^2 intersecções, $W_j \cap f^{-1}(W_k)$ onde $j, k \in \{1, 2, \dots, m\}$ das quais se retêm as não vazias. A partição $\{W_i^{(3)}\}$ obtem-se intersectando os elementos de $\{W_i^{(2)}\}$ com as imagens inversas dos elementos de $\{W_i^{(1)}\}$, ou seja, os m^3 elementos, $W_j \cap f^{-1}(W_k) \cap f^{-2}(W_l)$ onde $j, k, l \in \{1, 2, \dots, m\}$ e assim sucessivamente. Para a partição $\{W_i^{(n)}\}$ os elementos são os conjuntos não vazios das m^n intersecções da forma

$$W_{i_1} \cap f^{-1}(W_{i_2}) \cap f^{-2}(W_{i_2}) \cap \dots \cap f^{-(n-1)}(W_{i_n}) \quad (3.13)$$

onde $i_j \in \{1, \dots, m\}$. A entropia associada à partição $\{W_i^{(n)}\}$ é,

$$H(W) = - \sum_{i=1}^m \rho(W_i) \ln \rho(W_i) \quad (3.14)$$

defina-se a quantidade,

$$\begin{aligned} h(\rho, \{W_i\}) &= \lim_{n \rightarrow \infty} \frac{1}{n} H(\{W_i^{(n)}\}) \\ &= \lim_{n \rightarrow \infty} \{H(W^{(n+1)}) - H(W^n)\} \\ &= \lim_{n \rightarrow \infty} \{H(W^n)\} \end{aligned} \quad (3.15)$$

a entropia K-S ou métrica corresponde à maximização de $h(\rho, \{W_i\})$ sobre todas as partições iniciais possíveis, $h(\rho) = \sup_{\{W_i\}} h(\rho, \{W_i\})$.

Rényi, [Rén71], generalizou o conceito definindo a informação de Rényi de ordem α . Dada uma distribuição de probabilidade, $p_1, p_2, \dots, p_m$, a informação de Rényi de ordem α é,

$$H_\alpha(p_1, p_2, \dots, p_m) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^m p_i^\alpha \right), \alpha \in \mathbb{R}^+ - \{1\} \quad (3.16)$$

Os números H_α são generalizações de H , isto é, $\lim_{\alpha \rightarrow 1} H_\alpha = H$. Para $\alpha = 2$ vem $H_2(W_i) = -\ln(\sum_{i=1}^m \rho^2(W_i))$, defina-se, $K_2(\rho) = \sup_{W^{(n)}} \lim_{n \rightarrow \infty} H_2(W^{(n)})$ caso o limite exista.

O uso de uma partição no intervalo real correspondente à variável de observação do

sistema permite simplificar as expressões. Escolhendo os elementos da partição como intervalos $\{I_j\}_{j=1}^m$ de igual amplitude da recta real e considerando $p_{i_1, i_2, \dots, i_m}$ como a probabilidade de, num tempo arbitrário, n , o sistema se encontrar em I_{i_1} e passar em $n+1$ para I_{i_2} , etc. Pode definir-se a entropia como função do diâmetro e do número de intervalos da partição,

$$H_\alpha(m, \varepsilon) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^m p_{i_1, i_2, \dots, i_m}^\alpha \right) \quad (3.17)$$

e as entropias de ordem α definem-se como

$$\begin{aligned} h_\alpha &= \sup_P \lim_{m \rightarrow \infty} 1/m H_\alpha(m, \varepsilon) \\ &= \sup_P \lim_{m \rightarrow \infty} (H_\alpha(m+1, \varepsilon) - H_\alpha(m, \varepsilon)) \end{aligned} \quad (3.18)$$

onde P designa o conjunto de todas as partições da recta real. A entropia h_2 , que costuma-se designar por K_2 é particularmente interessante porque tem a propriedade, $K_2(\rho) \leq h(\rho)$.

O limite de H_α quando $\alpha \rightarrow 0$ é a entropia topológica, h_0 .

Para ilustrar este conceito, considere-se o exemplo da órbita periódica, S , estado estacionário do sistema referido na secção anterior. Seja $\rho(x)$ a medida associada às p funções- δ , número de estados da órbita periódica. Desta forma de todas as partições que se possam formar com subconjuntos de S , no máximo, apenas p têm probabilidade não nula. Assim a entropia de Rényi é,

$$H_\alpha = \frac{1}{1-\alpha} \ln \left(\sum_{j=1}^p p_j^\alpha \right) = \ln(p) \quad (3.19)$$

expoentes de Lyapunov em circunstâncias gerais. Se ρ for ergódica os expoentes de Lyapunov, $h_i(x_0)$ são os mesmos quase por toda a parte relativamente à medida ρ (q.t.p- ρ). Se a medida ρ for absolutamente continua significa que os expoentes são independentes do ponto inicial, excepto para um conjunto de medida de Lebesgue nula. Num célebre artigo, [Rue78] enunciou o seguinte teorema,

Teorema 3.3.1 *Seja $f : M \rightarrow M$ uma função de classe C^1 onde M é uma variedade compacta que deixa invariante uma medida ergódica, ρ com suporte em M , então,*

$$h(\rho) \leq \sum_{\lambda_i > 0} \lambda_i \quad (3.20)$$

Pesin(1977) mostrou ainda que a relação é de igualdade se f for um difeomorfismo e se a medida ρ possuir uma densidade diferenciável relativamente à medida de Lebesgue.

Se a entropia, h , se relaciona com os expoentes característicos da medida do atrator então pode ser usada para quantificar a predictabilidade. Seguindo [Kan99], se se conhecer num dado instante, com precisão finita, o estado de um sistema, que se traduz em possuir um ou mais números com representação binária finita, sabemos que a entropia permite calcular a taxa a que os números (*bits*) sofrem um *shift* para a esquerda. Com base nessa informação pode estimar-se o número de iterações necessárias para que o numero de *bits* conhecidos na iteração inicial desapareçam. Esse tempo, t_p é dado por $\varepsilon_0 \exp ht_p = 1$ sendo ε_0 a precisão da medição. Em média após um tempo $t_p = \ln \varepsilon_0 / h$ perde-se toda a previsibilidade e um aumento linear de previsibilidade implica um aumento exponencial da precisão.

No capítulo anterior fez-se referência ao facto de as variedades instáveis fazerem parte do conjunto atrator pelo que este conjunto ser uma espécie de união (disjunta) de variedades instáveis. A existência no conjunto atrator de diferentes tipos de intersecções, homoclínicas e heteroclínicas corresponde a dinâmicas complexas. Smale, [Sma67], mostrou que a existência de uma intersecção homoclínica implica uma dinâmica em

ferradura conhecida como (*Smale's horseshoe*).

O atractor de Hénon, [Hén76], é um exemplo de um atractor cuja forma apresenta dobras, esse efeito está relacionado com o facto de o atractor possuir um ponto fixo sela associado a um expoente de Lyapunov positivo sendo a linha visível o fecho da variedade instável associada.

Eventualmente as variedades instáveis poderão formar uma estrutura descontínua acumulando-se ou empilhando-se desordenadamente.

Isso significa que as medidas invariantes geradas pelo atractor possam ter também densidades com formas complexas. Por outro lado, espera-se que, devido à expansão e contracção na direcção instável, a medida possua uma densidade diferenciável relativamente aquelas direcções, [Rue89b]. A estas medidas dá-se o nome *SRB* (Sinai-Ruelle-Bowen) que se definem da seguinte forma. Seja S um conjunto ρ -mensurável da forma $\bigcap_{\alpha \in A} S_{\alpha}$ onde os S_{α} 's são conjuntos disjuntos.

Dada uma medida ergódica, ρ , com suporte compacto, as variedades W_x^u estão definidas *q.t.p.*- ρ . Uma vez que as variedades instáveis não se intersectam entre si, se $y \in W_x^u$ então também, $x \in W_y^u$ daí que as variedades instáveis particionem o espaço em classes de equivalência.

Como indicam, [Eck85], a demonstração da existência de medidas *SRB* para um dado sistema é uma tarefa bastante difícil e não se sabe se elas existem em geral.

O conceito de medida *SRB* surge pelas boas propriedades que uma medida com densidade diferenciável nas direcções instáveis poderia ter e uma vez que estas particionam o espaço poder-se-ia tentar usar a decomposição na sua definição, porém, convém lembrar que a propriedade ergódica da medida tem que se manter, pelo que se terá que recorrer a outra forma para a definir. Sendo S um conjunto ρ -mensurável da forma $\bigcap_{\alpha \in A} S_{\alpha}$ onde os conjuntos S_{α} 's são subconjuntos abertos e disjuntos das variedades instáveis, W^U , ou seja, cada S_{α} pode obter-se por intersecção de uma variedade instável local com S pode construir-se a medida ρ restrita a S como,

$$\rho(.) = \int \rho_\alpha m(d\alpha) \quad (3.21)$$

onde m é uma medida em A e ρ_α uma medida de probabilidade em S_α - medida de probabilidade condicionada associada à decomposição $S = \bigcap_{\alpha \in A} S_\alpha$. Naturalmente interessam apenas as medidas ρ_α absolutamente contínuas relativamente à medida de Lebesgue em W^u , isto é, [Eck85],

$$\rho_\alpha(d\xi) = \phi(\xi)d\xi, \text{ em } S_\alpha \quad (3.22)$$

$d\xi$ é um elemento de volume, S_α uma variedade com dimensão m^+ , soma das multiplicidades dos expoentes de Lyapunov positivos, e ϕ_α é uma função ρ -integrável. Diz-se que a medida ergódica é *SRB* se as probabilidades condicionais, ρ_α , forem absolutamente contínuas relativamente à medida de Lebesgue para alguma escolha de S com $\rho(S) > 0$ e uma decomposição $S = \bigcap_{\alpha} S_\alpha$. A definição é independente da escolha de S e sua decomposição.

Ledrappier e Young, [Led85] chegaram ao seguinte resultado,

Teorema 3.3.2 *Seja $f : M \rightarrow M$ um difeomorfismo- C^2 de uma variedade diferenciável, M , m -dimensional e ρ uma medida ergódica com suporte compacto. As seguintes condições são equivalentes,*

- a) *A medida ρ é SRB, ou seja, absolutamente contínua ao longo das direcções instáveis.*
- b) *A medida ρ satisfaz a igualdade de Pesin, $h(\rho) = \sum \{\lambda_i^+\}$, onde $\{\lambda_i^+\}_{i=1}^m$ é o conjunto de todos os expoentes de Lyapunov positivos.*

Se estas condições forem satisfeitas as funções densidade, ϕ_α , são diferenciáveis. ²

Como se pôde verificar, o comportamento que se designou por caótico compreende uma classe de sinais não lineares que estão entre o regular ou quasiperiódico e o imprevisível (próximo ao estocástico). Desde sempre foi visto como uma forma de ruído numa vertente física, ou como uma trajectória de um processo estocástico numa vertente estatística. No capítulo 1 fez-se uma descrição dos métodos e ferramentas mais comuns desta última abordagem. No entanto, como se referiu anteriormente, as características dos sinais caóticos, como o espectro, que apresenta valores elevados para toda a gama de frequências, não tem correspondência nos modelos estocásticos usuais (ver capítulo 1).

É claro que este facto aliado à necessidade de extrair informação física relevante acerca da dinâmica de sistemas físicos observados sob regime caótico, motivou o desenvolvimento de ferramentas que pudessem lidar com essas características. Não deixa de ser curioso o facto de que, desde a descoberta do essencial da dinâmica caótica, feita por Hadamard e Poincaré nos finais do século XIX e princípios do séc. XX, tenham decorrido mais de 80 anos até ao aparecimento de métodos de identificação dessas características em sistemas observados, facto a que não é alheio o extraordinário aumento verificado, nos últimos anos, na capacidade computacional.

Do ponto de vista da amostragem as características do caos colocam um sério problema. No caso dos fenómenos temporais, mais concretamente os que dizem respeito ao clima e à hidrologia, em que se usam séries históricas, é frequente não se poder ultrapassar uma dada frequência de amostragem e, geralmente, apenas se observa uma ou duas variáveis. Para os tipos de fenómenos que se podem recriar em laboratório e para os quais é possível observar várias variáveis, e, com possibilidade de haver uma frequência de amostragem bastante superior, surgem outros tipos de problemas.

Considere-se um sistema dinâmico puramente determinístico, se possuímos a regra de evolução e o estado presente, todos os estados futuros podem ser determinados. O conceito de estado de um sistema é bastante geral e é também usado em processos

²para mais detalhes ver [Eck85].

estocásticos. As cadeias de Markov são um exemplo de processos estocásticos onde o termo estado designa um vector de probabilidades, uma por cada estado considerado correspondendo à probabilidade da variável em estudo se encontrar em cada um dos estados.

Em qualquer situação experimental é inevitável lidar com funções discretas, toda a análise de sistemas tem lugar em tempo discreto uma vez que é impossível amostrar continuamente. Se se amostrar um sinal $s(t)$ em intervalos Δt começando em t_0 , os dados são da forma $s(n) = s(t_0 + n\Delta t)$ correspondendo a evolução à passagem de $s(k)$ para $s(k+1)$.

Sendo um sistema contínuo da forma (3.1), qualquer evolução amostrada tem a forma,

$$X(t_0 + (n+1)\Delta t) \approx X(t_0 + n\Delta t) + \Delta t F(X(t_0 + n\Delta t)) \quad (3.23)$$

assim as observações sucessivas têm a relação,

$$\begin{aligned} s(t_0 + k\Delta t) &\rightarrow s(t_0 + (k+1)\Delta t) \\ s(k) &\rightarrow s(k+1) \end{aligned} \quad (3.24)$$

A relação entre $X(t)$ e $s(n)$ vai ser aprofundada na secção seguinte.

3.4 Reconstrução do espaço de fase

A reconstrução de um espaço de fase de modo a que seja equivalente ao original a partir de uma série temporal é a base de todos os métodos que vão ser aplicados. É um passo fundamental para a aplicação de qualquer método. Uma maneira de levar a cabo tal reconstrução consiste no uso de funções de coordenadas desfasadas. Uma vez que se tem que converter uma série de valores em vectores de estado uma solução possível é recorrer aos valores da série observados em instantes igualmente espaçados

no tempo. Sendo a série uma sucessão de medidas escalares feitas no tempo de uma quantidade que depende do estado do sistema no instante em que é feita e supondo-a tomada em intervalos de tempo fixos vem,

$$s_n = s(X(n\Delta t)) + \varepsilon_n \quad (3.25)$$

De uma forma mais realista assume-se um erro de medição ε_n . Uma suposição usual na análise é a de que o processo de medição não se altera, ou seja se o sistema estiver no estado a e se corresponder a esse estado uma medição x então se mais tarde o sistema estiver novamente no estado a a medição do sistema será também igual a x . Uma reconstrução com dimensão m é formada pelos vectores,

$$\mathbf{s}_n = (s_{n-(m-1)\tau}, \dots, s_{n-\tau}, s_n) \quad (3.26)$$

onde m é o número de amostras. A diferença temporal em número de amostras, τ , ou em unidades de tempo, $\tau\Delta t$, entre os componentes do vector designa-se por desfasamento temporal.

O número de vectores no espaço de reconstrução será semelhante ao número de amostras consideradas para um desfasamento temporal correspondente a τ .

A pergunta incontornável que se coloca agora é a seguinte, até que ponto o objecto geométrico formado pelos vectores de reconstrução, $\mathbf{s}_n$, é equivalente à órbita original, X_n ?

Em que circunstâncias o atractor formado pelos vectores de reconstrução é equivalente ao original?

Relativamente a estas questões pertinentes convém esclarecer o que se pretende com o termo equivalente. Reconstrução é reconstituir um sistema original e o que é fundamental num sistema dinâmico é o seu determinismo, é esta propriedade que

se pretende explorar. Assim, a evolução do sistema dinâmico no espaço reconstruído deve depender única e exclusivamente do estado actual do sistema. Mas esta não é a única propriedade pretendida, uma vez que existe informação relevante na geometria do atractor, a reconstrução deve ser tal que mantenha as propriedades topológicas do atractor original. A manutenção da estrutura do espaço tangente, ou seja, garantir que a linearização da dinâmica em qualquer ponto está preservada pelo processo de reconstrução vai permitir obter as mesmas dimensões, expoentes característicos e entropias, etc.

Uma função, f , de classe C^1 em D é uma imersão se a sua derivada $D_f(x)$ for injectiva em todo o ponto de D ou, equivalentemente, se $D_f(x)$ tiver característica máxima em $T_x D$.

Um mergulho é uma aplicação, F , de uma variedade diferenciável compacta, X em Y é uma aplicação de classe C^1 e uma imersão injectiva e própria em X , ou seja, a imagem inversa de todo o compacto em Y é um conjunto compacto em X . Uma característica muito importante dos mergulhos é que tem as propriedades de um difeomorfismo entre o conjunto de partida e a sua imagem, $F(X)$ subvariedade de Y .

De facto, o mergulho é a situação ideal para uma aplicação de reconstrução uma vez que mantém todas as propriedades topológicas graças à sua bijectividade, ser de classe C^r e possuir inversa diferenciável pelo que os espaços tangentes e respectivas dimensões são preservadas. Coloca-se a questão, em que condições a projecção e consequente reconstrução se pode considerar um mergulho?

Whitney, [Whi36], mostrou que uma aplicação genérica diferenciável, F , de uma variedade diferenciável compacta de dimensão d , M , em $\mathbb{R}^{2d+1}$ é um difeomorfismo em M . Isto é, M e a sua imagem em $\mathbb{R}^{2d+1}$, $F(M)$ são difeomórficas.

O termo genérico significa que o conjunto de aplicações que verificam as condições de mergulho é aberto e denso no espaço das aplicações C^1 entre os respectivos conjuntos. Ser aberto em tal conjunto significa que uma perturbação arbitrariamente pequena sobre um mergulho não altera as suas características, isto é, a função perturbada continua a ser um mergulho. O facto de ser denso significa que qualquer aplicação está arbitrariamente próxima de um mergulho.

Uma explicação heurística do teorema de Whitney é apresentada em [Ott93] e [Aba96a], num espaço de dimensão d , um subespaço de dimensão d_1 e um de dimensão d_2 intersectam-se genericamente em subespaços de dimensão $d_1 + d_2 - d$ se este valor for negativo, não existe intersecção entre os subespaços. Quando os subespaços têm a mesma dimensão, d_A , como por exemplo uma intersecção da órbita com ela própria, as inteseccções não ocorrem quando $2d_A - d < 0$ ou seja, $d > 2d_A$.

Como indicam [Sau91] seria óptimo poder concluir-se do teorema de Whitney que $n = 2d + 1$ medições simultâneas fossem suficientes para reconstruir um estado da variedade A no espaço de medições $\mathbb{R}^n$. No entanto, subconjuntos abertos e densos mesmo em espaços euclidianos podem ser magros em termos de probabilidade. Um exemplo conhecido de uma união de conjuntos abertos e densos com medida de Lebesgue próxima de zero são as línguas de Arnold, [Rue89b].

É portanto de todo o interesse do experimentalista usar propriedades mais fortes e assim aumentar a probabilidade de se encontrar um mergulho. Assim em vez de conjuntos abertos e densos o experimentalista está interessado em garantir que a escolha de uma qualquer função resulte num mergulho com *probabilidade 1*.

Aqui reside um dos problemas que [Sau91] encontraram ao tentar a generalização é que o espaço de todas as aplicações de classe C^1 tem dimensão infinita e a noção de *probabilidade 1* não é uma generalização imediata daquela que existe para espaços de dimensão finita. Assim, [Sau91] introduziram para esse efeito um conceito novo, o de *prevalência*.

Definição 3.4.1 *Um subconjunto de Borel, S , de um espaço vectorial normado, V é prevalente se existir um subespaço, $E \in V$ de dimensão finita tal que para cada $\mathbf{v} \in V$, $\mathbf{v} + \mathbf{e}$ pertence a S para quase todo (no sentido de Lebesgue) o $\mathbf{e}$ em E .*

Em [Sau91] os autores apresentam uma generalização do teorema de Whitney substituindo o termo *genérico* por *com probabilidade 1* e substituindo a variedade, M , por um conjunto compacto invariante $A \in \mathbb{R}^k$ com uma dimensão capacidade ($\dim_K$) não inteira.

O teorema original de Whitney não permite obter uma estimativa da dimensão mínima para a qual quase toda a aplicação seja um mergulho. Suponha que se possui um conjunto invariante A de um fluxo em $\mathbb{R}^{100}$ com dimensão capacidade 1.4. Na ausência de informação acerca da possibilidade de compreender A numa variedade diferenciável de dimensão inferior a 100 o uso do teorema implicaria ter que se reconstruir a dinâmica usando pelo menos 201 dimensões. Assim, [Sau91] enunciaram o teorema,

Teorema 3.4.1 *Seja A um subconjunto compacto de $\mathbb{R}^k$ com dimensão capacidade, $\dim_k(A) = d$, e seja n um inteiro $> 2d$. Para quase toda a função C^1 $F : \mathbb{R}^k \rightarrow \mathbb{R}^n$,*

1. F é injectiva em A .
2. F é uma imersão de um subconjunto compacto C numa variedade diferenciável contida em A .

Apesar deste novo alento dado ao teorema de Whitney a verdade é que raramente se pode observar simultâneamente um grande número de variáveis independentes. De facto, na prática, é polémico falar-se de medidas ou observações independentes. Sugerido por [Rue78] posto em prática por [Pac80] mas apenas com o suporte teórico dado por [Tak80] o problema ultrapassa-se recorrendo às coordenadas desfasadas.

Definição 3.4.2 *Se F é um fluxo numa variedade, M , e T é um número positivo (o atraso) e, $h : M \rightarrow \mathbb{R}$ uma função de observação, de classe C^1 , define-se como função de coordenadas desfasadas,*

$$\Phi(h, F, T)(x) = (h(x), h(F^T(x)), h(F^{2T}(x)), \dots, h(F^{(d-1)T}(x))) \quad (3.27)$$

no caso de o tempo ser discreto e em vez de um fluxo, F , existir uma aplicação, f , a função de coordenadas desfasadas pode ser tal que não ocorra perda de iterações do

sistema, ou seja,

$$\Phi(h, f)(x) = (h(x), h(f(x)), h(f^2(x)), \dots, h(f^{(d-1)}(x))) \quad (3.28)$$

Uma vez que M é multidimensional e $h_n = h(x_n)$ é uma função unidimensional torna-se evidente que se terá que agrupar os elementos da série e o mais natural é fazê-lo respeitando a ordem temporal, ou seja, tomando valores sucessivos para construir os vectores,

$$y_n = (h_n, h_{n+1}, \dots, h_{n+d-1}) \quad (3.29)$$

O número de componentes d usados diz-se a dimensão do espaço de mergulho. Uma imposição coerente com o sistema original é que $\Phi(y_n) = y_{n+1}$.

O teorema enunciado a seguir é o da reconstrução de Takens para o caso do tempo ser discreto.

Teorema 3.4.2 *Seja D uma variedade compacta de dimensão d . Para os pares (S, h) , onde $S : D \rightarrow D$ é um difeomorfismo e $h : D \rightarrow \mathbb{R}$ é uma função de observação diferenciável. Ser um mergulho é uma propriedade genérica da função de coordenadas desfasadas, $\Phi(h, f) : D \rightarrow \mathbb{R}^{2d+1}$,*

$$\Phi(h, f)(x) = (h(x), h(f(x)), h(f^2(x)), \dots, h(f^{(2d+1)}(x))) \quad (3.30)$$

Considere-se a órbita periódica, A de um fluxo Φ . Na ausência de uma dinâmica, são necessárias 3 coordenadas independentes para reconstruir A , ou mais correctamente dizendo, quase toda a aplicação C^1 , $F = (f_1, f_2, f_3)$ de uma vizinhança de A em $\mathbb{R}^3$ é um mergulho. Suponha agora que apenas se possui o valor de uma função, h de $A : \mathbb{R}^3$ e a correspondente aplicação em coordenadas desfasadas. Pretende saber-se

se da mesma forma que acontece para a função genérica do enunciado do teorema (3.4.2), a aplicação $\Phi(h, F, T)$ é um mergulho.

Os autores em [Sau91] generalizaram este teorema substituindo variedades diferenciáveis por conjuntos compactos e em vez de uma dimensão inteira usou uma dimensão (fractal) capacidade qualquer,

Teorema 3.4.3 *Seja F um fluxo num subconjunto aberto, $U \in \mathbb{R}^k$ e seja A um subconjunto compacto de U com dimensão capacidade, d . Seja $n > 2d$ um inteiro e $T > 0$. Assumindo que A não contém ciclos limite com período T ou $2T$, apenas um número finito de pontos fixos, no máximo um número finito de pontos periódicos de período $3T, 4T, \dots, nT$ e que as linearizações desses pontos periódicos tenham valores próprios distintos. Então para quase toda a função C^2 , h em U a função de coordenadas desfasadas, $\Phi(h, F, T) : U \rightarrow \mathbb{R}^n$ é,*

1. Φ é injectiva em A .
2. Φ é uma imersão de cada subconjunto compacto C de uma variedade diferenciável contida em A .

Para o caso em que o tempo é discreto a extensão do teorema de Takens obtém-se substituindo D , variedade compacta por um conjunto fractal com dimensão capacidade, d_k e substituindo o termo genérico por *probabilidade 1* num espaço de probabilidade. O maior número de restrições relativamente ao teorema enunciado por Takens é justificado em [Sau91]. O principal resultado obtido pelos teoremas do mergulho é que não é a dimensão D do espaço de fase original que importa para a dimensão mínima do espaço de reconstrução mas sim a dimensão fractal da medida invariante gerada pela dinâmica do verdadeiro sistema [Kan99]. Sauer e Yorke, [Sau93], mostram que relativamente à dimensão correlação (D_C ou D_2) e caso se usem filtros lineares finitos sobre coordenadas desfasadas a dimensão do espaço de reconstrução apenas precisa de ser superior à dimensão correlação do atrator. Infelizmente o mesmo não se aplica à dimensão capacidade uma vez que o uso de filtros pode diminuir a dimensão

capacidade. As demonstrações dos teoremas enunciados nesta secção encontram-se em [Sau91].

Um dos aspectos relevantes dos teoremas de mergulho de fractais tem a ver com o facto de nada se mencionar quanto ao desfasamento temporal entre as observações pelo que qualquer um pode ser usado, no entanto, do ponto de vista do experimentalista as restrições com que se depara, especialmente no facto de não dispôr de uma órbita de tamanho infinito, torna a escolha do desfasamento temporal, τ , extremamente relevante. É esta questão que se vai abordar na secção seguinte.

3.4.1 Escolha do desfasamento temporal

O que se pretende atingir quando se escolhe um desfasamento temporal?

O valor tem que ser um múltiplo do intervalo de tempo amostral, Δ_t , uma vez que não existem mais dados e a interpolação é tão incerta como a estimação das derivadas de $s(t)$. Se for tão pequeno tal que não exista uma diferença significativa entre os valores sucessivos então verifica-se redundância, [Cas91b] e [Gib92]. Se as medições estiverem sujeitas a ruído, o que acontece quase sempre, e se o nível de ruído for superior à variação do sinal, $s(t)$, no intervalo de tempo coberto por um vector de reconstrução, $m\tau$, então os vectores assim formados não têm qualquer significado. Por outro lado, se τ for muito grande, acontece que as diferentes coordenadas estão pouco ou nada correlacionadas.

Existem na literatura, pelo menos uma dúzia de métodos diferentes para determinar τ . Tem-se verificado que alguns métodos para alguns sistemas dão melhores resultados que outros. Porém, como referem [Kan99] todos os métodos apontam valores da mesma ordem de grandeza. Por outro lado os resultados não devem depender excessivamente de τ dado que isso seria um sinal de que a reconstrução não preservou propriedades relevantes do atractor.

Uma vez que está em causa a correlação temporal entre as observações sucessivas, a *FAC* é a medida estatística mais antiga e mais utilizada no âmbito dos processos estocásticos, para quantificar a correlação temporal numa série.

Vários autores propõem diferentes critérios baseados na *FAC*, [Hol86] sugerem o desfasamento correspondente ao primeiro zero da função, [TE89] propõem o desfasamento correspondente ao nível 0.1 e [Sch88] aconselha o valor 0.5.

Como foi referido, uma das desvantagens da *FAC* é o facto de apenas quantificar a intensidade da ligação linear. Se a não linearidade estiver presente e dominar a relação entre as observações a *FAC* não é a medida mais indicada para caracterizar a relação, [Kan99].

Os autores em, [Fra86] propõem tomar, como estimativa de τ o desfasamento temporal correspondente ao primeiro mínimo da informação mútua média (ou redundância). A informação mútua não é mais do que a entropia entre várias medidas. Para o caso de duas medidas, representando por x, y as variáveis associadas às dimensões a entropia ou informação mútua média ou redundância média é,

$$\begin{aligned} I^{(xy)} &= \sum_{i,j} p_{i,j}^{(xy)} - \sum_i p_i^{(x)} \ln p_i^{(x)} - \sum_j p_j^{(y)} \ln p_j^{(y)} \\ &= \sum_{i,j} \frac{p_{i,j}^{(xy)}}{p_i^{(x)}} p_j^{(y)} \end{aligned} \quad (3.31)$$

caso se trate de uma medida de probabilidade, $p_{i,j}^{(xy)}$ correspondem a probabilidades conjuntas. Ou seja, se $p_i^{(x)}$ é a probabilidade de x se encontrar no conjunto i da partição do conjunto associado a esta variável e se $p_j^{(y)}$ a probabilidade de y se encontrar no conjunto j da partição do conjunto associado à variável y então $p_{i,j}^{(xy)}$ é a probabilidade de x estar em no conjunto i e y estar no conjunto j simultaneamente.

Num sistema determinístico, essas *probabilidades* calculam-se classificando os dados correspondendo os histogramas às partições do conjuntos. O uso da redundância na análise não linear de séries é em tudo análoga à utilização da *FAC*, no intervalo de valores correspondente à série constrói-se um histograma. Sendo p_i a probabilidade do sinal se encontrar na classe i e $p_{i,j}(\tau)$ a probabilidade de $s(t)$ estar na classe i e $s(t + \tau)$ na classe j a informação mútua para o desfasamento temporal, τ é,

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln p_{ij}(\tau) - 2 \sum_i p_i \ln p_i \quad (3.32)$$

para $\tau = 0$ as probabilidades conjuntas $p_{ij} = \delta_{ij}$, onde δ é o Delta de Kronecker corresponde à entropia de Shannon para a distribuição.

Numa situação de independência tem-se, $p_{i,j}^{(xy)} = p_i^{(x)} p_j^{(y)}$ e a informação mútua é zero. Não estando dependente, em termos de eficácia, da ligação funcional entre as variáveis a informação mútua parece ser a solução mais próxima do óptimo. No entanto colocam-se alguns problemas relacionados com este critério. Um deles diz respeito à classificação, apesar de numa situação ideal se esperar independência entre a redundância e a escolha das classes, o facto é que na prática, usando séries de tamanho limitado e dependendo do sinal (como se vai ver no capítulo 6) existe interacção entre os dois. Adicionalmente, o critério de escolha do desfasamento correspondente ao primeiro mínimo de $I(\tau)$ está sujeito à sua existência que como indicam [Kan99] ainda não se demonstrou que existe.

Não existe um método infalível para identificar o espaçamento óptimo entre as observações. As correlações temporais estão presentes em quase todo o tipo de dados. O problema da determinação do valor óptimo para τ foi resolvido por [PM92] fazendo uso do *space time plot*. Aqueles autores argumentam que, na presença de correlações temporais a probabilidade de um dado par de pontos estarem a uma determinada distância ε não depende apenas da distância mas também do intervalo de tempo entre eles, Δt . Esta dependência temporal e respectiva forma e pode ser estudada num gráfico que represente a evolução do número de pares de pontos em função do tempo e espaço. Esse gráfico pode fazer-se criando para cada separação temporal Δt um histograma acumulado de distâncias espaciais ou usar superfícies de linhas de contorno. No primeiro caso considera-se que se obteve uma independência entre as variáveis quando o gráfico, para os vários percentis, seja aproximadamente rectilíneo e paralelo ao eixo dos xx .

Como referem [Kan99] existem séries para as quais as curvas nunca estabilizam, é a situação usual em dados cujas baixas frequências do espectro são as mais relevantes, o que acontece por exemplo com o movimento browniano e com os processos *ARMA* fraccionários.

3.4.2 Escolha da dimensão do espaço de reconstrução

A escolha do desfasamento temporal é apenas o primeiro passo do processo de reconstrução do espaço de fase. De acordo com o teorema (3.4.3) é necessário conhecer a dimensão capacidade do atrator para se obter uma dimensão mínima suficiente para o espaço de reconstrução, convém não esquecer que um dos pressupostos do teorema é a inexistência de ruído. Pretende-se determinar a dimensão global inteira com o numero necessário de coordenadas de forma a que as órbitas geradas pela série não se autointersectem. A dimensão mínima para a qual não ocorrem autointersecções denomina-se dimensão de *embedding*. A possibilidade de se reconstruir a dinâmica do sistema num espaço euclidiano é um passo fundamental para inferir outras características relevantes sobre o sistema como por exemplo, o expoente máximo de Lyapunov, a entropia e as dimensões do atrator.

A maior parte dos sistemas caóticos conhecidos não têm dimensões muito grandes, poder-se-ia pensar que a utilização de espaços de grandes dimensões resolveria a questão, no entanto, como é sabido, a instabilidade do caos faz com que a ligação dinâmica entre o primeiro e o último elemento do vector de reconstrução diminua à medida que, m , dimensão do espaço, aumenta. Uma questão importante que se prende com a dimensão do sistema tem a ver com a relação que tem com o espaçamento temporal entre as observações. Como refere [Aba96a], se 2 quantidades, $s_A(n)$ e $s_B(n)$ medidas do mesmo sistema com diferentes espaçamentos temporais não existe garantia que a dimensão global, m , indicadas por cada uma delas seja idêntica, uma vez que cada medida corresponde a uma diferente aplicação não linear do verdadeiro espaço de fase para o espaço reconstruído. Esta afirmação estabelece um certo grau de dependência entre a dimensão de mergulho, m , e o espaçamento temporal, τ . Os

autores em, [AT93] referem ainda que o teorema de Takens indica uma dimensão suficiente que é baseada em considerações topológicas, mas uma dimensão suficiente nem sempre é necessária. Aqueles autores fazem uma distinção entre dimensão local e global indicando que em algumas situações as autointersecções deixam de existir em espaços de dimensão inferior à indicada pelo teorema. A vantagem em diminuir a dimensão tem a ver com a componente de ruído que, como é sabido, tem dimensão elevada e aumenta linearmente com o espaço em que se *mergulha* o atractor o que faz com que possam existir direcções contendo pontos espúrios sem relação com a dinâmica.

Na literatura encontram-se vários métodos que podem ajudar na escolha de uma dimensão. O mais utilizado e conceptualmente mais simples é o dos falsos vizinhos próximos, [Ken94].

3.4.3 Falsos vizinhos próximos

Sendo m_0 uma dimensão adequada para o espaço de reconstrução suponha que $m_0 - 1$ não é, o que acontece quando se diminui a dimensão? O conjunto de pontos no espaço com menos uma dimensão corresponde a uma projecção. Como o sistema é visto como uma sucessão de pontos que são simultâneamente objectos e imagens uns dos outros o que acontece é que pontos afastados em m_0 dimensões podem ficar próximos se projectados. Pode considerar-se que a projecção corresponde a uma quebra no determinismo, de facto, é quebrada a ligação dinâmica entre os pontos. O algoritmo verifica para cada ponto em dimensão m se o seu vizinho mais próximo continua a revelar uma proximidade métrica na dimensão, $m + 1$. Sendo,

$$S_n = (s_{n-(m-1)\tau}, \dots, s_{n-\tau}, s_n) \quad (3.33)$$

um ponto qualquer do espaço reconstruído existirá um outro ponto,

$$S^{NN} = (s_{p-(m-1)\tau}, \dots, s_{p-\tau}, s_p) \quad (3.34)$$

onde, $\{p - (m - 1)\tau, \dots, p\} \cap \{n - (m - 1)\tau, \dots, n\} = \emptyset$. Seguindo [Aba96a], o vizinho mais próximo é um falso vizinho se,

$$\frac{|s_{n+\tau} - s_{p+\tau}^{NN}|}{\|S_n - S_p^{NN}\|} > R_T \quad (3.35)$$

onde R_T é um limite que não sendo rígido não é totalmente arbitrário. Em [Aba96a] recomenda-se que $R_T = 15$. Na ausência de ruído, isto é, observando um sistema não perturbado, espera-se que a percentagem de falsos vizinhos decaia de cerca de 100% para 0% quando se atingir o valor de dimensão adequado e aí se mantenha para valores superiores. Na presença de ruído, como mostra [Aba96a], a situação pode ser diferente. Aquele autor refere ainda que se todas as regiões do atractor estiverem adequadamente amostradas a variação do número de falsos vizinhos com o número de amostras é pequeno.

Quando o número de pontos é limitado e a dimensão do espaço de mergulho é elevada os vizinhos mais próximos de cada ponto podem não estar muito próximos e tender a estar na periferia do espaço uma vez que quando volume o disponível aumenta, aumenta a distância entre os pontos. Assim, o facto de dois pontos serem os mais próximos um do outro não significa que estejam próximos relativamente à escala do atractor. De facto, se o vizinho mais próximo de S_n não estiver próximo vem que $\|S_n - S_p^{NN}\| \approx R_A$, onde R_A é o desvio-padrão dos dados, por conseguinte a distância em $m + 1$ dimensões é aproximadamente o dobro daquela verificada em m dimensões. Em [Ken94] propõem-se então que também se considere um vizinho próximo como falso se,

$$\frac{|s_{n+\tau} - s_{p+\tau}^{NN}|}{R_A} > 2 \quad (3.36)$$

onde $|s_{n+\tau} - s_{p+\tau}^{NN}|$ é a distância em $m + 1$ dimensões, [AT93].

Por detrás deste método está uma ideia geométrica simples, atinge-se uma dimensão adequada para um mergulho do atractor se a sua projecção por uma variável observada não mostrar auto-intersecções. Porém, como indica [Aba96a], apesar do método resultar bem na prática, inclusivé na presença de ruído, é importante referir que não existe nenhum teorema relativo a ele pelo que a escolha de uma dimensão adequada deve ser complementada com outras análises.

Uma vantagem do método dos falsos vizinhos é o facto de encarar o problema da estimação do espaço de mergulho directamente e não a partir de um funcional como o integral de correlação, quando se supõe que se observa um sistema determinístico. A aplicação deste método pressupõe o uso de uma distância, procedimento que encerra alguma arbitrariedade.

3.4.4 Filtros lineares sobre vectores de reconstrução

Uma importante ferramenta em análise de sistemas dinâmicos é a rotação dos eixos de coordenadas baseado no *singular value decomposition* (SVD), [Bro94], cuja contraparte em estatística tem o nome de análise em componentes principais (ACP). O objectivo deste método no contexto da análise de séries temporais é a caracterização da série usando apenas os seus componentes mais relevantes num espaço de mergulho $\mathbb{R}^m$ onde m está, provavelmente, sobredimensionado.

Os vectores de reconstrução formam uma nuvem de pontos em $\mathbb{R}^m$, e existem direcções em que a nuvem se expande e outras em que se contrai. A ACP permite determinar um conjunto de subespaços unidimensionais ordenado pela sua importância na explicação da variância da série.

Seguindo [Cha01], seja $\{X(t) : \alpha \leq \beta\}$, $(\alpha, \beta) \in \mathbb{R}$ um processo complexo em tempo

continuo e contínuo em média quadrática com $\text{Var}(X(t)) < \infty$ para todo o t . Seja $\gamma(s, t) = \text{cov}(X(s), X(t))$ a função de autocovariância pode mostrar-se, Yaglom(1986), que,

$$\gamma(s, t) = \sum_{j=1}^{\infty} \lambda_j \psi_j(s) \bar{\psi}_j(t) \quad (3.37)$$

é uma série uniformemente convergente onde os λ_j 's são os valores próprios positivos e as funções $\psi_j(s)$ são funções próprias tais que $\int_{\alpha}^{\beta} \psi_j(t) \bar{\psi}_k(t) dt = \delta_{jk}$, sendo δ_{jk} a função delta de Kronecker. Um outro resultado da teoria de Karhunen-Loève mostra que (3.37) é equivalente à representação espectral generalizada de $X(t)$ ou expansão de Karhunen-Loève,

$$\begin{aligned} X(t) &= \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j \\ \text{com} \\ Z_j &= \lambda_j^{-1/2} \int_{\alpha}^{\beta} \mathbf{X}(t) \bar{\psi}_j(t) dt \end{aligned} \quad (3.38)$$

tal que $\text{cov}(Z_j, Z_k) = \delta_{jk}$, isto é as variáveis $\{Z_i\}_{i=1}^{\infty}$ são ortogonais.

Em particular, se $t \in \{1, 2, \dots, M\}$ e $\mathbf{X}$ for real então trata-se de uma colecção finita de variáveis aleatórias $\mathbf{X} = (X(1), \dots, X(M))$ e a expansão de Karhunen-Loève não é mais do que uma ACP. Assim $\gamma(s, t)$ é a matriz de variância-covariância de $\mathbf{X}$ a que chamaremos S podendo escrever-se como $S = \sum_{j=1}^M \lambda_j \mathbf{e}_j \mathbf{e}_j^T$ onde os $\mathbf{e}_j$ são agora vectores próprios associados aos valores próprios λ_j 's. Uma vez que $\mathbf{e}_j^T \mathbf{e}_k = \delta_{jk}$ vem,

$$\begin{aligned}\mathbf{X} &= \sum_{j=1}^M \sqrt{\lambda_j(t)} \mathbf{e}_j Z_j \\ \text{onde} \\ Z_j &= \lambda_j^{-1/2} \mathbf{X}^T(t) \mathbf{e}_j\end{aligned}\tag{3.39}$$

tal que $\text{cov}(Z_j, Z_k) = \delta_{jk}$. Para uma amostra de tamanho N de $\mathbf{X}$ com matriz de covariância, S obtém-se valores e vectores próprios amostrais,

$$\begin{aligned}\hat{\mathbf{X}}_j &= \sum_{i=1}^M \sqrt{\hat{\lambda}_i} \hat{\mathbf{e}}_i \hat{Z}_i^{(j)} \\ \text{onde} \\ \hat{Z}_i^{(j)} &= \hat{\lambda}_i^{-1/2} \mathbf{X}_j^T \hat{\mathbf{e}}_i\end{aligned}\tag{3.40}$$

vem que $\hat{\mathbf{X}}_j = P \hat{\mathbf{X}}_j$ onde $P = \sum_{i=1}^M \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T$. A matriz P é uma matriz de projecção que projecta $\mathbf{X}$ no subespaço gerado pela base $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_M\}$. Supondo que os valores próprios, $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ onde $(k < M)$ são os mais significativos, ou seja, explicam uma parte significativa da variação de $\mathbf{X}$ então pode projectar-se $\mathbf{X}$ no subespaço gerado pelos vectores próprios correspondentes. Assim a projecção escreve-se como,

$$\hat{\mathbf{X}}_{j/k} = \left[\sum_{i=1}^k \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \right] \mathbf{X}_j\tag{3.41}$$

No contexto das séries temporais a *ACP* tem que ser adaptada. Seja $(X_1, \dots, X_N)$ a série observada subtraída da sua média. Sendo m a dimensão do espaço de mergulho pode construir-se a matriz de covariâncias,

$$C = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} S_i S_i^T \quad (3.42)$$

onde S_i é o i -ésimo vector de reconstrução tal como em (3.26). Sendo uma matriz simétrica real, os valores próprios de C são reais e os vectores próprios correspondentes são ortogonais. Os valores próprios, λ_i , correspondem ao quadrado dos semieixos do hiper-elipsóide que melhor se ajusta à nuvem de pontos e os correspondentes vectores próprios são ortogonais. As direcções mais relevantes correspondem aos maiores valores próprios. Sejam $\hat{\lambda}_1, \dots, \hat{\lambda}_m$ os valores próprios e $\hat{b}_1 \dots \hat{b}_m$ os vectores próprios correspondentes. Sendo k o número de valores próprios significantes no sentido que *explicam* uma percentagem da variância tida como suficiente. Os novos vectores de reconstrução, $\hat{S}_n$ são agora a projecção dos antigos, $S_n = (s_{n-m+1}, \dots, s_n)$, num espaço de dimensão k . No pressuposto que o atractor original existe num espaço de dimensão $k < m$ então $(m - k)$ das coordenadas são redundantes.

