



**SCHOOL OF  
ECONOMICS &  
MANAGEMENT  
LISBON**

**MESTRADO**  
MATEMÁTICA FINANCEIRA

**TRABALHO FINAL DE MESTRADO**  
DISSERTAÇÃO

SISTEMAS DINÂMICOS, ANÁLISE NUMÉRICA DE SÉRIES TEMPORAIS E  
APLICAÇÕES ÀS FINANÇAS

NUNO RICARDO MOURA FONTES

SETEMBRO - 2013



**SCHOOL OF  
ECONOMICS &  
MANAGEMENT  
LISBON**

**MESTRADO EM  
MATEMÁTICA FINANCEIRA**

**TRABALHO FINAL DE MESTRADO  
DISSERTAÇÃO**

**SISTEMAS DINÂMICOS, ANÁLISE NUMÉRICA DE SÉRIES TEMPORAIS E  
APLICAÇÕES ÀS FINANÇAS**

**NUNO RICARDO MOURA FONTES**

**ORIENTAÇÃO:**

**DOUTOR JOÃO LOPES DIAS**

**SETEMBRO - 2013**

## Abstract

Taken's theorem (1981) shows how the series of measurements from a given system can be used to reconstruct the original system's underlying dynamic process. In this work we start from this point and build a bridge between theoretical results and its practical application. Several algorithms are presented and then rebuilt in an effort to reach a middle ground between computer resources optimization and output accuracy. Among these algorithms, the biggest emphasis is put on the correlation dimension algorithm by Grassberger and Procaccia which allows for the deduction of the system's embedding dimension. The results derived are then used to build a forecast approach inspired by the analogues method. The purpose of this work is to show there is potential for dynamical systems' modelling tools to be used in financial markets, especially for intra-day purposes where decision and computational times need to be very small.

**Keywords:** *Dynamical Systems, Taken's Theorem, Correlation Dimension, Correlation Integral, Embedding Dimension, Analogues Method, Forecast, Financial Series*

## Resumo

O teorema de Takens (1981) mostra como uma série de medições obtidas de um dado sistema podem ser usadas para reconstruir o sistema dinâmico original. Neste trabalho, parte-se deste teorema e constrói-se a ponte entre conceitos teóricos e a sua aplicação numérica. Vários algoritmos são apresentados e depois reconstruídos com o objetivo de se atingir um compromisso entre otimização de recursos computacionais e rigor nos resultados. Entre esses algoritmos, a maior ênfase é colocada no do cálculo do integral de correlação de Grassberger-Procaccia que permite a dedução da dimensão de imersão de um dado sistema. Os resultados obtidos são usados na construção de um modelo de previsão inspirado pela abordagem dos pontos análogos, ou método dos análogos. O objetivo deste trabalho é mostrar que existe potencial na aplicação de ferramentas de modelação de sistemas dinâmicos caóticos no mercado financeiro, em especial em transações intra-diárias onde tempos de decisão e computação têm de ser muito reduzidos.

**Palavras-Chave:** *Sistemas Dinâmicos, Teorema de Takens, Dimensão de Correlação, Integral de Correlação, Dimensão de Imersão, Método dos Análogos, Previsão, Séries Financeiras*

## Agradecimentos

Ao meu orientador, o professor João Lopes Dias, por me dado a oportunidade de desenvolver a tese num tema tão interessante e ter aberto alternativas e novos caminhos nos momentos chave.

Ao Emanuel e ao Cristiano pelo apoio tão valioso ao longo do mestrado, ao Estrela pelo olhar *clínico* que emprestou à tese e pelo apoio, agradecimento que se estende ao Abel e ao Quintela.

Por fim, registar o agradecimento à minha Mãe, ao Meu Pai e à minha Mana e em particular ao Avô Fontes e à Avó Luísa e ao Avô Jorge e à Avó Marília e também ao meu Tio Paulo, sem os quais esta etapa académica não teria sequer começado.

*Beware the irrational, however seductive;  
Shun the 'transcendent' and all who invite you to subordinate or annihilate yourself;  
Distrust compassion; prefer dignity for yourself and others;  
Don't be afraid to be thought arrogant or selfish;  
Picture all experts as if they were mammals;  
Never be a spectator of unfairness or stupidity;  
Suspect your own motives, and all excuses.*

- Christopher Hitchens



# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Sistemas dinâmicos . . . . .	2
1.2	Teorema de Takens . . . . .	3
1.3	Sistemas Caóticos . . . . .	4
<b>2</b>	<b>Caracterização Qualitativa</b>	<b>6</b>
<b>3</b>	<b>Propriedades Quantitativas</b>	<b>9</b>
3.1	Dimensão Fractal; Algoritmo <i>Box-Counting</i> . . . . .	9
3.2	Dimensão de Imersão: Abordagem pelo Integral de Correlação . . . . .	11
3.2.1	Aplicação do Integral na Obtenção da Dimensão de Imersão . . . . .	14
3.2.2	Um Algoritmo Alternativo . . . . .	16
3.2.3	O Problema da Escolha da Distância Entre Pontos . . . . .	20
3.3	Expoente de Lyapunov . . . . .	23
<b>4</b>	<b>Previsão</b>	<b>25</b>
<b>5</b>	<b>Aplicação às Finanças</b>	<b>27</b>
5.1	Aplicação da metodologia sugerida . . . . .	28
5.2	Aplicação do Método dos Análogos . . . . .	31
5.3	Outros Ativos . . . . .	33
<b>6</b>	<b>Conclusão</b>	<b>35</b>
	<b>Anexo A Análise Gráfica e Analítica</b>	<b>39</b>
	<b>Anexo B Código Matlab</b>	<b>41</b>

# Lista de Figuras

1.1	Atrator de Lorenz . . . . .	2
1.2	Ilustração do teorema de Takens . . . . .	4
2.1	Evolução temporal das séries . . . . .	7
2.2	Representação das recorrências . . . . .	8
3.1	Relação entre caixas necessárias e respetiva área . . . . .	10
3.2	Dimensão de correlação   Série D . . . . .	14
3.3	Comparação entre algoritmos   Série A . . . . .	17
3.4	Comparação entre algoritmos   Série B . . . . .	18
3.5	Comparação entre algoritmos   Série C . . . . .	19
3.6	Problemas de vizinhança . . . . .	21
3.7	Dimensão de correlação em função de $k$ e $\epsilon$ . . . . .	22
3.8	Distância entre vizinhos . . . . .	24
4.1	Pontos análogos . . . . .	26
5.1	Séries temporais financeiras . . . . .	27
5.2	Dimensão <i>Box-Counting</i> vs $\epsilon$ . . . . .	29
5.3	Obrigações do tesouro . . . . .	29
5.4	Taxas de câmbio . . . . .	30
A.1	Visualização 3D . . . . .	39
A.2	Dimensão de correlação   Série D . . . . .	39
A.3	Autocorrelação e correlação parcial. . . . .	40



## Lista de Tabelas

3.1	Dimensão <i>Box-Counting</i> . . . . .	11
3.2	Comparação entre algoritmos . . . . .	20
3.3	Expoente de Lyapunov . . . . .	24
5.1	Séries financeiras recolhidas . . . . .	27
5.2	Dimensão <i>Box-Counting</i> . . . . .	29
5.3	Expoentes de Lyapunov   Séries financeiras . . . . .	30
5.4	Retornos obtidos através das diferentes estratégias . . . . .	32
5.5	Dimensão de correlação   Acções . . . . .	34
A.1	Algoritmo $\tilde{C}$   Série D . . . . .	39
A.2	Estatísticas das séries recolhidas . . . . .	40



---

# 1 Introdução

Uma série temporal é uma coleção de observações feitas de forma sequencial. Pode ser o número de manchas solares a cada minuto, o produto interno bruto de um país a cada ano, a temperatura média diária numa dada cidade ou o preço de um ativo registado ao dia, minuto ou segundo ou apenas uma sucessão de experiências de origem aleatória. A partir do momento em que temos uma série temporal gerada pelo processo ou modelo que desejamos estudar, seja ele um sistema solar, atmosférico, a economia ou um dado mercado financeiro, o objetivo passa primeiro por caracterizar e compreender tanto quanto possível o sistema que gerou a série e, numa segunda fase, prever futuros estados do mesmo.

Existem duas categorias de séries que interessa distinguir logo à partida: séries temporais que são apenas observações de variáveis aleatórias independentes e identicamente distribuídas; e séries temporais determinísticas em que, mesmo no caso de pouco ou nada sabermos sobre a mecânica que as gera, sabemos que os seus valores resultam de uma conjugação de variáveis que aspiramos determinar. Como vamos poder observar mais à frente, a linha que separa séries aleatórias de séries determinísticas pode não ser fácil de traçar.

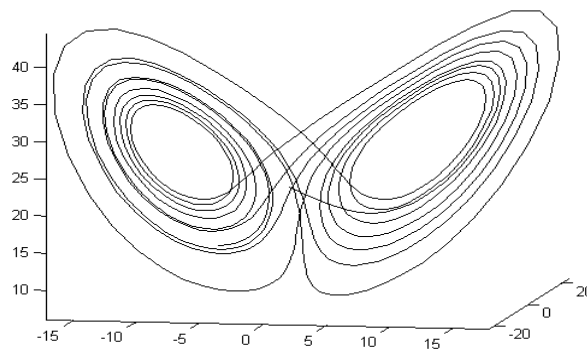
No caso em que estamos perante uma série temporal de origem determinística, ou seja, uma série em que os elementos não são mais do que medições feitas a um sistema dinâmico, existem várias ferramentas de análise que podem ser aplicadas na sua caracterização. Tipicamente recorrem-se a métodos lineares na modelação e previsão destas séries temporais. Neste trabalho tentar-se-à aprofundar a implementação de métodos numéricos de conceitos teóricos e metodologias desenvolvidas na área de estudo de sistemas dinâmicos não-lineares e, em particular, caóticos [1, 2, 3, 9, 15], com vista à sua aplicação aos mercados financeiros.

---

## 1.1 Sistemas dinâmicos

Um sistema dinâmico é um par  $(\Psi, T)$  onde  $\Psi$  é o espaço de fases, ou seja, o conjunto de todos os possíveis valores do sistema e  $T$  a regra de transição de estados tal que  $T: \Psi \rightarrow \Psi$ . Um sistema dinâmico pode ser linear ou não-linear, sendo não-linear se  $T(c_1X_1 + c_2X_2) \neq c_1T(X_1) + c_2T(X_2)$ .

Associado a um sistema dinâmico está um atrator. Um atrator é uma região contida no espaço de fases para a qual o sistema converge de tal forma que a partir de um certo momento, o sistema percorrerá infinitamente todos os pontos dessa região sem que a volte a abandonar. A geometria de um atrator está fortemente associada a características estruturais do sistema subjacente e pode variar desde um conjunto de pontos finitos, no caso de sistemas mais simples, até objetos de dimensão fractal, como é o caso de atratores da maioria dos sistemas caóticos. Na figura 1.1 encontra-se representada parte da trajetória de um sistema já depois da sua entrada no atrator.



**Figura 1.1:** Atrator de Lorenz

O sistema em causa é conhecido como Atrator de Lorenz e é um sistema dinâmico que revela comportamento caótico para uma certa conjugação de parâmetros, tendo portanto, um atrator de dimensão fractal. O objetivo deste trabalho passa por conseguir definir tanto quanto possível sistemas como o apresentado na figura 1.1 tendo apenas por base medições aplicadas sobre, pelo menos, uma variável do processo.

---

Séries temporais geradas por modelos dinâmicos podem ser difíceis de distinguir de séries aleatórias numa primeira análise e são modeladas ineficientemente usando métodos convencionais como modelos auto-regressivos lineares simples como o modelo ARMA, ainda bastante utilizado por analistas financeiros.