O resultado é então uma reconstrução num espaço de dimensão k . Em [Bro94] os autores mostram que quando $m \rightarrow N$ esta transformação tem as propriedades de uma transformada de Fourier. Para um m pequeno a ACP escolhe estruturas relevantes do espaço em vez das frequências relevantes. A ACP não permite, de forma única, a reconstituição da série escalar de tamanho Nk/m , correspondente à transformada inversa de Fourier, após truncagem das frequências que mais contribuem para a variância.

As transformações lineares ou pelo menos diferenciáveis sobre vectores de reconstrução não afectam a validade dos resultados [Kan99].

3.5 Medidas *SRB* e dimensões de medidas invariantes

As propriedades que as medidas *SRB* possuem fazem-nas boas candidatas para as medidas físicas referidas anteriormente. Pode mostrar-se que a média temporal de uma órbita, tende para uma medida *SRB*, ρ , quando $n \rightarrow \infty$ para todo o x num conjunto de medida de Lebesgue positiva. A medida de Lebesgue corresponde a uma noção de amostragem ou estatística mais natural do que a medida ρ inerente ao atrator que é, geralmente, singular.

Existe ainda o seguinte resultado para medidas *SRB* em sistemas do tipo Axioma-A, [Eck85],

Teorema 3.5.1 *Considere um sistema dinâmico determinado por um difeomorfismo- C^2 em tempo discreto numa variedade m -dimensional, M . Suponha que A é um atrator do tipo Axioma-A com base de atracção, U , então,*

- *Existe uma e uma só medida *SRB* com suporte em A .*
- *Existe um $S \subset U$ tal que $U \setminus S$ tem medida de Lebesgue nula e,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{f^k(x)} = \rho \forall x \in S \quad (3.43)$$

Pugh e Shub, [Pug88], mostraram que o requisito Axioma-A pode ser substituído por uma menor exigência bastando que não existam expoentes característicos nulos, [Eck85]. Provavelmente o aspecto mais importante de um conjunto é a sua dimensão. Usando uma linguagem corrente, a dimensão é a quantidade de informação necessária para especificar/localizar pontos de forma precisa. Em geral, a caracterização/descrição do tamanho de um conjunto de $\mathbb{R}^m$ implica naturalmente o uso da medida de Lebesgue. No estudo de sistemas não conservativos é usual a ocorrência de conjuntos de medida de Lebesgue nula, uma vez que a cada iteração do sistema

o volume *encolhe*, sem contudo desaparecer. É o caso dos atractores referidos na secção anterior e dos atractores estranhos em sistemas caóticos. Do ponto de vista do conceito clássico de dimensão, que consiste na possibilidade do estabelecimento de um difeomorfismo com um aberto de $\mathbb{R}^d$, dimensão do conjunto, o ponto fixo tem dimensão 0 e um ciclo limite, que é uma curva fechada, tem dimensão 1. O caso dos atractores estranhos é mais complicado devido à peculiaridade da sua geometria que, podendo ser descontínua e não numerável, sai fora da classe das variedades diferenciáveis. Mas porque é que, no contexto dos sistemas dinâmicos, se fala apenas dos atractores estranhos como conjuntos de determinação dimensional não trivial? O que caracteriza estes atractores é o facto de o sistema que os gera revelar *DSCI*. Numa secção anterior estabeleceu-se a relação entre esta propriedade e a existência de expoentes característicos positivos associados à medida natural do atractor. Os expoentes característicos, como invariantes da dinâmica, estão relacionados com a frequência com que as diferentes partes do atractor são visitadas, o que determina a sua geometria. A atribuição de uma dimensão a este tipo de conjuntos deve ser tal que caracterize quantitativamente a sua estrutura geométrica. Mais à frente, quando se falar da análise de sistemas observados, usando séries temporais, o leitor vai aperceber-se da importância da determinação experimental de um valor para a dimensão a partir dos dados.

Num contexto mais geral, uma definição de dimensão mais abrangente que a clássica permitirá, em qualquer situação, a comparação de conjuntos com medida de Lebesgue nula.

Seguindo [Rue89a], seja $A \neq \emptyset$ um conjunto munido de uma métrica no espaço cartesiano N -dimensional. Seja $N(r, A)$ o número mínimo de bolas abertas de raio r necessárias para cobrir A . Então a dimensão capacidade de A ($\dim_k(A)$) é,

$$\dim_k(A) = \lim_{r \rightarrow 0} \sup \frac{\ln(N(r, A))}{\ln(1/r)} \quad (3.44)$$

Exemplo 3.5.1 Um conjunto Λ diz-se de Cantor se for fechado, totalmente desconexo

e um subconjunto perfeito de um $I = \{x : 0 \leq x \leq 1\}$. A família quadrática de sistemas, $F_\mu(x) = \mu x(1 - x)$ (função logística) dá origem a conjuntos de Cantor. Definindo, $A_0 = \{x : F(x) \notin I\}$, isto é, A_0 é o conjunto de pontos que numa iteração saem fora de I e define-se $A_n = \{x \in I : F^n(x) \in A_0\}$, ou seja, A_n é o conjunto de pontos que saem fora de I na iteração $n + 1$. A sucessão de conjuntos A_n pode ser usada para designar o conjunto de pontos cuja evolução nunca sai de I que é, $\Lambda = I - (\cap_{n=0}^{\infty} A_n)$. Pode mostrar-se, [Dev89], que o conjunto Λ é um conjunto de Cantor, cuja construção é remanescente da construção do conjunto de Cantor dos meios-terços. Este conjunto obtém-se retirando sucessivamente os intervalos abertos dos meios-terços dos conjuntos que se obtêm de $I = [0, 1]$.

O conjunto de pontos que sobram na retirada sucessiva dos meios-terços são conjuntos de medida de Lebesgue zero uma vez que o comprimento total dos conjuntos remanescentes é $(2/3)^n$ que tende para zero quando $n \rightarrow \infty$. Além disso os conjuntos são não numeráveis uma vez que é possível estabelecer uma bijecção com os conjuntos de I . Cada ponto de um conjunto de Cantor pode ser localizado estabelecendo a seguinte relação, tendo em conta cada iteração, usar-se-à o número 1 se o conjunto em que ele se encontra na iteração seguinte estiver à direita e 0 se estiver à esquerda. Assim a localização de um ponto tem a forma, 10010101... que não é mais do que a representação binária dos números em I . O número de bolas necessário para cobrir os conjuntos de Cantor é da ordem,

$$N(r) \approx r^{-D_0} \quad (3.45)$$

onde $D_0 = \ln 2 / \ln 3$, ou seja, se se assumir que apenas se precisa, para localizar pontos no conjunto, com precisão r , então $N(r)$ dá, grosso modo falando, quanta informação se necessita para o fazer (a informação é a localização dos $N(r)$ conjuntos) e D_0 dá a taxa com que a informação necessária aumenta à medida que a precisão aumenta também. Outra característica importante é a auto-semelhança deste tipo de

conjuntos, no sentido de que, em qualquer iteração é possível obter por aumento de escala a estrutura visível em escalas maiores. Para o conjunto de Cantor $[2/3, 7/9]$ multiplicado por um factor de 9 resulta numa estrutura semelhante ao intervalo $[0, 1]$.

Seguindo [Rue89b], dado um conjunto $S \neq \emptyset$ munido de uma métrica e um raio, $r > 0$, seja σ uma cobertura de S para uma família numerável de subconjuntos σ_k de diâmetro $r_k \leq r$. Dado um real, $\alpha \geq 0$, defina-se,

$$m^\alpha(S) = \lim_{r \rightarrow 0^+} m_r^\alpha(S) \quad (3.46)$$

onde $m_r^\alpha(S) = \inf_{\sigma} \{\sum_{k=1}^{\infty} r_k^\alpha\}$.

Quando $r \downarrow 0$, $m_r^\alpha(S)$ aumenta para o limite $m^\alpha(S)$ a que se chama medida de Hausdorff de S na dimensão α . Define-se,

$$\dim_H S = \sup \{\alpha : m^\alpha(S) > 0\} \quad (3.47)$$

a que se chama dimensão de Hausdorff de S . A medida de Hausdorff, $m^\alpha(S) = +\infty$ para $\alpha < \dim_H(S)$ e $m^\alpha(S) = 0$ para $\alpha > \dim_H(S)$. Assim α é a dimensão de Hausdorff de S . As 2 propriedades básicas da dimensão de Hausdorff são, [Rue89a],

$$\begin{aligned} \dim_H(S) &\leq \dim_H(S') \text{ se } S \subset S' \\ &\text{e} \\ \dim_H(\cap_n S_n) &= \sup_n \dim_H S_n \end{aligned} \quad (3.48)$$

Para todo o compacto, S , vem $\dim_H(S) \leq \dim_K(S)$. A dimensão-informação de uma medida define-se à custa da dimensão de Hausdorff de um seu conjunto de medida 1. Dada uma medida, ρ , a sua dimensão-informação, $\dim_h \rho$ é a mais pequena dimensão de Hausdorff de um conjunto S de medida ρ igual a 1. Como indicam [Eck85], uma vez que S não é em geral um conjunto fechado, a dimensão de Hausdorff do suporte de ρ , $\dim_H(\text{supp} \rho)$ pode ser estritamente maior do que $\dim_H \rho$.

Para o conjunto de Cantor, a dimensão de Hausdorff é igual à dimensão capacidade. As dimensões de Hausdorff e capacidade têm em comum o facto de se basearem em conceitos geométricos baseados na métrica de conjunto e não em conceitos estatísticos ligados às medidas de probabilidade associadas ao atractor. Mandelbrot, [Man77], usou o termo fractal para designar os conjuntos com dimensão de Hausdorff não inteira. A importância da dimensão de um conjunto vai depender da maior ou menor facilidade com que se pode calcular experimentalmente. A dimensão informação como se vai ver a seguir pertence a uma classe de medidas que se baseiam na medida do atractor e não na sua métrica como acontece com a dimensão de Hausdorff.

Uma dimensão fractal simples para ρ é dada em termos da densidade de probabilidade de uma bola de raio r . Seja, $B_r(X) = \{Y : d(X, Y) \leq r\}$ a bola onde d é uma distância em $\mathbb{R}^p$ (Euclidiana ou outra).

Teorema 3.5.2 *Seja ρ uma medida de probabilidade numa variedade compacta de dimensão finita, M . Suponha que,*

$$\lim_{r \rightarrow 0} \frac{\ln(\rho[B_X(r)])}{\ln(r)} = \alpha \quad (3.49)$$

q. t. p.- ρ . Então $\dim_H \rho = \alpha$. Young(1982).

No caso de a medida ρ ser ergódica, a existência do limite (3.49) implica que α seja independente de X , [Ott93], isso define uma dimensão única, α que é a dimensão



informação de ρ .

Em particular, Young mostrou que várias outras dimensões como as dimensões generalizadas de Rényi são iguais no caso em que ρ é ergódica. Quando um atrator tem uma dimensão pontual diz-se um fractal homogêneo. No entanto, pode acontecer que,

$$\rho[B_{\mathbf{x}}(r)] \approx r^{\alpha(\mathbf{x})} \quad (3.50)$$

tal que α dependa de $\mathbf{x}$, nesse caso, o atrator é um fractal heterogêneo ou multifractal. A medida associada a um atrator desse tipo não é ergódica. Nesta situação a frequência com que algumas partes do atrator são visitadas varia drasticamente de parte para parte e pode acontecer que uma pequena região do atrator contenha quase todas as permanências do sistema, ou seja, uma concentração da medida natural do atrator, [Ott93]. Impõe-se, então, uma definição de dimensão que permita distinguir esses diferentes tipos de atrator e que generalize a dimensão capacidade. Devido ao trabalho de [Rén71], já mencionado quando se falou na secção anterior da entropia de uma medida, existe uma família de dimensões que diferem na forma como diferentes regiões com diferentes densidades são ponderadas fazendo *médias* das dimensões locais, $\alpha(\mathbf{x})$. No contexto da dinâmica caótica as dimensões de Rényi foram introduzidas por [GP] e [Hen83]. Originalmente, as dimensões generalizadas de Rényi, D_α foram definidas como,

$$D_\alpha = \frac{1}{1-\alpha} \lim_{\varepsilon \rightarrow 0} \frac{\ln I(\alpha, \varepsilon)}{\ln(1/\varepsilon)} \text{ onde,} \quad (3.51)$$

$$I(\alpha, \varepsilon) = \sum_{i=1}^{N(r)} p_i^\alpha \quad \alpha > 0$$

e $N(r)$ é o número de cubos de aresta r necessários para cobrir o atrator. O seu cálculo implica o estabelecimento de uma grelha onde se seleccionam os cubos de aresta r e

a probabilidade de o sistema se encontrar no cubo i é, $p_i(\mathbf{x}) = \int_{\mathbf{x}} d\rho(\mathbf{x})$. Tratando-se de um conjunto fractal, o centro dos cubos pode não ser um ponto do atrator pelo que alguns deles terão associada uma probabilidade de serem visitados pelo atrator. Assim, a probabilidade de o sistema se encontrar em A que admite uma partição, $A = \cup A_i$, é, $\langle A \rangle = \sum_i p_i A_i$. Usando esta notação pode reescrever-se a equação (3.51) como $\langle p^{\alpha-1} \rangle_{\rho} \approx r^{(\alpha-1)D_{\alpha}}$. A definição de D_{α} pode ainda generalizar-se usando uma escolha aleatória relativamente à medida, ρ , cubo de raio r_i e probabilidade p_i . Desta forma, pode definir-se indirectamente,

$$\langle \frac{p_i}{r_i^{D_{\alpha}}} \rangle \approx \text{constante } \forall r_i \rightarrow 0 \quad (3.52)$$

Se o resultado não depender da escolha dos dominios podem usar-se bolas de raio r com centro em pontos seleccionados aleatoriamente segundo ρ . Desta forma, a equação (3.50) escreve-se,

$$\langle p_r(\mathbf{x})^{\alpha-1} \rangle_{\rho} = \int_{\mathbf{x}} p_r(\mathbf{x})^{\alpha-1} d\rho(\mathbf{x}) \equiv C_{\alpha}(r) \quad (3.53)$$

também se chama a esta expressão, integral de correlação generalizado. Para um conjunto auto-semelhante, $C_{\alpha}(r) \propto r^{(\alpha-1)D_{\alpha}}$ quando $\alpha \rightarrow 0$. As dimensões generalizadas podem escrever-se como,

$$D_{\alpha} = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha - 1} \frac{\ln C_{\alpha}(r)}{\ln r} \quad (3.54)$$

A natureza multifractal do atrator estranho está relacionada com o espectro de valores de α associados à medida invariante, ρ , transportada por ele. Alguns autores, [Gra88] consideraram uma partição do atrator em conjuntos, Σ_{α} , onde a medida de

probabilidade, ρ tem um comportamento de acordo com um expoente α no intervalo $\{\alpha, \alpha + d\alpha\}$:

$$\Sigma_\alpha = \{x \in A : \rho[B_x]\} \approx r^\alpha \text{ quando } r \rightarrow 0 \quad (3.55)$$

O caso particular em que $\alpha = 1$ obtém-se usando a regra de L'Hopital, o que dá a dimensão informação,

$$d_1 = \lim_{r \rightarrow 0} \frac{\langle \ln p_r \rangle_\rho}{\ln r} \quad (3.56)$$

Desta expressão advém o facto de se chamar informação à dimensão D_1 uma vez que $\langle \ln p_r \rangle_\rho$ é, como se fez referência anteriormente, a informação média (de Shannon) necessária para localizar um ponto x com precisão r . Saliente-se o facto de a dimensão informação pelo facto de se basear na informação de Shanon verifica a propriedade aditiva.

Uma outra propriedade de D_α é o seu não crescimento com α . Pode mostrar-se, [Ott93] que,

$$D_{\alpha_1} \leq D_{\alpha_2} \text{ se } \alpha_1 > \alpha_2 \quad (3.57)$$

Sejam $\lambda_1 \geq \dots \lambda_p$ (o espectro de Lyapunov) os expoentes de Lyapunov de uma função f relativamente a uma medida invariante ρ , define-se a dimensão de Lyapunov da medida como,

$$\dim_\Lambda \rho = k + \frac{\lambda_1 + \dots + \lambda_k}{|\lambda_{k+1}|} \quad (3.58)$$

onde $k = \max \left\{ i : \sum_{j=i}^i \lambda_j \right\}$. Uma vez que os volumes de dimensão superior a $k + 1$ são contraídos pela evolução sugere que a dimensão de ρ deva ser inferior a $k + 1$, [Ily83]. Se ρ for uma medida *SRB* então, em geral, $\dim_H \rho = \dim_A \rho$. A relação entre a dimensão de Lyapunov com outras dimensões fractais são discutidas em [Eck85] e [Rue86].

A propriedade *mixing* é importante no contexto dos sistemas ergódicos. Um sistema ergódico pode ter alguma propriedade *mixing* associada à sua medida invariante, existem vários tipos de propriedades *mixing*. O essencial desta propriedade é que a evolução do sistema vai-se aproximando de uma situação de independência do passado.

3.6 Estimação de invariantes do sistema

Um sistema que exiba *DSCI* é um sistema caótico e sendo dissipativo possui um atractor estranho. Estando confinado a uma região de volume finito no espaço de fase e sendo uma região de onde nenhuma órbita poderá sair que geometria se espera para o atractor e de que forma as órbitas evoluem uma vez que não se podem intersectar? Considere um sistema do tipo (3.2) com as propriedades (3.3). Um facto essencial é a existência de uma medida de probabilidade, π , f -invariante em $\mathcal{M}$, conjunto atractor onde as órbitas evoluem. O conjunto $\mathcal{M}$ pode ser um espaço vectorial ou uma variedade diferenciável compacta.

A teoria ergódica constitui uma abordagem estatística ao estudo da dinâmica no atractor. Uma distribuição de probabilidade estacionária, ρ em $\mathcal{M}$ é f -invariante se

$$\forall A \subseteq \mathcal{M} : \pi(f^{-1}(A)) = \pi(A) \quad (3.59)$$

Se f for invertível então é equivalente a dizer que $\pi(f(A)) = \pi(A)$, desta forma f preserva uma distribuição estacionária. Suponha que X_0 é escolhido aleatoriamente em $\mathcal{M}$ usando uma distribuição de probabilidade estacionária, π . Então os iterados

$$f(x) = \begin{cases} 2x & x \in [0, \frac{1}{2}) \\ 2x - 1 & x \in [\frac{1}{2}, 1) \end{cases} \equiv 2x(\text{mod } 1) \quad (3.65)$$

A função f pode ser vista como uma aplicação no círculo unitário, $S^1 \rightarrow S^1$ por parametrização dos pontos do intervalo $[0, 1)$. Sabendo que $x_0 \in [0, 1)$ tem uma representação decimal binária da forma,

$$x_0 = 0.a_1a_2a_3\ldots = \sum_{j=1}^{\infty} 2^{-j}a_j, \quad a_j \in \{0, 1\} \quad (3.66)$$

A função f transforma x_0 em, $x_1 = f(x_0) = 0.a_2a_3, \ldots$ e assim sucessivamente. Uma diferença no dígito a_40 é pouco significativa no entanto este algarismo será o mais significativo após 40 iterações do sistema o que significa que uma pequena perturbação do valor inicial torna a evolução do sistema totalmente diferente. Este sistema apresenta também vários pontos periódicos e ciclos limite, basta ver que $x = 0.01010101010$ se repete em 2 iterações. Este tipo de operações numa cadeia de símbolos denomina-se shift. Um shift é uma aplicação, $\sigma : \Omega \rightarrow \Omega$ onde o elemento genérico é da forma $\omega = (\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots)$ de elementos de um conjunto finito, $\mathbf{V}$ e $(\sigma(\omega))_n = \omega_{n+1}$, o elemento sucessor, [Rue86].

Se a cada elemento de $\mathbf{V}$ for associada uma probabilidade, p_i obtém-se uma medida de probabilidade, ρ em $\mathbb{R}$ tal que $P(\omega_n = i) = p$ não depende de n . Esta medida diz-se de Bernoulli e o shift diz-se de Bernoulli do par (ρ, σ) , [Mañ83]. A medida ρ é shift-invariante. No caso $V = \{0, 1\}$ a escolha $p = 1/2$ é particularmente interessante uma vez que é preservada por f , diz-se f -invariante. Para estes sistema o conjunto dos pontos periódicos é denso e numerável. Para este sistema e também para a aplicação tenda, $x_{n+1} = 1 - 2|x_n - 1/2|$, um histograma da fracção de tempo que uma órbita com valor inicial em $(0, 1)$ está em cada um dos intervalos $[(m-1)/N, m/N]$, $m = 1(1)N$ converge para $1/N$. Para qualquer intervalo $[a, b]$ em $[0, 1]$ pode definir-se uma função que dá a fracção de tempo que uma órbita está nesse intervalo,

$$\rho([a, b]) = \int_a^b \rho(x) dx \quad (3.67)$$

para as aplicações indicadas, $\rho(x) = 1$ em $[0, 1]$. Desta forma, $\rho(x)$ diz-se uma densidade natural invariante. É uma medida produzida automaticamente pelo tempo que o sistema está nas várias partes do atrator. Se estiverem verificadas mais algumas condições obtém-se medidas que alguns autores [Rue89a], [Eck85] chamaram de medidas físicas.

As únicas medidas invariantes de interesse prático são as medidas físicas uma vez que são as únicas observáveis, por exemplo, a medida δ -Dirac num ponto fixo instável, é invariante mas não observável.

3.6.1 Estimação dos Expoentes de Lyapunov

Seguindo [Lu94], vamos introduzir os expoentes de Lyapunov num contexto mais geral, o das sucessões estacionárias de matrizes aleatórias e depois particularizamos para os sistemas determinísticos.

Considere uma sucessão estacionária de matrizes $p \times p$, $A_1, A_2, \dots$ e seja $\log^+ x = \max\{0, \log x\}$, onde $x > 0$, assume-se que, $E [\log^+ \|A_1\|] < \infty$.

A sucessão é ergódica se todos os conjuntos da σ -álgebra F_∞ tem probabilidade 0 ou 1, onde $F_\infty = \bigcap_n F_n$ e $F_n = \sigma \{A_n, A_{n+1}, \dots\}$. Defina-se o produto de matrizes,

$$T^n = A_n \cdots A_2 A_1 \quad (3.68)$$

O conceito de expoente de Lyapunov permite uma caracterização das taxas de crescimento de T^n quando $n \rightarrow \infty$. O teorema de Oseledec, [Ose68], generaliza um resultado anterior em normas de produtos de matrizes aleatórias estacionárias obtido

por Furstenberg e Kesten(1960). O teorema ergódico multiplicativo será enunciado no contexto das sucessões estacionárias de matrizes aleatórias,

Teorema 3.6.2 *O limite,*

$$\lim_{n \rightarrow \infty} (T^n T^n)^{(1/2n)} = \Lambda \quad (3.69)$$

existe com probabilidade 1. Os logaritmos dos valores próprios de Λ dizem-se expoentes de Lyapunov. Tendo as matrizes dimensão $p \times p$, os expoentes de Lyapunov são, $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p$. Se a sucessão $\{A_i\}$ for ergódica os expoentes são constantes embora o limite (3.69) seja em geral uma matriz aleatória. Sejam, $\lambda^{(1)} > \lambda^{(2)} > \dots$ expoentes de Lyapunov distintos e $d^{(i)}$ a multiplicidade de $\lambda^{(i)}$. Seja $E^{(i)}$, em geral, aleatório o subespaço de $\mathbb{R}^p$ gerado pelos vectores próprios associados aos valores próprios $\leq e^{\lambda^{(i)}}$ de Λ . Vem que, $\mathbb{R}^p = E^{(1)} \supset E^{(2)} \supset \dots$ então o seguinte resultado verifica-se com probabilidade 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|T^n X\| = \lambda^{(i)}, \text{ se } X \in E^{(i)} / E^{(i+1)} \quad (3.70)$$

em particular, para todos os vectores X que não estão em $E^{(2)}$, o limite é o maior expoente de Lyapunov.

Facilmente se verifica que, para o caso limite da matrizes, os escalares e assumindo que são positivos o teorema reduz-se ao teorema ergódico. A originalidade e dificuldade da generalização deriva do facto de se lidar com matrizes não comutativas.

Na aplicação ao caso de um sistema dinâmico do tipo (3.2) a adaptação consiste em, tomar para a sucessão de matrizes a matriz jacobiana da enésima iteração de f^n ,

$$T_X^n = T(f^{n-1}X) \cdots T(f(X))T(X) \quad (3.71)$$

onde $T(X)$ é matriz jacobiana de f em X . Assumindo que existe uma medida f -invariante, ρ com suporte compacto, a sucessão $\{T(f^{n-1}(X))\}$ é estacionária e satisfaz as condições do teorema (3.6.2) para um X escolhido aleatoriamente com distribuição, ρ . Desta forma o enunciado do teorema para um sistema determinístico é, [Rue89a],

Teorema 3.6.3 *O limite*

$$\lim_{n \rightarrow \infty} (T_X^n T_X^n)^{(1/2n)} = \Lambda_X \quad (3.72)$$

existe para q.t.p- ρ . Os logaritmos dos valores próprios de Λ_X dizem-se expoentes de Lyapunov do sistema. Se ρ for ergódica os expoentes de Lyapunov são independentes de X , isto é, não dependem de condições iniciais para o sistema. Sejam, $\lambda^{(1)} > \lambda^{(2)} > \dots$ expoentes de Lyapunov distintos e $d^{(i)}$ a multiplicidade de $\lambda^{(i)}$. Seja $E^{(i)}$ o subespaço de $\mathbb{R}^p$ gerado pelos vectores próprios associados aos valores próprios $\leq e^{\lambda^{(i)}}$ de Λ . Vem que, $\mathbb{R}^p = E^{(1)} \supset E^{(2)} \supset \dots$ então o seguinte resultado verifica-se com probabilidade 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|T^n X\| = \lambda^{(i)}, \text{ se } X \in E^{(i)} / E^{(i+1)} \quad (3.73)$$

em particular, para todos os vectores X que não estão em $E^{(2)}$, o limite é o maior expoente de Lyapunov. Para uma demonstração do teorema de Oseledec ver [Pol93].

A evolução de um sistema depende dos seus expoentes de Lyapunov. Se se perturbar o valor inicial por um factor aditivo $\delta X_0 \notin E_X^{(2)}$ vem que $\delta X_n \approx \delta X_0 e^{n\lambda_1}$. É claro que no caso de um atractor limitado a divergência exponencial não se mantém indefinidamente uma vez que δX_n não deve ultrapassar o diâmetro do atractor. No caso de se tratar de um sistema em tempo contínuo, o teorema é aplicado à função F de (3.1). Os limites que definem os expoentes de Lyapunov verificam-se substituindo $n \rightarrow \infty$ por $t \rightarrow \infty$, devido à continuidade não é preciso restringir t a valores inteiros. No caso do sistema dinâmico estar num espaço de dimensão infinita existe a necessidade de

garantir que o limite (3.73) existe, para isso requere-se que os $T_X^T = D_X F$ sejam operadores compactos e não apenas limitados como nos espaços de dimensão finita, Kolmogorov e Fomin(1958).

Numa situação como a descrita em (3.2) em que $M \equiv \mathbb{R}^m$, $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$, o operador $T(x) \equiv D_x f$ pelo que a matriz correspondente à n -ésima iteração obtém-se pela regra da cadeia. Caso exista uma medida ergódica, ρ , para f com suporte compacto está-se nas condições do teorema multiplicativo.

Num sistema dinâmico em tempo contínuo como o descrito em (3.1) com $F : \mathbb{R}^m \rightarrow \mathbb{R}^m$, os expoentes também se encontram definidos desde que exista uma medida F -invariante e ergódica com suporte compacto. Desta forma cumprem-se os pressupostos do teorema multiplicativo, de salientar ainda que os expoentes e os subespaços $E_x^{(i)}$ são independentes da norma usada, [Eck85].

No caso de o sistema dinâmico evoluir numa variedade diferenciável compacta de dimensão m vem que F é um difeomorfismo $M \rightarrow M$ e DF toma valores no fibrado tangente TM onde está definida uma métrica Riemanniana $\langle, \rangle_x$ em M definida à custa de um produto interno associado ao espaço tangente em cada ponto, $T_x M$, e os expoentes característicos são definidos em termos dessa métrica. Alternativamente podem usar-se parametrizações por subconjuntos de $\mathbb{R}^m$ para definir os expoentes característicos, os seus valores são independentes das parametrizações usadas [Eck85]. Que expoentes característicos correspondem aos atractores indicados anteriormente, ponto fixo, órbita periódica e o movimento quasiperiódico? Sendo atractores e com um conjunto bem definido de valores que o sistema pode tomar uma vez que se tratam de estados estacionários têm uma descrição probabilística simples usando medidas de Dirac.

- ponto fixo atrator.

Sendo um ponto fixo atrator e correspondendo a um estado estacionário, existe um ponto P tal que $F^t(P) = P$ para todo o t e $\lim_{t \rightarrow \infty} F^t(x) = P$. A medida δ -Dirac associada é $\rho = \delta_P$. Nesta situação uma vez que existe contracção do volume em todas as direcções todos os valores próprios da matriz $T(x)$ são menores que

1 o que faz com que todos os expoentes de Lyapunov sejam negativos. Caso o ponto fixo não seja atrator os sinais dos expoentes podem mudar, diferentes combinações dão origem aos pontos fixos hiperbólicos, pontos fixos repulsores e pontos sela.

- Ciclo limite atrator.

Sendo um estado periódico está associado a uma órbita periódica, $\Gamma = \{F^t a : 0 \leq t < T\}$ do sistema. Sendo atrator também se verifica que, sendo U uma vizinhança de Γ vem que $F^t x \rightarrow \Gamma$ quando $t \rightarrow \infty$ para todo o $x \in U$. Este atrator é descrito pela medida de Dirac,

$$\rho = \delta_\Gamma = \frac{1}{T} \int_0^T \delta_{F^t a} dt \quad (3.74)$$

No caso de uma órbita periódica pelo menos um valor próprio de $T(x)$ é igual a 1, uma vez que um elemento de volume não se contrai ao longo da órbita, consequentemente um expoente característico é zero. De forma análoga ao ponto fixo, o ciclo limite pode ter diferentes naturezas com a excepção de um expoente que tem que ser sempre nulo as combinações de sinal nos outros dão origem aos diferentes ciclos limites que se podem encontrar.

- Atrator quasiperiódico.

Esta situação foi descrita anteriormente. Tal como nos casos anteriores, irão existir tanto expoentes nulos quantas as frequências incomensuráveis que constituem o estado quasiperiódico. O estado estacionário é descrito por uma medida F -invariante e ergódica, ρ , que num objecto geométrico como um toro k -dimensional, produto de k círculos é uma medida de Haar, ρ em T^k , $(2\pi)^k d\varphi_1 d\varphi_2 \cdots d\varphi_k$ definida em termos das variáveis angulares. Tal como nos casos anteriores, a natureza do atrator quasiperiódico depende dos sinais dos restantes expoentes do espectro da medida.

E se o sistema possuir um atrator estranho, que espectro se espera? Esta situação está relacionada com um estado caótico do sistema e é descrita pela medida assintótica

e invariante gerada pelo atrator. Um exemplo de um atrator estranho num sistema contínuo em três dimensões é o atrator de Lorenz. Para este tipo de sistemas o caos apenas pode ocorrer para 3 ou mais dimensões, senão veja-se; considere-se um sistema a duas dimensões, existem as seguintes possibilidades para a medida ergódica, ρ :

- $\lambda_1 = \lambda_2 = 0$ e $\lambda_1 = 0$ e $\lambda_2 < 0$: trata-se de um ponto fixo ou órbita periódica.
- $\lambda_1 > 0$ e $\lambda_2 = 0$: esta situação corresponde à inversão temporal da anterior, é um ponto fixo ou uma órbita periódica repulsora.
- $\lambda_1 \neq 0$ e $\lambda_2 \neq 0$: ρ está associada a um ponto fixo.

nenhum destes casos corresponde a um atrator com expoente característico positivo, logo o caos só poderá ocorrer em dimensões iguais ou superiores a 3, este resultado é corolário de um teorema, para mais detalhes ver, [Eck85] e [Cam80].

No caso de um sistema em tempo discreto, o caos pode ocorrer em apenas uma dimensão. O espectro de expoentes críticos de uma medida ergódica é uma ferramenta de classificação de atractores e fornece uma indicação do comportamento global do sistema, a taxa de crescimento de um vector infinitesimal, $\delta x(t)$ é dado pelo maior expoente, λ_1 , a taxa de crescimento de um elemento de superfície é dado por $\delta \sigma(t) = \delta_1 x(t) \wedge \delta_2 x(t)$ é dada, em geral, pela soma dos dois maiores expoentes característicos, $\lambda_1 + \lambda_2$. Para um volume k -dimensional, a taxa de crescimento é dada, geralmente pela soma $\lambda_1 + \lambda_2 + \dots + \lambda_k$.

A classificação das séries observadas constitui uma parte importante da análise de séries temporais. Como foi referido em secções anteriores, séries temporais geradas por sistemas lineares gaussianos ou equivalentemente por processos estocásticos lineares gaussianos diferem entre si nas propriedades de segunda ordem. A classificação consiste na identificação do modelo particular que gerou a série com base na informação obtida a partir da *FAC* amostral e do *periodograma* localizando as frequências que mais contribuem para a variação total. Para aqueles modelos, a transformada de Fourier além de um invariante do sistema, serve como uma *impressão digital* do mesmo. No caso dos processos não lineares e especificamente o caso caótico que

apresenta, como se referiu anteriormente, um espectro de valor elevado para uma gama continua de frequências (espectro de banda larga) faz com que o espectro e respectivas características não possam ser usados como ferramenta de caracterização. Assim para aqueles sistemas ter-se-á que substituir a transformada de Fourier e a *FAC* por outras características do sistema. Como foi referido, o conhecimento dos expoentes de Lyapunov (espectro de Lyapunov) permitem conhecer a evolução das órbitas do sistema e o seu comportamento a longo prazo. Pode dizer-se que os expoentes de Lyapunov *medem* a predictabilidade média do sistema dinâmico.

Em análise não linear de séries temporais é de grande relevância a estimação dos expoentes característicos uma vez que sendo algum deles positivo é forte indício da presença de caos.

Para um sistema observado cujas observações formam uma série $\{x_i\}_{i=1}^N$ obtida num tempo finito, T , existem certamente limitações no que se pode afirmar sobre os expoentes característicos. Uma série de medições de um fenómeno corresponde sempre a uma projecção do espaço de fase do sistema num espaço unidimensional. Assim, a revelação de invariantes do sistema passa obrigatoriamente por uma reconstrução usando coordenadas desfasadas com todas as limitações, já referidas, que aquela técnica encerra. Seguindo [Kan94], sendo $\mathbf{x}(t)$ a evolução de uma órbita com valor inicial $\mathbf{x}(0)$ num espaço de fase apropriado, então o expoente máximo de Lyapunov é, com probabilidade 1,

$$\lambda_{\max} = \lim_{t \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \frac{1}{t} \ln \left(\frac{|\mathbf{x}(t) - \mathbf{x}_\varepsilon(t)|}{\varepsilon} \right) \quad (3.75)$$

para quase todos os vectores $|\mathbf{x}(0) - \mathbf{x}_\varepsilon(0)| = \varepsilon$. Estima-se $\lambda_{\max}$ aplicando (3.75) directamente no espaço de fase usando a evolução de órbitas próximas ou calculando 3.75 no espaço tangente.

Seja $\mathbf{u}(t) = \mathbf{x}(t) - \mathbf{x}_\varepsilon(t)$ e $J(t)$ a matriz jacobiana da aplicação de $\mathbf{x}(t)$ a $\mathbf{x}(t+1)$. O cálculo dos maiores expoentes, onde D é a dimensão do espaço de fase, passa por iterar

n vectores $\{\mathbf{u}_i(t), i = 1, \dots, n\}$, ortonormais, usando $\mathbf{u}_i(t+1) = J(t)\mathbf{u}_i(t)$. Convém re-ortogonalizar os vectores com alguma periodicidade de forma a que as somas dos logaritmos das expansões e contracções dos vectores quando divididos pelo número de iterações sejam os n expoentes.

O primeiro algoritmo proposto para calcular o expoente máximo de Lyapunov foi proposto por [WA]. O algoritmo consiste no mergulho da série num espaço de construção de dimensão adequada usando funções de desfasamento temporal, depois procurar para cada ponto o seu vizinho mais próximo cuja evolução se considera como uma órbita próxima. Sendo $\mathbf{s}_{n_1}$ e $\mathbf{s}_{n_2}$ dois pontos a uma distância $\|\mathbf{s}_{n_1} - \mathbf{s}_{n_2}\| = \delta_0$ calcula-se a distância entre a evolução das duas órbitas, δ'_1 , valor que representa a propagação do segmento de comprimento δ_0 pelo atractor. No entanto a geometria do atractor pode aproximar a evolução temporal sem que tenha uma relação com a dinâmica e por esse motivo Wolf e seus colegas acrescentaram um requisito adicional, a necessidade de o ângulo formado pelo segmento que une o ponto $\mathbf{s}_{n_1+1}$ a $\mathbf{s}_{n_2+1}$ de comprimento δ'_1 e pelo segmento que une $\mathbf{s}_{n_1+1}$ ao seu vizinho mais próximo de comprimento δ_1 seja menor do que um valor previamente fixado. Essa é a direcção do vector próprio associado ao expoente característico máximo. Se não for encontrado nenhum ponto nessas condições então mantém-se o ponto respectiva órbita em uso. O expoente maximo é calculado por,

$$\lambda_{\max} = \frac{1}{n_M - n_0} \sum_{k=1}^M \ln \frac{\delta'_k}{\delta_{k-1}} \quad (3.76)$$

onde M é o número total de substituições de pontos.

Infelizmente o método descrito não é robusto, como indica [Kan94], os resultados obtidos são pouco precisos. Por outro lado, uma vez que o método não permite testar se de facto existe uma divergência exponencial que é assumida corre-se o risco de calcular um expoente para uma série gerada por um processo estocástico cujo expoente característico máximo é infinito.

O algoritmo de Wolf usa apenas informação do ponto mais próximo porém é sabido

que os sistemas reconstruídos a partir de séries podem apresentar órbitas com grandes flutuações que têm a ver com a função de observação, além do mais mesmo num verdadeiro espaço de fase e em presença de expoentes positivos pode haver localmente uma contracção. O expoente de Lyapunov deve ser encarado como uma média das taxas de divergência locais. A somar a tudo isto acresce como é hábito o feito nefasto do ruído que pode ser minimizado através do uso de uma estatística de tendência central.

Um método proposto por [San85] e melhorado por [Eck86] consiste na utilização do espaço tangente. Uma vez que os expoentes de Lyapunov são consequência directa das derivadas da transformação do sistema, em simulações numéricas os expoentes são calculados a partir da derivada, $D_x f^t$.

O algoritmo proposto por aqueles autores consiste no seguintes passos principais,

- reconstruir a dinâmica num espaço de dimensão finita.
- obter as transformações tangentes usando o métodos dos mínimos quadrados sobre um conjunto de vizinhos.
- Calcular os expoentes à custa das funções lineares ajustadas.

Para um algoritmo detalhado ver [Eck86]. Outros autores foram mais longe e em vez de ajustes lineares fizeram ajustes polinomiais, [Bro91b]. Outros autores propõem o uso de ajustes não lineares globais usando *radial basis functions* e redes neuronais. Como indica [Kan94] todas estas técnicas obrigam a um *mergulho* da série num espaço euclidiano a que se chamou espaço de reconstrução que deve ter uma dimensão igual à do sistema original que é desconhecida. Se por um lado a dimensão escolhida para o espaço de reconstrução for demasiado pequena os expoentes calculados não correspondem aos verdadeiros uma vez que as órbitas podem divergir pelo facto de serem falsas vizinhas, por outro lado, se for demasiado grande aparecem expoentes espúrios uma vez que o espaço terá uma dimensão maior o menor inteiro maior que duas vezes a dimensão fractal do atractor mais 1. Assim entre os m expoentes

calculados no espaço reconstruído existem alguns que não estão definidos no verdadeiro espaço de fase. Os autores em, [Kan99], referem ainda que um atractor fractal de dimensão m_0 pode ser mergulhado numa variedade de dimensão $\leq m_0$ sendo $m_0 = \text{int}(D_F) + 1$. Daí que os vizinhos no espaço de fase reconstruído não fornecem informação suficiente para calcular mais do que m_0 expoentes.

Se adicionarmos a estas dificuldades dos métodos os problemas comuns associados às séries de dados observadas como a limitação na amostragem e o ruído não é de estranhar o resultado obtido por [Eck86] quando contaminaram com 4% de ruído uma série experimental e verificaram que não era possível calcular, sequer, o expoente característico máximo.

Para muitos sistemas observados como aqueles que se relacionam com os fenómenos climáticos onde está sempre presente o ruído, identificar um expoente máximo de Lyapunov positivo significa um grande avanço na sua compreensão uma vez que é um forte índice da presença de caos.

Da mesma forma que [WA], [Kan94] propõe um método que se apoia no facto de 2 órbitas vizinhas divergirem a uma taxa dada pelo expoente máximo de Lyapunov tendo em conta o facto de esse facto acontecer assintoticamente. De acordo com o exposto na secção anterior, a divergência exponencial ocorre após um período de tempo transiente uma vez que um qualquer vector diferença tem que evoluir na direcção mais instável. A taxa de divergência é dada pelo espectro de expoentes. O expoente máximo pode também definir-se como,

$$\lambda_\tau = \lim_{\epsilon \rightarrow 0} \frac{1}{\tau} \ln \left(\frac{|\mathbf{x}(t + \tau) - \mathbf{x}_\epsilon(t + \tau)|}{\epsilon} \right)$$

$$\mathbf{x}(t) - \mathbf{x}_\epsilon(t) = \epsilon \mathbf{w}_u(t) \quad (3.77)$$

onde $\mathbf{w}_u(t)$ é o vector próprio local associado ao expoente máximo $\lambda_{\max}$. O valor de λ_τ depende da estrutura do espaço tangente e é aproximadamente o mesmo apenas para uma pequena vizinhança. Por definição a média de λ_τ ao longo da órbita é o verdadeiro expoente máximo de Lyapunov.

Traduzida em pontos no espaço de fase a proposta de [Ros93] e [Kan94] consiste num algoritmo que permite testar directamente a divergência exponencial. Baseia-

se nos seguintes passos, seleccionar um ponto $(s)_{n_0}$ no espaço de fase reconstruído e procurar todos os vizinhos a uma distância menor ou igual a ε , esses pontos e respectivas evoluções consideram-se trajectórias vizinhas. Na situação ideal, isto é, se se tratasse do verdadeiro espaço de fase usando as distâncias entre os pontos das trajectórias podia-se medir com exactidão as flutuações da taxa de divergência. Para uma situação experimental a flutuação mede-se calculando a distância entre a coordenada referente ao desfasamento temporal τ das trajectórias de referência e vizinhas, $\text{dist}(\mathbf{s}_n, \mathbf{s}_{n_0}, \tau) = |\mathbf{s}_{n+\tau} - \mathbf{s}_{n_0+\tau}|$. Justifica-se este procedimento uma vez que essas distâncias são projecções não-lineares dos vectores de diferença no verdadeiro espaço de fase num espaço unidimensional gerado pela observação. Como projecção tem associado um módulo, $|\cos(\phi)|$ onde ϕ é o ângulo entre o vector próprio com a direcção associada a $\lambda_{\max}$ e a direcção local do subspaço correspondente ao do espaço da observação dos vectores no verdadeiro espaço de fase. O ângulo ϕ depende da posição o espaço de fase, tal como o expoente de Lyapunov, e será semelhante para todos os pontos da vizinhança do ponto de referência se as distâncias no espaço de fase não forem muito grandes. Calcula-se então a média das distâncias ao ponto de referência para todos os pontos,

$$S(\Delta n) = \frac{1}{N} \sum_{n_0=1}^N \ln \left(\frac{1}{|U(\mathbf{s}_{n_0})|} \sum_{\mathbf{s}_n \in U(\mathbf{s}_{n_0})} \text{dist}(\mathbf{s}_n, \mathbf{s}_{n_0}, \tau) \right) \quad (3.78)$$

onde $\mathbf{s}_{n_0}$, pontos de referência são os vectores de reconstrução, $U(\mathbf{s}_{n_0})$ é o número de pontos na vizinhança de $\mathbf{s}_{n_0}$ de raio ε .