## 1.2 Teorema de Takens

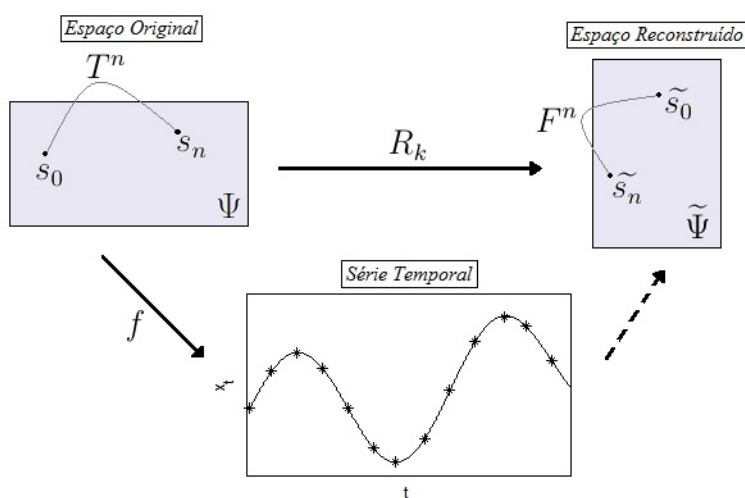
Se pensarmos em alguns dos sistemas enumerados no início, como por exemplo a economia ou o mercado financeiro, compreendemos que aquilo a que temos acesso não é mais do que medições em forma de índice ou preço. Por exemplo, o sistema que gera o movimento de uma taxa de câmbio envolve variáveis económicas, políticas e comportamentais como a taxa de juro dos respetivos bancos centrais, estabilidade política e expectativas dos investidores. A influência de todas essas variáveis sobre o sistema resulta numa série temporal unidimensional: a cotação da moeda.

A abordagem ao estudo de séries temporais neste trabalho terá como base o Teorema de Takens [14], que prova a existência de condições, explicitando-as, segundo as quais um sistema dinâmico pode ser reconstruído tendo apenas por base uma sequência de observações geradas por esse mesmo sistema. Este teorema permite construir novo um espaço de fases que replica o espaço de fases original. Este novo espaço designar-se-à por espaço reconstruído devido a apresentar apenas uma suave mudança de coordenadas em relação ao espaço original preservando as características essenciais.

Para reconstruirmos um dado espaço teremos de encontrar um mergulho desse mesmo espaço. Diz-se que  $B$  é um mergulho de  $A$  quando existe uma aplicação  $g$  injetiva e diferenciável (classe  $C^\infty$  ou *suave*) tal que  $g: A \rightarrow B$ . Por ser diferenciável garante-se uma transformação que preserva as características estruturais do sistema original. Desta forma sabemos que o espaço original pode ser estudado analisando apenas o seu mergulho.

**Teorema 1 (Takens [14])** *Seja  $(\Psi, T)$  um sistema dinâmico com espaço de estados  $\Psi \in \mathbb{R}^d$  de órbitas limitadas  $\{T^n s\}_{n \geq 0}$  e seja  $f: \Psi \rightarrow \mathbb{R}$ . Então, sempre que  $k > 2d$ , existe um subconjunto aberto e denso  $U$ , tal que se  $(T, f) \in U$ , o mapa de reconstrução  $R_k(s) = (f(s), f(Ts), \dots, f(T^{k-1}s)) \in \mathbb{R}^k$  define um mergulho.*

O que teorema de Takens acima anunciado garante é que através de medições  $x = f(s)$  efetuadas no espaço original  $\Psi \in \mathbb{R}^d$ , se consegue construir um espaço  $\tilde{\Psi} \in \mathbb{R}^k$  de tal forma que  $\tilde{\Psi}$  é um mergulho de  $\Psi$ .



**Figura 1.2:** Ilustração do teorema de Takens

Provada a existência do mapa  $R_k$ , e a partir das medições que resultam na série temporal  $(x_1, x_2, \dots, x_n)$ , tentaremos reconstruir o espaço original de modo a tirar conclusões relativas á características do seu atrator e prever futuros estados do sistema. Referir-nos-emos à dimensão  $k$ , a dimensão do espaço reconstruído, por dimensão de imersão.

### 1.3 Sistemas Caóticos

Sistemas caóticos são um sub-conjunto de sistemas dinâmicos não-lineares. São sistemas caracterizados por um comportamento de tal forma errático que as suas observações são facilmente confundidas com séries aleatórias. Vamos centrar atenções

---

neste sub-conjunto dos sistemas dinâmicos por ser aquele que apresenta um maior desafio em termos de análise numérica e também por ser aquele que mais se assemelha ao comportamento de séries obtidas em mercados financeiros.

Um sistema diz-se caótico se apresentar aperiodicidade e alta sensibilidade às condições iniciais. Uma forma de medir essa sensibilidade é através da estimação do expoente de Lyapunov. Dados dois pontos arbitrariamente próximos, cuja distância  $\delta$  é modelada ao longo das suas órbitas por uma função do tipo  $\delta_n = e^{\lambda n}$ , define-se  $\lambda$  como o expoente de Lyapunov. Se  $\lambda \leq 0$ , o sistema diz-se estável podendo ou não convergir para um ponto no espaço de fases. No caso em que  $\lambda > 0$ , estamos perante órbitas que exibem um comportamento instável e caótico. Significa isto que, independentemente da proximidade entre dois pontos no instante inicial a sua distância aumentará de forma exponencial, de tal modo que apesar de o sistema ser determinístico torna-se muito complicado prever futuros estados a partir da observação corrente.

Ao longo deste trabalho vamos enunciar conceitos teóricos de análise de sistemas dinâmicos cujo expoente seja positivo e desenvolver métodos que permitam a aplicação de algoritmos. O objetivo é conseguir fazer a ponte entre teoria e prática de forma rigorosa com vista à otimização do tempo de computação e do consumo recursos computacionais como memória virtual e processamento.

No Capítulo 2 faz-se uma primeira análise a séries temporais geradas por sistemas dinâmicos não lineares onde é focada a sua aparente aleatoriedade e é feita uma primeira abordagem com vista à distinção entre séries origem determinística e séries de origem aleatória. No Capítulo 3 introduzem-se métodos analíticos que permitem quantificar a complexidade do sistema onde se exploram os conceitos de dimensão fractal, dimensão de correlação, dimensão de imersão e expoente de Lyapunov. Neste capítulo são desenvolvidas novas abordagens computacionais que possibilitam a aplicação algoritmos complexos num tempo de computação bastante inferior. O Capítulo 4 é dedicado à previsão. É apresentado um método de pre-

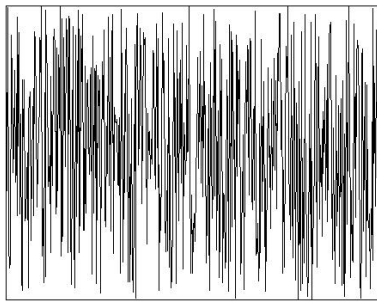
visão derivado do método dos análogos, também parcimonioso e leve em recursos computacionais. Por fim, no Capítulo 5, todos os procedimentos apresentados e desenvolvidos são aplicados. São criados portfólios com ativos financeiros do mercado de dívida e do mercado cambial que são geridos com base em estratégias de previsão de retornos. O desempenho do método desenvolvido no Capítulo 4 e que assenta em toda a análise desenvolvida nos capítulos anteriores é comparado com um modelo regressivo simples mas bastante popular entre as ferramentas de análise técnica.

## 2 Caracterização Qualitativa

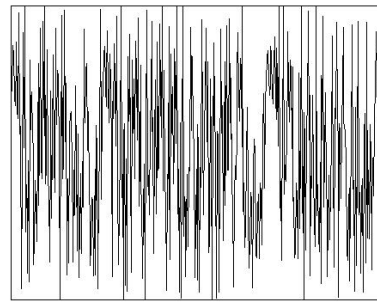
Dada a proximidade de comportamento entre séries geradas por certos modelos dinâmicos e séries de origem aleatória, torna-se antes de mais essencial distinguir entre uma sucessão de números arbitrária e uma sucessão gerada por modelos determinísticos. O primeiro passo consiste em observar representações gráficas das séries em estudo. Esta primeira abordagem, apesar de ingénua, poderá ajudar a delinear o caminho a seguir no estudo analítico mas não permitirá nunca tirar conclusões definitivas. Por exemplo, se numa situação de existência de padrões demasiado evidentes a hipótese de aleatoriedade pode ser afastada com relativa segurança, o contrario é absolutamente errado como vamos ver de seguida. Abaixo estão representadas séries temporais geradas por quatro sistemas diferentes. Dois são determinísticos não-lineares, um é determinístico linear com ruído e por fim temos uma sucessão de observações de uma variável aleatória. Fazendo o exercício de ligação entre as sequências descritas acima com as suas representações gráficas (figura 2.1) tornam-se evidentes as limitações desta abordagem.



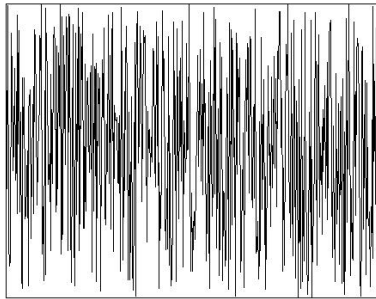
Figura 2.1: Evolução temporal das séries



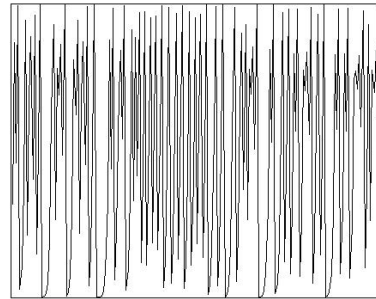
(a) Série A



(b) Série B



(c) Série C

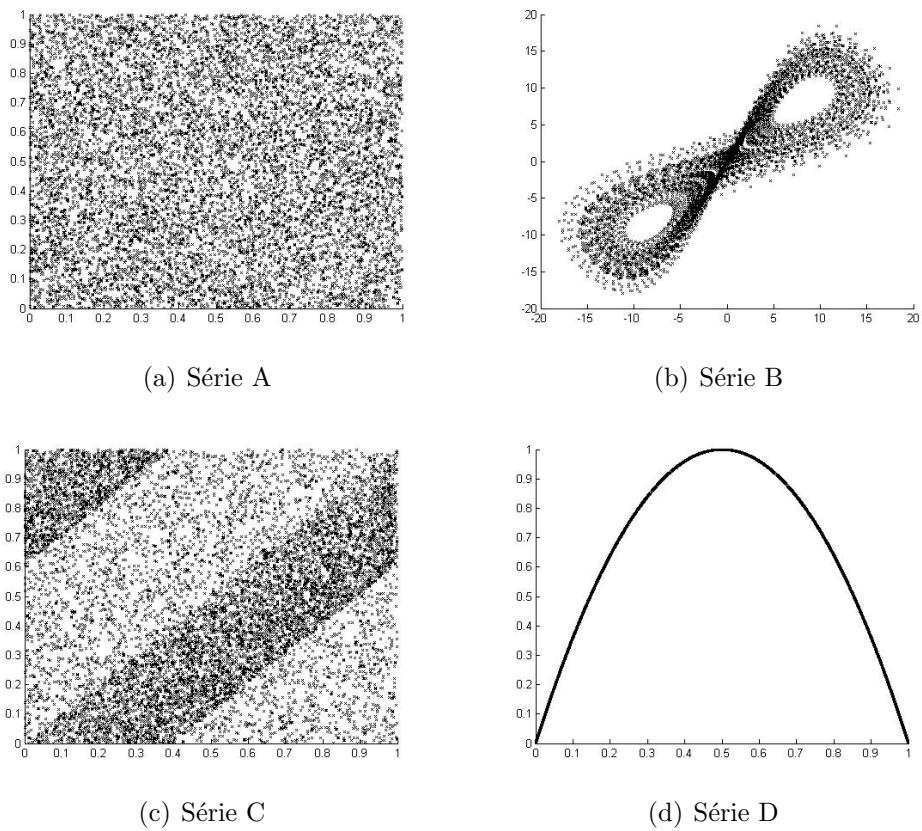


(d) Série D

Apesar de a figura 2.1 não parecer conter mais do que 4 visualizações da mesma série ou de 4 séries distintas mas aleatórias, os sistemas na origem das representações acima são bastante distintos e 2 dos sistemas em causa são mesmo determinísticos sem ruído. No entanto, apesar de uma observação gráfica simples não permitir dar qualquer passo em frente, existe algum potencial na análise gráfica que é possível explorar. Por exemplo, desenhando o gráfico de recorrências, isto é, desenhando os pares de pontos  $(x_t, x_{t-l})$  da partir da série recolhida  $(x_1, x_2, \dots, x_n)$ , em que  $l$  é o desfasamento temporal entre observações<sup>1</sup> (figura 3.5).

<sup>1</sup>A representação pode ser estendida a 3-dimensões. No caso das séries em estudo existem duas que produzem visualizações bastante interessantes, ver Anexo A.

Figura 2.2: Representação das recorrências



Com esta simples abordagem a natureza das séries fica mais clara, sendo já possível responder com sucesso ao desafio lançado, de associar a descrição dos modelos com as observações registradas. Torna-se de certa forma evidente que os sistemas B e D são determinísticos, C é determinístico com ruído enquanto a sequência A é aleatória.

Os sistemas usados para gerar os processos  $\{X_t\}_{t \in \mathbb{N}}$  mencionados acima serão usados ao longo deste trabalho, pelo que se impõe uma apresentação mais formal:

**Processo A** - Sucessão de variáveis aleatórias uniformemente distribuídas:

$$X_t \sim U[0, 1] \quad (2.1)$$

**Processo B** - Sucessão de observações feitas a partir da projeção em  $x(t)$ , do sistema gerado pelo Atrator de Lorenz [11]:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z, \end{aligned} \tag{2.2}$$

com  $\sigma = 10$ ,  $\rho = 28$  e  $\beta = 8/3$ . Foi usado um salto temporal  $\Delta t = 0.03$  na recolha das observações.

**Processo C** - Sucessão obtida de um sistema que alterna uma componente linear auto-regressiva com ruído e outra aleatória:

$$X_t = \begin{cases} (X_{t-1} + e^{-1/Y_t}) \bmod 1 & \text{se } Y_t < 3/4 \\ Z_t & \text{c.c.} \end{cases} \tag{2.3}$$

onde  $\{Y_t\}_{t \in \mathbb{N}}$  e  $\{Z_t\}_{t \in \mathbb{N}}$  são v.a.i.i.d. com distribuição  $U(0, 1)$  e  $X_0 \in [0, 1]$ .

**Processo D** - Sucessão obtida de um processo conhecido como Mapa Logístico:

$$X_t = \mu(1 - X_{t-1})X_{t-1}, \tag{2.4}$$

onde  $\mu = 4$  e  $X_0 \in ]0, 1[$ .

## 3 Propriedades Quantitativas

### 3.1 Dimensão Fractal; Algoritmo *Box-Counting*

A dimensão fractal de um conjunto [5, 12], é uma medida da sua forma geométrica e pode ser vista como uma abstração da dimensão Euclidiana já que os seus valores podem não pertencer a  $\mathbb{N}$ . Como já foi referido, tipicamente atratores caóticos têm dimensão fractal e esta é interpretada como uma contagem dos seus graus de

liberdade, ou seja, dos fatores que determinam a sua evolução. Pode ser feita uma analogia com uma partícula que num espaço Cartesiano é representada pelas suas 3 coordenadas e portanto tem 3 graus de liberdade.

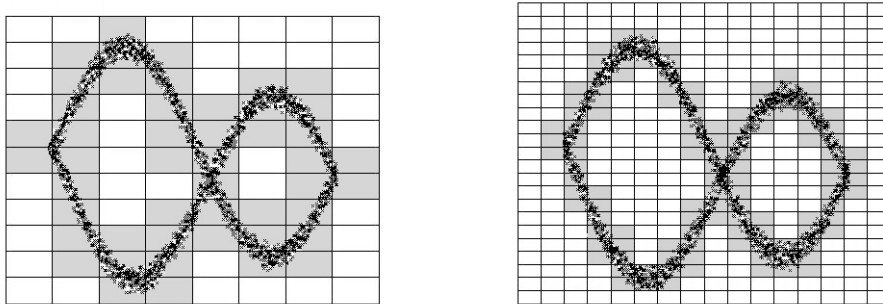
A dimensão de *Box-Counting* [8] é uma forma de aproximar a dimensão fractal de um atrator e é obtida através do algoritmo *Box-Counting*. O método devolve um valor associado à quantidade de caixas de lado  $\epsilon$ ,  $N_b(\epsilon)$ , necessárias para cobrir o conjunto de pontos (ver figura 3.1). Ou seja, obtém-se uma noção da complexidade do conjunto, da sua geometria, mas não da sua *densidade* já que não é feita nenhuma distinção entre uma caixa que contenha um ponto de outra que englobe centenas.

Definimos a dimensão *Box-Counting* de um dado conjunto  $A$ ,  $\dim_B(A)$ , da seguinte forma:

$$\dim_B(A) = \lim_{\epsilon \rightarrow 0} \frac{\log(N_b(\epsilon))}{\log(1/\epsilon)}, \quad (3.1)$$

À medida que a área de caixa ( $\epsilon^2$ ) diminui, tanto o número de caixas necessárias para conter o conjunto como  $\log(1/\epsilon)$  vão aumentar. Quando a série não é aleatória, ou seja, provém de um sistema determinístico, é de esperar que a dimensão *Box-Counting* convirja para um número inferior àquele obtido quando o algoritmo é aplicado a séries aleatórias ou com comportamento caótico.

**Figura 3.1:** Relação entre caixas necessárias e respetiva área



Embora esta medida tenha as limitações descritas acima, a sua implementação prática (através da função '*BoxCount.m*', ver anexo B) é bastante simples e não levanta quaisquer problemas a nível computacional. Importa, portanto, ficar com

esta abordagem em mente pois será útil na próxima secção quando for apresentado um novo algoritmo mais rápido e leve em recursos computacionais, com o objetivo de determinar a dimensão de imersão de um dado conjunto (capítulo 3.2). Por agora, vamos calcular a dimensão *Box-Counting* para as séries introduzidas previamente:

**Tabela 3.1:** Dimensão *Box-Counting*

	$\epsilon = 10^{-1}$	$\epsilon = 10^{-2}$	Dim.Frac.
Série A	2.01	1.89	-
Série B	1.53	1.52	2.05 [6]
Série C	1.98	1.85	-
Série D	1.49	1.22	0.54 [7]

Apesar de as medições obtidas se encontrarem um pouco longe dos valor conhecidos, para as séries em que esses valores foram calculados, os resultados obtidos permitem organizar as séries por ordem crescente de complexidade da seguinte forma: D - B - C - A. Esta conclusão vem em linha com as suposições feitas pela análise gráfica da figura 3.5. No entanto, a discrepância entre o valor devolvido por este algoritmo e o valor das dimensões dos atratores em estudo deixa evidente a necessidade de encontrar um método mais assertivo.

### 3.2 Dimensão de Imersão: Abordagem pelo Integral de Correlação

A dimensão de correlação,  $\alpha$ , de um dado conjunto é também uma medida da dimensão fractal do mesmo. Ao contrário da dimensão *Box-Counting* apresentada na secção anterior, a dimensão de correlação distingue entre um espaço de acumulação de pontos densamente populado de outro espaço também preenchido mas com uma menor quantidade de elementos. Esta medida permite portanto quantificar de forma mais completa o grau de complexidade do sistema em estudo, mas mais importante, esta medida é essencial na dedução do mergulho ótimo de um dado sistema. A dimensão de correlação surge como aproximação da dimensão  $d$  introduzida no Teo-

rema de Takens. Ou seja, conseguiremos reconstruir o sistema original num espaço de dimensão  $k$  da mesma ordem que  $2\alpha + 1$ .

Torna-se do maior interesse determinar a dimensão de imersão com um elevado grau de fiabilidade dado que esse valor permitirá construir uma nova série temporal, em que cada ponto será um vetor de reconstrução, que replique o sistema original sem conter demasiada informação. De facto, estamos numa situação em que informação inútil se torna parte do problema por não ser parte da solução, senão vejamos: para um conjunto que tenha dimensão de imersão ótima 4 e vetores de reconstrução,  $V_i^4 = (x_i, x_{i+1}, x_{i+2}, x_{i+4})$ , os vetores  $V_i^5 = (x_i, x_{i+1}, x_{i+2}, x_{i+4}, x_{i+5})$  contêm toda a informação de  $V_i^4$ . Ou seja, o conjunto dos pontos que compõe  $V_i^4$  está contido no conjunto dos pontos que compõe  $V_i^5$  para um mesmo  $i$ . No entanto, a informação adicional, por ser desnecessária, funciona como ruído que afeta os resultados obtidos por algoritmos aplicados posteriormente, tal como a estimação do expoente de Lyapunov ou métodos de previsão, não mencionando os elevados custos computacionais que daí advêm.

O integral de correlação de um conjunto, tal como introduzido por Procaccia e Grassberger [6], é uma medida essencial no cálculo da estimativa numérica da dimensão de correlação. O integral de correlação pode ser usado na dedução da dimensão  $k^*$  de duas formas. A primeira consiste em aproximar a dimensão de correlação, através do cálculo integral de correlação usando vetores de reconstrução cuja dimensão vai sendo incrementada até se atingir o ponto de *saturação* de informação. O menor valor  $k$  para o qual o valor de  $\alpha^{(k)}$ , a dimensão de correlação, estabiliza ou tem um maior abrandamento de crescimento será o valor que usaremos como o valor ótimo para análise e previsão. A segunda forma consiste em repetir o procedimento anterior e registar o valor de  $\alpha$  no qual a evolução estabiliza calculando-se então  $k^* = 2\alpha + 1$ . Em teoria os dois métodos produziriam o mesmo resultado e é isso que se verifica na grande maioria das aplicações, no entanto, pode existir uma

discrepância entre as duas indicações como vamos ver mais adiante.

Antes de partir para aplicações interessa definir de forma formal o integral de correlação, também conhecido por algoritmo Grassberg-Procaccia.

Seja a norma:

$$\|X\|_{\infty} = \max(|x_1|, |x_2|, \dots, |x_n|)$$

,

e a função de *Heaviside*:

$$H(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Seja  $X_n^k = (x_n, x_{n+1}, \dots, x_{n+k-1})$  um vetor de reconstrução de dimensão  $k$ , e  $N$  o número de pontos no espaço de fases reconstruído, então define-se  $C^{(k)}$  o integral de correlação de um conjunto imerso numa dimensão  $\mathbb{D}^k \subset \mathbb{R}^k$ :

$$C^{(k)}(A) = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} C_{N,\epsilon}^{(k)}(A)$$

onde

$$C_{N,\epsilon}^{(k)}(A) = \frac{1}{N(N-1)} \sum_{i=0}^N \sum_{j=0}^N H(\epsilon - \|X_i^k - X_j^k\|_{\infty}), \quad (3.2)$$

com  $i \neq j$ .

Fundamentalmente, é o número de pontos que distam entre si  $\epsilon$  ou menos dividido pelo número total de combinações de pares de pontos,  $N(N-1)$ , no conjunto. Ou seja, o integral de correlação é a probabilidade de dois pontos tomados ao acaso estarem a uma distância inferior a  $\epsilon$  um do outro. Em termos práticos, estes limites poderão apenas ser aproximados já que tanto um conjunto infinito de pontos como uma distância infinitamente pequena são requisitos incompatíveis com análise numérica. Iremos portanto trabalhar com o máximo número de pontos que for possível agregar e a distância mínima para a qual os algoritmos numéricos

conseguem correr *suavemente*. De facto, a escolha do  $\epsilon$  ótimo é uma questão importantíssima na computação do integral de correlação e no final do corrente capítulo é feita uma pequena discussão sobre a influência que a escolha do  $\epsilon$  tem sobre os resultados obtidos.