No verdadeiro espaço de fase, os vectores de diferença (que unem os pontos vizinhos) podem apontar em qualquer direcção pelo que a distância se comporta como, $\text{dist} = \sum_i a_i \exp(\lambda_i t)$, sendo os λ_i os verdadeiros expoentes correspondentes às diferentes direcções estáveis e instáveis. Desta forma, como refere [Kan94] não se pode esperar, para um pequeno τ , que ocorra um aumento linear. No entanto, esse crescimento linear ocorrerá para $S(\tau)$, com declive λ para valores de τ superiores.

A coordenada $s_{n_0+\Delta n}$ diz respeito a um ponto que não $\mathbf{s}_{n_0}$, uma vez que se desconhece

à partida o valor ideal para o espaço de reconstrução nem o raio ideal para a vizinhança será uma boa ideia fazer o cálculo (3.78) para vários valores das duas variáveis m e ε . Se se verificar um crescimento linear robusto de $S(\Delta n)$ então o seu declive é a estimativa do expoente máximo de Lyapunov expresso em valor/unidade de tempo. Por robustez entenda-se que a relação não deve ser uma característica de apenas alguns valores particulares de m e ε .

3.6.2 Estimação da dimensão-correlação

Em análise não-linear de séries o conceito de dimensão que recebe maior atenção é o de dimensão correlação, também designado por D_2 , isto é, dimensão generalizada de Rényi com $q = 2$. Esse facto deve-se em grande parte ao seu cálculo que é relativamente fácil. Sendo π_m uma medida de probabilidade ergódica de um sistema dinâmico em tempo discreto, $\{X_t, t \in \mathbb{Z}_+\}$ com atrator, $\Lambda \subseteq \mathbb{R}^m$. A correlação espacial do sistema é descrita pelo integral-correlação, $C(r) = E(\pi(B(X, r)))$ onde X e Y são i.i.d. com distribuição π_m ,

$$\begin{aligned} P(\|X - Y\| \leq r) &= E(P(\|X - Y\|) \setminus Y) \\ &= E(\pi(B(Y, r))) = C(r) \end{aligned} \quad (3.79)$$

$$(3.80)$$

De acordo com a definição a dimensão correlação, D_2 é o expoente que aparece na lei de escala, $C(r) \approx r^{D_2}$, que implica que

$$\lim_{r \rightarrow 0} \frac{\log C(r)}{\log r} = D_2 \quad (3.81)$$

assim num gráfico $\log - \log$, $C(r)$ é aproximadamente linear para um r que na prática não pode ser tão grande que não existam pontos fora da esfera de raio r e não tão

pequeno que não existam pontos dentro dela. Como indica [Aba96a], *pequeno* deve ser interpretado de forma independente da dimensão e a única *escala* de referência é o tamanho do atrator, R_A . Para aquele autor, um *pequeno* r é, $0 \ll \frac{r}{R_A} \ll 1$.

Recordamos que as dimensões generalizadas e a dimensão correlação em particular, são uma característica da medida invariante do atrator, caracterizam a forma como π se distribui no atrator. Se π for uma medida concentrada num ponto $D_2 = 0$, se π for absolutamente contínua relativamente à medida de Lebesgue então D_2 é igual à dimensão topológica do atrator. Na prática, considerando a série reconstruída num espaço euclidiano de dimensão m a dimensão D_2 varia de 0, dimensão de um conjunto singular, a m , número de variáveis de estado do sistema.

3.6.2.1 Integral de correlação amostral

Os autores Grassberger e Procaccia num célebre artigo, [Gra83], propuseram um algoritmo para estimar a dimensão correlação de um sistema a partir de uma série numérica observada, $\{S_n\}_{n=1}^L$ de tamanho L e $N = L - (m - 1)\tau$ vectores de reconstrução, $S_n = (s_{n-(m-1)\tau}, \dots, s_{n-\tau}, s_n)$. O conjunto de pontos em cada espaço m -dimensional considerado é uma aproximação amostral à distribuição teórica desconhecida, π_m^4 . A medida de probabilidade empírica construída atribuindo o mesmo peso, $1/N$, a cada vector de reconstrução designa-se por medida de reconstrução empírica (voltaremos a este assunto mais à frente). A proposta de Grassberger consiste na utilização desta medida empírica no cálculo do integral de correlação,

$$\hat{C}(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \Theta(r - \|S_i - S_j\|) \quad (3.82)$$

O integral de correlação amostral é a fracção de pares de vectores de reconstrução distintos que se encontram a uma distância menor ou igual a r . Para um raio, r , fixo o integral correlação (ou soma - correlação) pertence à classe das estatísticas-U introduzidas por Hoeffding, [Hoe48] como estimadores de funcionais da forma,

⁴o índice m designa a versão da medida invariante para um espaço de m dimensões

$$\theta(F) = \int_{\mathbb{R}^{km}} h(x_1, \dots, x_m) dF(x_1) \dots dF(x_m) \quad (3.83)$$

onde a função núcleo (*kernel*), $h : \mathbb{R}^{k \times m} \rightarrow \mathbb{R}$ é simétrica nos seus argumentos. As estatísticas-U têm a importante propriedade de serem invariantes por permutação dos dados. As estatísticas-U apresentam algumas propriedades para processos estacionários. Hoefdding mostrou que no caso de $\{X_n\}_{n \in \mathbb{N}}$ ser um processo estocástico independente e com função de distribuição marginal, F desconhecida, o estimador não enviesado e de variância uniformemente mínima (UMVU) de $\theta(F)$ é da forma,

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}) dF(x_1) \dots dF(x_m) \quad (3.84)$$

A estatística U_n denomina-se estatística-U de grau m correspondente ao núcleo, h . Existem outros exemplos de estatísticas-U, é o caso da variância amostral ($k = 1$, $m = 2$, $h(x, y) = 1/2(x - y)^2$) e as estatísticas de posto-sinal de Wilcoxon ($k = 1$, $m = 2$, $h(x, y) = 1 \{x + y > 0\}$). Para o integral de correlação vem, $m = 2$ e $h(x, y) = 1 \{\|x - y\| \leq r\}$.

Recentemente tem havido um interesse considerável em estatísticas-U em processos estocásticos estacionários com dependência. Uma das inúmeras aplicações é a estimação da dimensão de um atrator. Uma vez que a observação no tempo de um sistema dinâmico sujeito a ruído se pode considerar um processo estocástico, regra geral, não independente, as propriedades assintóticas do estimador da dimensão correlação podem estudar-se usando as propriedades correspondentes das estatísticas-U de processos estacionários dependentes.

Os autores em, [Aar96], demonstraram um resultado extremamente importante relativo à consistência das estatísticas-U de processos estocásticos estacionários e ergódicos,

Teorema 3.6.4 *Seja $\{X_n\}_{n \in \mathbb{N}}$ uma sucessão estacionária e ergódica com distribuição marginal F e seja $h : \mathbb{R}^{k \times m} \rightarrow \mathbb{R}$ uma função mensurável e limitada e continua quase por toda a parte relativamente à medida, F^m (q.t.p.- F^m). Então,*

$$U_n \xrightarrow{q.c.} \Theta(F) \text{ quando } n \rightarrow \infty. \quad (3.85)$$

No caso de o processo estocástico ser absolutamente regular o requisito da continuidade de F é dispensável. Um processo absolutamente regular ou com a propriedade *mixing* fraca tem um comportamento assintótico próximo do de um processo independente então apenas para os processos absolutamente regulares se aplica o teorema do limite central para as estatísticas-U. Assim o teorema (3.6.4) pode estender-se a sucessões estacionárias com boas propriedades *mixing*.

Pesin, [Pes93] mostrou que sob a suposição de ergodicidade verifica-se que o integral de correlação amostral é um estimador consistente. No entanto, como indicam [Cha01] apesar de ser consistente é necessário ter uma amostra de um tamanho bastante grande para que o integral de correlação amostral se aproxime do teórico em termos de forma, especialmente se os dados apresentarem fortes correlações temporais. Theiler, [The86], mostrou que na presença de séries correlacionadas e com amostras de média dimensão passa a haver uma relação temporal nos pares de vectores a uma pequena distância espacial. Esse facto leva a uma distorção do integral de correlação amostral que não permitindo boas estimativas da dimensão correlação especialmente em espaços de mergulho de maior dimensão.

Para ultrapassar essa deficiência do integral de correlação amostral, [The86] sugere uma pequena alteração à forma do integral cujo efeito é *deixar de fora* as parcelas correspondentes a pares de vectores com proximidade temporal inferior a um determinado nível pré-fixado, corrigindo dessa forma o viés. O estimador mais geral tem a forma,

$$\hat{C}(r) = \frac{2}{(N - n_{\min})(N - 1 - n_{\min})} \sum_{i=1}^{N-n_{\min}} \sum_{j=i+n_{\min}}^N \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (3.86)$$

onde Θ é a função de Heaviside e $n_{\min}$ é a chamada correcção de Theiler. Para $n_{\min} = 1$ a expressão é o integral de correlação amostral *standard*. Coloca-se a questão de qual o valor usar para $n_{\min}$. O tempo de correlação associado a $n_{\min}$ é $t_{\min} = n_{\min} \Delta t$. Não

tem que existir necessariamente uma relação entre $n_{\min}$ e valores da função que mede a correlação da série, a *FAC*.

3.6.2.2 Estimador de Grassberger - Procaccia

O algoritmo proposto por Grassberger e Procaccia faz uso do integral de correlação amostral corrigido ou não para se obter uma estimativa da dimensão correlação. Uma vez que se espera que, $\log C(r)/\log r = D_2 + o(1)$ atendendo a (3.82) então, substituindo $C(r)$ pelo integral de correlação amostral fica $\log \hat{C}(r)/\log r = D_2 + o(1) + O(1/\sqrt{N})$. Ao contrário do que acontece para o integral de correlação o estimador para a dimensão correlação, $\log \hat{C}(r)/\log r$ é assintoticamente enviesado, uma forma de reduzir o enviesamento é usar uma sucessão de raios, r_N tal que $r_N \rightarrow 0$ e $Nr_N \rightarrow \infty$ quando $N \rightarrow \infty$. É claro que havendo uma limitação no tamanho da série a obtenção de uma sucessão r_N nessas condições é muito difícil. Em [Bor99] concluiu-se que sob a absoluta regularidade da série e em presença de uma lei de escala perfeita $\log \hat{C}(r)/\log r$ é um estimador fortemente consistente de D_2 . No entanto, na prática, é difícil verificar tais pressupostos. O procedimento consiste em usar a regressão linear,

1. Calcular o integral de correlação para um conjunto de valores do raio, $r, r_1, r_2, \dots, r_k$ que [Cut01] propõe que sejam igualmente espaçados numa escala logaritmica.
2. Ajustar uma recta aos valores numa escala $\log - \log$ de $\hat{C}(r)$, r A estimativa dos mínimos quadrados do declive da recta é o estimador de Grassberger e Procaccia (*G-P*) da dimensão correlação, [GP].

Como indicam [Kan99] a atribuição de uma dimensão a uma série compreende uma prévia interpretação cuidadosa das curvas em causa. Uma vez que se trabalha com conjuntos finitos de dados que, contrariamente ao valor bem definido para o integral de correlação, têm dimensão nula, a estimação da dimensão a partir de uma série envolve obrigatoriamente uma extrapolação.

A auto-semelhança típica da fractalidade desaparece em regimes macrocópicos uma

vez que em grandes escalas são as estruturas macrocópicas do atrator que dominam o comportamento do integral de correlação, [Kan99]. Para pequenas escalas é possível observar o fenómeno da auto-semelhança que se verifica para espaços com dimensão $n_{\min} > D$ onde D é a dimensão do atrator. De acordo com (3.81) e usando a regra de *L'hôpital*, a existência de uma região de escala onde o gráfico $\log - \log$ de $\hat{C}(r)$, r tem um comportamento quase linear pode ser melhor observada num gráfico do declive $d(N, r)$, $\log(r)$. Neste gráfico a região de escala aparece como um linha paralela ao eixo dos xx .

3.6.3 Estimação da Entropia

O conhecimento da entropia de um sistema é importante para a sua caracterização. O seu inverso é como se viu atrás um valor cujo significado está directamente relacionado com a predictabilidade. Fornece igualmente informação topológica acerca do atrator. A entropia é, no entanto, um invariante difícil de estimar a partir de uma série uma vez que a computação requerida envolve mais dados do que a estimação de dimensões ou expoentes, [Kan99]. Existe ainda o problema da definição envolver o cálculo de um supremo de um limite sobre todas as partições possíveis, que não é praticável. Além disso, o limite, (3.18) constitui outro problema.

A estimação da entropia, h_α , requiere o uso de *caixas* análogo ao usado para determinar a dimensão-capacidade ou *Box-counting*. Para $q > 1$, usando uma partição cujos elementos têm o mesmo raio, ao atribuir-se-lhes a mesma probabilidade, [Kan99], chega ao resultado (3.53). Aplicando o integral de correlação amostral generalizado,

$$\hat{C}(r) = \frac{2}{N(N-1)^{\alpha-1}} \sum_{i=1}^N \left[\sum_{i \neq j} \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|) \right]^{\alpha-1} \quad (3.87)$$

espera-se que, para uma dimensão do espaço de *mergulho*, m ,

$$C(m, \varepsilon) \propto \varepsilon^{(1-\alpha)D_\alpha} e^{(\alpha-1)H_\alpha(m)} \quad (3.88)$$

O procedimento correcto passaria agora por determinar o supremo sobre todas as partições o que implicava possuir um mecanismo gerador de partições e que estas existissem em número finito. Depois sobre esse conjunto se fizesse o cálculo quando $\varepsilon \rightarrow 0$. O procedimento é de todo impossível com uma série finita de dados. Em vez disso [Kan99] indicam que a entropia se possa calcular se existir uma região para $\ln(\varepsilon)$ (*scaling range*) na qual os resultados sejam independentes de ε . Tal independência de escala existe se ocorrer uma estabilização dos expoentes locais,

$$d_\alpha(m, \varepsilon) = \frac{\partial \ln C_\alpha(m, \varepsilon)}{\partial \ln(\varepsilon)} \quad (3.89)$$

Na região para ε onde ocorre *scaling* o factor $\varepsilon^{(1-\alpha)} D_\alpha$ é aproximadamente constante e usando (3.18) estima-se a entropia fazendo,

$$h_\alpha(m, \varepsilon) = H_\alpha(m+1, \varepsilon) - H_\alpha(m, \varepsilon) = \ln \frac{C_\alpha(m, \varepsilon)}{C_\alpha(m+1, \varepsilon)} \quad (3.90)$$

O *scaling* terá que ocorrer para vários ε e para várias dimensões, m . Para um m suficientemente grande deverá haver convergência para um h_α . Claramente, como acontece para a dimensão, a entropia mais fácil de calcular é a h_2 , usando igualmente, o integral de correlação amostral.

3.7 Medidas de reconstrução

O conceito de medida de reconstrução de um sistema observado é essencial para poder estudar o fenómeno da intermitência do ponto de vista da reconstrução da dinâmica. Como se referiu no capítulo 1, uma série temporal diz-se estacionária se a sua estrutura não mudar ao longo do tempo, ou seja, se as suas propriedades probabilísticas não dependerem do tempo em que são medidas. Em muitos sistemas dinâmicos colocam-se condições sobre os parâmetros de forma a que o seu comportamento não seja divergente. Como se viu no capítulo 2, alguns tipos de não-estacionaridade podem facilmente detectar-se e eliminar-se da série.

Um outro conceito relacionado com a não estacionaridade é a da detecção de pontos/níveis onde ocorre uma mudança estrutural da dinâmica do sistema.

A dimensão correlação de uma série determinística é um exemplo de um invariante do sistema mas ao mesmo tempo é também um parâmetro de escala da medida invariante dos vectores de reconstrução. Da mesma forma que os processos lineares gaussianos geram séries cujas medidas de reconstrução são normais multivariadas, também as séries determinísticas fazem parte de uma classe de séries temporais cujas medidas de reconstrução podem ter uma estrutura fractal.

Para este tipo de séries é necessário um novo tipo de definição de estacionaridade, [Tak96]. Uma série limitada e infinita, $\{X_n\}_{n=1}^{\infty}$ é estacionária se as médias,

$$\bar{g} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n g(X_k, X_{k+1}, \dots, X_{k+m-1}) \quad (3.91)$$

existirem para todo o m e para cada função $g : \mathbb{R}^m \rightarrow \mathbb{R}$ continua. Sob esta condição, o teorema de representação de Riez, (ver [Fel71]), diz que existe uma medida de probabilidade associada, μ_m em $\mathbb{R}^m$ a que daremos o nome de medida de reconstrução m -dimensional.

A suposição de que a série é ilimitada pode ser contornada, [Dik99]. Como refere este autor existem outras propostas para definir estacionaridade de uma série, no entanto, a de Takens é a que está mais próxima do que é comum em estatística, baseia-se na série temporal e não no processo gerador. [Dik99] refere ainda que as séries estritamente estacionárias no sentido probabilístico geram séries que são estacionárias no sentido de Takens com probabilidade 1.

As medidas de reconstrução têm propriedades probabilísticas. Dado um subconjunto $A \subset \mathbb{R}^m$ o sistema tem uma medida associada dada por,

$$\mu(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n I_A(X_k, X_{k+1}, \dots, X_{k+m-1}) \quad (3.92)$$

onde $I_A(\cdot)$ é a função indicatriz. O valor $\mu_m(A)$ é a medida relativa do tempo que os vectores de reconstrução vão passar em A a longo prazo.

Entre os vectores de reconstrução para diferentes dimensões existem relações de consistência, por exemplo, a projecção dos vectores de reconstrução em m dimensões devem corresponder aos vectores de reconstrução para a dimensão $m - 1$. Este facto significa que nem todas as medidas de probabilidade são possíveis medidas de reconstrução, basta, por exemplo, que tenham medidas marginais diferentes. As medidas condicionais obtidas de medidas de reconstrução dizem-se medidas de previsão de ordem $m - 1$. Como refere [Dik99], o estudo de séries temporais não lineares do ponto de vista dos sistemas dinâmicos não é mais do que o estudo de medidas de reconstrução e previsão de séries temporais. Formalmente, de acordo com o teorema de Takens, para uma série determinística de dimensão finita, a medida de reconstrução, para um m suficientemente grande, está confinada a um conjunto de pequena dimensão (possivelmente fractal) e a medida de previsão tem a sua massa concentrada num ponto que corresponde a uma função das m observações anteriores. As estimativas da dimensão são formas de caracterização da medida de reconstrução. As séries geradas por processos puramente estocásticos encaixam-se igualmente nesta definição.

O conceito de medida de reconstrução é unificador das diferentes abordagens em análise linear estocástica e não linear caótica. As estimativas dos invariantes da série determinística são meios pelos quais a medida de reconstrução é caracterizada, expoentes de Lyapunov, dimensões e entropia generalizadas que têm como contra parte nos lineares estocásticos os parâmetros do modelo que caracterizam completamente a f.d.p. multivariada associada e consequentemente as medidas de reconstrução de todas as ordens.

Os conceitos como a reversibilidade podem escrever-se em termos das medidas de reconstrução. Uma série é reversível se a medida de reconstrução m -dimensional, μ associada aos vectores de reconstrução, [Dik99],

$$\mathbf{X}_n = (X_n, X_{n+\tau}, \dots, X_{n+(m-1)\tau}) \quad (3.93)$$

for, para todo o m , invariante para a transformação $P : \mathbb{R}^m \rightarrow \mathbb{R}^m$ definida por,

$P(x_1, \dots, x_m) = (x_m, \dots, x_1)$. Define-se a medida $P\mu$ como a imagem de μ sob P ,

$$P\mu(A) = \mu(P^{-1}A) \forall A \subset \mathbb{R}^m \quad (3.94)$$

a série é reversível se $\mu = P\mu$, ou seja, se a medida de reconstrução for invariante sob uma reversão do tempo. O conceito de reversibilidade para sistemas dinâmicos relaciona-se com a reversibilidade da série gerada, a relação não é, contudo, trivial veja-se, [Dik99]. Uma vez que os processos lineares gaussianos dão origem a séries reversíveis, a construção de um teste estatístico para testar a hipótese,

$$H_0 : \mu = P\mu \quad (3.95)$$

é de grande importância especialmente se se considerar o facto de alguns processos puramente estocásticos poderem gerar séries com integrais de correlação convergentes. Uma vez que a não-linearidade (e por consequência a irreversibilidade) é uma condição necessária para a ocorrência de caos, a construção de um teste para averiguar dessa característica é extremamente importante. O teste desenvolvido por [Dik99] testa a hipótese (3.95) para vectores de reconstrução independentes distribuídos de acordo com uma medida, μ . Esta restrição pode relaxar-se utilizando o método de blocos como se vai descrever mais à frente. O estimador que Diks propõe para testar a hipótese nula constrói-se da seguinte forma. Sendo d um escalar positivo, e 2 medidas μ_1 e μ_2 define-se a forma bilinear,

$$\begin{aligned} (\mu_1 - \mu_2) &= \iint K(\mathbf{r} - \mathbf{s}) d\mu_1(\mathbf{r}) d\mu_2(\mathbf{s}) \\ K(\mathbf{r} - \mathbf{s}) &= \exp - \|\mathbf{r} - \mathbf{s}\| / (2d_2) \end{aligned} \quad (3.96)$$

a forma bilinear é definida positiva o que faz de $(\mu_1 - \mu_2)$ um produto interno de μ_1 com μ_2 . Desta forma Diks mostra que $R = (\mu_1 - \mu_2, \mu_1 - \mu_2)$ é positivo e nulo sse μ_1 e μ_2 forem iguais. Esta propriedade pode ser usada para construir uma estatística

Entre os vectores de reconstrução para diferentes dimensões existem relações de consistência, por exemplo, a projecção dos vectores de reconstrução em m dimensões devem corresponder aos vectores de reconstrução para a dimensão $m - 1$. Este facto significa que nem todas as medidas de probabilidade são possíveis medidas de reconstrução, basta, por exemplo, que tenham medidas marginais diferentes. As medidas condicionais obtidas de medidas de reconstrução dizem-se medidas de previsão de ordem $m - 1$. Como refere [Dik99], o estudo de séries temporais não lineares do ponto de vista dos sistemas dinâmicos não é mais do que o estudo de medidas de reconstrução e previsão de séries temporais. Formalmente, de acordo com o teorema de Takens, para uma série determinística de dimensão finita, a medida de reconstrução, para um m suficientemente grande, está confinada a um conjunto de pequena dimensão (possivelmente fractal) e a medida de previsão tem a sua massa concentrada num ponto que corresponde a uma função das m observações anteriores. As estimativas da dimensão são formas de caracterização da medida de reconstrução. As séries geradas por processos puramente estocásticos encaixam-se igualmente nesta definição.

O conceito de medida de reconstrução é unificador das diferentes abordagens em análise linear estocástica e não linear caótica. As estimativas dos invariantes da série determinística são meios pelos quais a medida de reconstrução é caracterizada, expoentes de Lyapunov, dimensões e entropia generalizadas que têm como contra parte nos lineares estocásticos os parâmetros do modelo que caracterizam completamente a f.d.p multivariada associada e consequentemente as medidas de reconstrução de todas as ordens.

Os conceitos como a reversibilidade podem escrever-se em termos das medidas de reconstrução. Uma série é reversível se a medida de reconstrução m -dimensional, μ associada aos vectores de reconstrução, [Dik99],

$$\mathbf{X}_n = (X_n, X_{n+\tau}, \dots, X_{n+(m-1)\tau}) \quad (3.93)$$

for, para todo o m , invariante para a transformação $P : \mathbb{R}^m \rightarrow \mathbb{R}^m$ definida por,

$P(x_1, \dots, x_m) = (x_m, \dots, x_1)$. Define-se a medida $P\mu$ como a imagem de μ sob P ,

$$P\mu(A) = \mu(P^{-1}A) \forall A \subset \mathbb{R}^m \quad (3.94)$$

a série é reversível se $\mu = P\mu$, ou seja, se a medida de reconstrução for invariante sob uma reversão do tempo. O conceito de reversibilidade para sistemas dinâmicos relaciona-se com a reversibilidade da série gerada, a relação não é, contudo, trivial veja-se, [Dik99]. Uma vez que os processos lineares gaussianos dão origem a séries reversíveis, a construção de um teste estatístico para testar a hipótese,

$$H_0 : \mu = P\mu \quad (3.95)$$

é de grande importância especialmente se se considerar o facto de alguns processos puramente estocásticos poderem gerar séries com integrais de correlação convergentes. Uma vez que a não-linearidade (e por consequência a irreversibilidade) é uma condição necessária para a ocorrência de caos, a construção de um teste para averiguar dessa característica é extremamente importante. O teste desenvolvido por [Dik99] testa a hipótese (3.95) para vectores de reconstrução independentes distribuídos de acordo com uma medida, μ . Esta restrição pode relaxar-se utilizando o método de blocos como se vai descrever mais à frente. O estimador que Diks propõe para testar a hipótese nula constrói-se da seguinte forma. Sendo d um escalar positivo, e 2 medidas μ_1 e μ_2 define-se a forma bilinear,

$$\begin{aligned} (\mu_1 - \mu_2) &= \iint K(\mathbf{r} - \mathbf{s}) d\mu_1(\mathbf{r}) d\mu_2(\mathbf{s}) \\ K(\mathbf{r} - \mathbf{s}) &= \exp - \|\mathbf{r} - \mathbf{s}\| / (2d_2) \end{aligned} \quad (3.96)$$

a forma bilinear é definida positiva o que faz de $(\mu_1 - \mu_2)$ um produto interno de μ_1 com μ_2 . Desta forma Diks mostra que $R = (\mu_1 - \mu_2, \mu_1 - \mu_2)$ é positivo e nulo sse μ_1 e μ_2 forem iguais. Esta propriedade pode ser usada para construir uma estatística

usando a propriedade da reversibilidade. A quantidade,

$$Q = (\mu - P\mu, \mu - P\mu) \quad (3.97)$$

tem o valor zero se $\mu = P\mu$. Uma vez que P é uma isometria, vem,

$$Q = 2(\mu, \mu) - 2(\mu, P\mu) = \iint g(\mathbf{r}, \mathbf{s}) d\mu(\mathbf{r}) d\mu(\mathbf{s}) \quad (3.98)$$

onde

$$g(\mathbf{r}, \mathbf{s}) = 2 \left(e^{-\|\mathbf{r}-\mathbf{s}\|^2/(2d^2)} - e^{-\|\mathbf{r}-P\mathbf{s}\|^2/(2d^2)} \right) \quad (3.99)$$

A estatística Q pertence à classe das estatísticas introduzidas por Hoeffding, (3.83), como tal um seu estimador constrói-se facilmente atendendo à sua invariância por permutação,

$$\hat{Q} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N W_{ij} \quad (3.100)$$

onde

$$W_{ij} = g(\mathbf{X}_i, \mathbf{X}_j) \quad (3.101)$$

[Dik99] mostra que o estimador (3.101) é centrado e, uma vez que $g(\cdot, \cdot)$ é limitado, vem que, $E(g^2(\mathbf{X}_i, \mathbf{X}_j)) < \infty$) o que implica a consistência de $\hat{Q}$ e assim $\hat{Q} \rightarrow Q$ com probabilidade 1 para $N \rightarrow \infty$. O facto de se verificar $Q > 0$ para qualquer alternativa faz com que um teste unilateral baseado em $\hat{Q}$ que rejeite a hipótese nula para valores elevados de Q tenha potência relativamente a qualquer alternativa fixa. [Dik99] obteve a distribuição de nulidade para $\hat{Q}$, mais uma vez, com o pressuposto de os vectores de reconstrução serem indendentemente distribuidos. Sob essa suposição existe uma estatística suficiente mínima para a classe de medidas de reconstrução, dessa forma

e condicionalmente a média e o desvio padrão podem calcular-se facilmente, [Dik99]. Sob a hipótese nula, obtém-se para a estatística $\hat{Q}$,

$$\begin{aligned} E_c(\hat{Q}) &= 0 \\ V_c(\hat{Q}) &= \frac{4}{N^2(n-1)^2} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N W_{ij}^2 \right)^{(1/2)} \end{aligned} \quad (3.102)$$

o que faz com que $S = \hat{Q}/\sigma_c\hat{Q}$ seja uma variável aleatória de média zero e desvio unitário. A teoria das estatísticas-U mostra que a distribuição assintótica de S para vectores independentes não é necessariamente normal [Ser80].

Como refere [Dik99], o valor do parâmetro d da função núcleo pode influenciar os resultados, os valores de $m = 3$ e $d = 0.2$ resultaram para uma grande variedade de processos.

Como referimos anteriormente, a dependência entre os vectores de reconstrução provoca enviesamentos no teste. Existem dois tipos de dependência, o primeiro é introduzido pelo processo de reconstrução e o segundo deve-se à ligação dinâmica entre os vectores. A primeira está sempre presente provocando enviesamentos em $\hat{Q}$ que, no caso i.i.d., são da ordem de $1/N$. Recorde-se que este problema afecta igualmente uma outra estatística-U, o integral de correlação amostral. Tal como aconteceu para o I.C., pode aplicar-se a correcção de Theiler eliminando os vectores que estão temporalmente próximos. A estatística corrigida tem a forma,

$$\hat{Q} = \frac{1}{\binom{N-T+1}{2}} \sum_{i=1}^{N-T} \sum_{j=i+T}^N W_{ij} \quad (3.103)$$

onde T é a constante de correcção de Theiler. O valor esperado de W_{ij} é igual a Q se $\mathbf{X}_i$ e $\mathbf{X}_j$ forem mutuamente independentes com desfasamento temporal, $\tau = 1$ o que significa que basta tomar $T \geq m$ no caso i.i.d.. Existe ainda uma fonte de correlação que não é eliminada pelo uso da correcção de Theiler. A contribuição dos pares (i, j) e $(i + k, j + l)$ pode estar correlacionada para pequenos valores de k e l provocando

enviesamentos em $\hat{Q}$. Tendo em conta a simetria da estatística a correlação quer seja positiva ou negativa terá o mesmo efeito, provocará percentagens de erro do tipo I superiores ao nominal, isto é, rejeitar-se-á mais frequentemente a hipótese nula do que o esperado para a situação sem correlação temporal. Esta correlação não tem um peso significativo no caso de séries i.i.d., no entanto, em séries reversíveis com lento decaimento da correlação o efeito é bem visível como comprova [Dik99]. A correcção que [Dik99] propõe consiste em fazer médias dos valores de W_{ij} a partir de uma grelha $i - j$ de índices de dimensão $l \times l$,

$$W'_{ij} = \frac{1}{l^2} \sum_{p=1}^l \sum_{q=1}^l W_{(il+p), (jl+q)} \quad (3.104)$$

A ideia por detrás do procedimento é que, para um l suficientemente grande, os blocos ou segmentos da série, não sobrepostos, de tamanho l podem, na prática, considerar-se independentes. Para $l = 1$ obtém-se o teste estatístico usual com a correcção de Theiler. Na prática, perante séries que exibam dependência deve usar-se a correcção de Theiler e uma dimensão de blocos adequada. Obtém-se bons resultados quando T e l são da mesma escala temporal do decaimento da correlação da série.

O teste de reversibilidade proposto por [Dik99] tem em conta a possibilidade de o processo gerador da série ser uma transformação não linear de um processo linear gaussiano. A reversibilidade de uma série pode manter-se quando sobre ela actua uma transformação estática, mesmo quando não é injectiva. Por exemplo, a aplicação de filtros lineares mantém a reversibilidade se existir uma simetria temporal. Para a classe dos modelos lineares, Weiss(1975) mostrou, dentro da classe dos modelos lineares, que se um processo é reversível então tem inovações gaussianas, este é o recíproco, na classe dos modelos lineares, de um resultado conhecido. Excluindo o caso trivial de processos lineares com inovações i.i.d. todo o processo linear reversível é obrigatoriamente gaussiano. Esta é uma vantagem clara sobre os testes de simetria e gaussianidade de Rao-Gabr e Hinich que, como se indicou, não têm potência sobre alternativas lineares com ruído simétrico não gaussiano. Diks, [Dik99] apresenta ainda métodos alternativos para lidar com a dependência temporal, o problema é colocado

num âmbito mais geral, a estimação da variância de um estatística-U em observações dependentes.

Um outro teste proposto por [Dik96] relevante para as aplicações é o de diferenças em medidas de reconstrução.

Sejam μ_1 e μ_2 duas medidas de probabilidade em $\mathbb{R}^m$. Tal como o teste anterior, a distância pode ser estimada a partir de um conjunto de vectores amostrado independentemente de μ_1 e μ_2 . Considerando duas séries $\{\mathbf{X}_i\}_{i=1}^{N_1}$ e $\{\mathbf{Y}_i\}_{i=1}^{N_1}$ de μ_1 e μ_2 respectivamente. A quantidade $Q = (\mu_1 - \mu_2)$ onde a forma bilinear é a (3.96) tem o valor zero se $\mu_1 = \mu_2$ e $Q > 0$ para a situação oposta. Os autores, [Dik96] propõem como estimador de Q a variável aleatória, $Q = Q_{11} + Q_{22} - 2Q_{12}$ onde,

$$Q_{kl} = \iint h(\mathbf{s}, \mathbf{t}) d\mu_k(\mathbf{s}) d\mu_l(\mathbf{t}) = , k, l \in \{1, 2\}$$

$$h(\mathbf{s}, \mathbf{t}) = \exp(-\|\mathbf{s} - \mathbf{t}\|^2 / (4d^2)) \quad (3.105)$$

$$(3.106)$$

$h(\mathbf{s}, \mathbf{t})$ é a função núcleo. Os estimadores de Q_{11} , Q_{12} e Q_{22} são,

$$\hat{Q}_{11} = \frac{1}{\binom{N_1}{2}} \sum \sum_{1 \leq i < j \leq N_1} h(\mathbf{X}_i, \mathbf{X}_j)$$

$$\hat{Q}_{22} = \frac{1}{\binom{N_2}{2}} \sum \sum_{1 \leq i < j \leq N_2} h(\mathbf{Y}_i, \mathbf{Y}_j)$$

$$\hat{Q}_{12} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} h(\mathbf{X}_i, \mathbf{Y}_j) \quad (3.107)$$

Sendo uma estatística-U, o estimador de Q é não enviesado. O mesmo autor obtém ainda uma expressão para a variância de $\hat{Q}$ condicionada ao conjunto de vectores observados. No entanto, essa expressão não se aplica a vectores de reconstrução uma

vez que a dependência temporal obriga à eliminação de alguns vectores invalidando a expressão.

Tal como acontece no teste de reversibilidade há necessidade de, tendo em vista a redução do efeito da correlação temporal, usar a correcção de Theiler para o cálculo de $h(\mathbf{X}_i, \mathbf{X}_j)$ e $h(\mathbf{Y}_i, \mathbf{Y}_j)$ e a média de blocos para o cálculo de $h(\mathbf{X}_i, \mathbf{X}_j)$ e $h(\mathbf{Y}_i, \mathbf{Y}_j)$ e $h(\mathbf{X}_i, \mathbf{Y}_j)$.

A estatística do teste é obtida padronizando a estatística Q , isto é, $S = Q/\sqrt{\sigma_c Q}$. Uma vez que a função núcleo usada é definida positiva então o critério de rejeição é unilateral à direita. Assumindo uma distribuição unimodal para S , [Puk94], a probabilidade de encontrar um valor de S maior que 3 é menor do que 0.05.

3.8 Previsão no espaço de fase

A previsão é um procedimento diferente de outros já abordados noutras secções e que visavam uma melhor compreensão do fenómeno. A previsão implica algum tipo de modelação, como tal é uma tarefa de criação diferente da de estimar um invariante, por exemplo.

Prever o valor futuro de uma variável é um problema clássico em análise de séries temporais. É claro que para se fazer boas previsões ou para fazer melhores previsões ter-se-á obrigatoriamente que usar métodos que permitam extrair mais informação sobre o sistema.

Todos os métodos de análise não-linear de séries que foram estudados até aqui são medidas indirectas de previsibilidade.

As fontes de previsibilidade nos dados podem ser de diferentes naturezas, um sinal constante no tempo é fácil de prever, um sinal periódico tem o mesmo grau de dificuldade após a observação de um ciclo. Os números aleatórios encontram-se no lado oposto em termos preditivos, a melhor previsão é o valor médio. Os sinais interessantes encontram-se entre estes dois, não são puramente periódicos nem puramente aleatórios contendo algum tipo de relação entre os seus valores que pode ser usada para

previsão. A maior ou menor previsibilidade vai depender da natureza e intensidade das correlações dos dados. O capítulo dois foi dedicado ao modelos associados a séries cujas únicas correlações temporais são as mais bem estudadas, as correlações lineares. Refira-se ainda que a característica mais marcante das séries lineares gaussianas é o facto de o espectro e a *FAC* retirarem toda a informação dos dados possibilitando a existência de um preditor óptimo, [Pri81].

Um sistema determinístico não linear apresenta outros tipos de correlação. A evolução no tempo é determinada sem ambiguidade pelo estado do sistema em cada instante o que implica da mesma forma uma função de previsão determinística. Acontece que em sistemas observados o conhecimento do estado nunca é exacto e como se mostrou anteriormente na presença de *DSCI* as diferenças entre o valor exacto e o observado ou aproximado aumentam exponencialmente a uma taxa finita. A maior ou menor intensidade depende do número e valor dos expoentes positivos de Lyapunov. A existência de relações não lineares entre as variáveis promove o aparecimento de correlação não linear que não é captada na totalidade pelas ferramentas lineares usadas para estimar modelos lineares.

Identificam-se então duas fontes de previsibilidade, a existência de ligação linear entre as variáveis mesmo sobre perturbação aleatória e o determinismo. A generalidade dos sistemas caóticos conhecidos apresentam um decaimento exponencial da autocorrelação e um espectro de banda larga pelo que a construção de qualquer modelo predominantemente determinístico para esse tipo de dados não se pode basear na autocorrelação. Os métodos com mais sucesso e aqueles que de uma forma óptima retiram informação dos dados são os que exploram a correlação linear ou o determinismo não linear puro.

Dada uma série $\{X_n\}_{n \in \mathbb{N}}$ deve procurar-se o preditor de $\{X_{n+s}\}$, $s > 1$ como função dos valores passados,

$$\begin{aligned}\hat{X}_{n+s} &= F(x_n, x_{n-1}, \dots, x_{n-k+1}) \\ &= E(X_{n+s} | X_n = x_n, \dots, X_{n-k+1})\end{aligned}\tag{3.108}$$

No caso dos modelos lineares, F tem a forma indicada em (2.9). Usando, por exemplo, a modelação Box-Jenkins empregam-se autoregressões lineares com inovações gaussianas para estimar os parâmetros de F . Se F não possuir uma forma linear, o modelo linear não é apropriado. Como se referiu, as séries resultantes da observação de sistemas dinâmicos caóticos satisfazem, em geral, autoregressões não lineares. Considerando o exemplo mais simples, seja $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função caótica, a série obtida por $x_{n+1} = f(x_n)$ para algum $x_0 \in \mathbb{R}$ satisfaz um modelo de autoregressão linear de 1ª ordem com função não linear de autoregressão, f .

Considerando uma autoregressão não linear de uma ordem genérica com função caótica, T , construam-se vectores de reconstrução a partir da série, $X_i = (x_{i-k+1}, \dots, x_i)$ o teorema da reconstrução implica que, genericamente, para um k suficientemente grande, os vectores X_i acumulam-se numa região de $\mathbb{R}^k$ difeomórfica ao atractor original assim como a existência de uma função dinâmica difeomorfa à original função caótica, T . Na ausência de ruído, existe uma função $G : \mathbb{R}^k \rightarrow \mathbb{R}^k$ difeomorfa a T tal que $X_{i+1} = G(X_i)$. Em termos da série original a relação anterior mostra que existe uma função $F : \mathbb{R}^k \rightarrow \mathbb{R}$ tal que,

$$x_{i+k} = F(x_i, x_{i+1}, \dots, x_{i+k-1}) \quad (3.109)$$

A expressão (3.109) mostra que a série $\{x_i\}_{i \in \mathbb{N}}$ satisfaz um modelo de autoregressão não linear onde F é uma função não linear desconhecida e não restrita a qualquer forma paramétrica, [Bor98].

Na perspectiva dos métodos que exploram o determinismo não linear, qualquer fonte de ruído coloca problemas à previsão, conseqüentemente ao falar-se em previsão não linear ter-se-à que referir a redução não linear de ruído.

O primeiro método de previsão não linear é proposto por [Lor69]. Supondo que os dados são gerados por um sistema dinâmico da forma (3.1) e que F , desconhecida, é continua para prever o estado do sistema, $\mathbf{X}_{N+1}$ dado $\mathbf{X}_N$ procura-se no registo dos valores passados, $\mathbf{X}_n$ com $n < N$ o que estiver mais próximo de $\mathbf{X}_N$. Se n_0 for o instante correspondente a tal estado então por continuidade da aplicação, F , $\mathbf{X}_{N+1}$ estará próximo de $\mathbf{X}_{n_0+1}$. O artigo de Lorenz é anterior ao teorema de reconstrução

por isso tem que ser adaptado para se poder usar num espaço de fase reconstruído. Continuando com a suposição de que os dados provêm de um sistema dinâmico possui-se uma série de valores que podem não ser observações de uma variável do sistema mas sim de uma função de observação de uma ou mais variáveis do sistema que é, geralmente, desconhecida mas que se supõe diferenciável, s , tal que a série é $s_n = s(\mathbf{X}_n)$, $n = 1, \dots, N$. Usando funções de desfaseamento reconstrói-se o sistema num espaço de dimensão m com desfaseamento temporal, τ obtendo-se vectores da forma,

$$\mathbf{s}_n = (s_{n-(m-1)\tau}, \dots, s_{n-\tau}, s_n) \quad (3.110)$$

procedendo da mesma forma procura-se o vector, $\mathbf{s}_{n_0}$ mais próximo no sentido de uma norma a fixar e a previsão para $s_{N+\Delta N}$ é $s_{n_0+\Delta N}$. Neste contexto o método foi, como indicam [Kan99] proposto por [Pik86]. Este método tem a desvantagem de poder reproduzir porções de órbitas passadas como previsões além de que qualquer medida envolve um dispositivo com precisão limitada o que faz com que os dados contenham flutuações que se esperam aleatórias. Sendo σ o nível de resolução implica que ao procurar-se o valor mais próximo não se possa distinguir pontos que estejam numa esfera de raio σ . A média aritmética desses preditores é um procedimento mais razoável do que utilizar apenas o mais próximo.

Supondo que a reconstrução se fez num espaço m -dimensional, considerando uma esfera de raio ε em torno de $\mathbf{s}_N$ o preditor de $\mathbf{s}_N$ é,

$$\hat{s}_{N+\Delta n} = \frac{1}{|U_\varepsilon(\mathbf{s}_N)|} \sum_{\mathbf{s}_n \in U_\varepsilon(\mathbf{s}_N)} s_{n+\Delta n} \quad (3.111)$$

onde $|U_\varepsilon(\mathbf{s}_N)|$ é o número de elementos da vizinhança de raio ε de $U_\varepsilon(\mathbf{s}_N)$. Obviamente que a qualidade da previsão vai depender do número de vizinhos e respectiva distância ao ponto de referência o que está relacionado com o número de vectores de reconstrução, N . O valor ε terá que ser tal que existam sempre um ou mais pontos na vizinhança do ponto de referência.

Outros autores propõem métodos semelhantes para previsão, em [Sug] os autores usam como preditor uma média ponderada de exactamente $m + 1$ vizinhos para um espaço de mergulho de dimensão m tal que eles formem um *simplex* contendo o ponto. Uma variante estocástica deste método foi proposta em [Pap97], consiste em escolher um vizinho usando uma distribuição cuja probabilidade é mais baixa quanto mais distante se estiver do ponto de referência. Os autores em [Far87] propõem uma aproximação polinomial minimizando uma métrica imposta no espaço de fase e usando pelo menos $m + 1$ vizinhos.

Em [Heg99] os autores propõem que todos os vectores de reconstrução sejam tidos em conta para previsão alegando que na presença de ruído os pontos a uma distância de duas ou três vezes o nível de ruído têm relevância na previsão. Dessa forma propõem como preditor,

$$\hat{s}_{N+\Delta n} = \frac{\sum_{j \neq n} s_{j+k} \omega(|\mathbf{s}_n - \mathbf{s}_j|)}{\sum_{j \neq n} \omega(|\mathbf{s}_n - \mathbf{s}_j|)} \quad (3.112)$$

onde ω é o *kernel*⁵, se se fizer $\omega(z) = \Theta(\varepsilon - z)$ onde Θ é a função de Heaviside obtém-se o usual método da média dos vizinhos mais próximos na vizinhança de raio ε . Aqueles autores propõem para função peso, $\omega(|\mathbf{s}_n - \mathbf{s}_j|) = \exp \left[\frac{-1}{\sigma_N^2} \sum_{j=0}^{m-1} (s_{n-i} - s_{j-i})^2 \right]$ onde σ_N^2 é o desvio padrão do ruído. Uma vez que é quase sempre desconhecido os resultados podem ser muito sensíveis ao seu valor.