É expectável que o integral de correlação tenha uma evolução exponencial do tipo  $C_\epsilon^{(k)} \sim \epsilon^{\alpha^{(k)}}$  para  $\epsilon$  numa vizinhança de zero. Nesse caso,  $\alpha^{(k)}$  será a dimensão de correlação do conjunto em análise. Seja  $\alpha^{(k)}(A)$  a dimensão de correlação, tal que

$$\alpha^{(k)} = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \alpha_{N,\epsilon}^{(k)}(A), \quad (3.3)$$

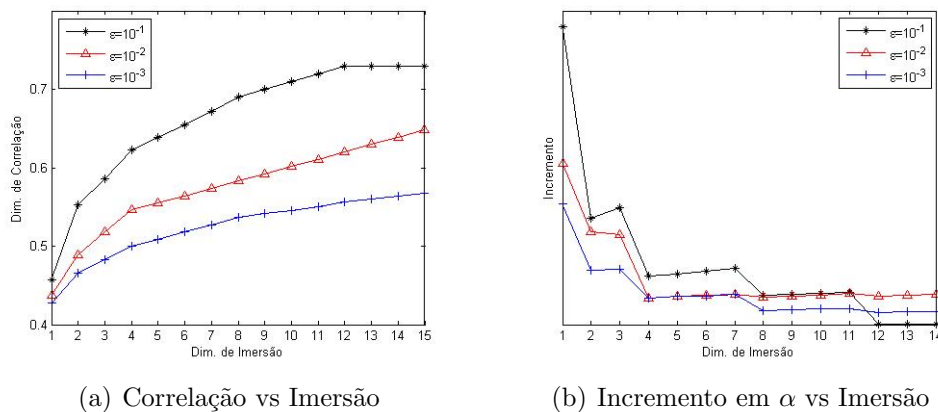
onde

$$\alpha_{N,\epsilon}^{(k)}(A) = \frac{\log(C_{N,\epsilon}^{(k)}(A))}{\log(\epsilon)}.$$

### 3.2.1 Aplicação do Integral na Obtenção da Dimensão de Imersão

De seguida aplica-se o algoritmo apresentado atrás (função '*DimCorrOld.m*', ver anexo B) à série D e exemplifica-se como se obtêm as respostas pretendidas através dos seus resultados.

**Figura 3.2:** Dimensão de correlação | Série D



À medida que a distância entre pontos  $\epsilon$  diminui e se aumenta a dimensão de imersão, a dimensão de correlação  $\alpha$  estabiliza e tende a aproximar-se do seu valor teórico de 0.5 [9]. Recordando o teorema de Takens, temos garantida a existência de



uma imersão para o espaço de fases original sempre que os vetores de reconstrução tenham dimensão  $k > 2d$ . Neste caso, para um  $d$  estimado um pouco acima dos 0.5 teríamos  $k > 1$ , e, como estamos interessados na menor dimensão que reconstrua o espaço original, tomaríamos  $k^* = 2$ . Importa no entanto referir que, a condição  $k > 2d$  é-nos apresentada como condição suficiente mas não necessária, ou seja, o teorema não proíbe a existência de mergulhos menores. Outra abordagem, como foi referido anteriormente, é observar a evolução da dimensão de correlação em função da imersão e escolher a dimensão  $k$  para a qual o crescimento de  $\alpha$  desacelere significativamente.

Como se pode observar na figura 3.2, com o crescimento de  $\epsilon$ , tanto o valor de  $\alpha$  como a sua variação, ou seja,  $\alpha^{(k+1)} - \alpha^{(k)}$ , tende a diminuir mantendo-se o ponto  $k = 2$  como momento onde existe uma mudança de comportamento mais evidente. Este comportamento torna-se ainda mais evidente na figura 3.7 onde se regista esta evolução para valores de  $\epsilon$  inferiores. O facto de não apresentarmos nesta análise valores  $\epsilon$  menores prende-se com a particularidade de o esforço computacional ser inversamente proporcional a  $\epsilon$ , o que torna necessário encontrar um compromisso entre rigor nas conclusões e eficiência computacional. Na maioria dos casos, como o apresentado acima, as duas análises convergem na dimensão  $k^*$ , mas não é de excluir a hipótese de se obter um  $\alpha$  que sugira uma dimensão de imersão superior àquela sugerida pela observação visual do seu comportamento face a  $k$ . Neste caso tomar-se-ia a menor dimensão de entre as duas como a ótima.

Ao aplicar o algoritmo Grassberger-Procaccia a um conjunto de pontos (mesmo conjuntos relativamente pequenos como os 2 000 pontos do caso descrito atrás), devido à necessidade de comparação entre todos os pares de pontos o tempo de computação facilmente cresce de minutos para horas. Isto representa uma grande limitação quando pretendemos analisar diversos conjuntos de dados como por exemplo a cotação do euro-dólar registado ao segundo durante o ano de 2012 e analisando a sua resposta a diferentes valores de  $\epsilon$  e diferentes dimensões de imersão  $k$ . Neste

cenário, os tempos de computação facilmente se transformam em dias, semanas e até mesmo meses. Ou seja, a técnica que temos em posse, por mais assertiva e relevante que seja, não é suscetível de servir interesses práticos onde seja necessário obter respostas *no minuto*.

### 3.2.2 Um Algoritmo Alternativo

O algoritmo apresentado em seguida surge como tentativa de resposta às limitações descritas acima. Este método aspira a estar no meio termo entre a mecânica do algoritmo *Box-Counting* e o algoritmo de Procaccia-Grassberger no que aos métodos de implementação e computação diz respeito. Vantagens e desvantagens serão discutidos mais adiante, apresentemos antes de mais uma definição formal do algoritmo.

Seja  $\mathbf{1}_A(x)$  a função indicatriz:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{se } x \in A, \\ 0 & \text{se } x \notin A. \end{cases}$$

Seja  $B_i(\epsilon)$  uma *caixa* de lado  $\epsilon$  análoga à caixa  $N_b(\epsilon)$  introduzida na definição do algoritmo *Box-Counting* mas generalizável a mais dimensões. Seja  $V(B_i)$  o número de pares existente na *caixa*  $i$ , tal que:

$$V_k(B_i) = \left( \sum_{j=1}^N [\mathbf{1}_{B_i}(X_j^k)] \right)^2 - \sum_{l=1}^N [\mathbf{1}_{B_i}(X_l^k)],$$

e seja a medida  $\tilde{C}(\cdot)$  definida da seguinte forma:

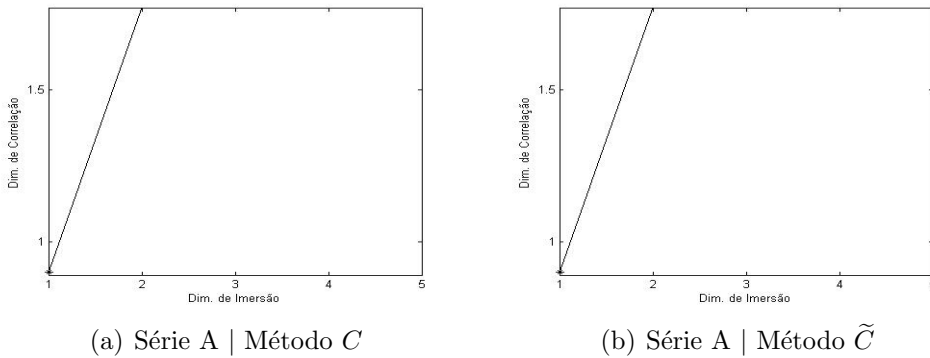
$$\tilde{C}_{N,\epsilon}^{(k)}(A) = \frac{1}{N(N-1)} \sum_{i=1}^{N^b} V_k(B_i(\epsilon)), \quad (3.4)$$

onde  $k$  é a dimensão de imersão,  $N$  é o número de pontos e  $N^b$  o número de caixas. Então, a nova estimativa de  $\alpha^{(k)}$ , chamada  $\tilde{\alpha}^{(k)}$  vem definida por:

$$\tilde{\alpha}_\epsilon^{(k)}(A) = \frac{\log(\tilde{C}_\epsilon^{(k)}(A))}{\log(\epsilon)} \quad (3.5)$$

É de esperar que  $\tilde{C}$  seja inferior a  $C$  para um mesmo conjunto devido à existência de pontos que são efetivamente vizinhos, ou seja, distam menos de  $\epsilon$  entre si, e no entanto são agrupados em caixas diferentes. Essa característica do método provocará certamente um desvio na estimativa da dimensão de correlação mas não na sua evolução com a dimensão de imersão. Ou seja, estaremos mais atentos à observação do crescimento de  $\tilde{\alpha}$  do que ao valor obtido. No entanto, não existirão grandes discrepâncias entre as duas abordagens. A grande vantagem do método é a sua rapidez de computação. O facto de apenas ser necessário analisar cada ponto uma única vez, de modo a decidir em que caixa o agrupar, em vez de fazer a comparação entre todos os pares que o algoritmo original tem implícita, faz com que a complexidade de processamento<sup>2</sup> diminua bastante, como vamos ver mais adiante. Em seguida, aplicam-se os dois algoritmos às restantes séries comparando-se as conclusões que ambas permitem tirar.

**Figura 3.3:** Comparação entre algoritmos | Série A

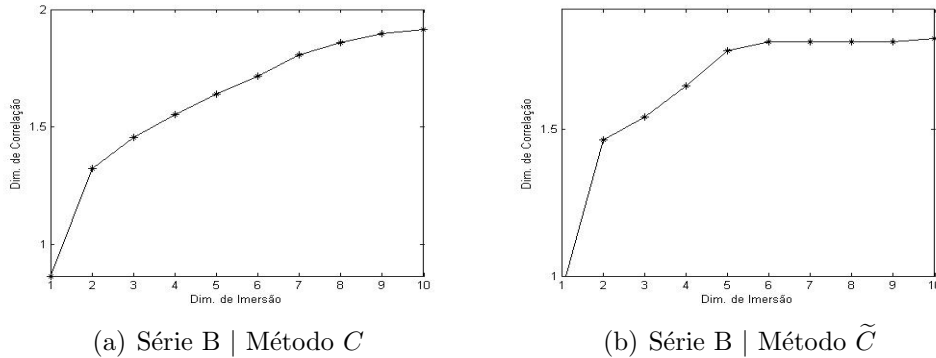


O que se observa no caso em que as observações são aleatórias e independentes entre si é que a dimensão de correlação rapidamente *explode* para infinito. Isto deve-se ao facto de o número de vizinhos em regiões pequenas tender para zero a

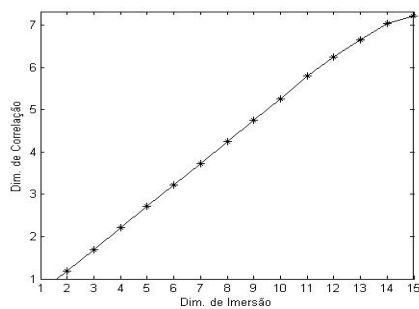
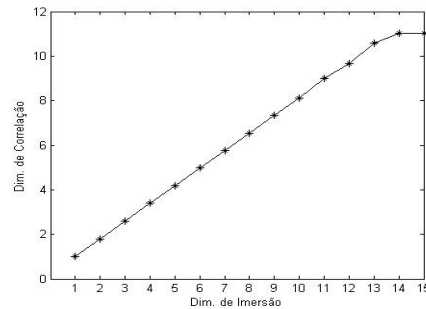
<sup>2</sup>A função '*CorrIntOld.m*' que é a aplicação do primeiro algoritmo tem uma complexidade  $O(n^2)$  devido aos dois ciclos necessários para fazer a comparação entre pares, enquanto a função '*CorrIntNew.m*' tem complexidade  $O(n)$ .

grande velocidade. Recordando (3.2) e (3.4), esta escassez de vizinhos faz com que os integrais estimados sejam também eles bastante próximos de zero (ou mesmo nulos) o que por sua vez resulta num crescimento exponencial de (3.3) e (3.5).