O conjunto de métodos de previsão que se baseiam em relações de vizinhança no espaço de fase [Kan99] chamaram *almost model free* que se inscrevem numa classe mais vasta de métodos locais no espaço de fase.

⁵Um *kernel* é uma função peso que depende apenas da distância ao ponto de referência

Capítulo 4

Janelas de correlação

Designa-se por *data-mining* um vasto conjunto de técnicas de extracção ¹ de informação [Fra92] em grandes volumes de dados tais como transformadas de Fourier, análise em componentes principais, *wavelets*, *genetic algorithms* entre outros. Em séries temporais a existência de correlação temporal obriga a que algumas técnicas não possam ser usadas e outras tenham que ser adaptadas daí que use o termo *Temporal data-mining* (TDM) em séries temporais. Vários autores usaram técnicas de TDM em séries, [Das97], [Agr93], [Ber96], [Pov99b, Pov98, Pov99a] entre outros. Em [AO01] encontra-se uma boa revisão do estado da arte em TDM. Um dos objectivos em TDM é descobrir relações entre sequências ou segmentos. Este procedimento envolve três passos principais, a representação e modelação dos dados, o estabelecimento de uma função ou medida de semelhança entre sequências e por último a aplicação de modelos para os dados.

4.1 Temporal data-mining

Nos parágrafos seguintes irá descrever-se sucintamente as principais ferramentas da TDM e as implicações para os dados que a sua aplicação possa causar. Do ponto de vista da TDM, a manipulação directa dos dados na forma original de forma eficiente é muito difícil pelo que a representação dos dados constitui um problema fundamental. Para os métodos de representação é muito relevante a descoberta e representação

¹pressupõe uma redução dos dados

de sequências. Um fenómeno temporal sujeito a ruído que apresente comportamentos semelhantes e recorrentes pode apresentar padrões. Usa-se o termo padrão (ou evento) para designar a ocorrência de um número significativo de sequências semelhantes. O reconhecimento desses comportamentos padronizados tem, obviamente, relevância preditiva. Quando se lida com grandes volumes de dados, a manipulação torna-se morosa e dispendiosa e por isso se recorre a transformações através de, por exemplo, aproximações lineares locais. Uma outra alternativa é transformar os dados para um espaço de dimensão inferior, que pode feita no domínio do tempo ou no domínio da frequência. Uma outra alternativa consiste em usar funções lineares por troços para aproximar a série, que é assim dividida em segmentos e a cada um deles é atribuída uma função linear. A representação baseada em transformações consiste em transformar os segmentos da série original em pontos num novo domínio onde se passa a representar a série. A semelhança entre segmentos passa a corresponder uma distância no espaço das frequências. Se utilizarmos a transformada discreta de Fourier basta reter os valores de potência referentes às primeiras k frequências e tomar esse ponto k -dimensional como a sua representação. A análise em componentes principais (*ACP*) corresponde, igualmente, a uma representação dos dados num espaço de dimensão inferior usando para isso o espectro dos valores próprios. Aos maiores valores próprios, em módulos, estão associados os vectores próprios cujas direcções são as que mais contribuem para a variância da série. Supondo que a série está reconstruída num espaço m -dimensional, se o analista concluir que os m_0 valores próprios mais representativos, $m_0 < m$, são valores próprios suficientes para representar a série então truncam-se todos os vectores e obtém-se uma projecção num espaço m_0 -dimensional. O intervalo de tempo correspondente a esta representação passa a ser, m/m_0 .

Os autores em, [Bro94], mostraram que quando $m \rightarrow N$ esta transformação tem as propriedades de uma transformada de Fourier, no entanto, para um m pequeno a (*ACP*) escolhe as estruturas relevantes no espaço em vez das frequências mais relevantes. Numa transformada de Fourier também se pode truncar de acordo com a potência o que seria mais razoável que o procedimento anterior. A razão pela qual não se recorre a esta operação reside no facto de não ser possível inverter o conjunto de

vectores m_0 dimensionais numa série de tamanho Nm_0/m , operação correspondente a uma transformada inversa, [Kan99].

Após representar cada sequência numa forma adequada é necessário definir a medida de semelhança entre sequências de forma a determinar se estão suficientemente próximas para se considerarem semelhantes.

Numa comparação de segmentos colocam-se sempre as questões relacionadas com os *outliers*, ruído e problemas de *scaling*. As medidas de semelhança a usar devem ser resistentes a esses efeitos.

No domínio do tempo, quando não existe qualquer transformação dos dados, a medida de semelhança pode ser a comparação elemento a elemento, ou distância euclidiana em $\mathbb{R}^n$, considerando a sequência de tamanho n como um vector.

Nas operações de minagem propriamente ditas, as de maior relevância para previsão são as que dizem respeito à descoberta de regras de associação. As sequências comuns ou semelhantes podem ser usadas para estabelecer regras de associação com relevância preditiva.

Na secção seguinte apresentamos uma aplicação de técnicas de *TDM* a séries hidrológicas.

4.2 Aplicações de *Temporal data-mining*

Como se referiu na secção anterior, a abordagem à análise de séries temporais via TDM levantam questões pertinentes do ponto de vista probabilístico e dinâmico, isto é, supondo que a série é uma realização de um processo estocástico, ou uma medição de uma variável dinâmica. O uso de representações com transformações simplifica os dados, porém, pode deixar-se de se perceber a relação entre o fenómeno e os dados transformados.

O reconhecimento de padrões passa pelo estabelecimento de uma medida de semelhança entre segmentos que permite, assim, identificar segmentos semelhantes. A obtenção de uma boa medida de semelhança não é tarefa fácil. O uso da correlação como medida de semelhança tem como vantagem sobre as distâncias euclidianas o facto

de conseguir reconhecer relações entre segmentos de diferentes ordens de grandeza (*scaling*), além disso, revela alguma resistência ao ruído e permite uma interpretação física directa. Porém, tem como desvantagem o facto de ser sensível a *outliers*.

Um regime de um caudal descreve um comportamento sazonal médio que reflecte as condições climáticas e físicas da bacia hidrográfica. Os estados preferenciais do sistema gerador dos caudais correspondem aos padrões dos regimes. O padrão está relacionado com o lado mais previsível do rio que ocorre geralmente na ausência de precipitação. O uso de métodos não-paramétricos de previsão em séries temporais como o [Kar87a] e [Kar87b] está relacionado com a procura e uso de padrões na previsão. No artigo, [Yak87], o autor propõe-se inferir a função de regressão não-linear², $m(x) = E[X(n+1) | X(n) = x]$, a partir dos valores iniciais de realizações de processos estacionários markovianos ergódicos com função densidade de probabilidade (f.d.p) de transição, $f(y|x)$ da variável aleatória, $X(n+1) | (X(n) = x)$. Assume-se que o processo tem uma f.d.p. estacionária para a distribuição incondicional de $X_i(n)$. O autor mostra que, sob a suposição da propriedade *mixing* fraca, o estimador da função de regressão converge em média quadrática. O autor refere igualmente a relevância do resultado no âmbito da identificação de padrões e aplica o método no processo hidrológico da precipitação-caudal. A relevância do método e a relação que tem com a metodologia que se vai aplicar merece uma breve descrição.

Considerando uma série nas condições indicadas acima, define-se $\{k(n)\}$, $n > 1$, como uma sucessão de inteiros não decrescente positiva onde $k(n) < n$. Sendo $\bar{R}(n)$ o raio da mais pequena bola centrada no estado x que contém $k(n)$ dos estados $X(j)$, $j < n$. Seja $B(r)$ a bola de raio r centrada em x , tome-se RO como um número relativamente grande, por exemplo a amplitude dos dados, e $R(n) = \min\{\bar{R}(n), RO\}$. Adoptando a regra,

Definição 4.2.1 *Seja $B_n = B(R(n))$ a bola de raio $R(n)$ centrada em x . A estimativa k -NN (nearest neighbours de $m(x)$) é dada por,*

$$m_n(x) = \frac{1}{k(n)} \sum X(j+1) : j \in S(n) \quad (4.1)$$

onde $S(n) = \{j : |X(j) - x| \leq R(n)\}$.

²ver capítulo 2 última secção

o índice j varia de 1 a $n - 1$ pelo que $m_n(x)$ fica completamente determinado pelo primeiro segmento inicial de tamanho n de $\{X(i)\}$.

Seja RO um qualquer número positivo, quase certamente que existe um inteiro $N = N(RO)$ tal que para todo o $n > N$ vem $\bar{R}(N) < RO$. Tendo em conta que o processo é ergódico, sendo $N(RO, n)$ o número de observações $X(i)$, $1 \leq i < n$ que fica na bola $B(RO)$ então com probabilidade 1, [Doo48], $N(RO, n)/n \rightarrow \pi(B(RO))$, onde π é a distribuição de probabilidade estacionária.

Este método está próximo do tipo de métodos locais *almost model free* referidos no capítulo seguinte. Aqui o valor seguinte tem como preditor a média dos sucessores dos valores passados da série que se encontram a uma distância inferior ao raio, r do valor actual. No caso do processo ser ergódico mostra-se que a proporção de pontos nessas condições é proporcional à probabilidade estacionária da bola.

Nos trabalhos de [Yak87] e [Kar87a] o método desenvolvido pelo primeiro é aplicado em processos hidrológicos. A série de referência é constituída por dados diários de caudal e de dados de precipitação numa frequência de 6 horas. O vector de referência usado é, $Z(n) = (Y(n), Y(n - 1), Y(n - 2), X(n), X(n - 1))$ onde os primeiros três valores dizem respeito ao caudal e os dois últimos à precipitação. Após escolha de uma norma adequada, a estimativa é encontrada após minimização de uma soma de quadrados cujos parâmetros são os pesos da métrica. Escolhe para esse efeito uma conjunto inicial de observações. Os resultados obtidos revelaram-se melhores que aqueles obtidos pela modelação *ARMAX* e pelo modelo físico para a relação precipitação-caudal (*Sacramento model*) de [Bur80].

O estimador de Yakowitz inclui-se na classe das estatísticas *k-NN* cuja caracterização é difícil razão pela qual, até à data, os métodos baseados em estatísticas sobre vizinhos não têm as suas propriedades estudadas, [Cha01].

Em aplicações deste tipo de métodos fora do contexto hidrológico, devemos fazer-se referência ao trabalho de [Mei98] em séries de teores de $CaCO_3$ ³. O objectivo era prever o teor de $CaCO_3$ que sendo baixo inviabilizaria a exploração do minério. Para isso os autores usaram um esquema de previsão semelhante ao de Yakowitz com a

³carbonato de cálcio, essencial para a produção de cimento.

diferença de que em vez da série numérica, que estava, segundo os autores, sujeita a grande contaminação, usavam uma versão classificada. Assim, relativamente ao teor de CaCO_3 os dados passavam de uma escala absoluta para uma escala ordinal.

Na secção seguinte vamos apresentar uma técnica de previsão baseada no reconhecimento de padrões usando a correlação como medida de semelhança.

4.3 Previsão usando Janelas de correlação

A nossa proposta de técnica de previsão inclui-se no conjunto de técnicas de TDM. Como critério de escolha dos vizinhos recorreremos, tal como em [Yak87] a uma medida de proximidade entre pontos num espaço n -dimensional. Os vizinhos encontrados são usados para produzir a previsão. Não se pretende aqui estabelecer quais as propriedades do preditor sugerido, para os diversos cenários possíveis de processos geradores de séries temporais, os processos estocásticos ou sistemas dinâmicos. No entanto, se a série apresentar um comportamento não estacionário ou extremamente irregular o preditor dará maus resultados. A motivação para a metodologia que se vai descrever de seguida encontra-se na existência, nos caudais do rio, de comportamentos próximos do determinismo e que são recorrentes, como se justifica no capítulo 6.

Em vez de se utilizar a distribuição de $X(n+1)$ condicionada a $X(n)$ usam-se janelas de tamanho w (variável) e em vez da distância euclidiana entre pontos usa-se a correlação linear.

O valor a prever é encarado como fazendo parte de uma subsucessão (consecutiva) recorrente que pode ser estimado procurando eventos correlacionados no passado. Dessa forma fixando a subsucessão de tamanho n cujo último valor é o actual procura-se no registo histórico segmentos que apresentam com ele uma correlação significativa. Sendo X e Y dois segmentos de tamanho n (não gaussianos), e assumindo que a população que lhes deu origem tem correlação ρ diferente de zero, pretende-se saber se a correlação entre eles é significativa. Como regra de determinação de significância optou-se por usar o limite superior do intervalo de confiança tendo como variável fulcral a do teste de Fischer para o caso de populações de variável bivariada normal,

[Sap90], [Mur90]. É um facto que a normalidade de (X, Y) não pode ser assumida, no entanto, a utilização da estatística correspondente a essa situação deve ser encarada como uma aproximação.

Os segmentos com correlação significativa são candidatos a produzir a previsão. O procedimento é efectuado para janelas de dimensões crescentes até que já não existam segmentos com correlação significativa. O conjunto de candidatos a previsão é obtido usando a regressão linear, ou seja, estabelecendo a ligação funcional entre os segmentos. O preditor será uma estatística, o mais vulgar deverá ser uma de localização central, no entanto, o preditor escolher-se-á de entre um conjunto de estatísticas onde se encontrará a média e estatísticas de ordem. A dificuldade desta escolha está, como se irá ver capítulo seguinte, nas distribuições dos valores extrapolados das regressões efectuadas nos segmentos correlacionados, podem ser muito diferentes dependendo da dinâmica dominadora no rio. No caso de ocorrerem muitos valores extremos a opção pela média pode não ser muito aconselhada.

Uma vez que, como já foi referido, a suposição de normalidade não é válida, é desejável a aplicação de métodos alternativos de regressão mais robustos, tais como a regressão *LAD* e *M*. Estes métodos costumam fornecer melhores resultados especialmente quando a suposição de normalidade não é válida, [Bir96].

O método proposto por nós é não paramétrico e quase sem pressupostos, excepto no uso da regressão linear simples. Pode ser adaptado tendo em conta as características da série, por exemplo, se o *scaling* não estiver presente então pode-se eliminar a contribuição de segmentos correlacionados em diferentes escalas.

Este método de previsão aplicado a trajectórias de processos lineares gaussianos não dá bons resultados o que mostra que se o ruído tiver um grande peso na variação ocorrerão poucos segmentos de valores consecutivos correlacionados o que faz com que a predictabilidade obtida seja pequena. Este facto parece indicar que segmentos correlacionados estão associados à presença de determinismo.

A previsão consiste em usar o conjunto de valores extrapolados a partir da recta de regressão estimada. O método é muito simples e interpretável. Em sistemas dinâmicos caóticos e na ausência de ruído, trajectórias com pontos iniciais próximos têm uma

evolução próxima, pelo menos em alguma vizinhança futura, o que implica que existam segmentos de trajetórias fortemente correlacionados, [Siv],[Kan99],[Aba96a].

O sistema de Hénon é usado para exemplificar o método de previsão apresentado usando a mediana dos valores obtidos por extrapolação.

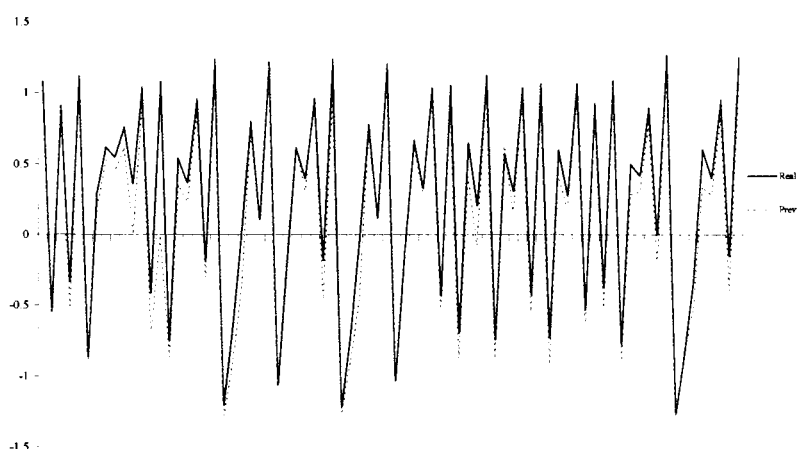


Figura 4.1: Previsão a um passo para o atrator de Hénon usando janelas de correlação.

O resultado obtido está de acordo com as características de sistemas caóticos, ou seja, a possibilidade de previsão razoável a curto prazo. O resíduo médio quadrático obtido para a previsão a um passo foi de 0.0197 para dados com média de 0.26 e desvio padrão de 0.72. Uma simples análise do cronograma dos valores reais e previstos mostra que a previsão é quase sempre inferior ao valor real mostrando que o uso da mediana do conjunto de valores extrapolados não é adequado, é aconselhável tentar um quantil de ordem superior.

Existe um paralelismo na abordagem ao problema da previsão de uma série entre o método vizinhos mais próximos e o das janelas de correlação. No primeiro a série é dividida em vectores de reconstrução num espaço de dimensão fixa. Os vectores apresentam sempre correlação temporal, até pela forma como são construídos. A previsão é obtida usando a informação de órbitas próximas, para isso usa-se a métrica euclidiana ou a norma do supremo. Nas janelas de correlação, usa-se informação de segmentos passados de dimensão variável. A métrica usada é a correlação linear. Para

os segmentos correlacionados calculam-se os parâmetros da regressão sobre o segmento fixado cujo valor futuro se pretende prever. Neste último passo está a maior diferença entre as duas abordagens.

Capítulo 5

Análise dos caudais diários do rio Paiva.

Este capítulo consiste na análise e caracterização e previsão da série dos caudais médios diários do rio Paiva. Iniciamos com algumas considerações gerais sobre o rio Paiva e sobre a precipitação na sua bacia. Comentamos, também, vários aspectos sobre a medição dos caudais .

A necessidade da utilização de métodos e técnicas não lineares, como por exemplo as provenientes de sistemas dinâmicos, é indicada pelos resultados dos testes de linearidade, gaussianidade e reversibilidade no tempo (tabelas (5.5), (5.6), (5.7) e (5.8)). O fenómeno dinâmico que melhor se perfila para modelar os caudais diários do Paiva é a intermitência, introduzida por Pomeau e Manneville, fazendo-se a analogia entre a fase laminar e as descidas de caudais e entre a fase irregular e as subidas de caudais. Na intermitência dinâmica a duração de ambas as fases depende da dinâmica do sistema; em contraste, conclui-se que a intermitência dos caudais diários se deve à chuva cuja dinâmica não é captada na reconstrução obtida com os caudais diários. Este facto comprova-se com os resultados de previsão obtidos (5.43). Desta forma, a duração da fase regular do comportamento dos caudais diários do rio depende exclusivamente do tempo entre as ocorrências de precipitação.

A diferença no comportamento dos caudais na ausência de chuva e na presença de chuva justifica uma análise dinâmica e estatística separada. Para isso começamos por fazer uma reconstrução de Takens, a partir da série, e calculamos os declives da curva-correlação (5.28). Para o regime de descidas de caudais consegue-se obter uma

previsão razoável para o valor no dia seguinte e a informação mais importante para essa previsão é o caudal do próprio dia (5.18). Conhecido o caudal no dia anterior, obtemos ainda uma distribuição de probabilidade que aproxima a distribuição dos diferenciais de caudais para o regime de descidas de caudais. Descobre-se igualmente que existe uma relação aproximadamente linear entre a média do diferencial e o valor do caudal, que é mais intensa nos caudais mais baixos. O desvio padrão tem um valor próximo da média que é aproximadamente 5-6% do valor do caudal.

Utiliza-se o método dos vizinhos mais próximos para fazer previsão pontual dos caudais médios diários, a um dia, obtendo-se resultados análogos para dimensões do espaço de reconstrução entre 1 e 4. A título comparativo, utilizaram-se as janelas de correlação e métodos lineares de previsão de série temporais (modelo ARIMA) também para produzir previsões pontuais. Estas técnicas resultaram em piores resultados de previsão quando comparados com o método dos vizinhos mais próximos (5.62). Notamos que as energias da análise em componentes principais, realizada após a reconstrução de Takens, apontam na mesma direcção dos resultados anteriores.

Mostramos a existência de uma curva das médias dos caudais, dado o caudal no dia anterior, que apresenta dois declives bem distintos, um para valores pequenos e outro para valores grandes de caudais, podendo assim ser aproximada por uma aplicação afim com dois troços (5.6).

Um outro aspecto da auto-semelhança encontrado na série temporal é revelado pelos três primeiros harmónicos da função de autocorrelação (5.5). O primeiro harmónico revela a sazonalidade anual do sistema com grande precisão, enquanto que os segundos e terceiros harmónicos mostram uma variação aproximada de 45 dias em torno de dois anos.

5.1 Algumas notas sobre o rio Paiva e sua bacia hidrográfica.

O rio Paiva nasce na serra de Leomil próximo de Moimenta da Beira e a sua bacia tem uma área de cerca de $700Km^2$. A disposição e localização da bacia hidrográfica e a

altitude poderão explicar as grandes diferenças encontradas nos postos de observação de precipitação.

O rio Paiva não é um rio intermitente e o período estival raramente faz desaparecer o caudal de superfície, pelo menos assim acontece, em Fragas da Torre, nos 53 anos de registo.

Segundo indicação da empresa responsável¹, o registo do nível de caudal é efectuado com um limnigrafo², assim os valores estimados do caudal são obtidos operando o nível medido de água com a curva de vazão.

Nos registos desde 1946 existe um ano com ocorrências de caudal inferiores a $0.05m^3/s$ (limite inferior de medição) em Castro Daire e em Fragas da Torre o valor mínimo é igual ou superior a esse valor nos 53 anos de observação. O valor $0m^3/s$ corresponde a um caudal abaixo do limiar de detecção do limnigrafo mas poderá não corresponder a um caudal de superfície inexistente. Esta evidência significa que a inexistência de um caudal superficial em Fragas da Torre é um acontecimento raro numa escala de tempo humana.

Os perfis escolhidos para estabelecimento e aferição da curva de vazão correspondem a zonas rochosas sendo muito pouco provável a ocorrência de alterações significativas do leito.³

O processo de medição envolve sempre um erro. É natural supor, tendo em conta o que dissemos sobre o limiar de detecção do caudal que o erro de medição depende da altura de água. Para quem analisa os caudais médios diários pode contar então com um erro que varia consoante o valor do caudal atingido. Os caudais mais elevados, pelo facto de a corrente ocupar zonas de mais e maior vegetação, uma vez que, quase sempre, o rio nunca lá passa, provoca maior turbulência da corrente podendo daí resultarem erros mais significativos.

Os dados relevantes para esta tese são os caudais médios diário medidos em Fragas da Torre e Castro Daire. O autor teve acesso a dados de precipitação⁴ respeitantes

¹Hidrorumo

²instrumento de medição do caudal

³informação da empresa Hidrorumo na pessoa do Sr. Eng. Aquino Paredes.

⁴gentilmente cedidos pela DROAT na pessoa do Sr. Engenheiro Rodrigues

aos totais diários medidos às 9h00 relativos a 7 postos de observação no período de 1/10/1983 a 30/09/01, incompletos em 2 estações estando uma delas situada em Castro Daire. Os registos de precipitação são recolhidos por um observador e não por um processo automático como acontece com os caudais.

Procedendo a um levantamento das subidas de caudal sem que ocorra chuva, constata-se que o erro de medição de caudal em Castro Daire tem um peso relativo maior do que em Fragas, talvez devido ao menor volume de água em causa. Em 16 anos de observação verifica-se que o caudal aumenta, sem que tenha ocorrido precipitação no próprio dia e no anterior, 192 vezes em Fragas e 566 em Castro Daire, correspondendo a 6.17% do período de observação em Fragas e 18.2% em Castro.

Exceptuando o dia 3 de Outubro de 1986⁵, a média do erro relativo nas subidas sem precipitação é cerca de 3% em Fragas da Torre e 33% em Castro Daire. No entanto, para este último conjunto de dados, a média do erro relativo passa para metade a partir de 1986 enquanto que em Fragas a média do erro é mais regular. Em ambos a evolução do erro relativo na ausência de precipitação diminui com o aumento de caudal, tal como seria de esperar.

Pelo exposto é natural escolher-se os caudais medidos em Fragas como principal conjunto de dados em detrimento dos caudais medidos em Castro.

Ao contrário do que acontece com os caudais, os dados diários de chuva são insuficientes para caracterizar a precipitação na bacia uma vez que a Organização Meteorológica Mundial recomenda o cálculo de módulos pluviométricos para um número de observações superiores a 30 anos. No caso da bacia do Paiva, dispomos apenas de 18 anos de precipitação diária observada pelo que os dados de chuva são usados apenas como indicadores.

Outros dados relevantes para um conhecimento mais aprofundado da bacia do Paiva tais como; informação de mais estações de precipitação, temperatura e evaporação, e uma completa descrição dos declives, tipos de solo e respectivas capacidades de impermeabilidade e cobertura vegetal não estão disponíveis. Assim a análise a desenvolver tem necessariamente limitações na metodologia não sendo possível, por exemplo, usar

⁵que parece ter sido um erro grosseiro de medição

modelos que integrem muitas variáveis descritoras como nos trabalhos [Lae97],[Bur80] que usam modelos muito abrangentes.

5.2 Caracterização dos caudais médios diários e da precipitação na bacia.

Nesta secção caracterizar-se-ão, tanto quanto possível, os caudais médios diários do rio Paiva medidos em Fragas da Torre e Castro Daire. Para o efeito dispõe-se de um registo de 53 anos, 1946-99 perfazendo 19358 observações⁶, figuras (5.1,5.2,5.3) respectivamente. Os dados de precipitação dizem respeito a 7 postos de observação na bacia do rio Paiva: Pendilhe, Mezio, Sobrado, Touro, Mosteiro, Vila Nova de Paiva e Castro Daire, referentes aos anos 1983/2001 estando incompletos em duas estações.

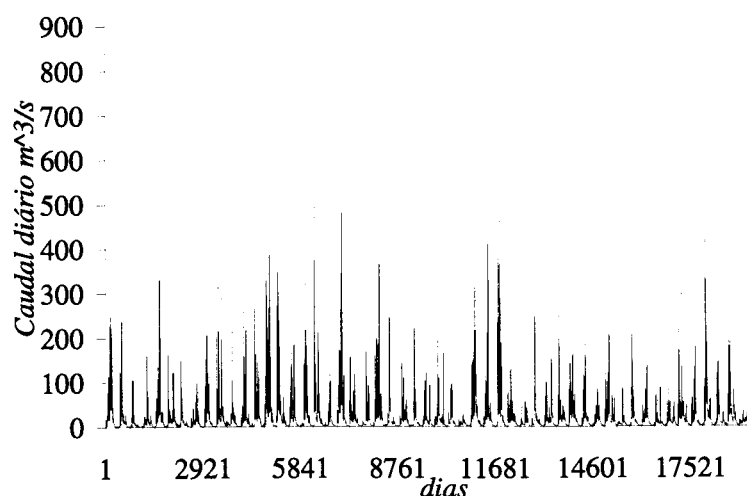


Figura 5.1: Cronograma dos caudais do rio Paiva em Fragas da Torre de 1/10/1946 a 30/09/1999.

Os cronogramas totais e parciais, figuras (5.2,5.3) mostram que os caudais apresentam grande variabilidade. É uma consequência das características da bacia hidrográfica, os 16 anos de valores irregulares de precipitação assim o comprovam.

Na tabela, (5.1), pode verificar-se que o desvio padrão é o dobro do valor médio, e os

⁶dados gentilmente cedidos pela empresa *Hidrorumo*.

máximos anuais superam quase sempre valores que são 20 vezes mais elevados do que o valor médio indicando também que existe um período de estio com caudais muito baixos.

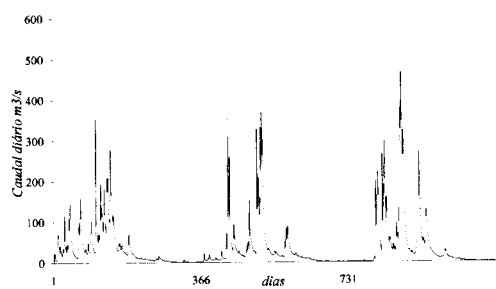


Figura 5.2: Cronograma dos caudais do rio Paiva em Fragas da Torre de 1/10/1976 a 30/09/1979.

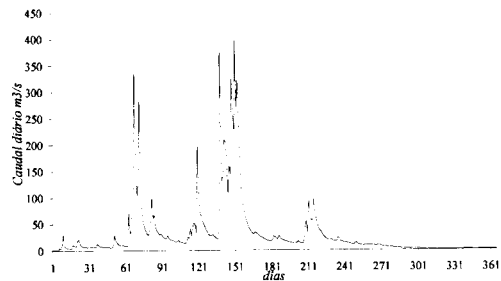


Figura 5.3: Cronograma dos caudais do rio Paiva em Fragas da Torre de 1/10/1977 a 30/09/1978.

Estatística	Fragas da Torre	Castro Daire
média	20.73 m³/s	6.62 m³/s
mediana	5.66m³/s	1.3 m³/s
desvio padrão	39.3m³/s	13.61m³/s
Coef. de Variação	1.9165	2.06
Assimetria	5.3	5.7
Curtose	45.98	45.56
Máximo	920.0 m³/s	200.0m³/s
Mínimo	0.06m³/s	0.0 m³/s

Tabela 5.1: Várias estatísticas para os caudais diários do rio Paiva.

O histograma dos caudais em Fragas da Torre e a mediana do conjunto mostram a predominância de caudais mais baixos, 75% dos caudais observados estão abaixo do nível médio. O estudo da autocorrelação e do periodograma fornecem as primeiras indicações da aperiodicidade da série.

Como se pode verificar, apesar do comportamento global indicar uma sazonalidade como um comportamento recorrente, a curva apresenta grandes irregularidades. Deve salientar-se a proximidade entre os valores da *FAC* estimada dos caudais medidos em Castro Daire e Fragas da Torre, a diferença absoluta máxima é de 0.0246 em que o

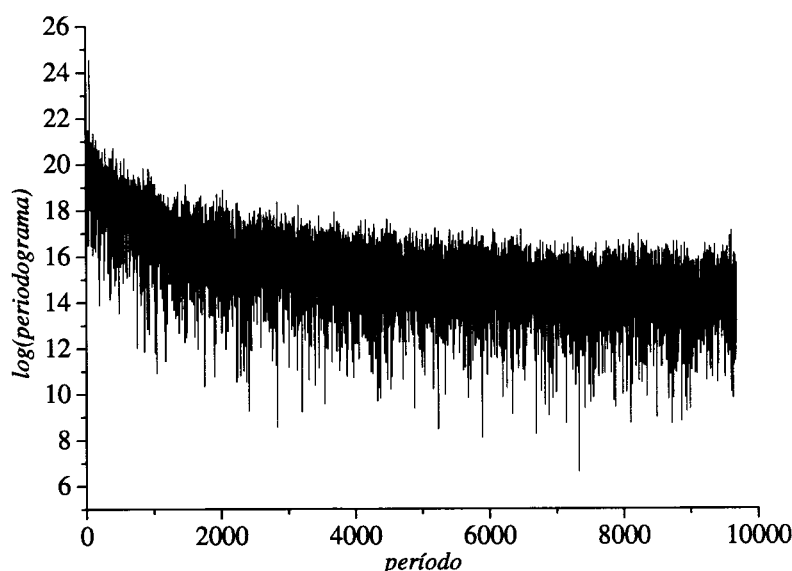


Figura 5.4: Logaritmo do periodograma dos caudais do rio Paiva em Fragas da Torre.

valor da correlação é de 0.15.

Este facto comprova que a componente linear do sistema é aproximadamente a mesma nas duas estações.

O logaritmo do periodograma, figura (5.4), indica um comportamento que reforça a suspeita do afastamento da linearidade/normalidade, apresenta um valor elevado para uma banda larga de frequências e com um decaimento lento, aproximando o comportamento dos caudais do comportamento de outros fenómenos não lineares [Ton90]. Uma vez que o espectro de potência reflecte a contribuição de todos os períodos possíveis, desde o intervalo de amostragem até ao intervalo de tempo correspondente ao tamanho da série, os valores elevados para frequências baixas significam que poderão existir oscilações importantes que não são tidas em conta devido à frequência de amostragem utilizada. O máximo da potência é atingido na frequência $1/365.245 \text{ dias}^{-1}$ o que corresponde ao ciclo anual ou sazonalidade. No entanto, registam-se valores elevados nas frequências; $1/691.3571 \text{ dias}^{-1}$, 774.32 dias^{-1} , 967.9 dias^{-1} , $182.6226 \text{ dias}^{-1}$ este último valor corresponde ao meio ano.

A significância dos picos da estimativa do periodograma só pode ser estudada num contexto muito restrito de séries. Os caudais do rio Paiva não se enquadram, como

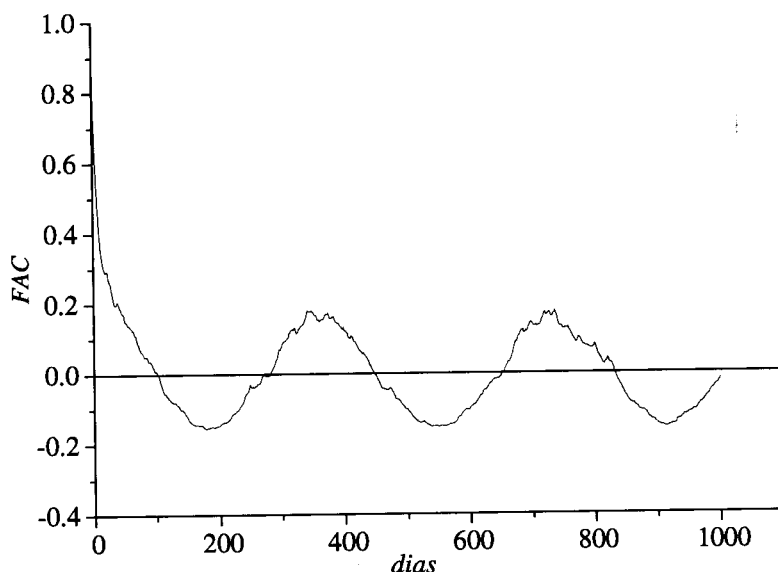


Figura 5.5: *FAC* dos caudais médios diários do rio Paiva em Fragas da Torre.

vamos ver a seguir, nos modelos gaussianos e por isso os picos da estimativa não se podem estudar quanto à sua significância estatística. No entanto, podemos fazer as seguintes considerações, se interpretarmos um ano hidrológico como a observação de um ciclo do rio os dados que possuímos representam apenas 53 ciclos. Os valores elevados de potência em períodos superiores a 1 ano, 691 e 774 dias, respectivamente podem estar a revelar um aspecto de auto-semelhança no comportamento dos caudais. Mostram uma variação aproximada de 45 dias em torno de 2 anos. Esta bifurcação revela a variabilidade dos períodos de seca e de chuva ao longo dos anos.

Os gráficos de curvas de correlação do tipo $E(X_t/X_{t-j}), j \in \mathbb{Z}$ permitem, como se indicou no capítulo 2, verificar se existe um afastamento da linearidade, figuras (5.6) e (5.7). Verifica-se que o afastamento é mais acentuado quanto mais elevado for o desfaseamento temporal considerado e especialmente para valores do caudal mais elevados.

Um aspecto importante que é característica em dados de caudais, [Kra02], é a existência de comportamentos recorrentes. No trabalho de [Cal01] faz-se uma análise dos comportamentos recorrentes dos caudais do rio Paiva e conclui-se que esse com-

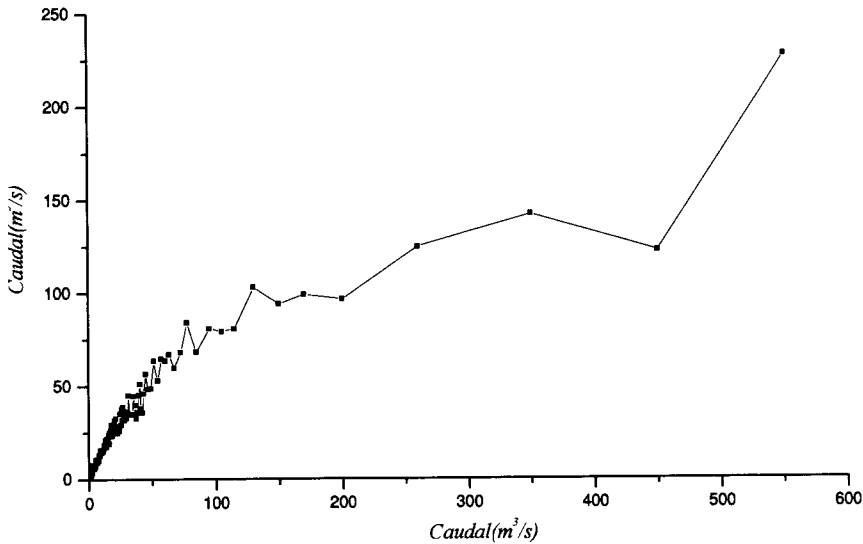


Figura 5.6: Curva de regressão amostral de X_t sobre X_{t-5} .

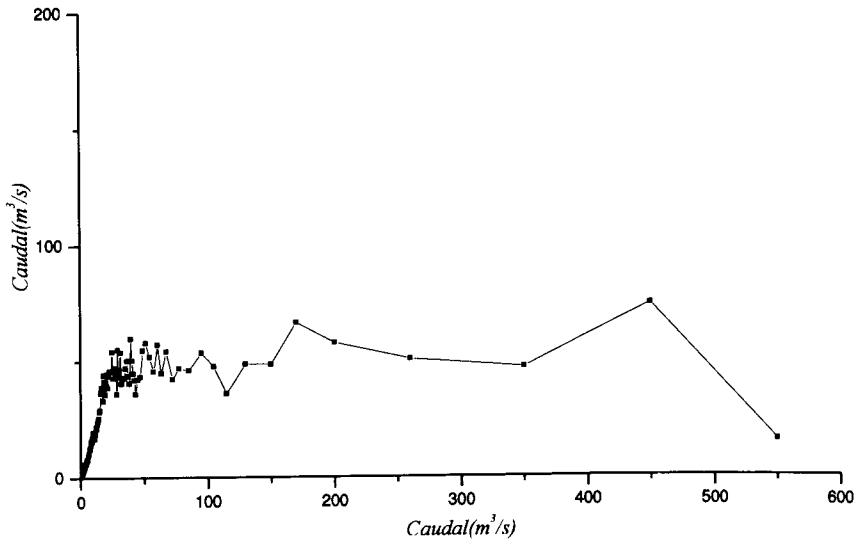


Figura 5.7: Curva de regressão amostral de X_t sobre X_{t-20} .

portamento fornece informação passível de ser usada para previsão. É, álias, esse comportamento que levou o autor a explorar as técnicas de *temporal data mining*. A figura (5.8) representa as correlações lineares entre segmentos, um dos quais fixo, de tamanho 68 dias, dos caudais diários medidos em Castro Daire. Apesar de o segmento considerado ter uma dimensão de 68 dias encontram-se segmentos cuja correlação linear é elevada. Se se usar um segmento mais pequeno encontram-se ainda mais segmentos correlacionados.

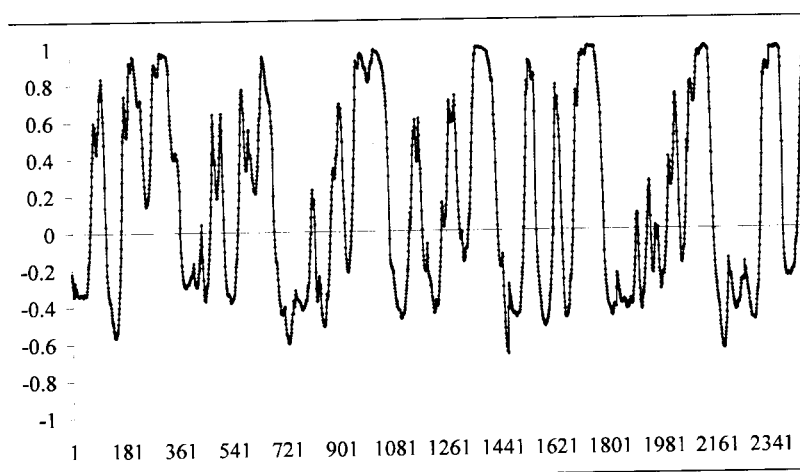


Figura 5.8: Evolução do coeficiente de correlação linear com janelas de tamanho 68.

Este comportamento mostra que têm que existir segmentos *mais ou menos* paralelos nos diagramas de fase obtidos com os mesmos caudais, figuras (5.9,5.10). O uso de diagramas de fase no âmbito das sucessões cronológicas é preconizado por [Ton90]. O autor indica que os diagramas de fase são uma ferramenta útil uma vez que põem a descoberto características próprias de sistemas determinísticos, (ver capítulo 2). As órbitas de alguns sistemas determinísticos mantêm-se próximas se os valores iniciais estiverem próximos, este facto faz com que o diagrama de fase apresente linhas paralelas e por vezes padrões. Esta é uma característica atípica de séries geradas por processos estocásticos o que significa que a série dos caudais apresenta algum determinismo que se está associado à ocorrência de valores semelhantes para as variáveis que dão origem ao caudal do rio.

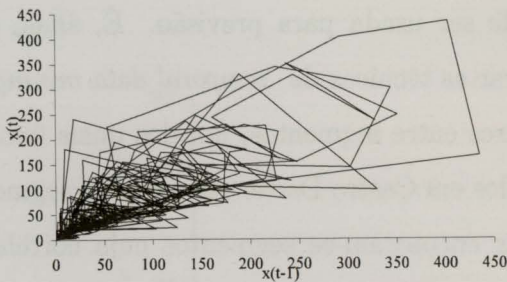


Figura 5.9: Diagrama de fase dos caudais em Fragas com desfaseamento temporal de 1 dia.

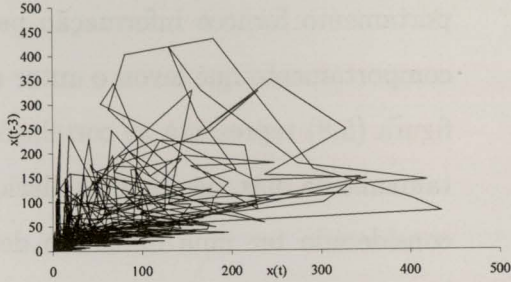


Figura 5.10: Diagrama de fase dos caudais em Fragas com desfaseamento temporal de 3 dias.

Nas figuras (5.11,5.12) representam-se alguns segmentos correlacionados alinhando-os no tempo. Representados desta forma observa-se mais facilmente o comportamento recorrente. A ocorrência destes comportamentos está relacionada, como vamos ver, com o determinismo.

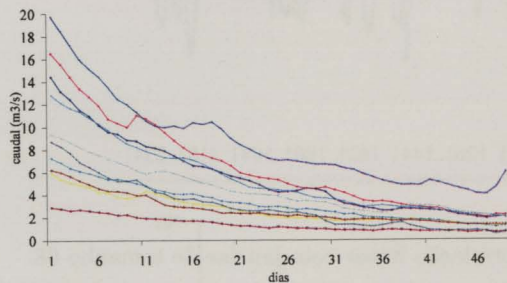


Figura 5.11: Alinhamento de seqüências de tamanho 45 (dias).

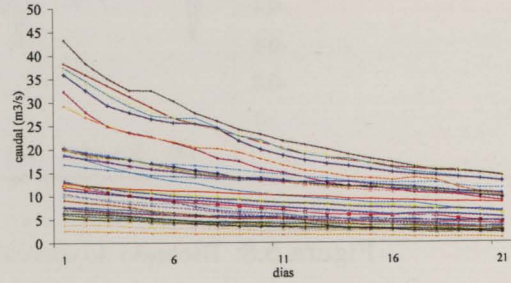


Figura 5.12: Alinhamento de seqüências de tamanho 21 (dias).

Estação	Média	Desvio padrão	dias s/ chuva ⁷
Pendilhe	3204.76mm	1650.71mm	3208
Mezio	1789.84mm	610.67mm	3119
Sobrado	1351.23mm	506.39mm	3681
Touro	1469.82mm	486.55mm	3678
Mosteiro	1520.57mm	446.47mm	3255
V.N.Paiva	1416.76mm	435.34mm	3143
C. Daire	1566.27mm	366.41mm	3375

Tabela 5.2: Algumas estatísticas para a precipitação anual nas 7 estações.

Sendo a precipitação o sistema alimentador do rio será na sua relação com o caudal que se encontrará a justificação para esse determinismo. Embora não seja significativo

do ponto de vista da caracterização climática, as precipitações anuais, para o período observado, nas várias estações ilustram bem as diferenças de precipitação entre as 7 estações, tabela (5.2). Apesar da pequena distância espacial, outros factores como a altitude e a disposição geográfica são extremamente relevantes. As estatísticas da tabela (5.2) dizem respeito à precipitação diária verificada desde 1/10/83 a 30/09/2001 para as estações de Pendilhe, Mezio, Sobrado, Touro, Mosteiro de Cabril, e desde 1/10/1983 a 30/09/1996 para as estações de Vila Nova de Paiva e Castro Daire. As séries diárias referentes a estas duas últimas estações estão incompletas. No número de dias sem chuva⁸ as diferenças são de 18% mais dias sem chuva em Sobrado do que em V. N. de Paiva, no entanto, os valores médios e a dispersão são semelhantes.

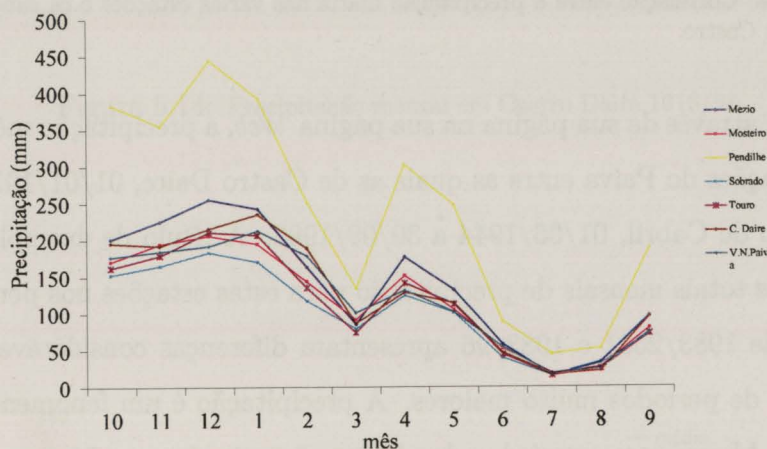


Figura 5.13: Médias da precipitação dos totais mensais 1983/96.

A estação de Pendilhe regista a maior precipitação, embora se encontre a uma altitude inferior (750m) a várias outras estações. A estação de Mezio é a que se encontra a maior altitude (950m). Este facto mostra que são várias as variáveis que intervêm na determinação da intensidade da precipitação em cada local. A tabela (5.3) representa a correlação entre as séries de precipitação de cada estação, Castro Daire e Mosteiro e Mosteiro com Mezio são as mais correlacionadas. As menos correlacionadas são Touro e Sobrado e Vila Nova de Paiva com Sobrado e ainda Touro com V.N. de Paiva. No mesmo período, a correlação entre as duas séries de caudais é de 92.6% o que mostra

⁸os dias indicados dizem respeito ao período 1/10/1983 a 30/11/1996

Estação	Pendilhe	Mezio	Sobrado	Touro	Mosteiro	V.N.Paiva	C. Daire
Pendilhe	1.0	0.66	0.46	0.50	0.59	0.47	0.59
Mezio	0.67	1.0	0.56	0.57	0.72	0.60	0.72
Sobrado	0.46	0.56	1.0	0.36	0.55	0.39	0.47
Touro	0.50	0.57	0.36	1.0	0.46	0.44	0.47
Mosteiro	0.59	0.72	0.55	0.46	1.0	0.51	0.62
V.N.Paiva	0.47	0.60	0.39	0.44	0.51	1.0	0.53
C. Daire	0.59	0.72	0.47	0.47	0.62	0.53	1.0

Tabela 5.3: Correlação entre as estações de precipitação.

que os caudais são, apesar de tudo, um fenómeno menos complexo. O INAG disponibi-

Estação	Pendilhe	Mezio	Sobrado	Touro	Mosteiro	V.N.Paiva	C. Daire
Castro	0.55	0.87	0.69	0.87	0.83	0.63	0.89
Fragas	0.46	0.92	0.70	0.93	0.87	0.71	0.93

Tabela 5.4: Correlação entre a precipitação diária nas várias estações e os caudais médios diários em Fragas e Castro.

liza ainda, através da sua página na sua página *Web*, a precipitação média mensal para várias estações do Paiva entre as quais as de Castro Daire, 01/01/1916 a 30/09/1998 e Mosteiro de Cabril, 01/06/1944 a 30/09/1995. A título de exemplo compare-se as médias dos totais mensais de precipitação para estas estações nos períodos indicados. Os anos de 1983/2001 e 1983/96 apresentam diferenças consideráveis relativamente às médias de períodos muito maiores. A precipitação é um fenómeno extremamente irregular. Mesmo para períodos de observação superiores a 50 anos, verifica-se que para alguns meses as diferenças entre os percentis 25 e 75, figura (5.14). A média do período 83/99 em Mosteiro apresenta um mínimo local em Março enquanto que os quartis e a média do período 1944/00 não revelam essa tendência. o primeiro quartil 1944/00 revela um máximo para o mês de Abril que não é acompanhado pelos outros quartis. No entanto, a média 1944/00 teve um comportamento mais próximo do primeiro quartil, ao atingir um mínimo local em Março e um máximo local em Abril. Estes conjuntos de médias são significativos uma vez que o período de observação é bastante extenso.

O gráfico (5.16) contém as estatísticas dos caudais médios mensais medidos em Fragas da Torre. A característica mais interessante é o facto de a média se aproximar do terceiro quartil enquanto que a mediana está mais próxima do primeiro quartil. Outro

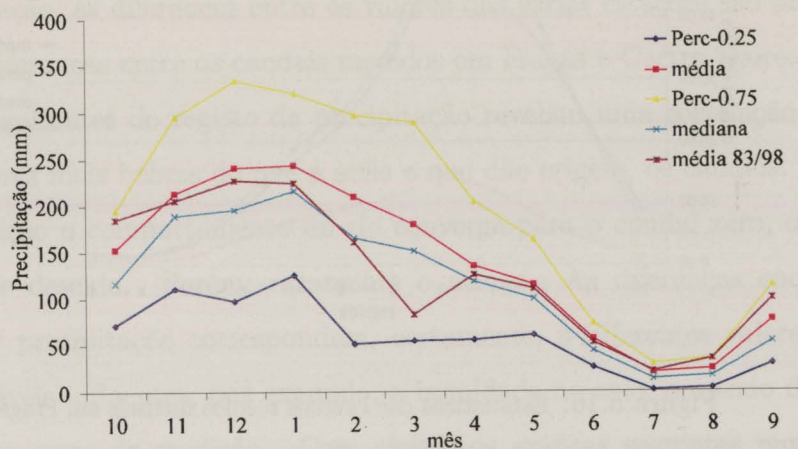


Figura 5.14: Precipitação mensal em Castro Daire 1916/98.

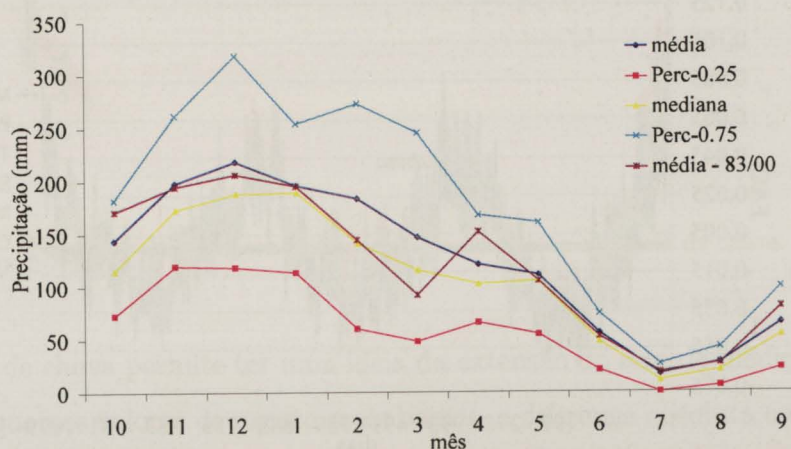


Figura 5.15: Precipitação mensal em Mosteiro 1944/00.

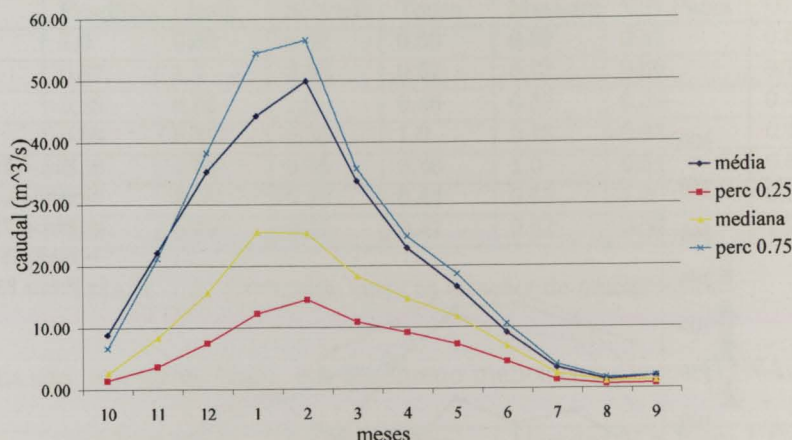
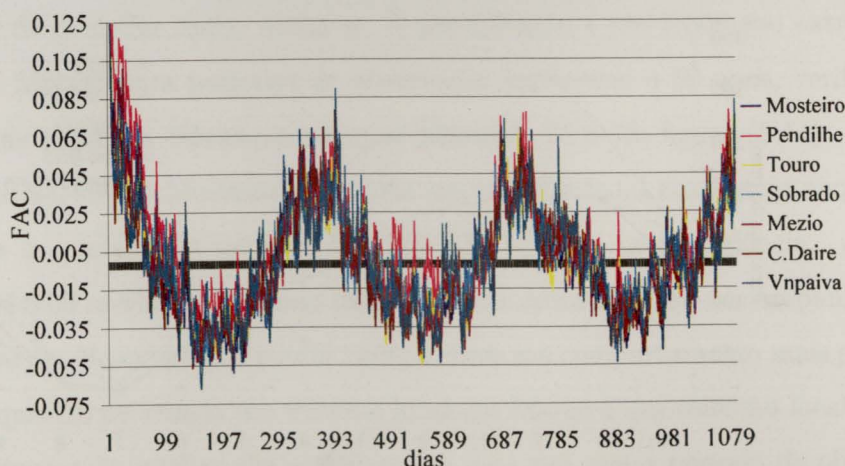


Figura 5.16: Estatísticas dos caudais médios mensais em Fragas.

dato relevante é o da mediana de Janeiro ser ligeiramente superior à de Fevereiro, para os outros quantis e para a média a situação é a inversa ainda que as diferenças possam não ser significativas no que diz respeito aos caudais. O estudo da precipitação diária coloca problemas numéricos e estatísticos devido à predominância de valores nulos. Um efeito estatístico dessa predominância de valores nulos verifica-se na diminuição da correlação. Na figura (5.17) apresenta-se a *FAC* das 7 estações nos períodos indicados.

Figura 5.17: *FAC* das 7 estações de precipitação.

Se compararmos a autocorrelação das séries de precipitação com a autocorrelação da série dos caudais facilmente se verifica que a correlação da precipitação é bastante mais

baixa. A estação cuja precipitação se relaciona menos intensamente com os caudais em Fragas é a de Pendilhe. A *FAC* amostral evidencia as diferenças entre o escoamento e a precipitação, as diferenças entre os valores das várias estações são muito maiores do que as diferenças entre os caudais medidos em Fragas e Castro Daire.

As séries resultantes do registo da precipitação revelam uma correlação linear entre os seus valores mais baixos do que a série a que dão origem, os caudais. Na ausência de precipitação o comportamento do rio converge para o caudal zero, o atractor da dinâmica de descida. Porém, raramente o atinge. As diferenças encontradas na ausência de precipitação correspondem, certamente, a diferentes valores iniciais de outras variáveis influentes, por exemplo, a humidade no solo, o estado da vegetação, e também a erros de medição. Com efeito, os gráficos seguintes mostram que o comportamento do rio pode ser, em determinadas situações, bastante previsível.

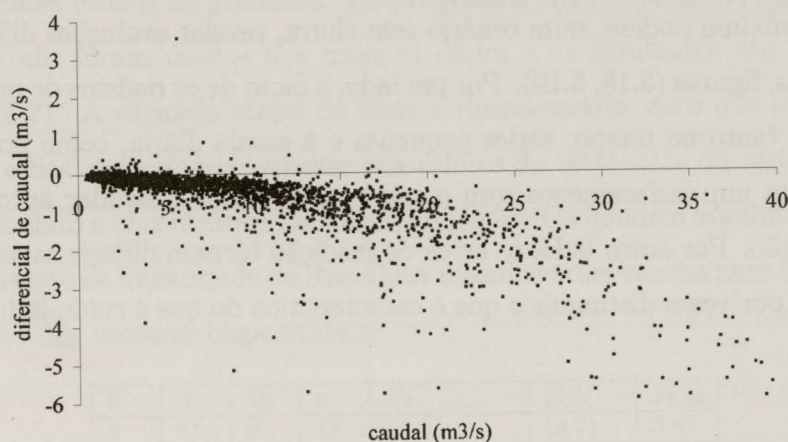


Figura 5.18: Diferencial de caudal a um dia na ausência de chuva.

A ausência de chuva permite ter uma ideia da extensão do erro de medição. De facto, verifica-se que nos valores de registo mais baixos, a diferença absoluta entre os valores verificados em dias consecutivos, sem chuva registada em nenhuma das estações, tem média próxima de zero, 0.026, e inclusive ocorrem frequentemente falsas subidas de caudal⁹ indicando que o ruído de medição tem influência visível.

⁹ou seja, o ruído de medição faz aumentar o caudal sem que na realidade tenha entrado mais água na bacia.

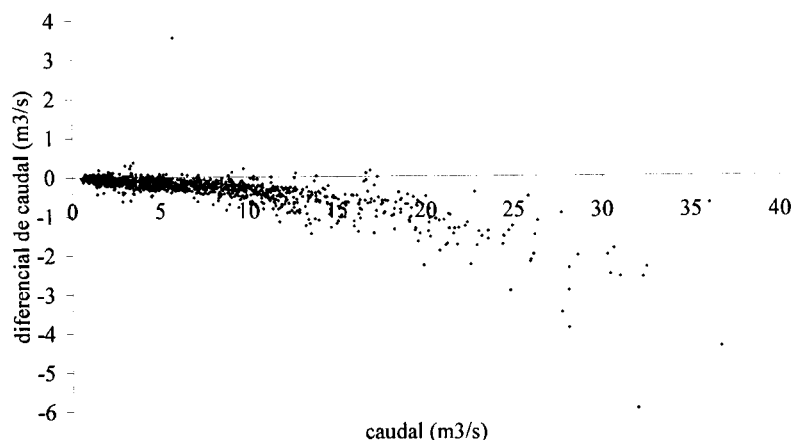


Figura 5.19: Diferencial de caudal a 5 dias na ausência de chuva.

Os gráficos do diferencial de caudal em dias sem chuva reflecte os diferentes comportamentos que este pode ter, ou seja, várias trajectórias com um nível caudal muito próximo podem, num cenário sem chuva, revelar evoluções diferentes ou muito próximas, figuras (5.18, 5.19). Por um lado, o facto de os registos de precipitação serem escassos tanto no tempo, séries pequenas e à escala diária, como no espaço, e ainda por terem um desfasamento com o caudal não permite estudar aprofundadamente a sua relação. Por outro lado, os erros de medição tornam difíceis as análises não sendo possível por vezes distinguir o que é característica do que é ruído induzido por erro de medição.

5.3 Detecção de não linearidade e não gaussianidade

Nesta secção vão efectuar-se alguns testes estatísticos tendo em vista a confirmação das suspeitas de falta de linearidade e gaussianidade. No capítulo 2 descreveram-se alguns testes estatísticos que podem ser usados para ensaiar hipóteses de gaussianidade e linearidade sobre a série observada. Os testes de gaussianidade e linearidade de Hinich, [Hin82] foram implementados em *MATLAB* por Swami, Mendel e Nikias¹⁰ quando

¹⁰*toolbox* disponível em www.mathworks.com.

aplicados na série dos caudais médios deram os resultados da tabela, (5.5).

série	c	g. l.	$\hat{S}$	valor-p
original	0.51	640	65869.98	0

Tabela 5.5: Parâmetros estimados e respectivo rácio-t.

O valor-p é zero o que significa que existe uma grande assimetria dos dados. Rejeitada a hipótese de gaussianidade seguimos para a etapa seguinte e ensaia-se a hipótese de linearidade, o quadro (5.6) resume os resultados obtidos.

série	$\hat{\lambda}$	$\hat{R}$	R	N
Fragas	200.05	132.78	38.20	271

Tabela 5.6: Resultados do teste de linearidade para a série de Fragas.

Rejeita-se liminarmente a hipótese de linearidade uma vez que o R amostral e teórico são de diferentes ordens de grandeza. Os programas em *FORTRAN77* desenvolvidos por Rao e Gabr foram usados nos mesmos dados e os resultados são apresentados no quadro (5.7). A segunda etapa do teste é desnecessária visto que o afastamento da simetria é bastante grande. Apesar dos valores da estatística do teste serem mais baixos, continuam a encontrar-se na zona de rejeição. Os quadros seguintes resumem a aplicação do teste de linearidade de Rao-Gabr usando vários valores para os parâmetros de construção dos vectores bispectrais.

k	d	Q	n	F_2	g. l.	$F_{0.05}$
6	61	6	13	41.0	(6,7)	3.87
7	47	9	4	745.2	(9,8)	3.39
8	34	13	21	5631	(13,8)	3.23

Tabela 5.7: Resultados do teste de gaussianidade de Rao-Gabr para vários valores dos parâmetros. Série dos caudais médios em Fragas.

Como se esperava a série dos caudais médios diários apresenta, para quaisquer valores dos parâmetros, um afastamento significativo relativamente à linearidade. Na abordagem pelos sistemas dinâmicos ao estudo de sucessões cronológicas fizemos referência no capítulo 3 ao teste de reversibilidade proposto por [Dik96]. O ensaio da hipótese de reversibilidade para a série original dos caudais médios diários do rio Paiva complementa os testes de simetria já usados. O quadro (5.8) contém os valores observados da

estatística de Diks para vários valores dos parâmetros correspondentes à reconstrução do espaço de mergulho, dimensão, m e desfasamento temporal, τ , e dos parâmetros de Theiler, T e tamanho do bloco, l .

m	τ	T	l	S
3	1	6	6	130.62
4	1	6	6	132.80
5	1	6	6	131.08
3	1	12	12	107.32
4	1	12	12	108.82
5	1	12	12	106.64
3	1	24	24	80.70
4	1	24	24	81.28
5	1	24	24	80.04

Tabela 5.8: Resultados do teste de reversibilidade para vários valores dos parâmetros. Série dos caudais médios em Fragas.

Como se pode verificar os valores são extremamente elevados quando comparados com os de séries reversíveis mesmo para grandes blocos e elevados valores do parâmetro de Theiler. O recurso à análise não linear faz então todo o sentido [Dik99].

5.4 Reconstrução do espaço de fase

Como se salientou na secção anterior, o sistema gerador de caudais é resultado da interacção não linear de muitas variáveis, e portanto, não deve ser de excluir a possibilidade de ocorrer, pelo menos em determinadas situações, um comportamento complexo com previsão difícil. Esse cenário designa-se comumente por caos determinístico. A possibilidade de os processos climáticos, dos quais os hidrológicos são uma parte, exibirem *DSCI*, um dos sintomas do caos determinístico, remonta ao trabalho de [Lor63]. O trabalho de [Pac80] e [Tak80] permitiram o desenvolvimento de ferramentas de *análise de séries* do ponto de vista da dinâmica com particular destaque para a estimação da dimensão-correlação de [GP] e mais tarde de [San85] e [Eck86]. Os trabalhos, [Nic84] e [TE89] são pioneiros na determinação de dimensões fractais em séries gerados por sistemas climáticos ou relacionados. No entanto, após [Pal94], [Pro91], e, [Os89] terem mostrado que alguns processos estocásticos, como o movimento Browniano, apresentavam um integral de correlação convergente sem

contudo serem caóticos iniciou-se uma busca de processos identificadores do movimento caótico a partir de dados como séries temporais que, aparentemente, ainda não tiveram sucesso. Assim surgiram novos métodos de diagnóstico, a previsão não linear, [Far87], [Far88], *surrogate data*, [The92] e [Sch98] diagramas *DVS*, [Cas91a] e o índice de Cao, [Cao97]. Se nenhum método, em particular, pode garantir que determinado conjunto de dados é gerado por um sistema caótico, o resultado da aplicação de várias técnicas pode permitir, com uma base mais firme que essa possibilidade seja ou não descartada. Como ficou bem patente na análise feita na secção anterior, é improvável que o sistema hidrológico gerador dos caudais do rio seja controlado apenas por uma dinâmica caótica o mais provável é que coexistam vários tipos de dinâmica não necessariamente caóticas. Como referem [Por97] trata-se de saber se os processos hidrológicos exibem uma forma fraca de determinismo que pode estar mascarada por erros de medição, por uma frequência de amostragem insuficiente ou até mesmo pela escassez de outros dados para além dos caudais. Nesta subsecção o tema principal é a análise da dinâmica que comanda o comportamento do rio do qual se possui duas séries de caudais e alguns dados de precipitação. Para o efeito usar-se-ão os métodos descritos no capítulo 4.

A reconstrução do espaço de fase compreende duas etapas, a escolha do desfaseamento temporal mais indicado, para instantes superiores ao mínimo disponível (diário), e a escolha da dimensão do espaço de mergulho. O teorema de Takens garante que, para um sistema observado sem ruído, sem erros de medição e limitação no tamanho, é possível *mergulha-lo* num espaço euclidiano com dimensão adequada. As propriedades de um *mergulho* deixam inalteradas as características dinâmicas e topológicas do sistema permitindo operar com o sistema reconstruído tal como se fosse o original. No entanto não existem sistemas observados nas condições do teorema. Na prática, em análise não linear de séries pressupõe-se que os erros e a limitação no tamanho são apenas obstáculos a ultrapassar.

A designação de vector de reconstrução é usada para séries provenientes de sistemas determinísticos, no entanto, é prática corrente usar generalizar essa designação para vectores obtidos pelo mesmo processo de séries observadas de sistemas que à partida, pelo menos, estão contaminadas por ruído. A figura (5.20) representa a curva da

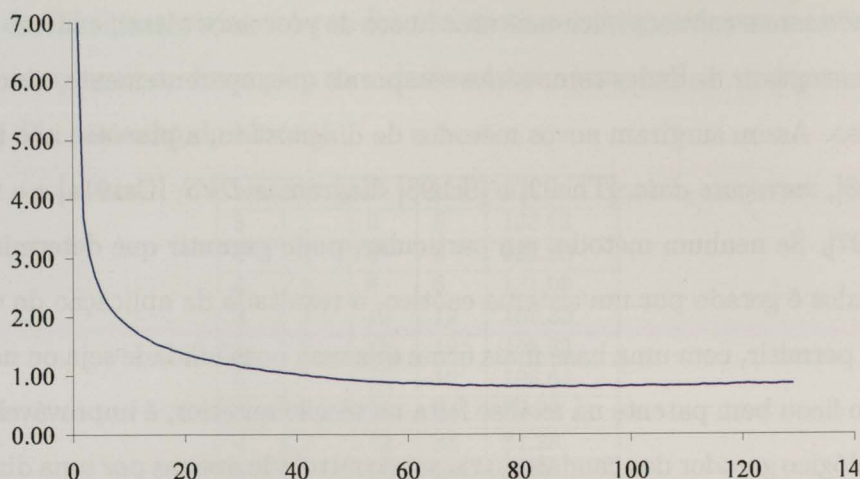


Figura 5.20: Informação mútua média para a série dos caudais médios diários em Fragas.

Informação mútua média para a série de caudais de Fraga da Torre onde a partição usada do espaço de resultados da variável tem 128 classes¹¹, como se referiu no capítulo 4, esta quantidade mede a informação, no sentido de Shanonn, que se tem acerca de $s(t + \tau)$ dado que se conhece $s(t)$. O primeiro mínimo da curva é atingido em 53 dias, um valor que a ser usado para desfasamento temporal colocaria em risco a análise por insuficiência de dados. Independentemente das considerações a fazer adiante, compararemos previsões obtidas com diferentes espaçamentos temporais. Verifica-se que a *FAC* desce pela primeira vez abaixo do nível $1/e$ em 11 dias e em apenas 7 para o nível 0.5. Porém, necessita de 69 dias para o nível 0.1. Se atendermos ao comportamento do rio, verifica-se que em muitos eventos os extremos visíveis podem corresponder a 7 dias ou até menos. De facto, eventos relevantes para o comportamento do escoamento do rio Paiva acontecem em períodos temporais de apenas alguns dias. Esses eventos não são tidos em conta se se tomarem espaçamentos temporais superiores ao diário. É um fenómeno análogo ao alisamento em análise de Fourier. Os autores em [PR96] indicam ainda que os critérios de escolha do desfasamento temporal baseados na *FAC* obrigam frequentemente à utilização de

¹¹dados obtidos com o pacote informático para *MATLAB*, *TSTOOL*, desenvolvido pelo grupo de Merkwitth et al. que se pode encontrar em <http://www.physik3.gwdg.de/>.

espaçamentos temporais superiores a 30 dias neste tipo de dados, facto que pode estar relacionado com uma componente sazonal e não estar directamente relacionado com a dinâmica. A referência daqueles autores à variação horária dos caudais é reforçada pelas subidas do caudal do rio que, quando amostrado diariamente, consistem em escassas 2 a 4 observações. Os mesmos autores em [Por97] referem ainda que a utilização de espaçamentos temporais superiores a 1,2 dias não são aconselhados caso se esteja perante um atractor heterogéneo. Se existir uma dependência do estado a solução ideal passaria pela utilização de espaçamentos não-uniformes. No entanto os métodos existentes para a estimação dos invariantes não são compatíveis com tal procedimento.

Para determinação da dimensão do espaço de mergulho vão usar-se o algoritmo dos falsos vizinhos mais próximos, e o algoritmo da dimensão correlação. O primeiro revela que a percentagem de falsos vizinhos próximos é elevada para dimensões do espaço inferiores a 5 e para qualquer espaçamento temporal de 1 a 7 dias. Para valores superiores da dimensão, a percentagem aproxima-se de zero não havendo uma diferença significativa entre os vários espaçamentos, figura (5.21). A que se deve essa

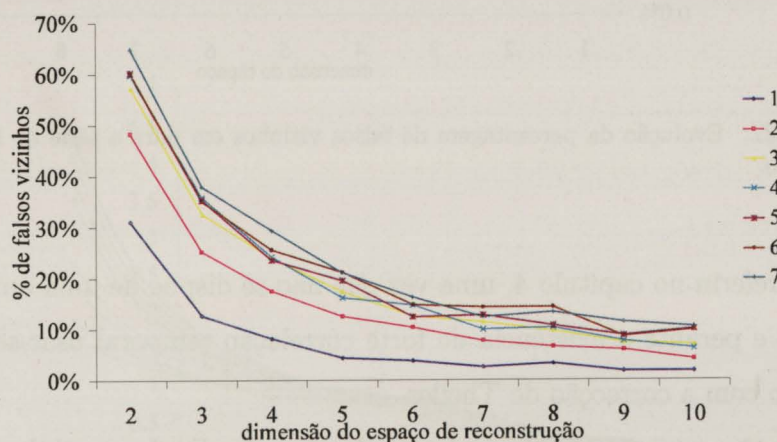


Figura 5.21: Evolução da percentagem de falsos vizinhos em função do espaço para espaçamentos temporais de 1 a 7 dias.

diferença no comportamento do algoritmo dos falsos vizinhos próximos? Uma vez que o rio reflecte aspectos dinâmicos da chuva, seria interessante, mesmo sabendo da limitação dos dados de precipitação, verificar qual o comportamento deste algoritmo

nas séries longas de precipitação.

A figura (5.60) representa a curva de percentagem de falsos vizinhos para as séries de caudais diários de Fragas e de precipitação em Mezio. Verifica-se uma correlação elevada entre as curvas embora, a percentagem de falsos vizinhos, no caso da precipitação, apresente um decaimento mais lento.

Da mesma forma que para o espaçamento temporal, usar-se-ão várias dimensões para o espaço de mergulho e comparar-se-ão os resultados obtidos em previsão.

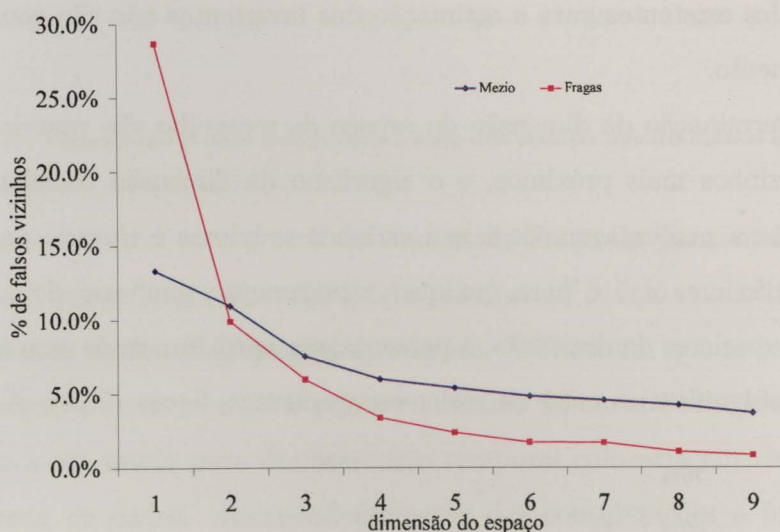


Figura 5.22: Evolução da percentagem de falsos vizinhos em para a série de Fragas e a série de precipitação.

Como se referiu no capítulo 4, uma vez que não se dispõe de uma amostra de grande dimensão e perante a existência de forte correlação temporal usar-se-á o integral de correlação com a correcção de Theiler.

As curvas têm oscilações estáveis de acordo com a oscilação anual do sistema pelo que não existe um valor que permita obter uma independência entre o espaço e o tempo. No entanto, uma vez que o decaimento da *FAC* é lento optou-se por usar o parâmetro de correcção com o valor 10. A evolução dos declives do integral de correlação, figura (5.4), mostram três comportamentos distintos para diferentes escalas. Para raios superiores até cerca de $30m^3/s$ os declives aumentam de zero até cerca de $0.9m^3/s$ de

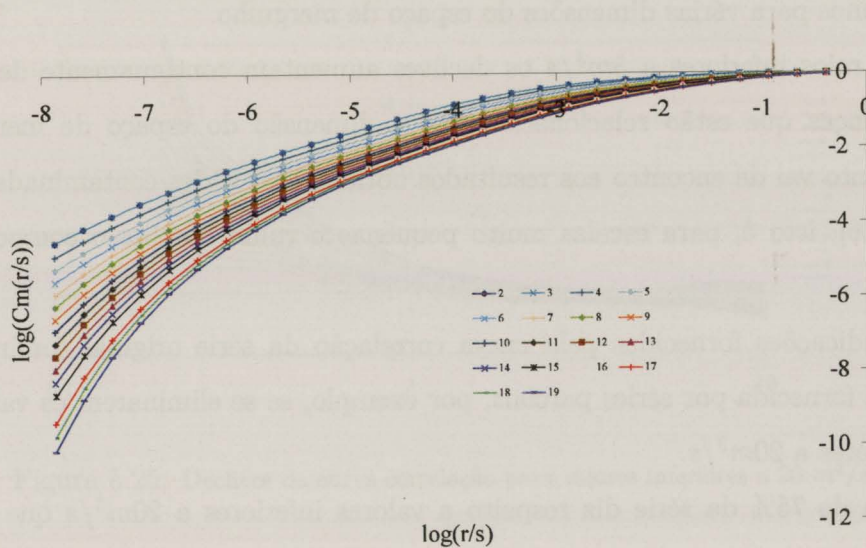


Figura 5.23: $\log(C_m(r)) - \log(r)$ para várias dimensões de 2 a 19

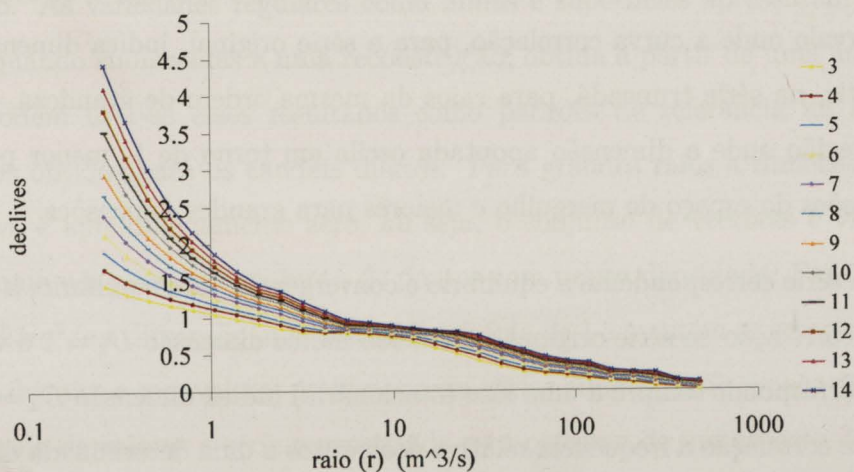


Figura 5.24: Evolução dos declives dos segmentos de recta do gráfico (5.23).

forma linear ficando estacionários nesse valor, entre os 5 e os $30m^3/s$ e com valores próximos para várias dimensões do espaço de mergulho.

Para raios inferiores a $5m^3/s$ os declives aumentam continuamente de valor e com diferenças que estão relacionadas com a dimensão do espaço de mergulho. Este aumento vai de encontro aos resultados obtidos com séries contaminadas com ruído, [The86], isto é, para escalas muito pequenas o ruído domina o comportamento da série.

As indicações fornecidas pela curva correlação da série original confirmam-se pela curva fornecida por séries parciais, por exemplo, se se eliminarem os valores da série inferiores a $20m^3/s$.

Cerca de 75% da série diz respeito a valores inferiores a $20m^3/s$ que corresponde, aproximadamente ao valor médio, que é apenas 2.25% da amplitude dos dados. Este é um indício claro da possibilidade da existência de diferentes escalas observadas às quais corresponderão certamente diferentes tipos de dinâmica. Mais adiante retomaremos este aspecto relevante para a compreensão do fenómeno.

Como se pode verificar, na curva correlação correspondente à série truncada, figura (5.60), deixa de existir um intervalo de valores para os quais a dimensão indicada é aproximadamente zero o que está de acordo a escala usada. Por outro lado, para o intervalo onde a curva correlação, para a série original, indica dimensão 1, passou a existir, na série truncada, para raios da mesma ordem de grandeza, 2 a $10 m^3/s$, uma região onde a dimensão apontada oscila em torno de 1, menor para pequenas dimensões do espaço de mergulho e maiores para grandes dimensões.

Para a série correspondente a equilíbrio e convergência para equilíbrio, a região onde a curva correlação na série original e truncada indica dimensão $D_2 = 1$ o valor indicado (que corresponde sempre a uma fase estacionária) indica dimensão $D_2 = 0$ a 5. Sendo a curva correlação a frequência relativa dos pontos a uma determinada distância o que parece acontecer é que, para a escala em questão, se obtém para uns segmentos de órbita, dimensão 1 e para outros dimensão 0. Ou seja, nos casos de convergência para equilíbrio, o valor inicial pode ser de diferentes ordens de grandeza. Temos então um

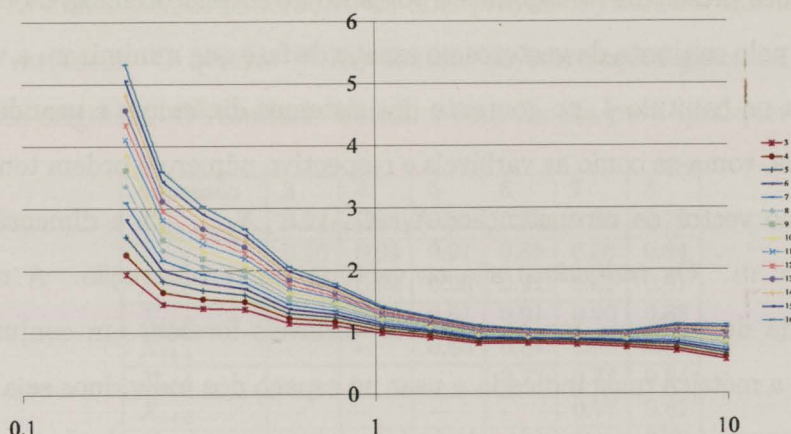


Figura 5.25: Declives da curva correlação para valores inferiores a $20 \text{ m}^3/\text{s}$

comportamento regular do fenómeno em diferentes escalas.

Este efeito coloca um problema à determinação da dimensão. As baixas dimensões apontadas para valores elevados do raio, r podem estar relacionadas com a escassez de informação sobre o comportamento do rio quando este atinge caudais elevados.

O valor da curva correlação interpreta-se como uma estimativa da probabilidade de dois vectores do sistema reconstruído estarem a uma distância inferior a um determinado raio. As variedades regulares como linhas e superfícies apresentam dimensões inteiras quando submetidas a uma reconstrução, obtida a partir de uma amostragem. Assim podem usar-se esses resultados como padrões de referência na análise dos resultados obtidos para os caudais diários. Para grandes raios a dimensão indicada pela curva é aproximadamente zero, ou seja, o conjunto de vectores é visto, a essa escala, como um ponto ou conjunto de pontos com pouca densidade. Para raios entre os 5 e os $25 \text{ m}^3/\text{s}$ a dimensão indicada está próximo de 1 e mantém-se estacionária nesse intervalo de raios e com valores próximos para várias dimensões do espaço de mergulho. Nessa escala de valores a série reconstruída está próximo de um objecto de dimensão 1. A ACP pode ajudar a perceber o resultado obtido pela curva de correlação. Como referem [Bro94] o número de valores próprios significativos da matriz de covariância corresponde à dimensão do subespaço que contém o atractor.

A análise em componentes principais referida no capítulo 5 permite ainda encontrar

as direcções principais (ortogonais e portanto de correlação nula) da nuvem de pontos formada pelo conjunto de vectores no espaço de fase que minimizam a variância. Como se referiu no capítulo 4, no contexto dos sistemas dinâmicos e usando a terminologia estatística, toma-se como as variáveis e respectivo número a ordem temporal, tal como surgem no vector de reconstrução, $X_t, X_{t+1}, \dots, X_{t+m-1}$ e a dimensão do espaço de mergulho, m . Os *indivíduos* são os vectores de reconstrução. A matriz de dados construída desta forma faz com que as variáveis formem um conjunto homogéneo, pelo que a métrica mais indicada a usar no espaço dos indivíduos seja a euclidiana. O facto de as variáveis construídas desta forma terem uma correlação positiva entre si faz com que a primeira componente principal defina um *factor de tamanho*, [Sap90]. Pelo teorema de Frobenius-Perron, uma matriz simétrica de termos positivos admite um vector próprio onde todos os componentes têm o mesmo sinal. Dado que se correlaciona positivamente com todas as variáveis e os indivíduos são ordenados por valor crescente do conjunto de variáveis (pela média), e se adicionalmente todas as correlações forem da mesma ordem então a primeira componente principal é proporcional à média das variáveis iniciais.

Usando a *ACP* determinam-se as direcções principais do conjunto de dados e os respectivos pesos que têm nos dados. Essa informação será relevante para a compreensão do comportamento da curva correlação.

Usando o conjunto de dados original, obtém-se para a matriz de dados resultante da série dos caudais medidos em Fragas, e para dimensões do espaço de mergulho de 3 a 8, os eixos ou factores principais¹² e respectivas percentagens da inércia inicial que explicam, tabela (5.9).

dimensão	1º v.próprio	%	2º v.próprio	%
3	4018.45	86.66	424.67	9.16
4	5116.84	82.76	617.28	9.98
5	6147.33	79.51	821.15	10.62
6	6963.53	75.06	1024.27	11.04
7	7793.24	72.00	1367.12	12.63
8	8612.60	69.62	1699.14	13.74

Tabela 5.9: Dois maiores valores próprios da matriz de covariância para a série de Fragas.

¹²uma vez que se usa a métrica euclidiana

Para todas as dimensões consideradas, os dois primeiros eixos explicam sempre mais de 80% da variância. O quadro (5.10) contém as correlações entre as componentes e as variáveis originais. O resultado obtido é semelhante para qualquer dimensão do

dimensão	3	4	5	6	7	8
X_t	0.92	0.88	0.85	0.83	0.81	0.78
X_{t+1}	0.96	0.94	0.91	0.89	0.86	0.84
X_{t+2}	0.92	0.94	0.93	0.91	0.89	0.87
X_{t+3}	-	0.88	0.91	0.91	0.90	0.88
X_{t+4}	-	-	0.85	0.87	0.88	0.87
X_{t+5}	-	-	-	0.77	0.82	0.84
X_{t+6}	-	-	-	-	0.77	0.81
X_{t+7}	-	-	-	-	-	0.76

Tabela 5.10: Correlações entre a primeira componente e as variáveis iniciais.

espaço considerada, a primeira componente principal tem correlação elevada com todas as variáveis. Os valores para a primeira componente mostram que a contribuição das variáveis é aproximadamente a mesma existindo uma simetria, na contribuição, entre as variáveis equidistantes do centro do vector de reconstrução. Tratando-se de vectores cujos componentes dizem respeito aos mesmos valores, mas, desfasados no tempo, o facto de a primeira componente resultar de uma contribuição semelhante de todas as variáveis mostra que em boa parte do tempo a série se mantém aproximadamente constante. De facto, o comportamento do caudal em perda de água, isto é, na ausência de chuva, é o de um lento decaimento, tanto mais lento quanto menor for o caudal. Se se juntar isso ao facto de as perdas de água representarem mais de 80% dos dados fica explicado o resultado para a primeira componente. Este resultado reforça o da curva-correlação no que diz respeito ao intervalo de raios que indica dimensão 1. A figura (5.26) mostra a série correspondente à projecção no primeiro eixo e a comparação com a série original.

A figura (5.27) mostra os eixos correspondentes aos 3 primeiros componentes principais, que explicam 91.4% da variância, no espaço dos indivíduos, para uma dimensão 7 do espaço de mergulho.

Em aplicações da *ACP* fora do contexto dos sistemas dinâmicos o critério habitualmente usado para determinar o número de componentes a reter é o da percentagem

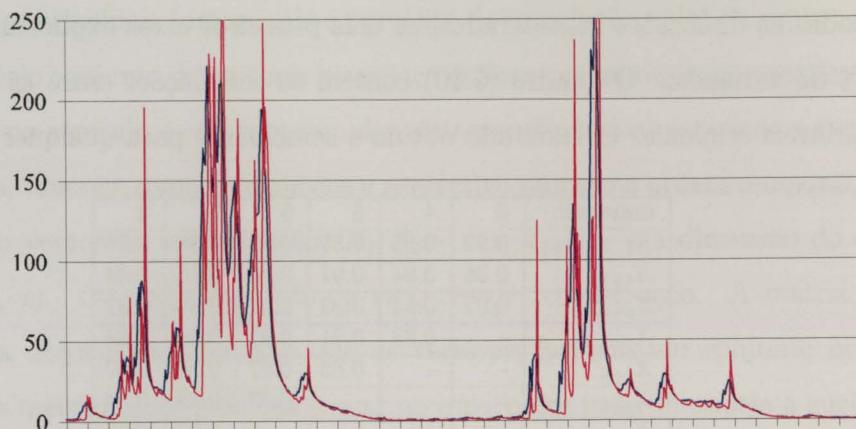


Figura 5.26: Série original e a obtida por projecção no 1º componente principal.

da inércia total explicada, [Sap90]. Por exemplo, se os dois primeiros componentes ultrapassarem 90% então, claramente, a nuvem de pontos estará bem representada num subespaço de dimensão 2. A avaliação da percentagem deve ter em conta o número de variáveis, uma percentagem de 10% tem uma relevância diferente quando se tem 10 variáveis em vez de 100.