**Figura 3.4:** Comparação entre algoritmos | Série B



Neste caso os valores estimados apresentam uma evolução bastante distinta. Com ambos os métodos se observa que a dimensão de correlação estabiliza um pouco abaixo de  $\tilde{\alpha} = 2$  (a dimensão de correlação deste atrator está estimada em  $2.05[6]$ ). Ou seja, tanto a aplicação de  $C$  como de  $\tilde{C}$  aponta para  $k^* = 5$ , o que é consistente com a observação visual do crescimento de  $\alpha$  em função da dimensão de imersão, em particular no segundo método. De facto, como se pode observar na figura 8(a), onde se aplica  $C$  a este conjunto, a quebra no crescimento de  $\alpha$  à passagem de  $k = 5$  torna-se mais evidente para valores menores de  $\epsilon$ . No entanto, essa exigência de menores distâncias de vizinhança não é extensível ao segundo método onde esse comportamento é já inequívoco.

**Figura 3.5:** Comparação entre algoritmos | Série C(a) Série C | Método  $C$ (b) Série C | Método  $\tilde{C}$ 

Neste caso, em que estamos a estudar uma série determinística com ruído, a dimensão de correlação cresce de forma aparentemente linear, estabilizando apenas perto de valores de  $k$  entre 13 e 15. De acordo com o primeiro método, a série tende para uma dimensão de correlação um pouco acima de 7, notando-se uma quebra a partir de  $k = 14$ . Na aplicação do segundo algoritmo observa-se essa desaceleração logo em  $k = 13$  apesar de a estabilização só ocorrer perto de valores de  $\tilde{\alpha} = 11$ . O primeiro método sugere portanto  $k^* = 14$  ou 15 enquanto o segundo indica  $k^* = 14$ .

Os valores obtidos para a série temporal D usando o método  $\tilde{C}$  replicam também as conclusões obtidas na secção anterior (ver Anexo B). Na tabela 3.2 apresenta-se os resultados obtidos pelos os dois algoritmos e as respectivas conclusões.

**Tabela 3.2:** Comparação entre algoritmos

	Série A (N = 2 000)		Série B (N = 3 000)		Série C (N = 7 000)	
	$C$	$\tilde{C}$	$C$	$\tilde{C}$	$C$	$\tilde{C}$
<b>K*</b>	$\infty$	$\infty$	5	5	14/15	13/14
<b>Pares Vizinhos</b>						
$K = 1$	4 081	3 906	5 111	4 978	1 999 332	1 799 430
$K = K^*$	-	-	430	234	1 423	192
<b>Dim. Corelação (<math>\alpha</math>)</b>						
$K = 1$	0.92	1.03	0.88	0.97	0.64	1.00
$K = K^*$	-	-	1.54	1.65	6.95	9.33
<b>Tempo Computação</b>	103.5s	24.6s	207.3s	50.7s	711.9s	114.9s

O primeiro facto a realçar é o de a dimensão ótima deduzida através do algoritmo alternativo sugerido ser consistente com aquela obtida usando o integral de correlação na sua formulação original. Também de acordo com as expectativas é o facto de o número pares de vizinhos contabilizados ser inferior. Outro registo digno de nota, é o de o tempo de computação necessário para produzir os resultados apresentados ser notoriamente inferior durante a aplicação do algoritmo  $\tilde{C}$ . Aliado à manutenção da capacidade em tirar conclusões corretas sobre a dimensão de imersão, esta melhoria em desempenho computacional confirma que o método alternativo atinge os objetivos a que se propõe.

### 3.2.3 O Problema da Escolha da Distância Entre Pontos

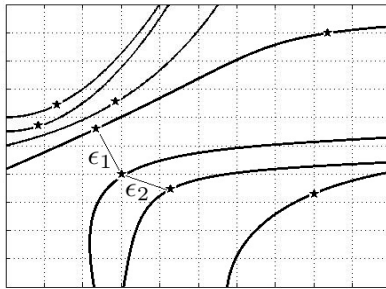
As principais limitações associadas à escolha da distância estão exemplificadas em baixo, são preocupações a ter sempre em mente quando se aplica o algoritmo. Importa, antes de mais, referir que todos os algoritmos usados no tratamento dos dados neste trabalho normalizam a série temporal no início do seu tratamento numérico. Ou seja, dado o vetor  $X = (x_1, \dots, x_n)$  de observações, passamos a trabalhar a partir

de um vetor  $\tilde{X}$  tal que cada elemento sofre a transformação linear:

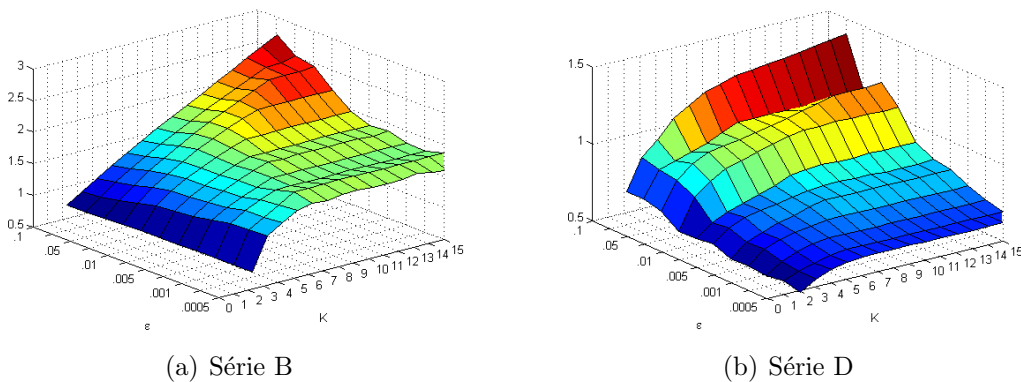
$$\tilde{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}.$$

Esta transformação mantém todas as propriedades da série intactas e permite-nos uniformizar o valor de  $\epsilon$  para o tratamento de qualquer série.

**Figura 3.6:** Problemas de vizinhança



No exemplo acima, um  $\epsilon$  superior a  $\epsilon_2$  significaria contar como vizinhos pontos que não deveriam ser considerados como tal, inflacionando o integral de correlação. Por outro lado, tomar um  $\epsilon$  inferior a  $\epsilon_1$  implicaria negligenciar pontos erradamente. Deve ser também tomada em consideração o facto de existirem limites para a precisão de uma medição de uma observação bem como da precisão com que o computador pode trabalhar. Isto significa que qualquer  $\epsilon$  inferior a estes limites resultaria numa contagem de zero vizinhos mesmo num conjunto em que todos os pontos devessem ser considerados como tal. Se atualmente os limites de precisão numérica computacional são praticamente inexistentes, a precisão das medições continua a ser um problema a ter em conta. A figura 3.7 mostra a evolução da dimensão de correlação estimada, em função de  $k$  e da distância de vizinhança.

**Figura 3.7:** Dimensão de correlação em função de  $k$  e  $\epsilon$ 

Pode-se ver que considerando uma distância demasiado grande o plano  $\tilde{c}Ok$  regista um comportamento mais próximo ao associado a séries com ruído ou aleatórias. À medida que  $\epsilon$  diminui a evolução das estimativas começa a aproximar-se daquela que sabemos ser a correta, mantendo-se relativamente inalterada a partir de certo ponto. É portanto importante não definir a distância de vizinhança com um critério demasiado largo quando estudarmos séries sobre as quais não conhecemos a *evolução correta*. Então o ideal seria tomar um  $\epsilon$  tão pequeno quanto possível? Sim, no entanto recorde-se que o algoritmo consiste na indexação de pontos, arrumando-os na respetiva caixa. Tomando, a título de exemplo,  $\epsilon = 10^{-1}$  e  $k = 5$ , para uma série já normalizada o número de caixas necessárias é  $10^5$ . No mesmo contexto, reduzindo  $\epsilon$  para  $10^{-2}$  e  $10^{-3}$ , seriam necessárias  $10^9$  e  $10^{14}$  caixas, respetivamente. Considerando que é possível fazer a contagem com um inteiro de 8 *bytes*, precisaríamos de 74 *Gigabytes* e 7260 *Terabytes*, respetivamente, de memória virtual o que é um requisito impossível de satisfazer. Estes cálculos são meramente indicativos do crescimento exponencial em  $\epsilon$  dos recursos computacionais consumidos. Uma das formas que usamos para minorar um pouco este problema é através de matrizes esparsas, que não requerem alocação de memória para elementos nulos.



### 3.3 Expoente de Lyapunov

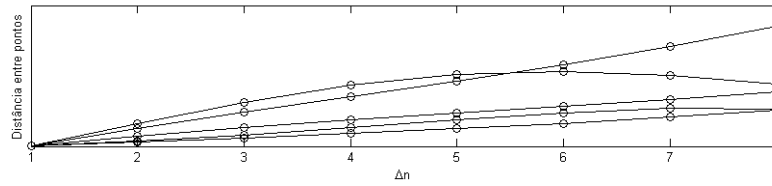
Depois de encontrar a imersão ótima para os nossos dados estamos em condições de partir para a fase de previsão. No entanto, antes é importante ter uma medida do nível de sensibilidade às condições iniciais da série. De certa forma, queremos quantificar a probabilidade de gerar previsões com um alto grau de aderência à realidade. As estimativas que obtivemos nas secções anteriores, para as dimensões *Box-Counting* e correlação, servem ambas para medir o grau de complexidade do sistema. A estimativa do expoente de Lyapunov vai, por sua vez, permitir quantificar o grau de caos no sistema, ou seja, a sua imprevisibilidade. Em termos formais, o expoente de Lyapunov é a medida de divergência de pontos cujas trajetórias, no momento inicial, distam  $\epsilon$  uma da outra. Assumindo que a distância entre esses dois pontos é descrita ao longo das suas órbitas por  $\delta_n = e^{\lambda n}$ , o expoente de Lyapunov  $\lambda$  vem dado por:

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \log(\delta_n) \quad (3.6)$$

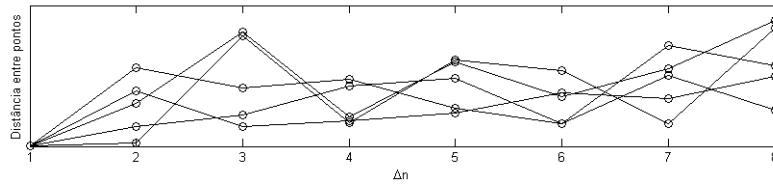
A abordagem ao cálculo numérico do expoente começa na pesquisa por pontos que existam dentro de uma vizinhança  $\epsilon$  um do outro. De seguida é registada a evolução da divergência de trajetórias,  $div_n$ , tal que  $div_n(X_i, X_j) = \|X_{i+n} - X_{j+n}\|_\infty$ . Faz-se um ajustamento, usando um método de regressão linear, de forma a estimar a dependência de  $\log(div_n)$  em relação ao salto temporal, tal que

$$\hat{\lambda}_i = \log(div_n) \sim \Delta n.$$

Por fim, apresenta-se a estimativa  $\hat{\lambda} = \frac{1}{t} \sum_{i=1}^t \lambda_i$ , a média dos expoentes obtidos pela regressão linear. Na figura 3.8 está representada a evolução da distância entre alguns pares de pontos vizinhos, que vai ajudar a perceber o porquê do procedimento de estimação do expoente de Lyapunov depender de uma regressão linear.

**Figura 3.8:** Distância entre vizinhos

(a) Série B



(b) Série C

Escolheram-se os primeiros 5 pontos vizinhos e registou-se a sua distância nos 8 primeiros saltos temporais. Apesar dos expoentes estimados (e conhecidos, ver tabela 3.3) destas duas séries serem semelhantes, o comportamento tipo da evolução da distância entre pontos vizinhos é bastante distinto. No caso da série B, a distância entre pontos evolui de forma relativamente regular enquanto a série C, devido à sua natureza parcialmente aleatória, apresenta um comportamento mais irregular. No entanto, à medida que  $\Delta n$  aumenta e em média, a distância entre dois pontos vizinhos de cada uma das séries aumenta de forma semelhante.