Para a série dos caudais diários do rio Paiva o número de componentes a reter vai de 2 a 3 para dimensões do espaço de 3 a 8.

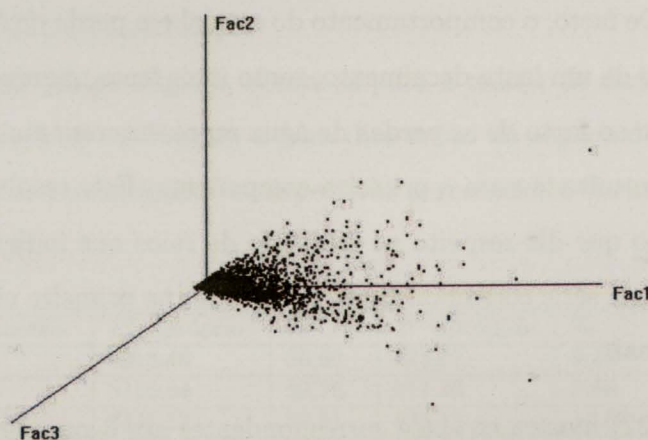


Figura 5.27: Representação dos indivíduos no espaço dos 3 eixos principais.

Os autores [Alb88] referem que a decomposição em valores singulares (ou ACP) conduz

a um conjunto de variáveis estatisticamente independentes e por isso deve fazer-se a análise da correlação sobre a reconstrução da série projectada pela primeira componente. As figuras (5.28) e (5.29) representam, respectivamente a curva-correlação e os declives para a série obtida por projecção sobre a primeira componente. As diferenças fundamentais relativamente à série original estão na maior proximidade entre as curvas das várias dimensões e na região de aparente estacionaridade de dimensão 1. Para a série projectada na primeira componente essa região começa para raios mais pequenos, $8m^3/s$ em vez de $20-25m^3/s$ para a série original mas acaba para raios mais pequenos, cerca de $1-2m^3/s$ em vez de $4-5m^3/s$.

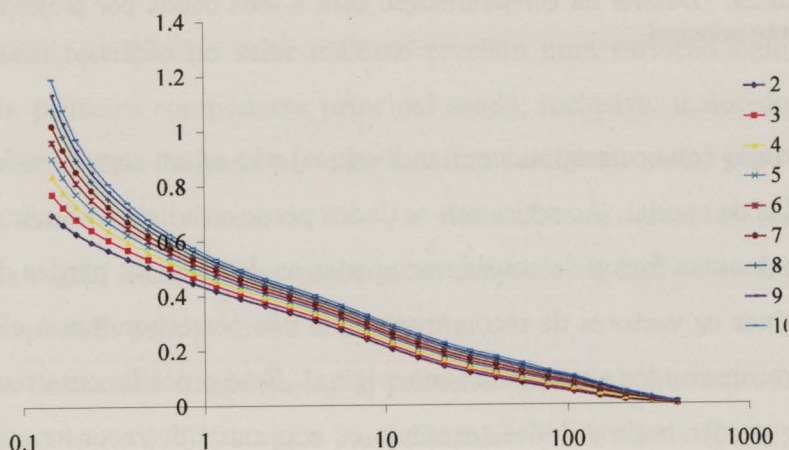


Figura 5.28: Curva-correlação para a série obtida por projecção sobre a primeira componente principal

Sendo o regime dinâmico de perda de caudal aquele mais observado e tendo em conta o resultado obtido da análise à curva-correlação é de todo o interesse usar a ACP apenas nos *indivíduos* pertencentes aquele regime. Porém, há que distinguir duas formas diferentes de considerar apenas perdas de caudal. Pode considerar-se uma série de segmentos de descidas com um tamanho mínimo, por exemplo, a dimensão mais pequena do espaço de mergulho, e a partir dessa série obter os vectores de reconstrução. Este procedimento tem a vantagem de se conseguirem conjuntos de reconstrução com o mesmo tamanho da série original, no entanto, é inevitável que alguns dos vectores de

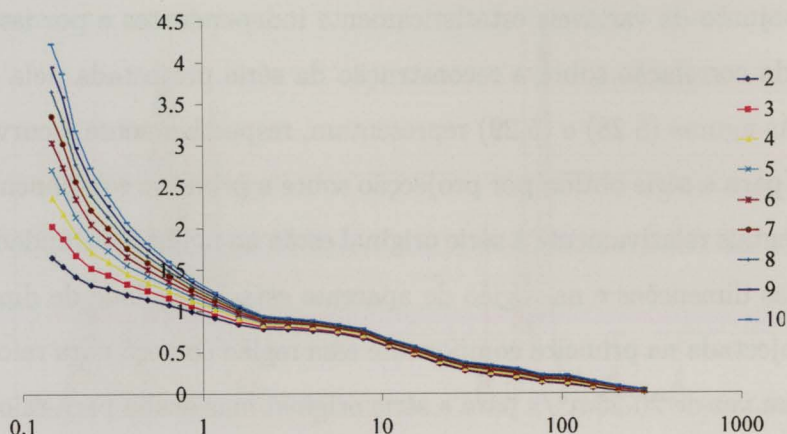


Figura 5.29: Declives da curva-correlação para a série obtida por projecção sobre a primeira componente principal

reconstrução (que correspondem a indivíduos) não sejam correspondam efectivamente a descidas de caudal. Ao retirarem-se dados perde-se a ligação dinâmica e a respectiva ligação. A outra forma de considerar apenas as descidas ou perdas de caudal consiste em eliminar os vectores de reconstrução que não correspondam a efectivas perdas de caudal reconstruídos a partir da série original. Este procedimento tem o inconveniente de fazer perder muitos dados tornando os conjuntos de reconstrução com tamanhos decrescentes para dimensões do espaço de mergulho crescentes, de tal forma que um mergulho de dimensão 4 fica com metade dos dados de um mergulho de tamanho 2. O primeiro procedimento parece ser o mais razoável tendo em conta as desvantagens. As perdas de caudal correspondem, *grosso modo*, à evolução do caudal na ausência de precipitação. Para efeitos de significância da *perda* de caudal considerar-se-ão apenas segmentos de tamanho igual ou superior a 3. Outra questão relevante tem a ver com a escala dos segmentos, o valor inicial do segmento de perda pode ser diferente, inclusive de diferentes escalas. Por isso vão considerar-se 3 diferentes conjuntos de dados, o primeiro deles sem restrição de valor, o segundo para valores inferiores ao caudal que limita a região de dimensão 1 da curva-correlação, $20m^3/s$ e o terceiro que corresponde aproximadamente ao ponto médio dessa região e também à mediana dos caudais, $8m^3/s$. Para evitar as repetições no texto chamar-lhes-emos *SP1*, *SP2* e *SP3*.

Para dimensões do espaço de mergulho de 3 a 8, o valor próprio associado ao eixo ou factor principal¹³ e respectiva percentagem da variância total explicada está no quadro (5.15).

dimensão	%SP1	%SP2	%SP3
3	76.83	94.09	94.53
4	72.13	92.12	92.92
5	68.63	94.43	91.52
6	65.71	97.55	90.23
7	63.21	93.93	89.02
8	61.02	93.44	87.88

Tabela 5.11: 1º valor próprio da matriz de covariância relativa às 3 séries *SP1*, *SP2* e *SP3* para várias dimensões.

As descidas sem restrição no valor máximo revelam uma variação significativa não explicada pela primeira componente principal sendo, inclusive, maior do que a série original. A mesma série limitada no valor máximo revela um maior peso do primeiro valor próprio, maior para o limite de $20m^3/s$. Em ambos os casos, limitar as descidas a um valor da região estacionária resulta num conjunto de indivíduos cuja variância está melhor explicada pela primeira componente.

Para todas as dimensões consideradas, o primeiro eixo explica sempre mais de 92% da variância, sendo de 97.6% o valor para dimensão 6. A tabela ((5.10)) contém as correlações entre as componentes e as variáveis originais.

Os vectores de reconstrução ou *indivíduos* resultantes da observação do rio na ausência de chuva formam um conjunto que está *contido* num subespaço de dimensão 1 dado que, de acordo o critério aplicado anteriormente, apenas um valor próprio é significativo. Este resultado vai de encontro à observação feita relativamente aos declives da curva-correlação uma vez que, quando se eliminam as *subidas* de caudais, isto é, se eliminam os segmentos da série associados à dinâmica de subida, a dimensão sugerida pelos declives na maior escala passa a ser 1.

Se tivéssemos optado somente por considerar vectores de reconstrução correspondentes a perdas de caudal os resultados seriam ainda mais contundentes no que diz respeito à tese da dimensão 1 para o regime de perda de caudal. No entanto, as curvas correlação

¹³uma vez que se usa a métrica euclidiana

para as várias dimensões são de difícil interpretação.

A existência de escalas e a relação que estas têm com as dinâmicas existentes no rio permite limitar o erro de previsão quando se decide qual a dinâmica que vai prevalecer. O erro de previsão vai depender desse facto, obviamente reflecte o maior ou menor peso relativo da componente determinística e o maior ou menor conhecimento que se tem da dinâmica contida na série histórica. A este assunto regressar-se-á mais à frente quando se falar em previsão no espaço de fase.

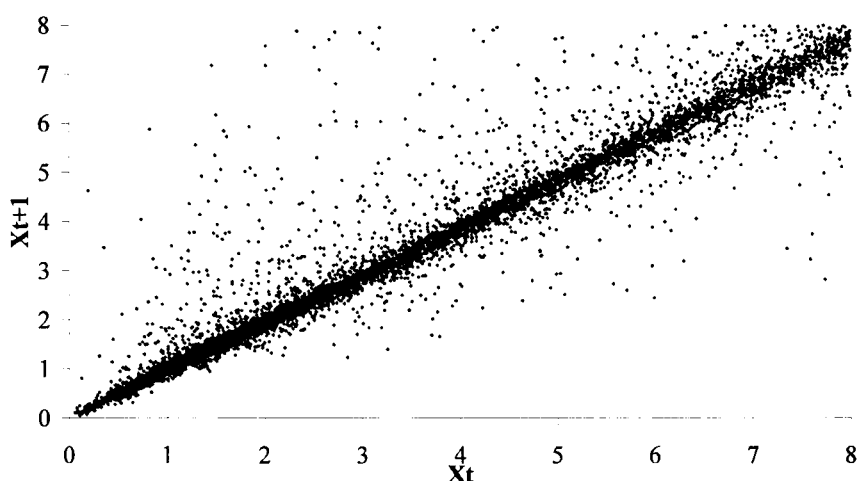


Figura 5.30: Diagrama de fase para valores entre 0 e $8m^3/s$.

As figuras (5.30) e (5.31) são diagramas de fase da série dos caudais do Paiva, correspondem aos mesmos dados observados em diferentes escalas. Para a maior escala (figura 5.31), a primeira característica que se destaca é a concentração de pontos na diagonal (e imediatamente abaixo) o que confirma a região de dimensão 1 para o referido intervalo de raios da curva-correlação.

Os decaimentos correspondem às perdas absolutas de água em dias consecutivos que, num rio como o Paiva, estão associados a dias sem chuva. Como se pode verificar no registo diário de Fragas, em 16 anos foram poucos (6.17%) os registos relativos a subidas de caudais sem ocorrência de precipitação e mesmo nestes os correspondentes a caudais mais elevados acontecem na véspera de um dia com registo de precipitação.

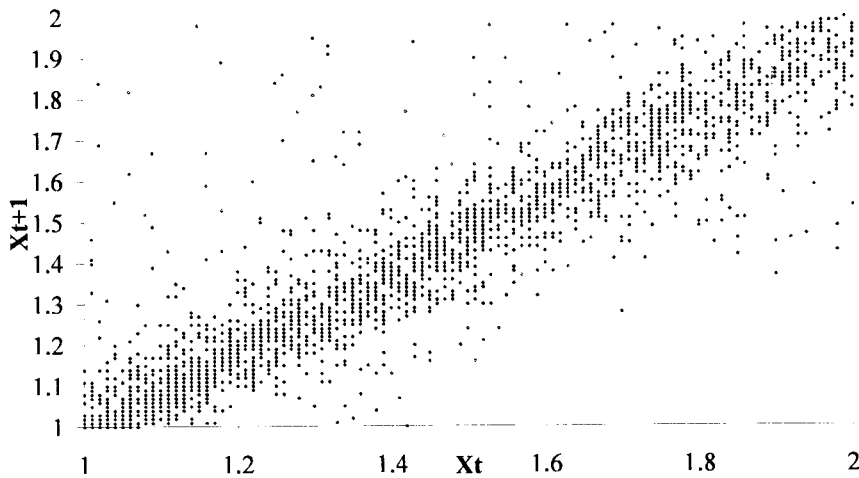


Figura 5.31: Diagrama de fase para valores entre 1 e $2m^3/s$.

Este facto poderá corresponder a uma sobreposição de períodos uma vez que a medição de precipitação está temporalmente desfasada com a do caudal.

Para as classes de caudais indicadas na secção anterior estão representadas nas figuras (5.32) e (5.33) as distribuições amostrais dos valores absolutos do decaimento obtidos a partir da diferença de caudais.

A tabela (5.12) resume as principais estatísticas dos decaimentos. Já se tinha mencionado os possíveis factores que influenciam a perda de água. Outra característica importante é o facto de se verificar decaimento nulo para valores elevados de caudal.

classe	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
centro	0.25	0.625	0.875	1.125	1.375	1.625	1.875	2.125	2.375	2.75	3.25	3.75
N	99	232	420	480	434	421	360	345	312	480	361	343
Média	0.034	0.048	0.042	0.050	0.074	0.073	0.101	0.107	0.124	0.154	0.215	0.236
Mediana	0.03	0.03	0.03	0.04	0.06	0.06	0.08	0.08	0.1	0.12	0.16	0.17
Mínimo	0	0	0	0	0	0	0	0	0	0	0	0
Máximo	0.18	0.33	0.31	0.26	0.49	0.42	0.55	0.91	0.59	1.61	1.85	2.12
1º	0.01	0.01	0.01	0.01	0.03	0.02	0.04	0.04	0.05	0.07	0.09	0.09
3º	0.05	0.065	0.06	0.07	0.1	0.1	0.13	0.14	0.17	0.19	0.26	0.27
AIQ	0.04	0.055	0.05	0.06	0.07	0.08	0.09	0.1	0.12	0.12	0.17	0.18
D.padrão	0.031	0.052	0.040	0.048	0.067	0.066	0.094	0.111	0.109	0.159	0.212	0.280
Assimetria	1.78	1.87	1.88	1.64	2.05	1.58	2.06	3.22	1.82	3.96	3.23	3.72
Curtose	4.75	4.76	6.12	3.30	6.53	3.21	5.27	15.51	4.24	25.27	15.63	17.31

Tabela 5.12: Estatísticas para o decaimento. Valores em m^3/s , quando aplicável.

classe	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23
centro	3.75	4.25	4.75	5.25	5.75	6.25	6.75	7.25	7.75	8.25	8.75	9.25
N	343	358	330	317	301	281	263	256	259	244	253	248
Média	0.236	0.216	0.298	0.28	0.336	0.356	0.377	0.386	0.497	0.458	0.541	0.573
Mediana	0.17	0.17	0.23	0.2	0.23	0.27	0.31	0.325	0.35	0.36	0.41	0.39
Mínimo	0	0	0	0	0	0	0.01	0	0	0	0	0
Máximo	2.12	1.98	2.27	2.55	3.54	3.09	2.13	3.9	4.92	4.91	5.11	5.77
1ºQ	0.09	0.08	0.12	0.12	0.14	0.17	0.18	0.19	0.22	0.21	0.21	0.24
3ºQ	0.27	0.28	0.34	0.33	0.37	0.4	0.45	0.46	0.53	0.55	0.63	0.625
AIQ	0.18	0.2	0.22	0.21	0.23	0.23	0.27	0.27	0.31	0.34	0.42	0.385
D.padrão	0.280	0.224	0.315	0.292	0.428	0.375	0.309	0.382	0.617	0.468	0.604	0.694
Assimetria	3.72	3.24	3.35	3.90	4.38	3.63	2.27	4.76	4.18	4.66	3.85	4.05
Curtose	17.31	16.38	14.82	21.51	24.03	17.62	7.13	34.45	21.38	35.82	20.13	20.92

Tabela 5.13: Estatísticas para o decaimento. Valores em m^3/s , quando aplicável.

Uma característica comum a quase todos os decaimentos está no facto de a curva da densidade exponencial que melhor se ajusta, ter, relativamente ao histograma e para quase todas as classes, um valor mais elevado para valores próximos da origem, mais baixo para valores intermédios e novamente mais alto para valores mais afastados à direita do centro.

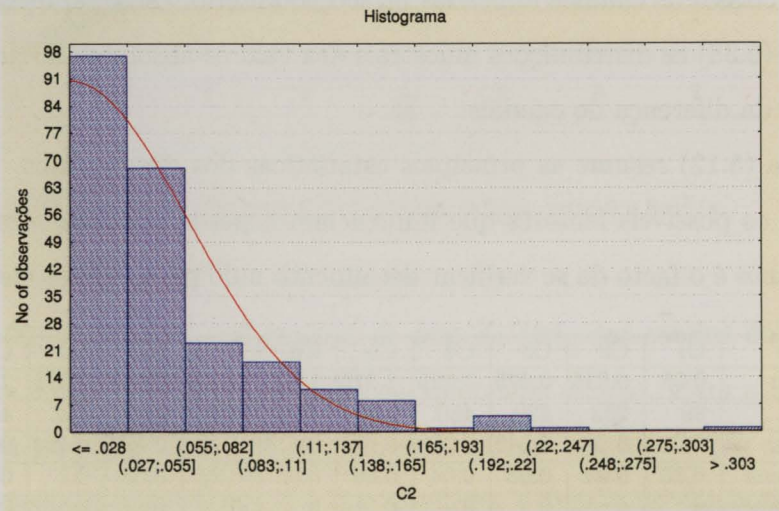


Figura 5.32: Histograma dos decaimentos como valor do caudal condicionado à classe C2 e densidade meia-normal ajustada.

Outras características importantes têm a ver com a relação entre a classe e o decaimento respectivo e à relação entre a média e o desvio padrão do decaimento. Como se pode ver na figura (5.34), a recta de regressão explica 97% da variância. A relação

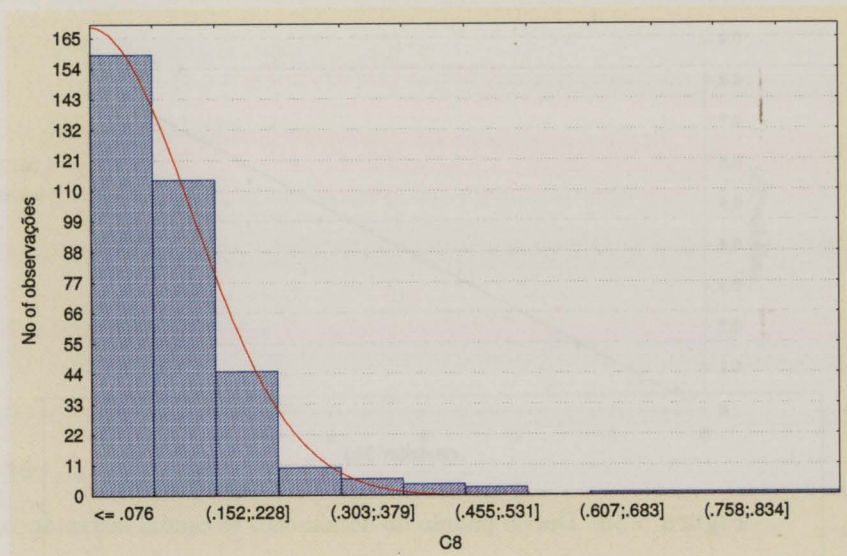


Figura 5.33: Histograma dos decaimentos como valor do caudal condicionado à classe C8 e densidade meia-normal ajustada.

entre o caudal e a média do seu diferencial a 1 dia é aproximadamente linear. Estes factos relativos às estatísticas dos decaimentos mostram que o desvio padrão é cerca de 5-6% do valor do caudal. Não será, portanto, de esperar que qualquer método de previsão pontual consiga, para o regime de descidas de caudal, fornecer uma previsão com um erro inferior.

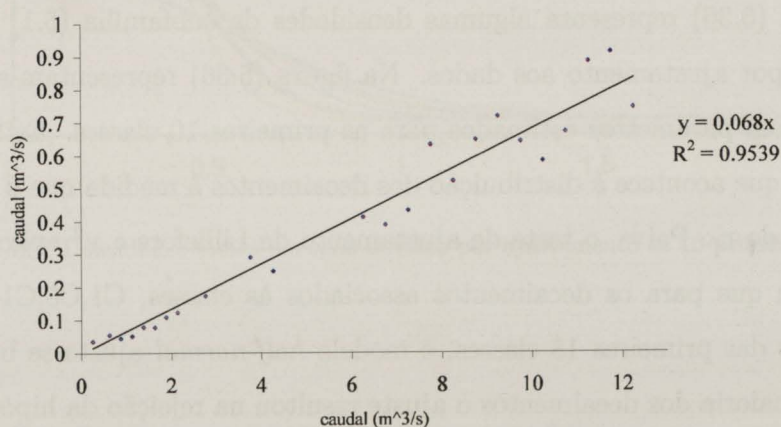


Figura 5.34: Média do decaimento vs caudal instantâneo.

Uma subfamília de distribuições que permite simultaneamente um afastamento à

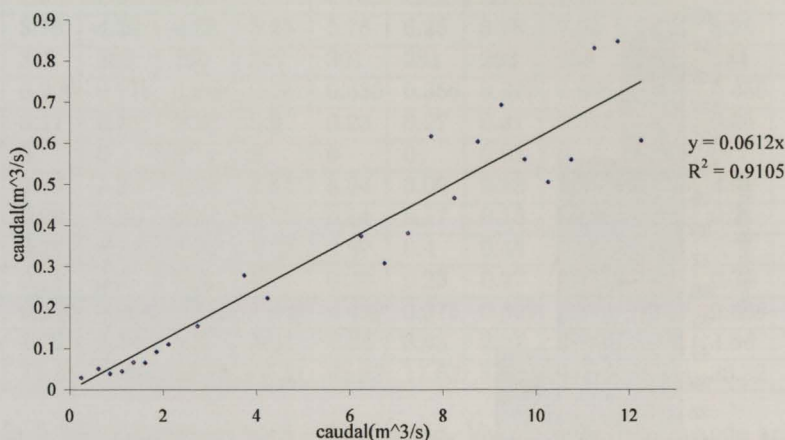


Figura 5.35: Desvio padrão do decaimento vs caudal instantâneo.

distribuição exponencial, tal como descrito acima, é a subfamília de densidades correspondentes à *meia normal*, isto é, a normal truncada a valores superiores à média.

$$f(x, \theta) = \frac{2\theta}{\pi} \exp \frac{-\theta^2 x^2}{\pi}, \theta \geq 0 \quad (5.1)$$

Esta distribuição é a de $|X|$ onde X tem distribuição gaussiana com média zero e desvio padrão $1/\theta$. O coeficiente de achatamento, γ_2 é zero e o de assimetria, γ_1 é $1/\theta^{14}$.

A figura (5.36) representa algumas densidades da subfamília (5.1) com parâmetros obtidos por ajustamento aos dados. Na figura (5.36) representam-se as curvas densidade com parâmetros estimados para as primeiras 10 classes. Este gráfico dá uma ideia do que acontece à distribuição dos decaimentos à medida que o caudal aumenta. No caso do rio Paiva, o teste de ajustamento de Lilliefors e χ^2 aplicados aos caudais mostram que para os decaimentos associados às classes, C1, C8, C10, C12, C13, no conjunto das primeiras 18 classes, o modelo *half-normal* ajusta-se bem. No entanto, para a maioria dos decaimentos o ajuste resultou na rejeição da hipótese de nulidade. Outro facto relevante diz respeito à relação entre o desvio padrão e a média dos decaimentos. A relação é aproximadamente linear, como seria de esperar. Verifica-

¹⁴<http://mathworld.wolfram.com/Half-NormalDistribution.html>

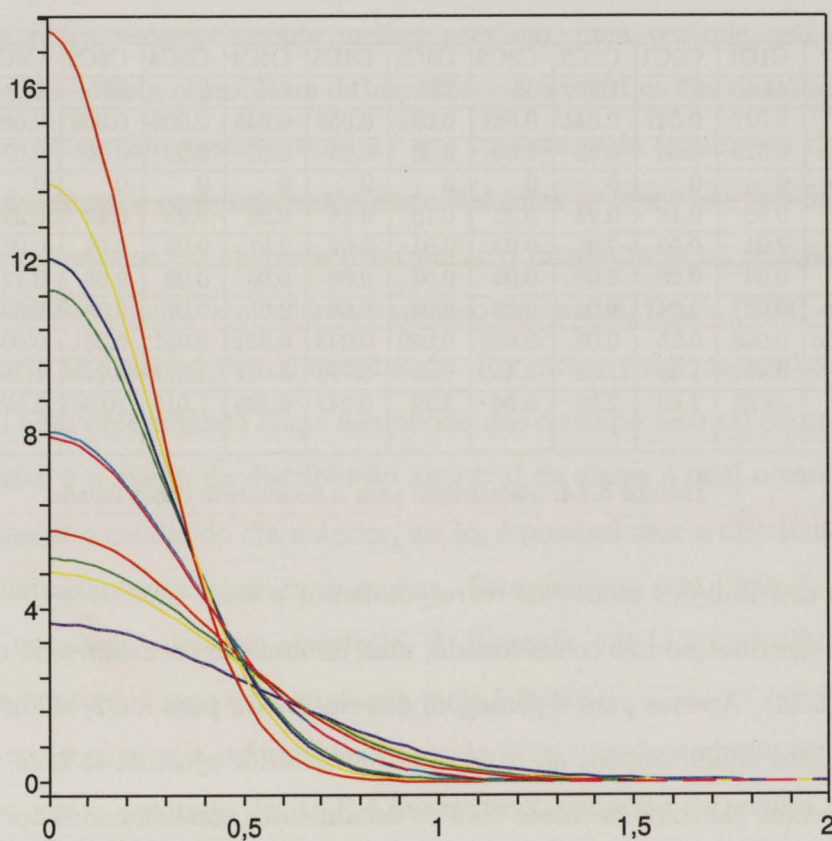


Figura 5.36: Densidades *meias-normais* obtidas por ajustamento às 10 primeiras classes.

se que, para os caudais mais baixos, a mudança de classe por perda de água em dias consecutivos é pouco provável e para esses casos o tamanho das amostras para a situação em que se dá a mudança de classe de caudal são pequenas colocando em causa a significância estatística. Até à classe 16 (de 5.5 a 6.0 m^3/s) a ocorrência mais provável é que o caudal se mantenha na mesma classe no dia seguinte. Por convenção designa-se a variável aleatória correspondente aos decaimentos que levam o caudal a passar para a classe j dado que se encontra na classe i , com $j \leq i$, por C_iC_j .

classe	C1C1	C2C1	C2C2	C3C2	C3C3	C4C3	C4C4	C5C4	C5C5	C6C5	C6C6	C7C6
N	44	15	130	37	235	71	235	91	231	91	224	95
Média	0.027	0.047	0.044	0.061	0.034	0.056	0.043	0.066	0.059	0.082	0.056	0.090
Mediana	0.025	0.03	0.03	0.05	0.03	0.05	0.03	0.06	0.05	0.07	0.05	0.07
Mínimo	0	0	0	0	0	0	0	0	0	0	0	0
Máximo	0.08	0.18	0.23	0.33	0.16	0.18	0.26	0.23	0.42	0.32	0.35	0.26
1ºQ	0.01	0.02	0.01	0.03	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.04
1ºQ	0.04	0.06	0.06	0.09	0.05	0.08	0.06	0.09	0.08	0.12	0.08	0.13
AIQ	0.03	0.04	0.05	0.06	0.04	0.06	0.05	0.07	0.06	0.09	0.06	0.09
D.padrão	0.022	0.05	0.05	0.062	0.029	0.043	0.041	0.051	0.051	0.064	0.049	0.066
γ_1	0.457	1.69	1.58	2.450	1.28	0.776	2.012	1.023	2.241	1.395	1.64	0.830
γ_2	-0.72	2.66	2.35	8.96	2.22	0.247	6.565	1.015	10.867	2.529	5.488	-0.032

Tabela 5.14: Estatísticas para o decaimento condicionado.

Para as distribuições amostrais correspondentes a estes caudais verifica-se, relativamente à distribuição não condicionada, uma diminuição da média e do desvio padrão, tabela (5.14). Apenas para a passagem das classes C3 para a C2, as distribuições dos decaimentos condicionados ao caudal no dia anterior ajustam-se bem à distribuição *meia-normal*, tal como acontece com os decaimentos condicionados apenas ao caudal do dia anterior.

De acordo com o que disse sobre as distribuições condicionadas, o conhecimento do valor do caudal no dia anterior ao presente deveria conduzir, na ausência de chuva, a uma previsão mais precisa. Porém, a diminuição do tamanho da amostra devido ao condicionamento, cada vez mais exigente em termos de dados, faz com que essa diminuição não ocorra. O que acontece, de facto, é que de acordo com a informação disponível, existe, para cada classe considerada, um número óptimo de caudais a usar no condicionamento, tendo em vista, maior precisão na previsão.

Esta informação, num conjunto de dados como os caudais do Paiva, vai influenciar alguns parâmetros da reconstrução e é usada pelos métodos de previsão local no espaço de fase, embora, de forma não supervisionada uma vez que se tratam de métodos não paramétricos. Explicam-se, assim, os diferentes resultados obtidos para diferentes dimensões do espaço de mergulho que é o mesmo que dizer para diferentes condicionamentos. Desta forma, em métodos que usam as distribuições amostrais, o uso de classificação para obter mais distribuições amostrais do comportamento do rio não significa necessariamente melhor previsão, uma vez que, em sistemas tão complexos, se levanta o problema da significância estatística das distribuições.

Não deixa de ser interessante verificar que na fase mais regular do rio, isto é, na ausência de chuva, o comportamento deste, tal como se manifesta nos dados, é essencialmente estocástico. O aparente determinismo manifesta-se na relação entre a distribuição do decaimento e o valor instantâneo do caudal, figura (5.34) e também na relação entre as distribuições do decaimento. Por outras palavras, conhecendo o valor do caudal num determinado dia, e assumindo que o tempo seco vai ocorrer a previsão mais sensata é a média da distribuição amostral da classe à qual o caudal pertence. Se se conhecer o caudal do dia anterior, então, é possível usar a distribuição amostral condicionada com uma média mais precisa. Este processo está limitado pelo número de dados nessas distribuições amostrais. A dimensão que fornece melhores previsões para o rio em tempo seco é determinada mais à frente.

A forma como se usou a informação fornecida pela curva correlação e a comparação que se fez com o resultado da *ACP* é incomum no contexto da análise não linear de séries. Em muitos dos trabalhos publicados sobre o tema usa-se exclusivamente a curva correlação para obter a estimativa da dimensão do atrator procurando estabelecer um intervalo de raios onde os declives da curva sejam estacionários de forma a poder-se usar regressão para estimar a dimensão. No entanto, este procedimento envolve sempre alguma subjectividade uma vez que não existem métodos matemáticos rigorosos para identificar e determinar esse intervalo de valores.

A série dos caudais do rio Paiva não é uma excepção à regra, a análise do integral de correlação não permite, a estimação da dimensão do atrator do sistema. A

intermitência é um desses fenómenos, corresponde a uma alternância irregular do sinal entre um comportamento laminar (regular) e um caótico (irregular). A duração da fase mais irregular é variável, pode ser longa ou consistir em apenas breves *picos*. Na literatura são propostos diferentes cenários que explicam um comportamento intermitente, a que se fez referência no capítulo 3. Convém referir que frequentemente, num sistema desse tipo o regime irregular ocorre numa escala de tempo e espaço diferente do regime laminar. É sabido que em rios de pequenas bacias as variações diárias do caudal são significativas em períodos de precipitação, [Aba96b], [Por97]. Isto é, regiões pertencentes ao atractor e por consequência relevantes para a caracterização da medida invariante não são observadas pelo que essa informação é perdida e não pode ser usada.

Este facto mostra que o caudal responde rapidamente às solicitações da chuva e o processo pode desenrolar-se em horas revelando um tempo de reacção curto. Do ponto de vista económico este facto é certamente relevante e possuir essa informação a uma frequência horária permitiria compreender melhor a relação caudal-precipitação.

Como referem [Kan99] a intermitência em análise de dados deve ser considerada num sentido mais amplo. A fase laminar do sistema intermitente não deve ser interpretada como um período mas sim como uma fase mais regular da série, por exemplo, apenas mais regular que a fase turbulenta.

No caso dos caudais do rio Paiva é razoável a suposição que os *saltos* da série estão associados a valores críticos de um parâmetro de controlo do sistema.

Para que a estimação de invariantes possa ocorrer é necessário que a medida natural do atractor, $\rho(\cdot)$, base para o cálculo da dimensão correlação, esteja estacionária, isto é, que não sofra alterações ao longo do tempo, caso contrário, não se pode estimar a dimensão fractal, caso esta exista. A fase irregular é relevante para a caracterização do atractor e da respectiva medida natural. Uma vez que tem duração variável é natural que quanto mais breve e maior a sua amplitude no espaço, maior terá de ser a série para permitir a sua caracterização com detalhe suficiente para que a estimação de invariantes seja possível.

Como vimos, a falta de homogeneidade da série coloca problemas à análise, o ele-

vado peso das componentes de baixas frequências no espectro levanta questões de estacionaridade nos dados. Com este cenário, torna-se difícil obter estimativas não enviesadas da dimensão uma vez que quase sempre se observa a fase laminar, [Kan99]. Esta argumentação pode ser rigorosamente posta à prova usando um teste estatístico. No entanto, como vamos ver a seguir, as correlações temporais limitam o espectro de testes utilizável.

A medida de reconstrução, [Tak96], surge, como se fez atrás referência, como uma aproximação natural à medida invariante do atrator. Para que a estimação de invariantes faça sentido deve-se investigar se alguma medida de reconstrução associada à série está estacionária. Existem, na literatura, vários testes estatísticos para testar a hipótese de estacionaridade de uma série. O procedimento usual consiste em particionar a série e comparar as propriedades estatísticas de primeira e segunda ordens de cada uma. Alguns desses métodos foram desenvolvidos para detectar tendências não-estacionárias em alguma característica da série tal como a média ou a variância, [Gra64], [Gre57], enquanto que outros foram desenvolvidos para avaliar diferenças na estrutura de covariância ou do espectro das várias secções, [Jen61].

Quando a não linearidade está presente as diferenças podem encontrar-se em momentos de ordem superior e por isso vai aqui encarar-se a não-estacionaridade num sentido mais abrangente, não nas propriedades estatísticas dos invariantes de séries geradas por modelos gaussianos mas em diferenças entre medidas de probabilidade que, como acima se indica, constituem a base para o cálculo dos invariantes do sistema.

Nas figuras, (5.37), (5.38) representam-se os histogramas correspondentes às séries 1946-99 e 1946-72. Formalmente a medida representada nos histogramas corresponde à medida de reconstrução para $m = 1$. Devido a requisitos de consistência, a medida de reconstrução para $m = 1$ coincide com a projecção de medidas para $m > 1$ por isso, a não estacionaridade desta medida implica a não estacionaridade de qualquer medida de reconstrução obtida com a série. O teste de Kolmogorov-Smirnov para duas amostras pode, verificados os seus pressupostos, ser aplicado às medidas de reconstrução unidimensionais uma vez que não é necessário, neste teste, conhecer a distribuição em causa. Para minimizar o grau de dependência entre os dados a

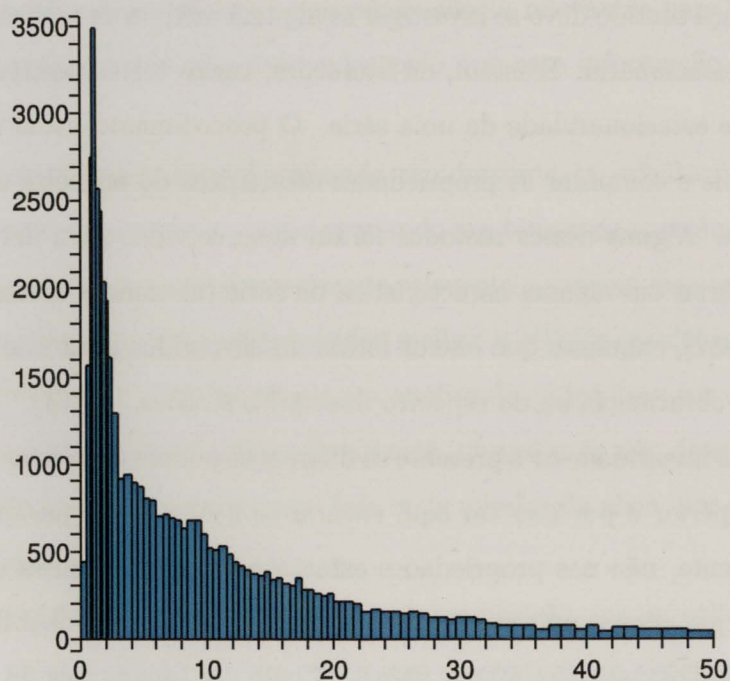


Figura 5.37: Histograma dos caudais médios para o período 1946/99.

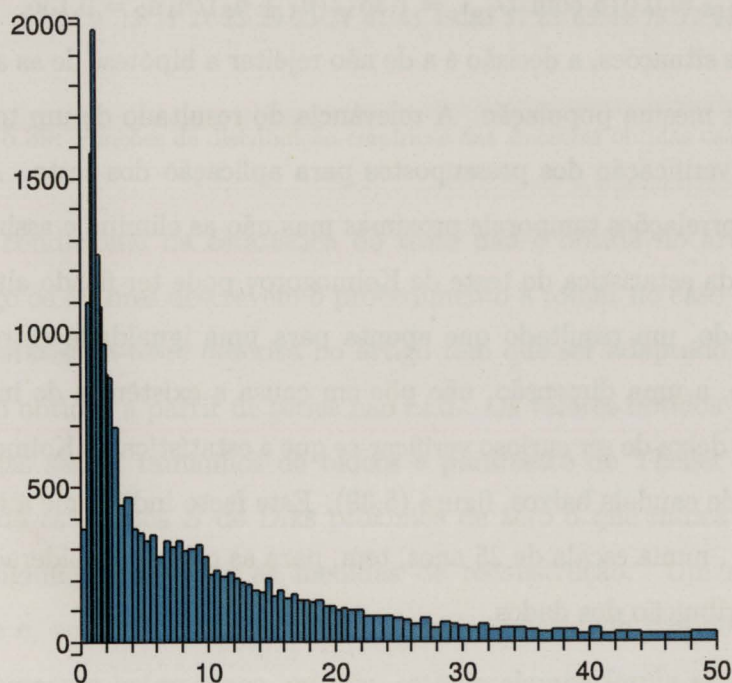


Figura 5.38: Histograma dos caudais médios para o período 1946/72.

usar no teste vão recolher-se amostras de tamanho 500 e 200 às duas *metades* da série correspondentes aos anos, 46-72 e 72-99. Este procedimento designa-se por *bootstrap*, [Efr79] ou reamostragem. O máximo do valor absoluto das diferenças entre as distribuições amostrais é 0.064 correspondente ao caudal $1.25m^3/s$. Dado que as amostras são de grande dimensão, o valor crítico aproximado para um teste bilateral e para um nível de significância de 0.05 é $D_{\text{tab}}^{0.05} = 1.36\sqrt{(n_1 + n_2)/n_1n_2}$ sendo $n_1 = 500$ e $n_2 = 500$ o tamanho das amostras para as distribuições F_1 e F_2 respectivamente. O valor observado da estatística, $D_{\text{obs}}^{0.05}$, é 0.086. Reamostrando com $n_1 = n_2 = 200$ obteve-se $D_{0.05} = 0.075$ com $D_{\text{tab}} = 1.36\sqrt{(n_1 + n_2)/n_1n_2} = 0.136$.

Em ambas as situações, a decisão é a de não rejeitar a hipótese de as amostras procederem de uma mesma população. A relevância do resultado de um teste de hipóteses depende da verificação dos pressupostos para aplicação dos teste. A reamostragem diminui as correlações temporais próximas mas não as elimina e assim, a distribuição de nulidade da estatística do teste de Kolmogorov pode ter ficado alterada.

Por outro lado, um resultado que aponta para uma igualdade entre as medidas de reconstrução, a uma dimensão, não põe em causa a existência de intermitência. No entanto, não deixa de ser curioso verificar-se que a estatística de Kolmogorov ser obtida com classes de caudais baixos, figura (5.39). Este facto indica que a irregularidade da precipitação, numa escala de 25 anos, tem, para as classes consideradas, um pequeno peso na distribuição dos dados.

Para além disso, a medida que está a ser testada corresponde à medida de reconstrução para um espaço de dimensão 1. Se o atractor tivesse, dimensão 1, que alias, vimos anteriormente que corresponde à dimensão deste sistema para uma determinada escala, então, segundo o teorema de Takens, o atractor estaria bem reconstruído num espaço de dimensão igual ou superior a 3. Pelo exposto, é conveniente usar um teste melhor adaptado a esta situação. O teste de diferenças em medidas de reconstrução de Diks, [Dik96] é um teste indicado para este propósito.

A estatística do teste foi calculada para comparar segmentos de subséries consecutivas, 46-51 e 72-77, 51-56 e 77-82 e por fim as séries 46-72 e 72-99. A expressão para

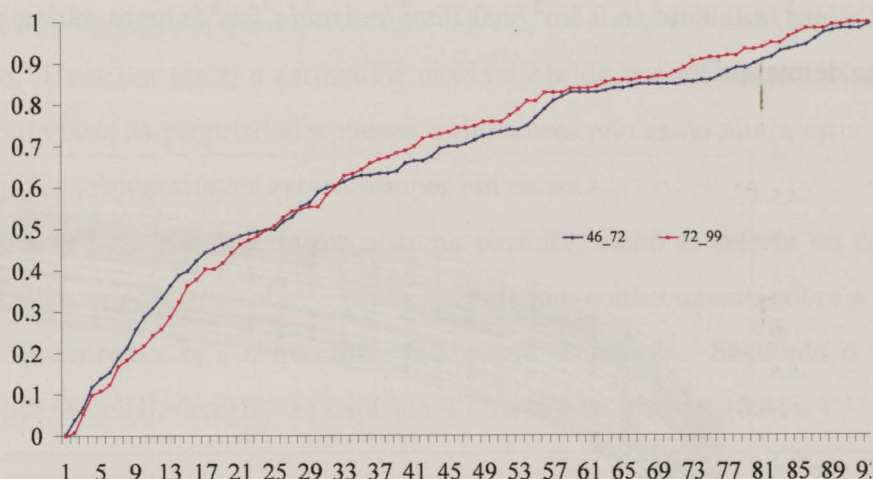


Figura 5.39: Funções de distribuição empíricas das amostras obtidas usando *bootstrap*.

a variância condicional da estatística do teste não é obtida no artigo [Dik96]. No mesmo artigo os autores descrevem o procedimento a tomar no caso de séries i.i.d.. O cálculo da estatística-teste descrita no artigo tem que ser adaptado para medidas de reconstrução obtidas a partir de séries não i.i.d.. Os valores obtidos para a estatística para as várias séries, tamanhos de blocos e parâmetro de Theiler resultam sempre em valores da estatística S de Diks próximos de zero o que indica que não existem diferenças significativas entre as medidas de reconstrução. Um sistema pode ser intermitente e, ao mesmo tempo, pode existir regularidade mesmo numa série com um comportamento heterogêneo, ou seja, está de alguma forma estacionário. Assim, o comportamento dos declives da curva correlação poderá estar relacionado com a insuficiente amostragem.

A dimensão-correlação não é, como se referiu, a única forma de caracterizar a dinâmica de um sistema. Caso a estimação da dimensão não seja satisfatória poderá tentar-se estimar outros invariantes tais como os expoentes de Lyapunov. A existência de um expoente positivo é um indício da existência de dinâmica caótica. Para tal recorrer-se-á ao algoritmo proposto por Kantz, [Kan94], apresentado no capítulo 4. O algoritmo propõe-se fornecer uma estimativa para o expoente máximo de Lyapunov. A figura (5.40) representa a curva $S(\tau)$ para vizinhos próximos nos raios,

$\varepsilon = 0.15m^3/s, 0.25m^3/s, 0.5m^3/s, 0.75m^3/s, 1m^3/s, 5m^3/s$ para várias dimensões do espaço de mergulho.