O expoente de Lyapunov foi estimado segundo esta abordagem e os resultados foram registados na tabela 3.3.

**Tabela 3.3:** Expoente de Lyapunov

	Série A	Série B	Série C	Série D
$\tilde{\lambda}$	2.721	0.814	0.157	0.744
$\lambda[\text{ref}]$	-	0.906[9]	-	0.693[4]

Todas as séries revelam comportamento caótico. Apesar disso a série C é significativamente mais estável que as anteriores. Isso pode ser justificado pelo facto da série ser apenas parcialmente aleatória e ter como componente determinística

uma função linear. A série A regista o comportamento mais errático, o que seria expectável à partida.

Os resultados acima permitem concluir que na fase de previsão dos futuros estados do sistema, com base nos registos passados, esperaríamos obter previsões de maior qualidade para a série C do que para a série A. Em seguida apresenta-se um método de previsão para o qual o expoente de Lyapunov é uma medida extremamente relevante.

## 4 Previsão

Neste capítulo vamos introduzir uma abordagem à previsão usando a mecânica subjacente ao método dos análogos e aplicado sobre o espaço reconstruído, ou seja, aos vetores de reconstrução obtidos através dos procedimentos descritos nas secções anteriores. A ideia é encontrar os pontos do espaço reconstruído mais próximos do ponto que contém a última observação,  $X_T$ . A partir desse conjunto de *vizinhos*, com base naquilo que foi a evolução desses mesmo pontos, estima-se por analogia o estado seguinte,  $X_{T+1}$ , do sistema. Designa-se o conjunto dos análogos de  $X_T$  por  $A_\epsilon(X_T)$  e define-se da seguinte forma:

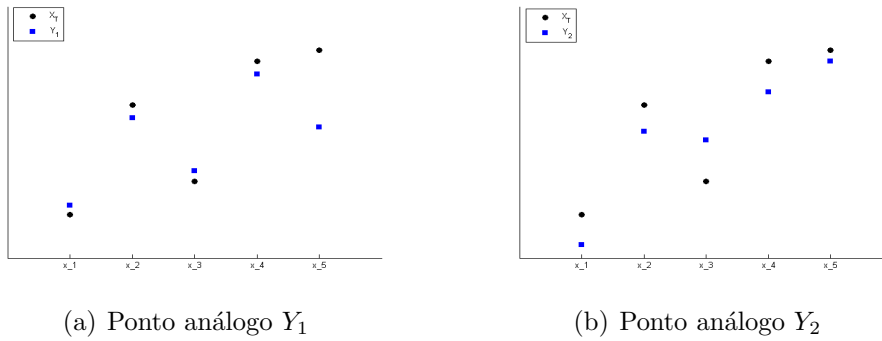
$$A_\epsilon(X_T) = \{X_i: \|X_i - X_T\|_\infty < \epsilon\} \quad (4.1)$$

Em seguida, para um dado análogo  $X_i$ , define-se que quanto mais próximo este se encontrar da observação  $X_T$ , maior será a contribuição de  $X_{i+1}$  para a previsão de  $X_{T+1}$ . Ou seja, tomar-se-à a média do estado seguinte dos pontos análogos ponderada pelo inverso da distância ao último ponto registado. Formalmente, seja  $\hat{X}_{T+1}$  a previsão para o instante  $T + 1$ :

$$\hat{X}_{T+1} = \frac{\sum_{i=1}^{T-1} [\mathbf{1}_{A_\epsilon(X_T)}(X_i) \cdot X_i \cdot (d_i)^{-1}]}{\sum_{i=1}^{T-1} [\mathbf{1}_{A_\epsilon(X_T)}(X_i) \cdot (d_i)^{-1}]} \quad (4.2)$$

Devido ao facto da norma usada ser  $\|\cdot\|_\infty$ , o algoritmo funciona como um mecanismo de reconhecimento de padrões. Ou seja, procurar por dois pontos análogos, nos moldes definidos atrás, equivale a procurar uma sequência de observações que evolua a uma distância tão pequena quanto possível da sequência de observações do último ponto registado (ver figura 4.1).

**Figura 4.1:** Pontos análogos



O ponto  $Y_1$  tem, em média e considerando as normas mais usuais, uma distância inferior a  $X_T$  apesar da sua última observação sair fora do padrão de comportamento de  $X_T$ . Já o ponto  $Y_2$ , apesar de os seus elementos terem uma distância superior em média, a sua sequência é bastante fiel à última registada. O que se pretende com a abordagem da previsão pelos análogos é precisamente encontrar os pontos do tipo  $Y_2$  mais próximos de  $X_T$ . Relembrando a definição de expoente de Lyapunov, esperar-se-à que a qualidade da previsão seja inversamente proporcional à estimativa de  $\lambda$ .

## 5 Aplicação às Finanças

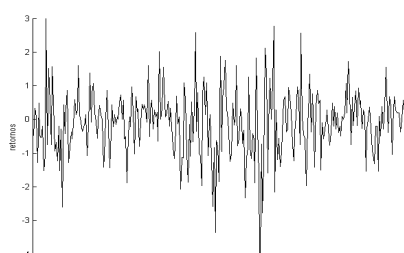
Vamos agora aplicar a metodologia apresentada e desenvolvida nas secções anteriores a dados reais obtidos no mercado financeiro. Foram recolhidas as seguintes séries de observações do mercado de dívida e do mercado cambial.

**Tabela 5.1:** Séries financeiras recolhidas

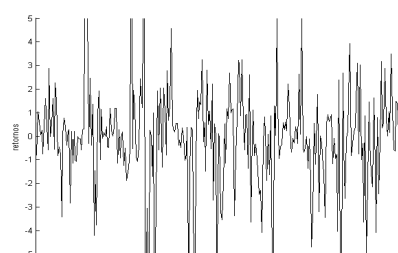
Ativo	Frequência	Início	Número observações	Notação
Obrigação Portuguesa (10 anos)	diária	Mai-1994	5 042	<i>pt10y</i>
Obrigação Britânica (10 anos)	diária	Jan-1980	8 764	<i>uk10y</i>
Obrigação Americana (10 anos)	diária	Jan-1966	12 171	<i>us10y</i>
Cotação euro vs dólar	ao minuto	Ago-2013	31 138	<i>eurusdm</i>
	diária	Set-1971 <sup>3</sup>	10 962	<i>eurusdd</i>
Cotação libra vs dólar	ao minuto	Nov-1971	31 215	<i>gbpusdm</i>
	diária	Ago-2013	10 920	<i>gbpusdd</i>

As séries foram depois transformadas de forma a se obter a variação percentual em relação ao período anterior, ou seja, construiu-se a série dos retornos desse ativo financeiro. A partir deste momento, sempre que nos referirmos a séries dos ativos em causa estaremos a falar da sua série de retornos. Na figura 5.1 encontra-se a representação gráfica de quatro dessas séries recolhidas.

**Figura 5.1:** Séries temporais financeiras

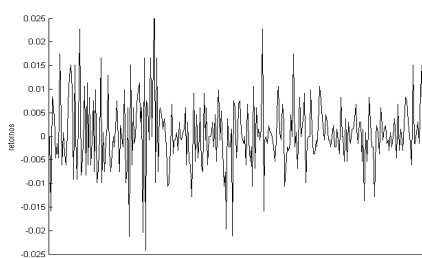


(a) Obrigações EUA

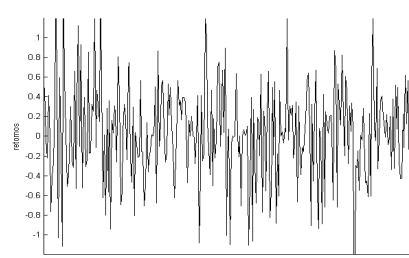


(b) Obrigações Portugal

<sup>3</sup>Apesar de o euro só ter entrado em vigor no dia 1 de Janeiro de 1999, o Banco Europeu de Investimento manteve desde o início dos anos 70 a cotação de um índice de agregação de várias moedas Europeias que veio a ser substituído, em paridade, pelo euro.



(c) EUR/USD ao minuto



(d) EUR/USD diário

São representações típicas de retornos financeiros: bastante erráticas e sem padrões reconhecíveis a olho nu. É de notar a semelhança entre as séries desenhadas acima e as representadas na figura 2.1. As medidas estatísticas mais relevantes destas séries vão ao encontro dos factos estilizados sobre séries financeiras [13], em particular: inexistente auto-correlação ou correlação parcial de qualquer ordem e *kurtosis* bastante superior a 3, ou seja, têm distribuição leptocúrtica (ver anexo A).

## 5.1 Aplicação da metodologia sugerida

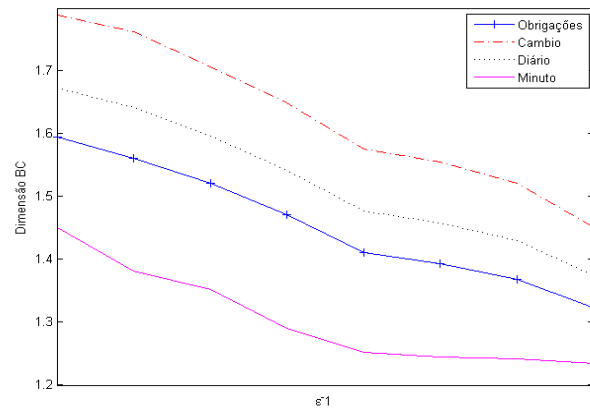
Vamos agora proceder à aplicação da metodologia desenvolvida nos Capítulos 2, 3 e 4 aplicando-a às séries financeiras de forma a quantificar esse sistema que assumimos dinâmico não-linear que é o mercado financeiro e de modo a obter os resultados necessários à computação do algoritmo de previsão.

### • Dimensão Box-Counting

Aplicou-se primeiro o algoritmo de *Box-Counting* de modo a ter uma primeira indicação sobre o grau de complexidade dos sistemas em análise. Calculou-se esta dimensão para as séries isoladamente e agrupadas pelas suas principais características e registaram-se os seguintes resultados:

**Tabela 5.2:** Dimensão *Box-Counting*

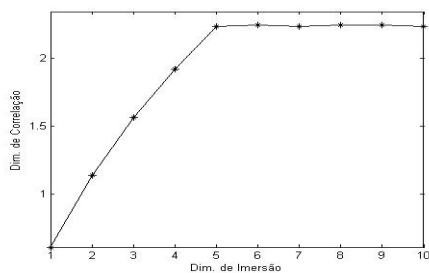
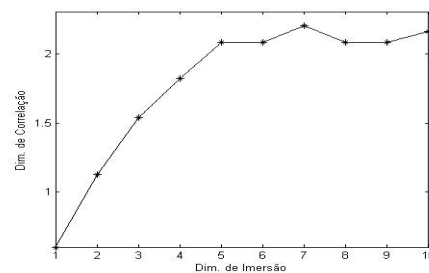
Série	$\epsilon = 10^{-1}$	$\epsilon = 10^{-2}$
<i>pt10y</i>	1.45	1.29
<i>us10y</i>	1.75	1.50
<i>uk10y</i>	1.69	1.47
<i>eurusdd</i>	1.78	1.57
<i>eurusdm</i>	1.40	1.27
<i>gbpusdd</i>	1.80	1.58
<i>gbpusdm</i>	1.50	1.24

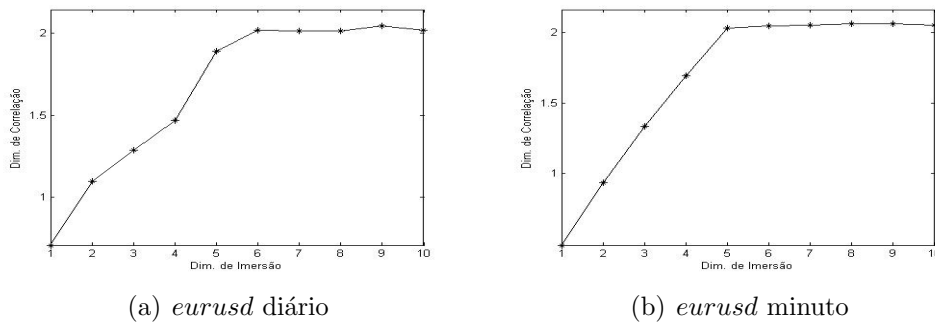
**Figura 5.2:** Dimensão *Box-Counting* vs  $\epsilon$ 

Na tabela 5.2 estão registados os valores obtidos tomando caixas de lado  $\epsilon = 0.1$  e  $\epsilon = 0.01$  e na figura 5.2 apresenta-se a representação gráfica da mesma evolução agrupando as observações por características. Os resultados sugerem que os sistemas são bastante semelhantes sendo apenas notória uma diferença entre as séries de frequência intra-diária e as restantes. O sistema gerador das séries dos títulos do tesouro Português apresenta uma dimensão de *Box-Counting* inferior à das restantes obrigações.