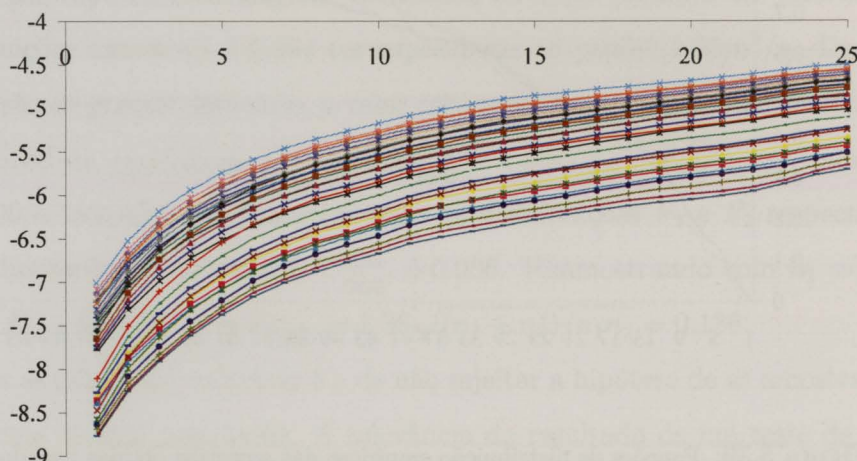


Figura 5.40: Curva $S(\tau)$, para várias dimensões de 2 a 19 e vários raios de vizinhança

As questões relacionadas com a intermitência e a estimação da dimensão correlação colocam-se também na estimação dos expoentes de Lyapunov. A coexistência de diferentes dinâmicas em diferentes escalas tanto espaciais como temporais mas observadas com a mesma frequência de amostragem coloca dificuldades na fixação de uma vizinhança ε fundamental para a estimação de invariantes em séries naturais. Se se tentar contornar o problema usando vizinhanças muito grandes acontece que deixa de fazer sentido falar-se em *vizinho próximo* uma vez que as dinâmicas envolvidas poderão ser muito diferentes e, assim sendo, qualquer estimação baseada nessas órbitas vizinhas não terá qualquer utilidade, figura(5.40).

A análise dos declives das curvas aponta para um crescimento linear para os valores de ε com destaque para os mais baixos e para várias dimensões de espaço (5-10) para um resultados obtidos com desfasamento de $\tau = 9, 10, 11$ dias. É um conjunto muito pequeno de valores onde ocorre a relação aproximadamente linear.

O expoente máximo estima-se como sendo, $\lambda_{\max} = 0.0986 \pm 0.003$, é um valor bastante baixo quando comparado com sistemas caóticos conhecidos, [Kan94]. No entanto é positivo, característica de sistemas com *DSCI*. A explicação para este valor poderá

residir no tipo de caudais que intervêm no cálculo, tal como acontece com a dimensão correlação. Uma vez mais, o estimador usado é obtido usando trajectórias vizinhas no espaço de fase, as propriedades desses estimadores não estão ainda estudadas pelo que a significância estatística estará sempre em causa.

O conhecimento da entropia de um sistema permite, como se referiu no capítulo 3, ter informação importante sobre a predictabilidade e conhecimento sobre a forma do atractor. A entropia k_2 é a que mais facilmente se calcula. Seguindo o algoritmo sugerido por [Kan94], descrito no capítulo 4 obteve-se o gráfico, (5.41).

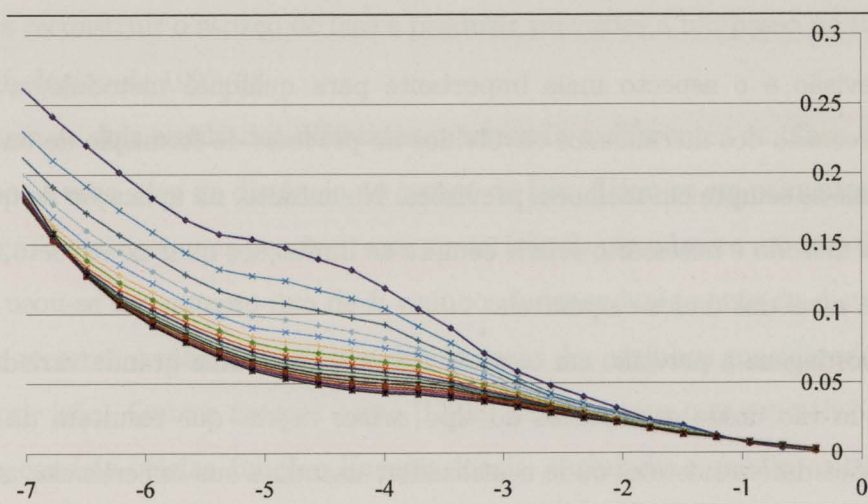


Figura 5.41: $\ln(C_2(m, \varepsilon/s)/C_2(m+1, \varepsilon/s))$ vs $\ln(\varepsilon)$ para várias dimensões de 3 a 20.

Tal como acontece para a dimensão correlação, ocorre, para valores mais elevados da dimensão do espaço, > 12 , o que parece ser uma *scaling range* para a curva indicada, para valores de $\ln(\varepsilon/s) \in (-4.5, -4)$ ou seja para $\varepsilon \in (9.75, 17)$. O valor indicado para a entropia-correlação é, $h_2 = 0.047 \pm 0.0028$. Os valores de $\ln(\varepsilon/s)$ para os quais ocorre esta relação linear ou quase linear entre as curvas são os que coincidem com os da curva de correlação, figura (5.23). Este facto não é de se estranhar uma vez que o integral de correlação amostral é a estatística usada em ambas estimações. Um outro pormenor da estimação da entropia correlação prende-se com a perda de *vizinhos* quando se aumenta a dimensão do espaço de mergulho. Tendo em conta a equação (3.90) verifica-se que $C_2(m+1, \varepsilon) = C_2(m, \varepsilon) \exp(-h_2)$, ou seja, a proporção de pares

de pontos diminui cerca de 4.5% em cada incremento unitário do espaço de mergulho usado. Se se atender ao facto de a *scaling range* ocorrer para dimensões superiores a 12, vem que, relativamente à dimensão 3, existe uma diminuição teórica em mais de 70% do integral de correlação para a mesma distância. Se acrescentarmos este facto ao que foi dito sobre o integral de correlação e a intermitência dos dados é de admitir a possibilidade de a *scaling range* não existir de facto.

5.5 Previsão por métodos locais

A previsão é o aspecto mais importante para qualquer metodologia uma melhor compreensão dos mecanismos envolvidos no processo de formação de um caudal deve traduzir-se sempre em melhores previsões. No entanto, na aplicação de qualquer modelo ou método é necessário referir sempre as limitações quer no aspecto metodológico quer a nível dos dados disponíveis.

As abordagens à previsão em caudais compreendem uma grande variedade de métodos que vão desde os modelos do tipo *caixas negras* que resultam da aplicação de equações determinísticas onde o utilizador, usando a sua experiência, atribui valores aos parâmetros desconhecidos, aos modelos conceptuais muito elaborados onde se faz uso de informação de muitas variáveis observadas temporalmente e espacialmente até aos modelos determinísticos-estocásticos que tentam combinar ambas as abordagens. Modelos mais complexos e exigentes em termos do conhecimento do fenómeno podem aplicar-se, se dispusermos de uma grande quantidade e qualidade de dados, que nem sempre estão disponíveis. Por exemplo, o estudo da relação entre precipitação e caudal implica o conhecimento de muitas variáveis relevantes como o comportamento dos solos, a temperatura, estado da vegetação. Além disso, as frequências de amostragem têm que ser elevadas e com intervalos inferiores ao diário. Como referem Porporato e Ridolfi(2001), das diferentes abordagens, fruto de diferentes propostas e âmbitos de aplicação, é importante reter que os métodos de previsão têm que ter em conta a componente determinística do fenómeno e descrever as suas principais não linearidades. No entanto, existem sempre mecanismos da dinâmica do caudal que ainda não

têm uma descrição matemática adequada. Os modelos extremamente parametrizados exigem uma grande quantidade de dados de modo a obter estimativas robustas e evitar efeitos cuja causa não se consegue explicar. Vários autores usaram métodos de previsão localmente no espaço de fase reconstruído, [Jay94],[Jay02], [PR96],[Por97],[Gio01], [Por01], [LL98] e [IM02] entre outros.

De forma a ilustrar a capacidade preditiva do método dos vizinhos mais próximos, cujo desempenho depende em grande parte do tamanho da série histórica, usou-se a série dos caudais médios diários medidos em Fragas nos anos hidrológicos de 1946 a 1997 para reconstruir o espaço de fase e produzir previsões a um passo no futuro para o ano hidrológico de 1997/98.

Apresentam-se dois resultados diferentes recorrendo a diferentes escolhas do número de vizinhos a usar e/ou da distância no espaço de fase. Usou-se em primeiro lugar uma vizinhança, o mais pequena possível, com a restrição de um número mínimo de vizinhos e depois usou-se um número fixo de k vizinhos independentemente da distância. Em ambos se recorreu à média aritmética das imagens dos vizinhos considerados, para se obter a previsão. Outros autores como [LL98] usaram para o mesmo tipo de dados, séries de caudais, várias funções lineares locais, incluindo a média aritmética dos vizinhos, e concluíram que esta tinha uma capacidade preditiva superior às outras funções locais.

A previsão não linear local sendo um método não paramétrico, no sentido de que não se usa um modelo com expressão funcional parametrizada caracteriza-se pelo facto de a obtenção do valor previsto dependa dos parâmetros de reconstrução e do número de vizinhos fixado, o que tem uma relação estreita com a distância entre o ponto de referência e os seus vizinhos mais próximos. Ou seja, embora não se esperem resultados muito díspares, uma vez que não existe uma razão dinâmica para tal, diferentes valores para aqueles parâmetros fornecem previsões diferentes e com erros médios quadráticos também diferentes.

As tabelas (5.15) e (5.16) contém um resumo da qualidade dos resultados da previsão.

Como seria de esperar, as previsões a um passo no futuro obtidas com diferentes opções

m/τ	3	4	5	6	7	8	9	10	11	12	13	14	15
1	11.01	9.58	9.54	9.82	10.37	10.26	9.65	9.59	10.64	9.88	10.00	10.89	11.58
2	14.55	13.64	14.61	14.27	14.90	14.84	14.08	13.08	14.68	15.06	15.35	15.58	16.29
3	12.89	13.16	17.45	19.19	16.30	14.98	14.16	17.84	16.00	17.02	16.08	13.93	14.30
4	11.83	15.71	20.33	19.25	17.37	17.15	18.06	20.25	17.10	18.23	18.99	18.35	17.79
5	20.84	19.56	22.30	22.67	22.61	21.13	17.59	18.33	17.97	18.26	18.66	18.84	21.04
6	25.42	27.04	23.02	22.40	13.33	15.84	15.01	15.18	15.87	17.38	18.30	18.82	19.18
7	24.04	25.43	31.34	27.26	29.89	27.96	26.68	28.48	29.24	28.33	27.80	28.09	30.03
8	26.23	22.21	23.40	20.09	19.24	17.20	19.99	16.97	20.06	20.15	20.41	20.46	19.52
9	13.29	17.75	13.11	13.52	15.18	13.07	17.20	19.01	19.14	19.75	19.35	18.72	19.88
10	39.90	38.31	34.61	33.47	32.83	33.63	32.66	37.80	37.81	38.91	40.47	40.52	40.63

Tabela 5.15: Resíduo quadrático médio para o ano hidrológico 1997/98 para diferentes dimensões do espaço de fase e espaçamentos temporais em dias. Vizinhança de raio variável.

m/τ	3	4	5	6	7	8	9	10	11	12	13	14	15
1	8.98	7.36	7.08	7.39	8.38	8.35	8.63	8.70	9.29	9.03	9.10	10.12	10.91
2	10.87	12.33	11.72	12.84	12.55	14.01	14.30	14.53	15.28	14.93	15.37	14.28	14.49
3	12.05	11.43	13.04	13.60	13.28	13.49	13.82	15.50	15.58	16.00	14.52	14.40	13.40
4	8.93	11.01	13.35	13.37	11.99	12.36	12.42	13.10	14.04	14.23	15.22	15.07	16.14
5	18.63	17.08	20.32	19.11	17.76	21.08	20.67	20.53	21.40	21.79	21.96	21.62	21.87
6	16.57	16.74	16.92	18.15	14.64	15.09	14.58	14.84	14.72	15.38	17.63	18.17	18.53
7	20.95	22.27	26.13	26.85	26.64	25.25	25.66	26.20	25.93	25.38	25.63	25.43	26.44
8	17.88	17.57	17.96	16.72	17.68	20.37	20.54	18.38	20.69	20.77	20.37	20.21	19.72
9	17.26	18.98	14.31	14.70	13.93	14.04	18.02	18.28	15.71	15.63	15.50	15.97	16.33
10	34.99	34.02	33.23	33.13	33.21	34.90	35.23	39.62	39.66	40.83	40.39	39.97	39.93

Tabela 5.16: Resíduo quadrático médio para o ano hidrológico 1997/98 para diferentes dimensões do espaço de fase e espaçamentos temporais em dias. Vizinhança de raio fixo.

de vizinhança têm um resíduo médio quadrático da mesma ordem de grandeza. Apesar disso a opção por uma vizinhança fixa parece fornecer melhores resultados. Deve destacar-se igualmente os melhores resultados obtidos com desfasamento temporal diário, tabela (5.16).

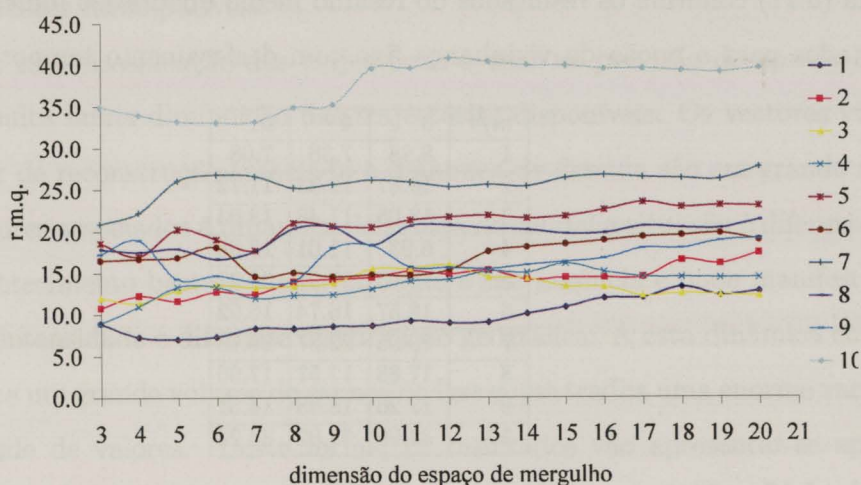


Figura 5.42: Resíduo médio quadrático para a várias dimensões e desfasamentos temporais usando a opção de número de vizinhos fixo.

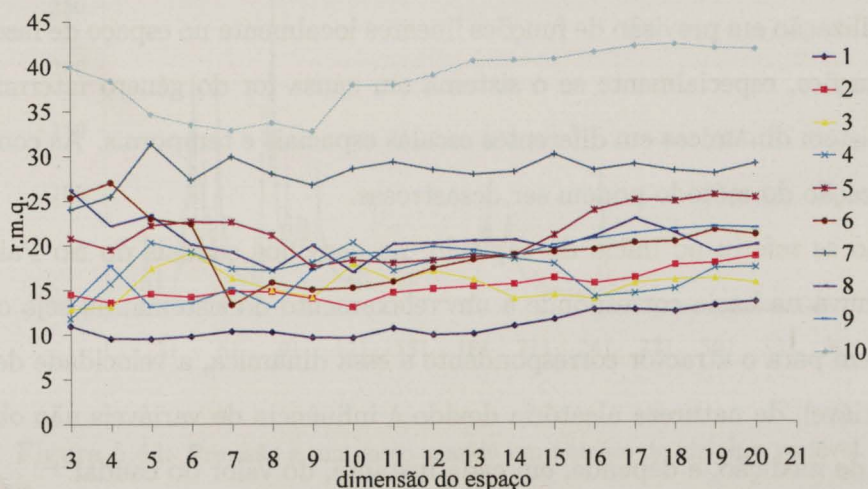


Figura 5.43: Resíduo médio quadrático para a várias dimensões e desfasamentos temporais usando a opção de número de vizinhos variável.

A previsão com melhor resíduo médio quadrático (RMQ) foi obtida com um espaço de

reconstrução de dimensão 6 ou 7 e com espaçamento temporal diário. No entanto, as diferenças encontradas para o resíduo não se podem considerar significativas uma vez que a ordem de grandeza é a mesma. A correlação entre trajectória real e a prevista pode também ser usada para quantificar como indicador da qualidade da previsão. A tabela (5.17) confirma os resultados do resíduo médio quadrático indicando melhores resultados para a opção da vizinhança fixa com desfaseamento temporal diário. Em

m/τ	3	4	5
1	8.98	7.36	7.08
2	10.87	12.33	11.72
3	12.05	11.43	13.04
4	8.93	11.01	13.35
5	18.63	17.08	20.32
6	16.57	16.74	16.92
7	20.95	22.27	26.13
8	17.88	17.57	17.96
9	17.26	18.98	14.31
10	34.99	34.02	33.23

Tabela 5.17: *RMQ* da previsão para o ano hidrológico 1997/98 e diferentes dimensões e espaçamentos temporais.

[IM02] também se obtiveram resultados da mesma ordem de grandeza de *RMQ* para diversas dimensões do espaço de fase em dados de caudais.

A utilização em previsão de funções lineares localmente no espaço de fase tem algumas limitações, especialmente se o sistema em causa for do género intermitente em que coexistem dinâmicas em diferentes escalas espaciais e temporais. As consequências da utilização do método podem ser desastrosas.

Como se referiu no início do capítulo, no caso dos caudais do rio Paiva a ausência de chuva na bacia corresponde a um relaxamento do sistema, ou seja os valores convergem para o atractor correspondente a essa dinâmica, a velocidade de convergência é variável, de natureza aleatória devido à influência de variáveis não observadas e ao erro de medição, e depende, em cada instante, do valor do caudal.

A optimização do processo de previsão local passa por, em primeiro lugar, distinguir diferentes cenários de previsão. A primeira decisão a tomar é se a previsão se faz contando com a ocorrência ou não de chuva. Esta separação é fundamental porque distingue duas dinâmicas totalmente diferentes e que de forma alguma poderiam

ocorrer em simultâneo.

Esta separação de trajectórias de subida e descida no espaço de fase é especialmente indicada para séries que exibam intermitência, não faz sentido usar médias de trajectórias do espaço de fase correspondentes à fase regular e turbulenta uma vez que são acontecimentos incompatíveis.

Claro que esta classificação das trajectórias a usar na previsão de acordo com o seu futuro resulta numa diminuição das trajectórias disponíveis. Os vectores vizinhos de um vector de reconstrução associado à dinâmica de descida são em grande número.

Para vectores associados à dinâmica de subida de caudal a situação é diferente, a chuva é um acontecimento bastante mais raro que a sua ausência e pode manifestar-se com diferente intensidade e diferente distribuição geográfica. A esta dinâmica corresponde igualmente um grande volume do espaço de fase o que traduz uma enorme variabilidade e amplitude de valores. Desta forma, os resultados vão apresentar-se aplicando o método de previsão local na sua versão original e adaptada às características do rio Paiva.

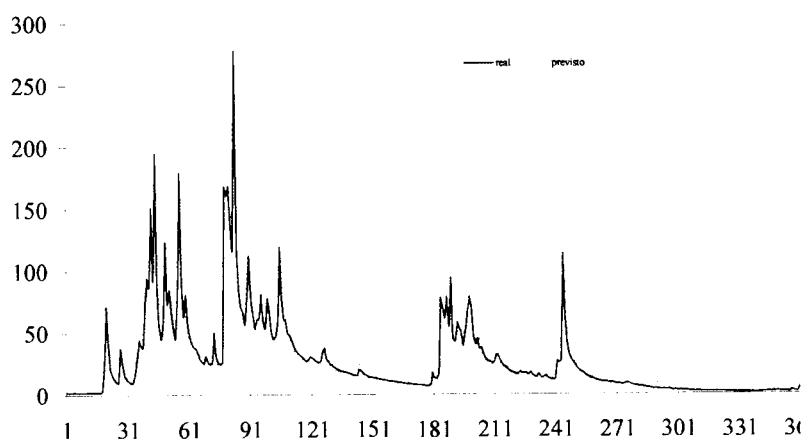


Figura 5.44: Previsão a um passo usando um número de vizinhos variável.

As previsões a 1 passo para o ano hidrológico 1997/98 para as duas opções de previsão a nível local estão representadas nas figuras (5.44,5.45). A característica mais impor-

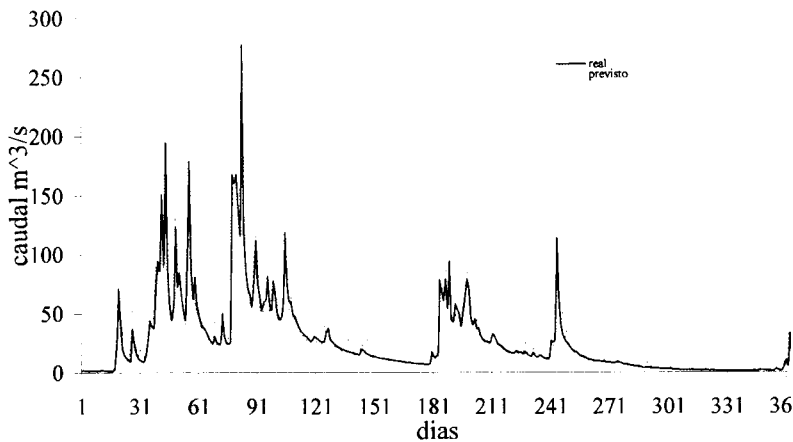


Figura 5.45: Previsão a um passo usando um número de vizinhos fixo.

tante dos cronogramas reside na heterogeneidade na precisão da previsão. Quanto maior for o caudal menor a capacidade preditiva.

Os resíduos do modelo são por definição a diferença entre o previsto e o observado. Caso o modelo em causa represente bem os dados espera-se que os resíduos tenham características de ruído branco. Deste modo pode pensar-se que a análise de resíduos pode dar indicações sobre a adequação do modelo estimado, isto é, se as propriedades dos resíduos se aproximarem das do ruído branco então pode admitir-se que o modelo representa bem os dados. Para um modelo não paramétrico como o dos vizinhos mais próximos isso equivale a dizer que o determinismo é a única característica relevante nos dados. Os resíduos estão representados na figura (5.46).

Uma simples análise visual é o suficiente para verificar que não se está em presença de ruído branco. A existência de valores pequenos e de *saltos* com uma grande amplitude não pode corresponder a uma série ruído. Uma vez que não se trata de um modelo linear gaussiano o máximo que se pode esperar é que a série residual seja i.i.d.. A FAC residual (figura (5.47)) confirma as suspeitas levantadas na observação do cronograma.

O teste de Bartlett, [Mur93], não é aplicável devido à suposição de normalidade, que aqui, como se referiu, não faz sentido. O teste de Kendall e Stuart é menos exigente

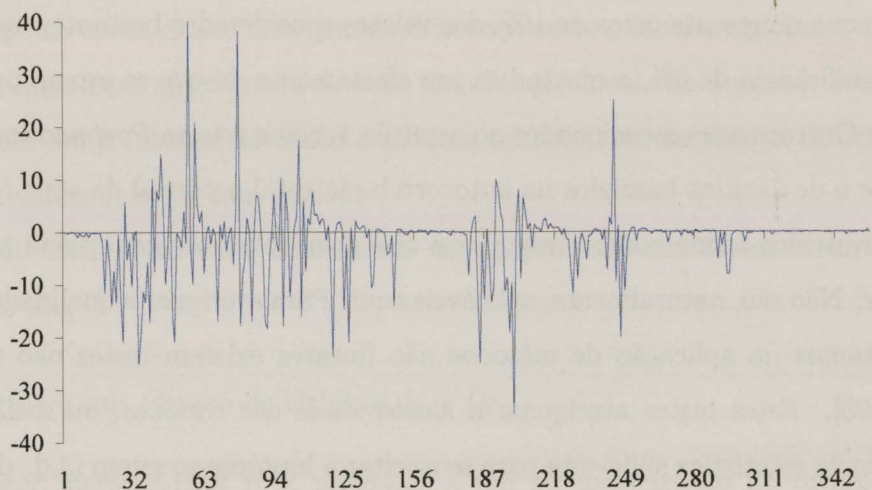


Figura 5.46: Série residual para a previsão a um passo do ano 1997/98 usando os vizinhos mais próximos

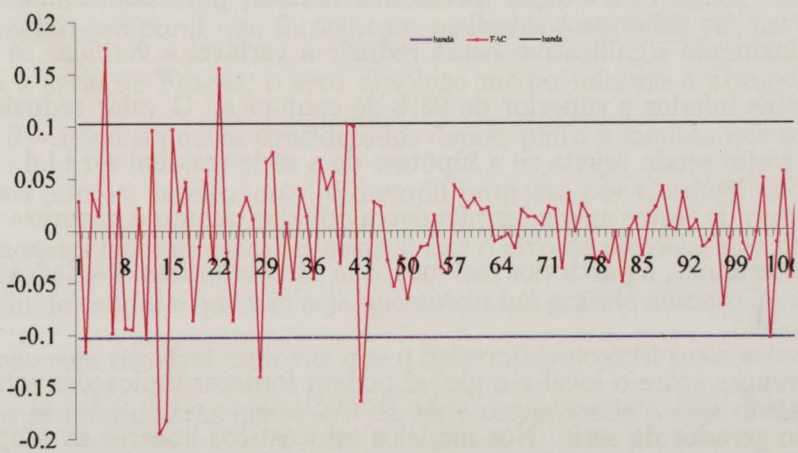


Figura 5.47: FAC da série residual para a previsão a um passo do ano 1997/98 usando os vizinhos mais próximos

nos seus pressupostos e é aplicável nesta situação , (ver capítulo 1).

A *FAC* amostral ultrapassa por 9 vezes as bandas inferior e superior, nos desfasamentos, o que representa cerca de 10% dos valores considerados bastante superior ao nível de significância de 5%, o que indica um afastamento do que se espera para uma série i.i.d.. Outros testes mencionados no capítulo 1, como o teste *Portmandeau*, o de *Ljung Box* e o de Jenkins baseados na autocorrelação total e parcial da série residual forem desenvolvidos tendo como pressuposto que o modelo que deu origem aos resíduos é linear. Não são, naturalmente, aplicáveis aqui. Para averiguar a qualidade dos resíduos resultantes da aplicação de métodos não lineares existem testes não paramétricos, [Mur93]. Estes testes averigam a aleatoridade dos resíduos, ou melhor, se existe evidência estatística suficiente para se rejeitar a hipótese se serem i.i.d. de média zero. O teste baseado nos pontos de viragem, apresenta, como valor observado para estatística, T , o valor 212, ao que corresponde o valor reduzido, -0.96. Este valor encontra-se no intervalo centrado de 95% de confiança para a v.a. normal reduzida pelo que não existe evidência estatística para se rejeitar a hipótese de dados não i.i.d..

A estatística do teste da diferença de sinais tem como valor observado, $S = 154$. Sob a suposição de aleatoridade espera-se que, $\mu_S = 1/2(n - 1)$, onde n é o tamanho da série. Tendo distribuição assintótica normal, para determinar se o resultado é estatisticamente significativo basta reduzir a variável e verificar se se encontra fora dos limites inferior e superior de 95% de confiança. O valor reduzido é observado é -2.09 e assim sendo rejeita-se a hipótese de a série residual ser i.i.d..

O conjunto de testes aplicados não deixa dúvidas quanto à natureza da série residual. Falta determinar, a partir dos resíduos, que características dos dados não são captadas pelo modelo.

As diferenças entre o local e o global podem fornecer indicações sobre a natureza do processo gerador da série. Nos modelos estocásticos lineares as propriedades locais e globais são as mesmas, em dinâmica não linear isso não acontece. Em [Cas91a] o autor propôs um algoritmo denominado *Deterministic versus stochastic* baseado no método dos vizinhos mais próximos, de modo a distinguir entre o determinístico e o estocástico. O procedimento depende de um parâmetro, k , que representa o número de vizinhos e

controla o carácter local da previsão. Para os modelos estocásticos lineares, a melhor previsão é obtida para um $k \approx n$, no entanto, se a melhor previsão for conseguida com um k pequeno então existe uma grande diferença entre o local e o global e a dinâmica local fica mais bem aproximada usando uma função local. Se o melhor resultado for obtido com um k intermédio então o sistema tem as duas naturezas, estocástico não linear ou determinístico com ruído.

O procedimento é em tudo semelhante ao da previsão não linear local, os passos principais são;

- Escolher um número de vizinhos, $\mathbf{k} = \{k_1, \dots, k_q\}$
- Obter, para cada $k \in \mathbf{k}$ previsões a um ou mais passos no futuro para um subconjunto adequado dos dados¹⁵ e uma função linear local.
- Calcular o erro absoluto médio.

Em [Cas91a] é usado o erro médio absoluto por ser mais resistente a *outliers*. A figura (5.48) mostra o comportamento do erro de previsão absoluto médio em relação a k como percentagem de todos os dados usados no modelo local. Usou-se a dimensão e o desfasamento temporal que forneceram melhores resultados em previsão, tabela (5.2). Para a série de Fragas, o erro absoluto médio mínimo é atingido quando se consideram 0 – 1% dos pontos considerados depois piora à medida que se consideram cada vez mais pontos. Infelizmente, o método proposto por Casdagli não tem grande utilidade em séries intermitentes. É natural que o erro de previsão seja menor no caso de diminuição de caudal e que não seja necessário um grande número de *vizinhos* para que a previsão seja razoável uma vez que o determinismo está mais acentuado. Além disso e como se referiu atrás quase 2/3 da série corresponde a essa dinâmica. Assim, os resultados obtidos apenas confirmam alguns resultados obtidos anteriormente.

No seguimento do que foi dito sobre a predictabilidade do sistema, espera-se que o número de vizinhos a uma determinada distância tenha uma relação com o caudal. O

¹⁵no caso dos processos hidrológicos faz todo o sentido usar um ano

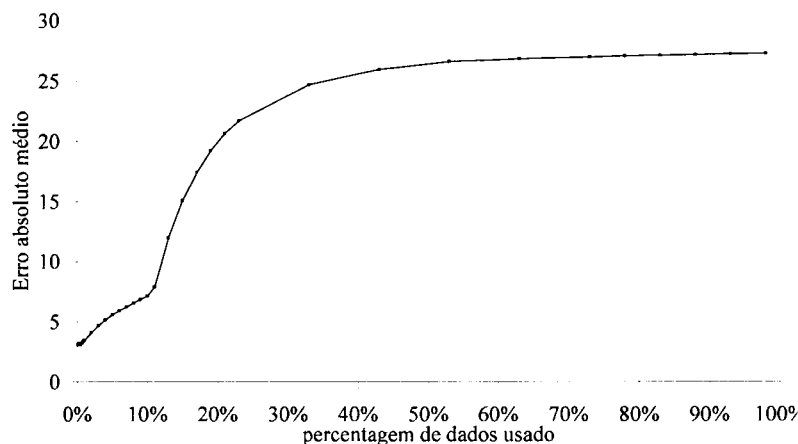


Figura 5.48: *DVS plot* para o rio Paiva em Fragas. Erro médio absoluto em função da percentagem de vizinhos usada.

desempenho do método de previsão pelos vizinhos mais próximos é bastante superior para caudais mais baixos como se pode comprovar pelas figuras, (5.49) e (5.50). Os vectores de reconstrução cujos elementos correspondem a caudais elevados têm poucos vizinhos pelo que as distâncias ao ponto de referência têm que ser aumentadas. Assim para produzir a previsão é necessário usar informação de órbitas mais distantes. Sendo os vizinhos mais próximos um método que cuja previsão pontual é uma média local procurando, assim, explorar o determinismo, o aumento das vizinhanças conduzirá a piores previsões.

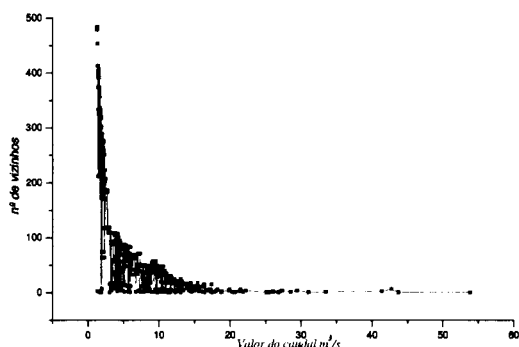


Figura 5.49: Vizinhos no espaço de fase em função do caudal do rio. Ano hidrológico 1997/8.

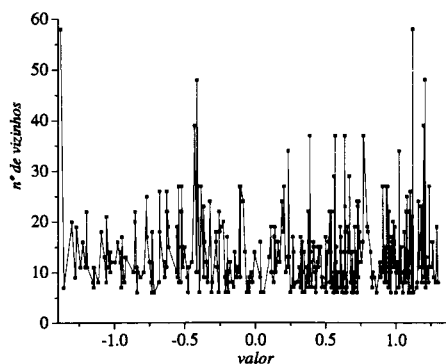


Figura 5.50: Vizinhos encontrados a uma distância ϵ no espaço de fase em função do valor em cada instante. Atrator de Hénon.

Para efeitos de comparação com os caudais do Paiva colocamos na figura(5.50) o

atractor de um sistema caótico conhecido, [Hén76]. As previsões a um passo que se apresentam de seguida foram filtradas de acordo com o cenário pretendido, apenas para o regime de descidas de caudais. A filtragem do conjunto de dados para produzir a previsão consiste em não usar informação dos vizinhos próximos cujo futuro é uma subida de caudal. Na tabela (5.18) apresentam-se os resíduos médios quadráticos e

m	RRM(A)	RMQ(B)
1	0.0486	2.915
2	0.0328	1.004
3	0.0301	0.888
4	0.0310	1.034
5	0.0341	1.155
6	0.0378	1.193
7	0.0427	1.375
8	0.0472	1.651
9	0.0523	1.605
10	0.0584	2.063

Tabela 5.18: Resíduo médio quadrático e resíduo relativo médio para os anos hidrológicos 1995-99 para o regime de descidas de caudais e filtragem de vizinhos.

resíduos médios relativos para a previsão a um passo para as descidas de caudais dos anos hidrológicos 1995-99 usando várias dimensões do espaço de fase. As previsões foram obtidas nas condições acima indicadas. As diferenças entre as várias dimensões do espaço de fase consideradas são pouco significativas, no entanto, a dimensão 3 fornece os melhores resultados. Os resultados mostram também que, no caso de não haver precipitação, e, sem qualquer restrição nos caudais considerados, a informação (local) ideal do fenómeno é de 3 dias. O uso de mais informação de caudais nos dias anteriores não resulta numa melhoria significativa na previsão.

5.6 Previsão usando janelas de correlação

Como foi referido no capítulo anterior a presença de algum determinismo faz com que ocorram comportamentos recorrentes cujos valores têm correlação elevada com segmentos passados. Para obter previsões para os caudais do rio Paiva segundo o método descrito no capítulo 5 e pelos motivos indicados usaram-se vários métodos de regressão. A regressão linear simples, (*LSR*), e a regressão robusta a que dizem respeito

as estimativas $M(RLR)$ com base na função de Huber com parâmetro 1.5 (A) e a de Hampel com pesos, $H1 = 1.5$, $H2 = 3.0$ e $H3 = 4.5$, (B), e também a regressão L_1 ou a dos desvios absolutos mínimos (LAD) [Bir96]. Na tabela (5.19) apresentamos os

regressão	RMQ
LSR	22.22
M-R(A)	22.45
M-R(B)	22.41
LAD-R	23.17

Tabela 5.19: Resíduo quadrático médio para o ano hidrológico 1997/98 para diferentes tipos de regressão.

valores do resíduo médio quadrático obtido para a previsão do ano hidrológico 1997/98 para cada um dos tipos de regressão referidos. Como se indicou em secções anteriores, a dinâmica correspondente ao comportamento mais regular do rio vai estar associada a um maior número de segmentos correlacionados o que significa maior predictabilidade. A dinâmica mais irregular correspondente a uma resposta à precipitação vai estar associada a menor número de segmentos correlacionados verificando-se, inclusive, a inexistência de correlação e, como consequência, a impossibilidade de obter uma previsão com o método.

Uma vez que se tratam de extrapolações, a partir de uma recta de regressão, existe a possibilidade de eventos de ordens de grandeza diferentes poderem contribuir para uma previsão. Por um lado, este facto pode ser prejudicial para a previsão uma vez que se a dinâmica pode ser diferente em escalas diferentes, por outro lado, a eliminação de segmentos, para efeitos de cálculo da previsão, limita por vezes o número de segmentos tornando as estatísticas mais pobres. A existência de uma forte correlação em segmentos de escalas de grandeza diferentes pode ser resultado da acção do ruído, uma vez que se consideram janelas de pequena dimensão. Desta forma experimentou-se eliminar alguns dos segmentos correlacionados cuja ordem de grandeza fosse muito diferente. O resultado foi que, para o ano hidrológico 1997/98, em vez de 7 dias sem possibilidade de fazer previsão ocorreram 34 dias sem que o resíduo quadrático médio melhorasse de forma significativa.

Os resultados obtidos para os diferentes métodos de previsão são muito próximos, o

que significa que o problema não está num hipotético afastamento da normalidade que poderia afectar os parâmetros da regressão linear simples.

As figuras (5.28,5.29) representam o cronograma do ano hidrológico e as previsões a um passo para cada um dos métodos de regressão. Para preditor escolheu-se uma estatística de ordem, o primeiro quartil do conjunto de previsões.

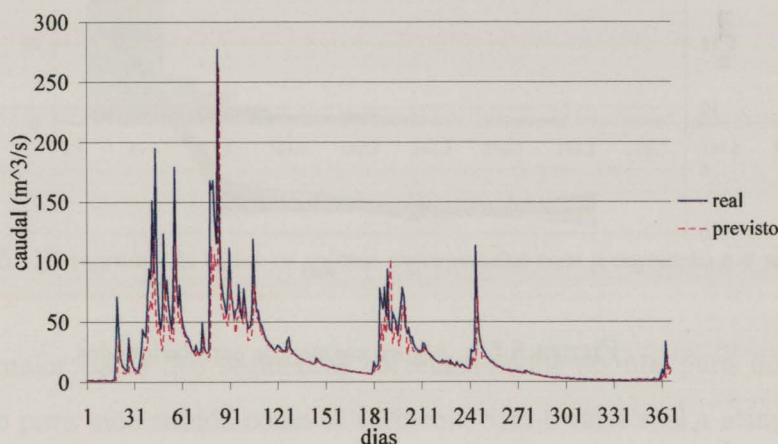


Figura 5.51: Previsão a um passo usando janelas de correlação e *LSR*. Ano hidrológico 1997/8.

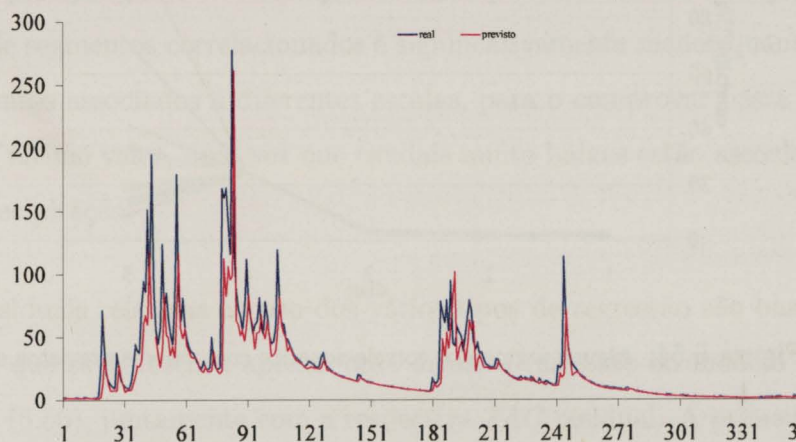


Figura 5.52: Previsão a um passo usando janelas de correlação e *RSR*. Ano hidrológico 1997/8.

As deficiências dos preditores baseados em estatísticas de ordem são as mesmas das estatísticas de ordem como medidas de localização. Na prática, para a previsão para o dia seguinte intervém apenas um segmento. É pouco provável que o segmento cujo

futuro está mais próximo à aquele que vai ocorrer corresponda sempre ao primeiro quartil do conjunto de previsão.

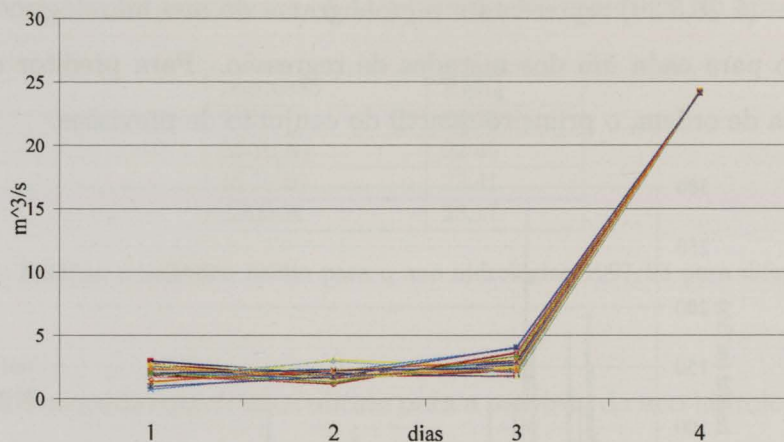


Figura 5.53: Alguns segmentos correlacionados.

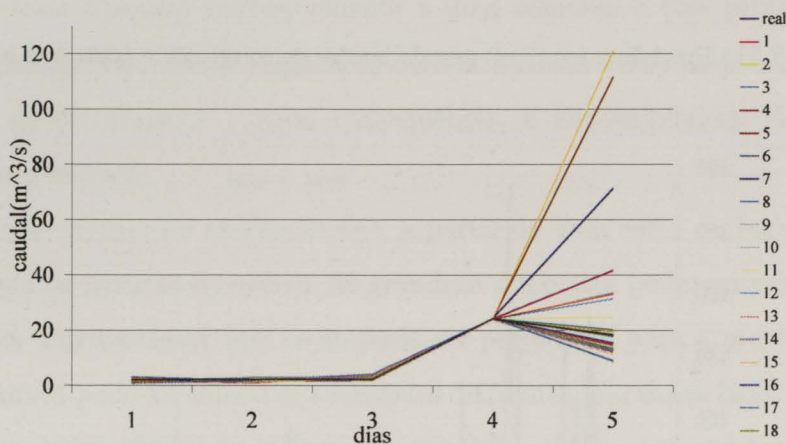


Figura 5.54: Alguns segmentos correlacionados com valores previstos e valor real.

A figura (5.53) representa alguns segmentos correlacionados. O histograma da figura (5.55) indica que a média dos valores obtidos por extrapolação, por ser sensível a valores extremos não é uma opção razoável para preditor.

Na figura (5.54) está representado o segmento para o qual se pretende fazer previsão e alguns segmentos correlacionados após correcção de escala por regressão. O último valor representado é o valor previsto usando cada um dos segmentos. Como se pode

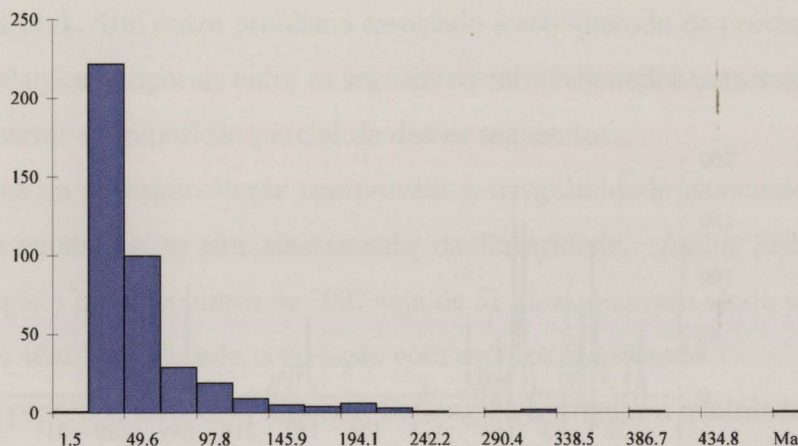


Figura 5.55: Histograma de todos os valores extrapolados com a regressão e o valor observado.

verificar a maior parte dos segmentos correlacionados aponta para uma descida do caudal e não para uma subida como se verificou, com o valor real a atingir $71.38m^3/s$. Assim, obteve-se um mau resultado em previsão. Esta situação vai verificar-se quase sempre que o último valor do segmento a prever esteja associado a um período de ausência de precipitação e o valor a prever corresponda a um dia com precipitação. O número de segmentos correlacionados é significativamente menor quando se consideram segmentos associados a diferentes escalas, para o comprovar basta tomar como referência o último valor, uma vez que caudais muito baixos estão associados a vários dias sem precipitação.