- **Dimensão de Imersão**

Em seguida procedeu-se à aproximação da dimensão de correlação de cada conjunto através do algoritmo  $\tilde{C}$  registando-se a sua evolução em função de  $k$  nas figuras 5.3 e 5.4.

(a) *us10y*(b) *pt10y***Figura 5.3:** Obrigações do tesouro



**Figura 5.4:** Taxas de câmbio

O primeiro facto que salta à vista é o da semelhança no comportamento transversal aos diferentes tipos de ativo e as diferente frequências (os resultados obtidos para a série *gbpusd* são em tudo semelhantes aos da série *eurusd*). De resto, à parte do facto de as obrigações estabilizarem com  $\alpha$  ligeiramente superior (característica que se repete dos resultados do algoritmo *Box-Counting*), ambas registam uma paragem súbita quando se passa de  $k = 5$  para  $k = 6$ . A partir destes resultados, as observações são agrupadas nos respetivos vetores de reconstrução que serão usados na fase de previsão.

- **Expoente de Lyapunov**

Antes de passar à previsão, registaram-se as estimativas para o expoente de Lyapunov, na tabela 5.3.

**Tabela 5.3:** Expoentes de Lyapunov | Séries financeiras

	<i>pt10y</i>	<i>us10y</i>	<i>uk10y</i>	<i>eurusdd</i>	<i>eurusdm</i>	<i>gbpusdd</i>	<i>gbpusdm</i>
$\tilde{\lambda}$	0.105	0.214	0.145	0.113	0.493	0.111	0.394

As séries de frequência intra-diária apresentam uma maior sensibilidade às condições iniciais. Taxas de câmbio recolhidas diariamente é a categoria que apresenta um expoente superior, sendo que o expoente estimado para as obrigações Portuguesas é aquele que indicia uma maior previsibilidade. Todos os expoentes estimados são positivos, confirmando-se a existência de comportamento caótico.



## 5.2 Aplicação do Método dos Análogos

A aplicação do método dos Análogos vai ser feita integrada numa estratégia de investimento cujo retorno será depois comparado com outras estratégias. As decisões de investimento são tomadas tendo em conta as indicações de 3 métodos diferentes: o método dos análogos, apresentado no Capítulo 4; um modelo auto regressivo simples, AR(1), bastante usado ainda hoje nos mercados e por último decisões tomadas de forma aleatória.

A implementação será simples: para um dado ativo, conforme o modelo indique um retorno positivo ou negativo para o próximo período será tomada uma posição longa ou curta, respetivamente. No caso de ser prevista uma mudança de direção, por exemplo, a passagem de um retorno positivo para um negativo, fecha-se a posição longa e assume-se uma curta. No caso em que o retorno previsto seja o mesmo que o corrente a posição mantém-se inalterada. Assume-se que não existem custos de transação e que o mercado é líquido em todos os momentos.

Para efeitos de comparação foram criadas 3 carteiras. A primeira carteira, Carteira O, é constituída pelas 3 obrigações de dívida soberana com peso igual no portfólio num total de €1 500 000 de investimento. A Carteira M, destinada a transações intra-diárias tem como ativos subjacentes a taxa de câmbio euro-dólar e libra-dólar em proporção igual num total de investimento de €1 000 000. Por fim, foi constituída a Carteira D, semelhante à Carteira M mas reservada a transações diárias. O desempenho será avaliado ao fim de cerca de 250 transações no caso de frequência diária (1 ano de investimento) e 1440 transações no caso de frequência ao minuto (1 dia de investimento). Os resultados obtidos foram registados na tabela 5.4.

Tabela 5.4: Retornos obtidos através das diferentes estratégias

	Análogos		AR 1		Aleatório		
	Saldo <sub>T</sub>	% <sup>4</sup>	Saldo <sub>T</sub>	%	Saldo <sub>T</sub>	%	
O	pt10y	697 400	+39.48	819 300	+63.86	541 000	+8.2
	us10y	479 050	-4.19	508 850	+1.77	508 500	+1.7
	uk10y	480 350	-3.93	514 800	+2.96	479 500	-4.1
	<b>Total</b>	<b>1 656 800</b>	<b>+10.5</b>	<b>1 842 950</b>	<b>+22.9</b>	<b>1 529 000</b>	<b>+2.0</b>
M	gbpusdm	500 180	+9.53	500 365	+19.90	499 710	-35.8
	eurusdm	499 980	-1.00	499 638	-19.89	500 090	+15.7
	<b>Total</b>	<b>1 000 160</b>	<b>+4.27</b>	<b>999 925</b>	<b>+0.01</b>	<b>999 390</b>	<b>-10.5</b>
D	eurusdd	519 100	+3.82	456 550	-8.69	521 500	+4.3
	gbpusdd	526 450	+5.29	491 150	-1.77	505 000	+1.0
	<b>Total</b>	<b>1 045 550</b>	<b>+4.56</b>	<b>947 700</b>	<b>-5.23</b>	<b>1 026 500</b>	<b>+2.7</b>
<b>Tempo de Decisão</b> <sup>5</sup>	0.548s		1.558s		-		

Apesar da boa rentabilidade, o desempenho da carteira O é altamente insatisfatório quando comparado com o registado usando o AR(1). De resto, mesmo apesar de o expoente de Lyapunov estimado para a série das obrigações Portuguesas ter sido o menor, a rendibilidade obtida, em qualquer dos modelos, acabou por ser bastante acima das expectativas. Isto pode dever-se a várias características associadas à realidade portuguesa que não estão presentes nos outros ativos. O país está sob intervenção externa, o que significa que não existem emissões no mercado

<sup>5</sup>Retornos anualizados (assumem-se 250 dias de transação por ano). Os retornos anualizados das carteiras têm em conta o desempenho cumulativo de cada activo em separado.

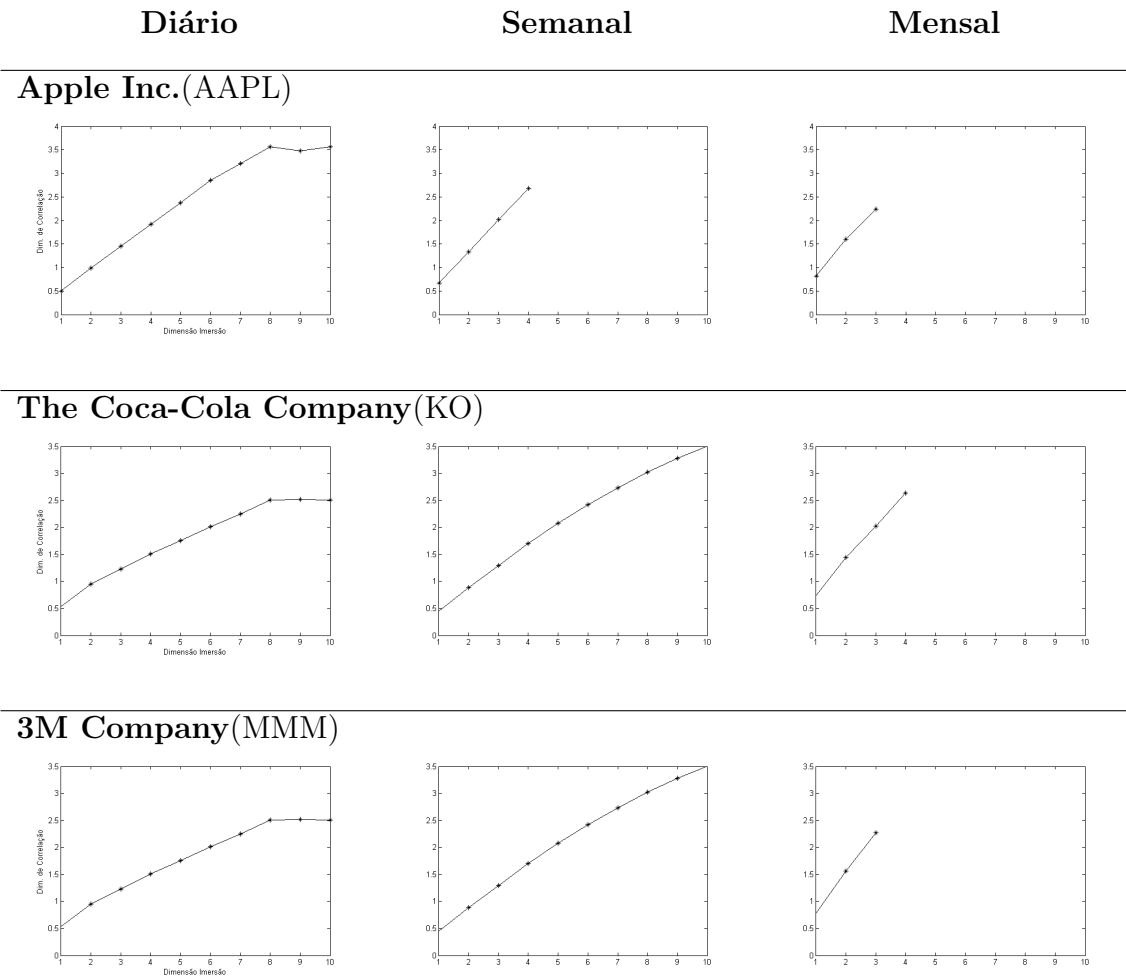
<sup>5</sup>Tempo médio de computação da função *'Analog.m'* no caso da primeira estratégia e tempo médio de estimação de um modelo AR(1), através da função *'garchfit'* disponível na biblioteca base do MatLab, para a segunda estratégia.

primário, há instabilidade política e existe menor liquidez quando comparando com as restantes observações. A carteira M regista um desempenho ligeiramente positivo, acima daquele obtido com a estratégia de previsão linear. Por fim, a Carteira D regista uma rentabilidade inequivocamente superior à obtida pelos restantes métodos e num valor já bastante razoável para um investimento financeiro. Outro facto digno de nota é de o tempo de computação necessário para a tomada de decisão ser bastante pequeno (apesar de não conseguir fazer concorrência ao tempo de *pensamento* nulo de um decisão aleatória).

### 5.3 Outros Ativos

Nesta secção fazemos uma pequena apresentação do que são os resultados obtidos usando a mesma metodologia de aproximação da dimensão de correlação a acções de empresas cotadas no NYSE (*New York Stock Exchange*). Escolheram-se as acções das maiores empresas, considerando a respetiva capitalização de mercado, de três sectores distintos: tecnológico, alimentício e industrial. Os valores obtidos encontram-se registados na tabela 5.5.