As séries residuais relativas ao uso dos vários tipos de regressão são bastante semelhantes pelo que se apresenta apenas uma delas, a referente ao modelo de regressão *LSR*, figura (5.56), juntamente com a respectiva *FAC* residual. A primeira análise do cronograma revela um afastamento daquilo que se espera de uma série i.i.d.. Tratando-se de uma série constituída pelos valores, $X_{obs} - X_{prev}$, existe um maior número de valores positivos o que corresponde a um enviesamento do estimador usado. Álias, o teste de pontos de viragem confirma isso mesmo com valores para a estatística estandardizada de -2.91 para a regressão linear simples e -2.66 para as restantes regressões, valores que indicam a rejeição da hipótese não paramétrica de se tratar

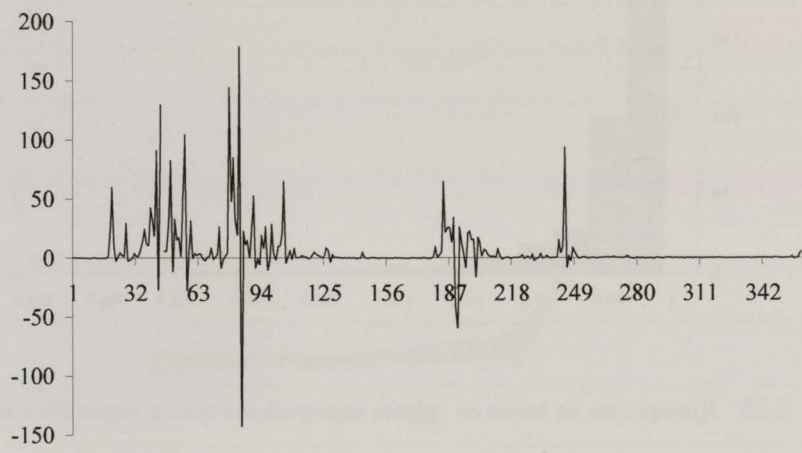


Figura 5.56: Série residual usando a regressão LSR.

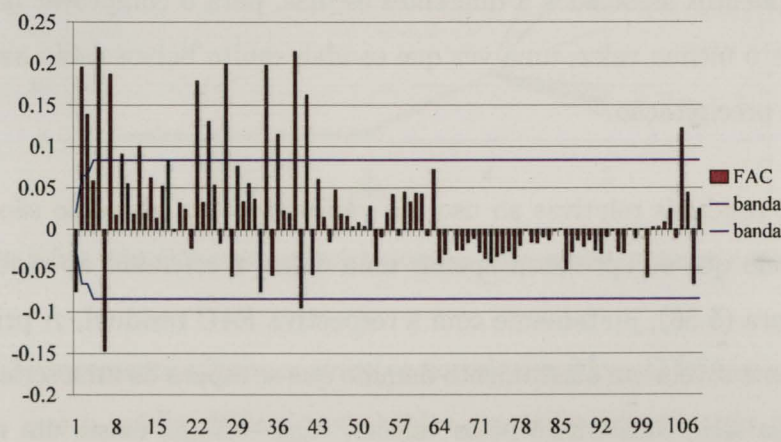


Figura 5.57: FAC da série residual.

de uma série i.i.d.. Um outro problema associado a este método de previsão tem a ver com as correlações temporais entre os segmentos correlacionados uma vez que poderá, inclusivé, ocorrer sobreposição parcial de desses segmentos.

As limitações da regressão linear comprovam a irregularidade associada ao sistema gerador dos caudais e ao seu afastamento da linearidade. Assim justifica-se que, embora o registo histórico antes de 1997 seja de 51 anos, ocorram ainda sequências de dias que não têm uma elevada correlação com eventos do passado.

Tal como se procedeu na secção anterior, falta agora averiguar a qualidade da previsão obtida, usando apenas as descidas de caudais e suprimindo as trajectórias correlacionadas cujo futuro é uma subida de caudal. Para efeito de comparação com o

preditor	R.m.q.	
	(A)	(B)
média	8.836	0.687
1ºQ	11.895	0.748
2ºQ	9.785	0.701
3ºQ	6.767	0.620

Tabela 5.20: Resíduo quadrático médio (coluna (A)) e resíduo relativo médio (coluna (B)) para os anos hidrológicos 1995/99 e apenas para as descidas de caudais usando diferentes preditores.

método da secção anterior usa-se o mesmo conjunto de dados e previsões na mesma situação, ou seja, filtrando as previsões de aumento de caudal. A tabela (5.20) resume os resultados,

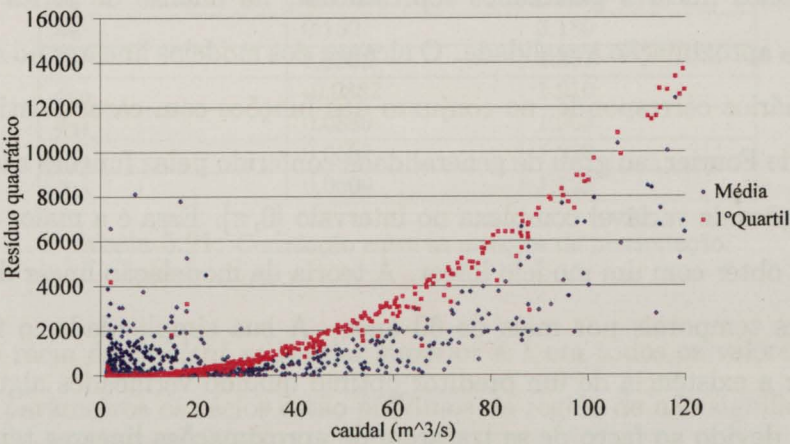


Figura 5.58: Comportamento do resíduo médio quadrático da previsão obtida usando como preditor a média e o primeiro quartil.

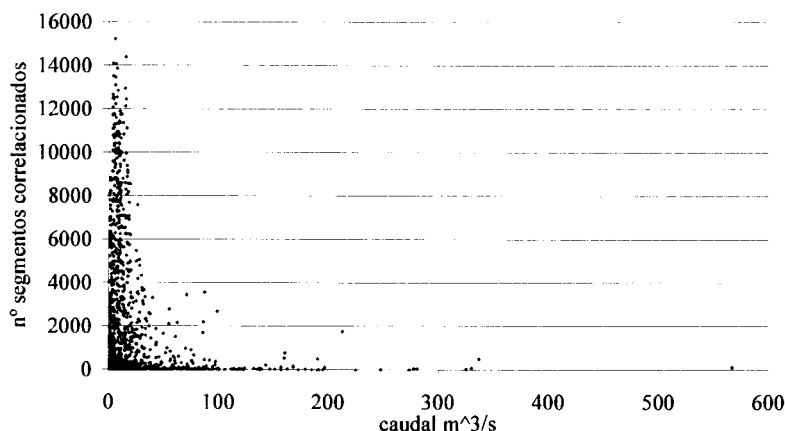


Figura 5.59: Número de segmentos correlacionados em função do valor do caudal.

Os resíduos verificados são maiores do que os obtidos com o método dos vizinhos mais próximos. A média dos valores extrapolados gera resíduos da mesma ordem do que os obtidos com as estatísticas de ordem embora o 3º quartil dê o melhor *RMQ* e o 1º quartil o melhor resíduo relativo médio. Este facto indica que para as descidas de caudais existam fortes correlações que certamente serão espúrias contaminando as distribuições de valores extrapolados.

5.7 Estimação de um modelo linear

Os modelos lineares gaussianos representam, na análise de séries temporais, uma primeira aproximação à realidade. O alcance dos modelos lineares no seio dos processos estacionários corresponde, no conjunto das funções com características das transformadas de Fourier, ao grau de generalidade conferido pelas funções racionais no espaço das funções de variável complexa no intervalo $[0, \pi]$. Essa é a maior generalidade que se pode obter com um modelo linear. A teoria da modelação linear dominou a análise de séries temporais por mais de 60 anos. A sua simplicidade, o facto de a teoria garantir a existência de um preditor óptimo quando verificados alguns pressupostos, e ainda devido ao facto de se tratarem de aproximações lineares tendo como critério de minimização a norma L_2 permitiu a caracterização completa destes modelos. Para mais informação sobre estes modelos ver; [Box76b], [Pri81], na língua portuguesa de

salientar a obra [Mur93]. O resultado obtido com a modelação linear será comparado com o obtido pelos outros dois métodos.

Usando o pacote *ITSM96*¹⁶, estimou-se um modelo linear $AR(15)$, após transformações que tornaram estacionária a série¹⁷. Não havendo critérios rigorosos para determinar se uma série está estacionária usa-se o bom senso, tendo em conta que todas as transformações têm sempre consequências indesejáveis tais como, o aparecimento de variância espúria, [Mur93].

A incapacidade do pacote informático *ITSM96* em analisar séries de dimensão superior a 2300 observações levou a que apenas se considerassem os anos hidrológicos 93/94 a 98/99. Naturalmente que um modelo estimado usando toda a série, isto é, a totalidade dos 53 anos hidrológicos resultaria num modelo diferente. Usando a linguagem de [Bro91a] o modelo ensaiado corresponde, relativamente à série original, sem transformações, a um $SARIMA(15, 1, 0) \times (0, 1, 0)_{365}$,

$$(1 - \sum_{i \in \{2,4,5,6,8,10,11,14,15\}} \phi_i B^i = \varepsilon_t \tag{5.2}$$

Os valores obtidos para os parâmetros no processo de estimação, e, respectivos rácios-*t* encontram-se na tabela (7.1).

Parâmetro	Estimativa	Rácio-t
ϕ_2	0.128	2.625
ϕ_4	0.0571	1.213
ϕ_5	0.0610	1.325
ϕ_6	0.150	3.189
ϕ_8	0.0651	1.397
ϕ_{10}	-0.0882	1.910
ϕ_{11}	0.0880	1.903
ϕ_{14}	0.0597	1.228
ϕ_{15}	0.0509	1.122

Tabela 5.21: Correlação entre as estações de precipitação.

Apesar de o rácio de *Student* se revelar superior a 1 em todos os valores estimados, para alguns parâmetros os rácios estão próximos da região de não significância.

¹⁶distribuído gratuitamente com o livro [Bro91a]
¹⁷o conceito de estacionaridade no sentido da modelação linear compreende a eliminação de tendência, estabilização da variância, eliminação de movimentos sazonais e do *drift* que corresponde à instabilidade da média.

A matriz de correlação entre estimadores revela a inexistência de correlação significativa entre as estimativas obtidas, tendo em conta os valores de referência defendidos por diversos autores, [Box76b].

1	0.185	0.0415	0.0733	0.172	0.0739	0.0432	0.1	0.0443
0.185	1	0.0385	0.155	0.078	0.152	-0.0238	0.133	0.0797
0.0415	0.0385	1	0.0367	0.0304	0.04	0.133	0.0649	0.0586
0.0733	0.155	0.0367	1	0.133	0.0423	0.0624	0.0793	0.0203
0.172	0.078	0.0304	0.133	1	0.0822	0.0294	0.0181	-0.0536
0.0739	0.152	0.04	0.0423	0.0822	1	-0.00457	0.0522	0.0426
0.0432	-0.0238	0.133	0.0624	0.0294	-0.00457	1	0.0463	0.0274
0.1	0.133	0.0649	0.0793	0.0181	0.0522	0.0463	1	0.0178
0.0443	0.0797	0.0586	0.0203	-0.0536	0.0426	0.0274	0.0178	1

A avaliação da qualidade do ajustamento de um modelo estimado pode medir-se através da análise de resíduos. As figuras (5.60,5.61) representam, respectivamente, os resíduos resultantes do ajuste do modelo aos anos hidrológicos de 1995/99 e a respectiva *FAC* residual.

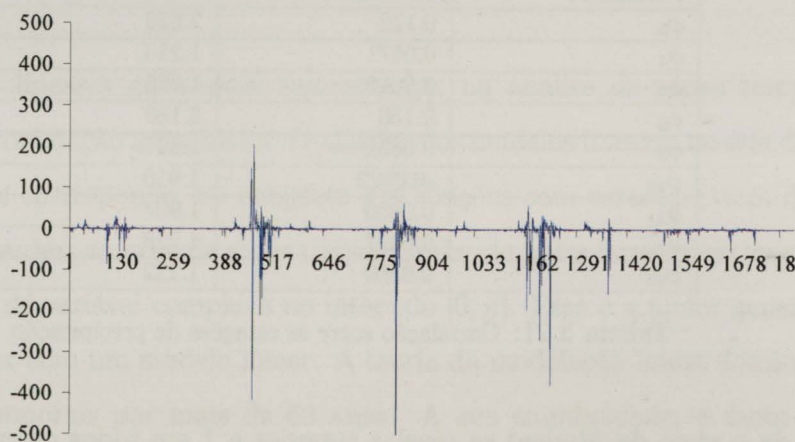


Figura 5.60: Série residual 1995/99.

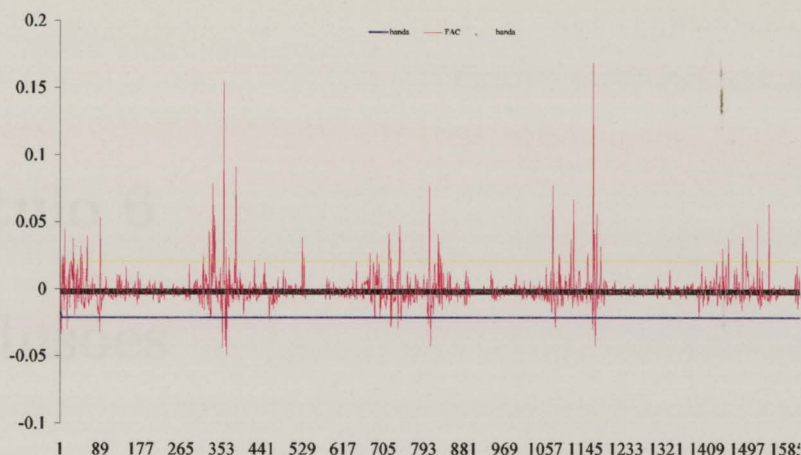


Figura 5.61: *FAC* da série residual 1995/99.

Os resíduos que se apresentam foram obtidos após inversão das transformações efectuadas a partir do modelo estimado para a série transformada. Os resultados obtidos, para os testes mostram a inadequação do modelo. Os resíduos médios quadráticos obtidos com o modelo foram de 30.74 para ambos os regimes e 9.17 para o regime das descidas de caudal. Para este último regime de caudais, o resíduo relativo médio foi de 0.16. Os testes estatísticos da avaliação da qualidade do ajustamento são efectuados sobre os resíduos correspondentes à série transformada. Como se pode prever pelo cronograma residual, as estatísticas dos resíduos estão longe do que se espera de uma série de ruído branco. A análise do cronograma e da correlação dos resíduos revelam que o modelo linear não capta ou não consegue explicar a variação associada às subidas de caudal. Esta situação é comum quando se pretende usar modelos lineares gaussianos em séries resultantes da observação de sistemas naturais, [Ton90]. Os modelos lineares dão bons resultados. A modelação linear *Box Jenkins* funciona bem quando os dados revelam estar próximos de o que se espera da normalidade. Um mau resultado obtido pode-se explicar por um afastamento da normalidade e/ou da linearidade. O gráfico (5.62) corresponde à previsão a um passo, usando o modelo estimado, para o ano hidrológico 1998/99.

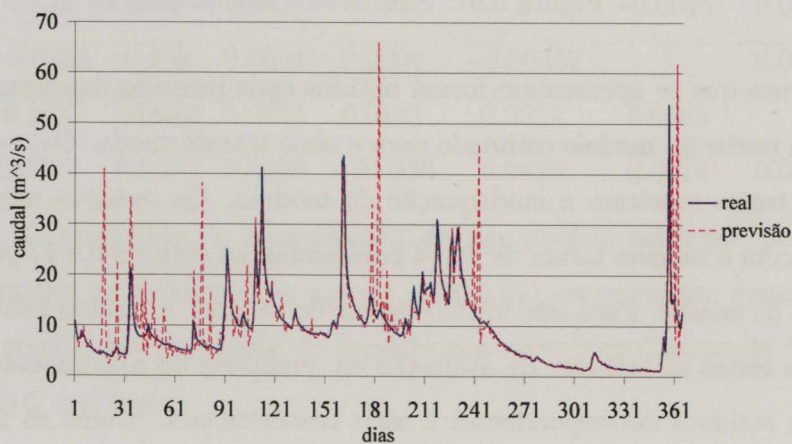


Figura 5.62: Previsão a um passo usando o modelo AR(15). Ano hidrológico 1998/99.

Capítulo 6

Conclusões

Os resultados que obtivemos mostram que os caudais médios diários do rio Paiva não constituem um sistema dinâmico determinístico caótico. Utilizamos a série dos caudais médios diários do Paiva para obter uma reconstrução da dinâmica através do método de Takens no sentido de estudar o todos os regimes de caudais e depois apenas o regime de descidas de caudais. Usamos, depois, o método dos vizinhos mais próximos para tentar prever a um passo no futuro. Os resíduos obtidos mostram um ganho significativo na qualidade de previsão quando apenas se usam as descidas de caudais (5.16, 5.18). Esta diferença de qualidade na previsão provém da acção da chuvas, que são imprevisíveis quando apenas se tem informação sobre os caudais diários e daí sentir-se a falta de determinismo nos dados. Fomos igualmente incapazes de obter estimativas para invariantes como a dimensão-correlação, o máximo expoente de Lyapunov e a entropia-correlação (figuras 5.23,5.40,5.41).

Os resultados de previsão obtidos para o regime de descidas de caudais revelam valores do resíduo médio quadrático e relativo próximos para espaços de reconstrução de diferentes dimensões, sendo que a dimensão 3 fornece os melhores resultados. Os resultados obtidos com dimensões 1, 2 e 4 são muito aproximados dos melhores resultados. De facto, o resíduo relativo médio obtido utilizando a dimensão 1 para o espaço de reconstrução foi de 4.8% que é óptimo no sentido de ter a mesma ordem de grandeza do desvio padrão das distribuições do decaimento (tabela 5.18, figura 5.37). Notamos que as energias da análise em componentes principais, realizada após a reconstrução de Takens, apontam na mesma direcção dos resultados anteriores.

Isto significa também que a informação adicional dos caudais nos dias anteriores cujos valores reflectem as variáveis não observadas, como a humidade dos solos, não melhoram significativamente a previsão. Este resultado não deriva do facto de o método dos vizinhos mais próximos não estar devidamente optimizado, mas sim, a uma limitação dos dados. Concluimos assim que, para as descidas de caudal, a informação mais determinante para prever o caudal do dia seguinte é o caudal do próprio dia.

Os resultados obtidos com a previsão usando a técnica das janelas de correlação forneceram resultados que vão de encontro ao que dissemos atrás. Os melhores resultados foram obtidos para as descidas de caudais onde se consegue obter alguma predictabilidade (figura 5.51). No entanto, os resíduos obtidos com esta técnica são piores do que os obtidos com o método dos vizinhos mais próximos. Para o mesmo conjunto de dados obteve-se um resíduo relativo médio de cerca de 62% enquanto que o resíduo médio quadrático foi de 6.7 utilizando como estimador o terceiro quartil do conjunto de valores extrapolados. Uma vez que os resultados obtidos pelo método dos vizinhos mais próximos, nas descidas de caudais, revelam ter um resíduo relativo próximo do desvio padrão das distribuições dos decaimentos, não se espera que a técnica das janelas de correlação possa vir a fornecer melhores resultados do que aquele método. O modelo ARIMA estimado revelou resultados que vão de encontro aos dois métodos anteriores, para o regime de descidas de caudais nos anos hidrológicos de 95/99, o resíduo quadrático médio foi de 9.17 e o resíduo relativo médio de 16% (figura 5.62).

Capítulo 7

Anexo

Neste anexo apresentamos os dados relativos aos caudais médios diários do rio Paiva medidos em Fragas da Torre relativos ao ano hidrológico de 1998/99, de 1 de Outubro de 1998 a 30 de Setembro de 1999.

10.95	4.39	25.5	15.03	11.35	11.05	11.26	31.07	10.04	3.57	3.87
8.39	7.96	20.31	13.45	11.02	10.29	10.6	26.81	9.66	3.29	4.71
7.23	21.46	17.93	12.03	10.68	9.73	10.23	21.36	9.33	3.1	4.82
7.48	19.95	16.34	11.31	10.4	12.29	9.84	18.86	8.78	2.87	4.2
8.93	12.09	15.39	10.58	10.1	12.78	9.41	17.42	8.26	2.75	3.32
8.06	9.73	14.23	10.02	9.74	42.76	9.06	16.35	7.85	2.66	3.04
7.17	8.77	13.29	9.85	9.42	43.72	8.72	15.42	7.41	2.53	2.81
6.55	8.12	12.72	11.58	11.67	25.5	8.52	14.47	6.89	2.42	2.7
6.05	7.79	12.1	12.5	13.3	20.31	8.37	25.79	6.87	2.24	2.4
5.7	7.81	11.62	11.03	11.38	17.93	8.1	28.55	6.53	2.17	2.27
5.49	7.73	11.23	10.12	10.56	16.34	8.23	21.46	6.21	2.23	2.16
5.28	7.99	10.83	9.6	10.12	15.39	9.26	26.02	6	2.24	2.01
4.9	8.6	10.48	9.44	9.88	14.23	11.58	29.47	5.76	2.19	1.91
4.57	10.25	11.25	9.23	9.5	13.29	10.51	27.38	5.52	2.31	1.85
4.45	8.8	13.05	12.35	9.13	12.72	9.5	22.22	5.22	2.18	1.77
4.29	8.02	16.24	29.41	8.92	12.1	9.58	19.8	5.12	2.01	1.74
4.58	7.72	15.17	20.53	8.87	11.62	14.18	18.25	4.98	2.08	1.76
4.83	7.54	13.23	17.42	8.7	11.23	17.88	17.1	4.52	2.01	1.66
4.44	7.3	12.23	27.13	8.51	10.83	14.7	15.73	4.5	1.97	1.59
4.23	7.02	11.63	41.4	8.48	10.48	12.97	14.62	4.29	1.86	1.59
4.17	6.81	13.06	26.16	8.48	11.25	12	13.9	4.21	1.92	1.59
3.96	6.47	13.28	20.5	8.35	13.05	15.94	13.44	4.28	1.93	1.57
3.76	6.37	12.28	18.11	8.45	16.24	20.78	13.06	4.33	1.83	1.49
4.27	6.14	11.71	16.59	8.82	15.17	16.61	12.2	4.22	1.75	1.46
4.62	6.22	11.26	15.67	8.48	13.23	16.99	11.76	3.86	1.72	1.46
5.74	6.25	10.6	17.29	8.46	12.23	17.68	11.16	3.9	1.57	1.46
4.84	5.9	10.23	14.99	8.47	11.63	17.55	10.78	3.31	1.46	1.54
4.51	5.79	9.84	13.97	8.02	13.06	18.76	10.81	3.25	1.46	1.52
4.32	5.78	9.41	13.21	7.74	13.28	16.4	11.03	3.64	1.56	1.49
4.23	5.77	9.06	12.51	8.39	12.28	15.59	10.75	4.16	1.63	1.41
4.23	5.45	8.72	11.69	10.36	11.71	21.81	10.24	3.77	1.98	1.6

Tabela 7.1: Caudais médios diários em m^3/s medidos em Fragas da Torre relativos ao ano hidrológico de 1998/99.



Referências

- [Aar96] Burton J. Dehling-H. Gilat D. Hil T. Weiss B. Aaronson, J. Strong laws for l- and u- statistics. *Transactions of the American Math. Society*, 348:2845–2865, 1996.
- [Aba96a] H. Abarbanel. *Analysis of Observed Chaotic Data*. Springer, 1996.
- [Aba96b] Lall U. Moon-Y. Mann M. Sangoyomi T. Abarbanel, H. Nonlinear dynamics and the great salt lake: A predictable indicator of regional climate. *Energy*, 21:655–665, 1996.
- [Agr93] Faloutsos C. Swami-A. Agrawal, R. Efficient similarity search in sequence databases. In *Foundations of data organization and Algorithms Proceedings*, pages 69–84, 1993.
- [Alb88] Muench J. Schwarz-C. Mees A. Rapp P. Albano, A. Singular value decomposition and the grassberger-procaccia algorithm. *Phys. Rev. A*, 38:3017, 1988.
- [All96] Sauer T. Yorke-J. Alligood, K. *Chaos - An introduction to dynamical systems*. Textbooks in Mathematical Sciences. Springer - Verlag, 1996.
- [AO01] C. Antunes and A. Oliveira. *Temporal data mining: an overview*. 2001.
- [AT93] Brown R.-Sidorowitch J. Abarbanel, H. D. and L. Tsimring. The analysis of observed chaotic data in physical systems. *R. of Mod. Phys.*, 65(4):1331–92, 1993.

- [Bar46] M. Bartlett. On the theoretical specification of sampling properties of autocorrelated time series. *J. Royal Stat. Soc. B*, 8:27–41, 1946.
- [Ber84] Pomeau Y. Vidal-C. Bergé, P. *L'ordre dans le Chaos*. Herman, 1984.
- ✓ [Ber96] Clifford J. Berndt, D. *Finding Patterns in Time Series: a dynamic programming approach*. Advances in knowledge discovery and Data Mining, 1996.
- [Bir96] Dodge Y. Birkes, D. *Alternatives Methods of Regression*. Wyley, 1996.
- ✓ [Bor98] S. Borovkova. *Estimation and Prediction for Nonlinear Time Series*. PhD thesis, Rijksuniversiteit Groningen, February 1998.
- [Bor99] Burton R. Dehling-H. Borovkova, S. Consistency of the takens estimator for the correlation dimension. *Ann. Appl. Probab.*, 9:376–390, 1999.
- [Box76a] Cox-D. Box., G. E. P. An analysis of transformations. *Journal of the Royal Statistical Society B*, 26:211–252, 1976.
- ✓ [Box76b] Jenkins G. Box, G. *Time Series Analysis*. Holden-Day, 1976.
- [Bri75] D. Brillinger. *Time Series, Data Analysis and Theory*. Holt, Rinheart and Winston, 1975.
- [Bro91a] Davis R. Brockwell, P. *Time Series: Theory and Methods*. Springer, 1991.
- [Bro91b] Bryant P. Abarbanel-H. Brown, R. Computing the lyapunov spectrum of a dynamical system from an observed time series. *Phys. Review A*, 43:2787–2806, 1991.
- [Bro94] King G. Broomhead, D. Extracting qualitative dynamics from experimental data. In Sauer T. York J. Ott, E., editor, *Coping with chaos*. Wyley Series, 1994.
- [Bur80] Ferral F. Burnash, J. *National Weather Service River Forecasting Manual User's Guide*. U. S. National Weather Service, 1980.

- [Cal01] Gonçalves R. Calheiros, F. Sampling in time series, 2001.
- [Cam80] M. Campanino. Two remarks on the computer study of differentiable dynamical systems. *Commun. Math. Phys.*, 74(15), 1980.
- ✓ [Cao97] L. Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*, 110:43–50, 1997.
- [Cas91a] M. Casdagli. Chaos and deterministic versus stochastic nonlinear modelling. *J. Royal Stat. Soc. B*, 54:303, 1991.
- [Cas91b] M. Casdagli. State space reconstruction in the presence of noise. *Physica D*, 51:52, 1991.
- [Cha86] Tong H. Chan, K. On tests of non-linearity in time series. *J. of Forecasting*, 5:217–228, 1986.
- [Cha01] Tong H. Chan, K. *Chaos: A Statistical Perspective*. Springer Series in Statistics. Springer, 2001.
- [Cut01] C. Cutler. Some results on the behavior and estimation of the fractal dimensions of distributions on attractors. *J. Statis. Physics*, 62, 2001.
- [Das97] Gonopoulos D. Mannila H. Das, G. Finding similar time series. In *Principles of Knowledge Discovery and Data-Mining*, 1997.
- [Dav80] H. David. *Order Statistics*. 1980.
- ✓ [Dev89] R. Devaney. *An Introduction to Chaotic Dynamical Systems*. Addison Wesley Publishers, 2nd edition, 1989.
- [Dik96] Zwet W. Takens F. DeGoede J. Diks, C. Detecting differences in vector delay distributions. *Phys. Rev. E*, 53:2169–2176, 1996.
- [Dik99] Cees Diks. *Nonlinear Time Series Analysis Methods and Applications*. Nonlinear Time Series and Chaos. World Scientific Publishing, 1999.

- [Doo48] Doob. *Stochastic Processes*. Wiley, 1948.
- [Eck81] J. Eckmann. Roads to turbulence in dissipative dynamical systems. *Rev. of modern Phys.*, 53:643, 1981.
- [Eck85] Ruelle D. Eckmann, J. Ergodic theory of chaos and strange attractors. *Reviews of modern Physics*, 57(3), July 1985.
- [Eck86] Oliffson Kamphorst Ruelle D. Ciliberto S. Eckmann, J. Lyapunov exponents form time series. *Phys. Rev. A*, 34:4971, 1986.
- [Efr79] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.
- ✓ [Far87] J. Farmer, J. Sidorowich. Predicting chaotic time series. *Phys. Review Letters*, 59(845), 1987.
- [Far88] Sidorowich J. Farmer, J. *Exploiting chaos to predict the future and reduce noise*. World Scientific, 1988.
- [Fei78] M. Feigenbaum. Quantitative universality for a class of nonlinear transformations. *J. of Statistical Phys.*, 19, 1978.
- [Fel71] W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 1971.
- [Fra86] Swinney H. Frazer, A. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33:1134–1140, 1986.
- [Fra92] Piatetsky-Shapiro G. MatheusC. Frawley, W. Knowledge discovery in databases: An overview. *AI magazine*, 13(3):57–70, 1992.
- [Gib92] Farmer-J. Casdagli M. Eubank S. Gibson, J. An analytic approach to practical state space reconstruction. *Physica D*, 57, 1992.

- [Gio01] Giornata di Studio: Metodi Statistici e Matematici per l'Analisi delle Serie Idrologiche. *Previsione di piene fluviali: confronto tra metodi non lineari*, 2001.
- [Gon96] R. Gonçalves. Análise linear e não linear da sucessão dos núcleos de oxigênio-18. Master's thesis, Faculdade de Ciências da Universidade do Porto, 1996.
- [Gon03] Lopes-N. Gonçalves, E. Séries temporais - modelações lineares e não lineares. XI Congresso Anual da Sociedade Portuguesa de Estatística, 2003.
- [GP] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica 9D*, pages 189–208.
- [Gra64] Hatanaka-M. Granger, C. *Spectral Analysis of economic time series*. Princeton University Press, 1964.
- [Gra83] Procaccia-I. Grassberger, P. On the characterization of strange attractors. *Phys. Review Letters*, 50:346–349, 1983.
- [Gra88] Badii-R. Politi A. Grassberger, P. Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors. *J. Stat. Physics*, 51(135), 1988.
- [Gre57] Rosenblatt-M. Grenander, U. *Analysis of stationary time series*. Wiley, 1957.
- [Heg99] Kantz-H. Schreiber T. Hegger, R. Practical implementation of nonlinear time series methods: the tisean package. *Chaos*, 9:413–35, 1999.
- [Hén76] M. Hénon. A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.*, 50:69, 1976.
- [Hen83] Procaccia-I. Hentschel, H. The infinite number of generalized dimensions of fractals and strange attractors. *Physica D*, 8:435–444, 1983.
- [Hin82] Melvin Hinich. Testing for gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3:169–176, 1982.

- [Hoe48] W. Hoeffding. A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.*, 19:293–325, 1948.
- [Hol86] Mayer-Kress G. Holzfuss, J. *An approach to error-estimation in the application of dimension algorithms*, pages 114–122. Springer, 1986.
- [Ily83] Y. Ilyashenko. On the dimension of attractors of k -contracting systems in an infinite dimensional space. *Vestn. Mosk. Univ. Ser. 1 Mat. Mekh.*, pages 52–58, 1983.
- [IM02] B. Islam. M, Sivakumar. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. *Advances in Water Resources*, 25:179–190, 2002.
- [Jay94] Lai F. Jayawardena, A. W. Analysis and prediction of chaos in rainfall and stream flow time series. *J. Hydrology*, 153:23–52., 1994.
- [Jay02] Li W. Xu P. Jayawardena, A. Neighbourhood selection for local modelling and prediction of hydrological time series. *Journal of Hydrology*, 258:40–57, 2002.
- [Jen61] G. Jenkins. *Spectral analysis of time series*. Wiley, 1961.
- ✓ [Kan94] H. Kantz. A robust method to estimate the maximal lyapunov exponent of a time series. *Phys. Lett. A*, 77:185, 1994.
- ✓ [Kan99] Schreiber T. Kantz, H. *Nonlinear Time Series Analysis*, volume 7. Cambridge Nonlinear Science Series, 1999.
- [Kar87a] Yakowitz S. Karlsson, M. Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. *Water Res. Research*, 23(7):1300–1308, July 1987.
- [Kar87b] Yakowitz S. Karlsson, M. Rainfall-runoff forecasting methods, old and new. *Stochastic Hydraulics*, pages 303–318, 1987.

- ✓ [Ken94] Brown R. Abarbanel H. Kennel, M. Determining embedding dimension for phase-space reconstruction using a geometrical construction. In Yorke Ott, Sauer, editor, *Coping with chaos*. Wyley, 1994.
- [Kol58] A. Kolmogorov. A new invariant of transitive dynamical systems. *Dokl. Akad. Nauk. SSSR*, 119:861, 1958.
- [Kra02] Gottschalk L. Krasovskaia, I. River flow regimes in a changing climate. *Hydrological Sciences*, 47(4):597–609, 2002.
- [Lae97] J. Laenen, A. Risley. Precipitation-runoff and streamflow-routing models for the willamette river basin. Water-resources investigations report 95-4284, U.S. Geological Survey, 1997.
- [Lan44] L. Landau. On the energy loss of fast particles by ionization. *Jour. Phys. USSR*, 8(201), 1944.
- [Led85] Young L. Ledrappier, F. The metric entropy of diffeomorphisms, parts i, ii. *Ann. Maths.*, 122:509, 1985.
- [LL98] Islam S. Rodriguez-Iturbe I. Liu, Q. and Y. Le. Phase-space analysis of daily streamflow: characterization and prediction. *Adv. water resources*, 21:463–475, 1998.
- [Lor63] E. Lorenz. Deterministic nonperiodic flow. *J. of Atmospheric Science*, 20:130–141, 1963.
- [Lor69] E. Lorenz. Atmospheric predictability as revealed by naturally occurring analogues. *Journal of Atmospheric Sciences*, 20:636–646, 1969.
- ✓ [Lu94] Zhan Qian Lu. *Estimating Lyapunov Exponents in Chaotic Time Series*. PhD thesis, University of North Carolina at Chapel Hill, 1994.
- [Man77] B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman and Company, 1977.

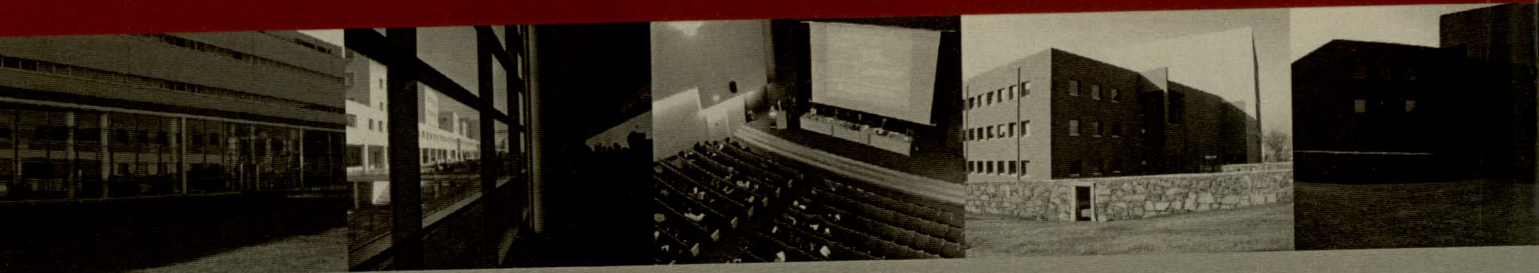
- [Mañ83] R. Mañé. *Introdução à teoria ergódica*,. Instituto de Matemática Pura e Aplicada, 1983.
- [Mei98] Silva J Calheiros-F. Meixedo, J. Data analysis in sedimentary. pages 105–110, 1998.
- [Mur90] B. Murteira. *Probabilidades e Estatística*, volume 2. 1990.
- ✓ [Mur93] Müller D. Turkman-K. Murteira, B. *Análise de sucessões cronológicas*. McGrawHill de Portugal, 1993.
- [Nic84] Nicolis G. Nicolis, C. Is there a climate attractor? *Nature*, 311:529–532, 1984.
- [Os89] Provenzale A. Osborne, A. Finite correlation dimension for stochastic systems with power-law spectra. *Physica D*, 35:357, 1989.
- [Ose68] V. Oseledec. A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Transactions of the Moscow Mathematical Society*, 19:197–221, 1968.
- [Ott93] E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, 1993.
- [Pac80] Crutchfield J. Farmer-J. Shaw R. Packard, N. Geometry from a time series. *Phys. Review Letters*, 45(9):712–716, 1980.
- [Pal94] Novotna D. Palus, M. Testing for nonlinearity on weather records. *Phys. Lett. A*, 193, 1994.
- [Pap97] Provenzale A. Smith-L. Paparella, F. Local random analogue prediction of nonlinear processes. *Physics Letters A*, 235:233–240, 1997.
- [Pes93] Y. Pesin. On rigorous mathematical definitions of correlation dimension and generalized spectrum for estimations. *J. Statistical phys.*, pages 529–547, 1993.

- [Pik86] A. Pikovski. Discrete-time dynamic noise filtering. *Sov. J. Commun. Technol. Electr.*, 31:81, 1986.
- [PM92] Smith L. Vio-R. Provenzale, A. and G. Murante. Distinguishing between low-dimensional dynamics and randomness in measure time series. *Physica D*, 58:31–49, 1992.
- [Pol93] M. Pollicott. *Lectures on Ergodic Theory and Pesin Theory on Compact Manifolds*. Lecture Note Series. Cambridge University Press, 1993.
- [Pom80] Manneville P. Pomeau, Y. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Mathematical Phys.*, 74, 1980.
- [Por97] Ridolfi L. Porporato, A. Nonlinear analysis of a river flow time sequences. *Water Resources Research*, 33:1353–1367., 1997.
- [Por01] Ridolfi L. Porporato, A. Multivariate nonlinear prediction of river flows. *Journal of Hydrology*, 248:109–122, 2001.
- [Pov98] Feng X. Povinelli, R. Temporal patterns identification in time series data using pattern wavelets and genetic algorithms. In *Artificial Neural Networks in Engineering*, 1998.
- [Pov99a] Feng X. Povinelli, R. Data mining of multiple nonstationary time series. Technical report, Marquette University, 1999.
- [Pov99b] R. Povinelli. *Time Series data-mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events*. PhD thesis, Marquette University, 1999.
- [PR96] A. Porporato and L Ridolfi. Clues to the existence of deterministic chaos in river flow. *International Journal of Modern Physics B*, 10(54):1821–1862, 1996.
- [Pri81] M. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.

- [Pri88] M. Priestley. *Non-linear and Non-Stationary Time Series*. Academic Press, 1988.
- [Pro91] Osborne A. Soj-R. Provenzale, A. Convergence of the k2 entropy for random noises with power law spectra. *Physica D*, 58:361, 1991.
- [Pug88] Shub M. Pugh, C. Ergodic attractors. *Trans. Am. Math. Society*, 308, 1988.
- [Puk94] F. Pukelsheim. The three sigma rule. *Am. Statist.*, 48:88–91, 1994.
- [Rao84] M. Rao, S. Gabr. *An Introduction to Bispectral Analysis and Bilinear Time Series*, volume 24 of *Lecture Notes in Statistics*. Springer, 1984.
- [Rén71] A. Rényi. *Probability Theory*. Noth-Holland, 1971.
- [Ros93] Collins S. De Luca Rosenstein, M. A practical method for calculating the largest lyapunov exponents from small data sets. *Physica D*, 65:117, 1993.
- [RT] D. Ruelle and F. Takens. On the nature of turbulence. *Commun. math. Physics*, 20:167–192.
- [Rue78] D. Ruelle. An inequality for the entropy of differentiable maps. *Bol. Soc. Bras. Mat.*, 9:83–87, 1978.
- [Rue81] D. Ruelle. Small random perturbations of dynamical systems and the definition of attractors. *Communications Math. Phys.*, 82:137–151, 1981.
- [Rue86] D. Ruelle. Resonances of chaotic dynamical systems. *Phys. Rev. Letters*, 56:405, 1986.
- [Rue89a] D. Ruelle. *Chaotic Evolution and Strange Attractors*. Cambridge University Press, 1989.
- [Rue89b] D. Ruelle. *Elements of differentiable dynamics and bifurcation theory*. Academic Press, 1989.

- [San85] Swada Y. Sano, M. Measurement of the lyapunov spectrum from a chaotic time series. *Phys. Rev. Lett.* 73, 1985.
- [Sap90] G. Saporta. *Probabilités Analyse des Données Et Statistique*. Éditions Technip, Rue Ginoux 27 75737 Paris, 1990.
- [Sau91] Yorke J. Casdagli M. Sauer, T. Embedology. *J. Stat. Physics*, 65:579–616, 1991.
- [Sau93] Yorke J. Sauer, T. How many delay coordinates do you need? *Int. J. Bif. and Chaos*, 3:737, 1993.
- [Sch88] H. Schuster. *Deterministic Chaos: An introduction*. Physik Verlag, 1988.
- [Sch98] T. Schreiber. Interdisciplinary application of nonlinear time series methods. Technical report, University of Wupperthal, 1998.
- [Ser80] R. Serfling. *Approximation theorems of mathematical statistics*. Wiley, 1980.
- [Sha48] C. Shannon. A mathematical theory of communications. *Bell Syst. Tech. Journal*, 27:379–423, 623–656, 1948.
- [Siv] B. Sivakumar. Chaos in hydrology: important issues and interpretations. *J. of Hydrology*, 227:1–20.
- [Sma67] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73:747, 1967.
- [Sug] May R. Sugihara, G. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734–41.
- [Tak80] F. Takens. *Detecting strange attractors in Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–81. Springer, 1980.
- [Tak96] F. Takens. Estimation of dimension and order of time series. In Van Gils I. Hoveijin Takens F. Broer, S, editor, *Nonlinear Dynamical Systems and Chaos*. Birkhäuser, 1996.

- [TE89] A. Tsonis and J. Elsner. Chaos, strange attractors and weather. *Bulletin of the American Metereological Society*, 70(1):14–24, 1989.
- [The86] J. Theiler. Spurious dimension from correlation algorithms applied to limited time series data. *Physical Review A*, 34:2427–2432, 1986.
- [The92] Eubank S Longtin A. Galdrikian B. Farmer J. Theiler, J. Testing nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.
- [Ton90] H. Tong. *Non-linear Time Series, A Dynamical System Approach*. Clarendon Press Oxford, 1990.
- [Tuc99] W. Tucker. *The Lorenz attractor exists*. PhD thesis, Dep. of Mathematics, Uppsala University, warwick@math.uu.se, 1999.
- [WA] J. B. Swinney H. Vastano J. Wolf A., Swift. Determining lyapunov exponents from a time series. *Physica D*, 16:285–317.
- [Whi36] H. Whitney. Differentiable manifolds. *Ann. Math.*, 37:645–680, 1936.
- [Wol38] H. Wold. A study in the analysis of stationary time series. Technical report, Univ. Uppsala, 1938.
- [Yak87] S. Yakowitz. Nearest-neighbour methods for time series analysis. *J. Time Series Analysis*, 2(8), 1987.
- [Yul27] G. Yule. On a method of investigating periodicities in disturbed series with special reference t wolfer’s sunspot numbers. *Philos. Trans. Royal Soc. London A*, 226:267–298, 1927.



FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

BIBLIOTECA



0000072322