Tabela 5.5: Dimensão de correlação | Acções



Á semelhança dos títulos do tesouro e das taxas de câmbio, existe evidência de natureza determinística nas séries de retornos diários das acções das empresas supracitadas. No entanto, a dimensão de imersão estimada é  $k^* = 8$  revelando a existência de mais graus de liberdade, ou seja, mais dependências, do que nos ativos usados nas carteiras de investimento. De certa forma surpreendente, tendo em conta os resultados obtidos para os retornos diários, é o facto de os retornos semanais exibirem um comportamento bastante semelhante ao da série C (determinístico mas com ruído) e no caso dos retornos mensais este comportamento ser em tudo semelhante ao da série A (série aleatória). A ser desenvolvida uma experiência nos moldes da efetuada na secção 5.2, fica de certa forma excluída a viabilidade de uma aplicação com decisões tomadas semanal ou mensalmente.

## 6 Conclusão

Neste trabalho tratámos de introduzir vários conceitos e procedimentos existentes no estudo de sistemas dinâmicos não lineares desenvolvendo métodos de aplicação prática com vista à modelação e previsão de séries temporais financeiras.

Foram introduzidos algoritmos computacionais para o cálculo de diversas medidas como dimensão fractal, dimensão de correlação e expoente de Lyapunov, que permitem uma obtenção dos resultados num tempo útil bastante inferior àquele necessário aquando da implementação *ipsis verbis* dos seus passos. Com isto, dado que se manteve inalterada a capacidade de tirar conclusões corretas, possibilita-se o tratamento de séries temporais financeiras com um elevado número de observações históricas e cujas transações podem ser feitas a uma frequência tão baixa quanto for possível ao mercado processar as ordens.

A interpretação económica dos valores estimados para as dimensões de imersão e correlação pode ser traiçoeira uma vez que estamos a tratar de dimensões fractais e portanto uma unidade pode dizer respeito a mais do que um fator de influência. Apesar disso, é possível fazer comparação entre as dimensões de cada série, e nesse campo não é de estranhar que as dimensões estimadas para as acções sejam superiores às das obtidas no caso das obrigações do tesouro e taxas de câmbio. De facto, dado que tanto os títulos do tesouro como as taxas de câmbio estão ambas ligadas à saúde económica e política dos respetivos países compreende-se que os seus graus de liberdade condigam. No caso das acções, pode-se argumentar que a sua flutuação é influenciada pelas mesmas variáveis que os ativos anteriores, aos quais acrescem fatores idiossincráticos dos mercados em que estão inseridas (no caso, os segmentos tecnológico, da alimentação e industrial).

Importa sublinhar que o trabalho empírico não é extenso o suficiente de modo a construir conclusões definitivas, no entanto, os resultados obtidos na secção 5.2 deixam a porta aberta a mais investigação na modelação de séries temporais financeiras com ferramentas de análise de séries dinâmicas não-lineares. Não só se obteve

*lucro* nas 3 carteiras previamente elaboradas como foi possível superar o algoritmo de teste em 2 desses portfólios.

O facto de se terem registado indícios fortes de relações não-lineares entre retornos intra-diários e diários, indícios menos relevantes no caso da frequência ser semanal, e indícios inexistentes no caso dos retornos mensais, sugere que é mais fácil encontrar determinismo quanto mais curtos forem os tempos da recolha de informação. Isto pode dever-se ao facto de o mercado estar segmentado em investidores de longo prazo e investidores de curto prazo. O que os resultados obtidos suportam que as ferramentas apresentadas modelam com maior sucesso o comportamento dos investidores de curto prazo.

## Bibliografia

- [1] Borovkova, S. (1998). *Estimation and Prediction for Nonlinear Time Series*. Groningen, Netherlands: University Library Groningen.
- [2] Cadasgli, M.,Stephan, E.,Farmer,J. and Gibson, J. (1991). State space reconstruction in the presence of noise. *Physica D* 51, 52-98.
- [3] Deyle, E. and Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *Plos one* 6(3):e18295. doi:10.1371/journal.pone.0018295.
- [4] Eckmann, J. and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* 57,617.
- [5] Falconer, K. (2003). *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley and Sons, Ltd.
- [6] Grassberger, P. and Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.* 50, 346-349.
- [7] Grassberger, P. (1981). On the Hausdorff Dimension of Fractal Attractors. *Journal Stat. Phys.* 26, 173-17.
- [8] Hentschell, H. and Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica*. 8D, 435.
- [9] Kantz, H., Schreiber, T. (1997). *Nonlinear Time Series Analysis*. Cambridge, UK: Cambridge University Press.
- [10] Leok, M. (1992). Estimating the attractor dimension of the equatorial weather system. *Phys. Rev. LettActa Phys. Pol.* 85 Supplement S-27.
- [11] Lorenz, E. (1963). Deterministic nonperiod flow. *Journal of Atmospheric Science* 20, 130-141.

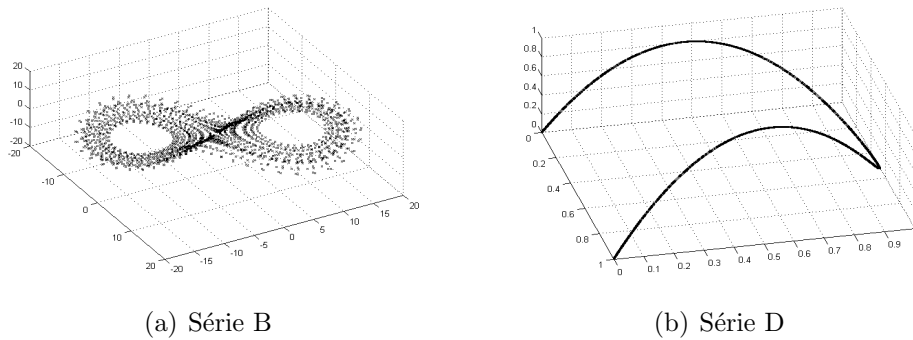
- [12] Mandelbrot, B. (1983). *The Fractal Geometry of Nature*. Henry Holt and Company.
- [13] Nicolau, J. (2011). *Modelação de Séries Temporais Financeiras*. ISEG/UTL e CEMAPRE.
- [14] Takens, F. (1981). Detecting strange attractors in turbulence. *Dynamical systems and turbulence. Lecture Notes in Mathematics* 898, 336-381. Springer-Verlag.
- [15] Thelier, J. (1990). Estimating fractal dimension. *The Lincoln Laboratory Journal* 7(6), 1055-1070.



## Anexo A Análise Gráfica e Analítica

### Representação Gráfica

Figura A.1: Visualização 3D



(a) Série B

(b) Série D

### Comparação entre algoritmos

Figura A.2: Dimensão de correlação | Série D

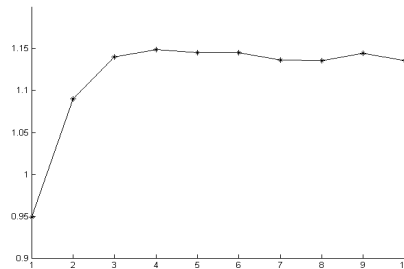


Tabela A.1: Algoritmo  $\tilde{C}$  | Série D

	$C$	$\tilde{C}$
$K^*$	2	2
<b>Pares Vizinhos</b>		
$K = 1$	6461	3769
$K = K^*$	2402	1642
<b>Dim. Correlação</b>		
$K = 1$	0.83	0.91
$K = K^*$	0.95	1.06
<b>Tempo Computação</b>	272.1s	68.3s

## Factos Empíricos

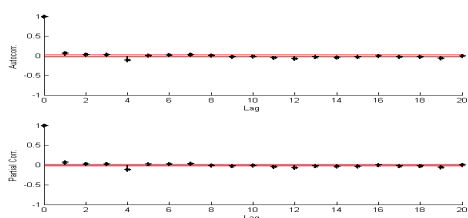
Tabela A.2: Estatísticas das séries recolhidas

	$r_A\%$	$\hat{\sigma}_A\%$	$\widehat{sk}$	$\widehat{k}$
<b>Cotações de Acções</b>				
Apple, Inc.	27.60%	47.19%	-0.39	12.56
3M Company	10.31%	23.26%	-0.11	9.87
The Coca-Cola Company	10.73%	29.19%	-2.91	18.01
<b>Taxas de Câmbio</b>				
EUR/USD	0.96%*	9.45%	0.12*	6.83
GBP/USD	-1,26%*	9.31%	-0.03*	7.35
<b>Títulos do Tesouro**</b>				
Portugal	-0.10%	16.51%	-0.25	45.61
Reino Unido	-1.08%	10.13%	-2.16	18.47
Estados Unidos	-0.01%	11.42%	0.32	10.92

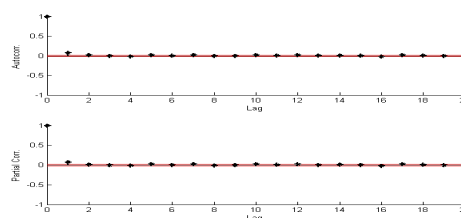
\* - indicadores não comparáveis com os restantes dada a natureza do ativo.

\*\* - toma-se a série dos preços da obrigação calculados através da série das *yields*.

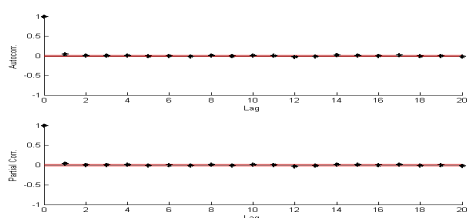
Figura A.3: Autocorrelação e correlação parcial.



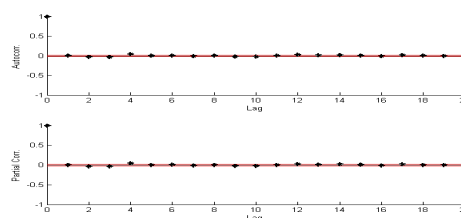
(a) Títulos Tesouro Portugal



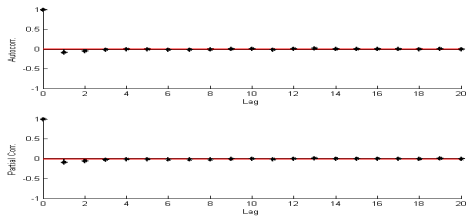
(b) Títulos tesouro EUA



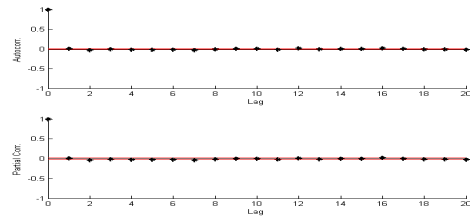
(c) Títulos tesouro Britânico



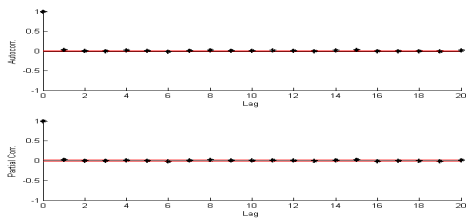
(d) Apple, Inc.



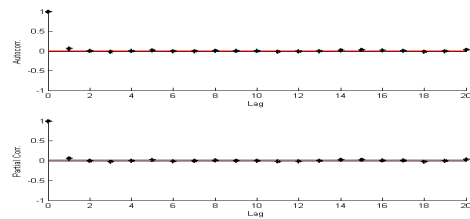
(e) Coca-Cola Company



(f) 3M Company



(g) Euro-Dólar



(h) Libra-Dólar

## Anexo B Código Matlab

### • Função BoxCount.m

```
function [bcdim numboxes totnumboxes] = BoxCount(serie,r)
import edu.stanford.covert.util.SparseMat;
%BoxCount Calcula a Dimensao Box-Count a Partir de:
%   -uma dada serie temporal (serie)
%   -um dado comprimento de lado das caixas (r)
%Devolve:
%   -bcdim: a dimensao
%   -numboxes: quantidade de caixas populadas
%   -totnumboxes: numero total de caixas
if nargin==1
    r=.1;
end
serie=floor(((serie-min(serie))/(max(serie)-min(serie)))*(1/r));
DIM= repmat(ceil(1/r)+1,1,2);
a(2,1)=uint32(0);
count=SparseMat(a);
count((max(DIM))^2,1)=1;
for i=1:(length(serie)-1)
    coor=num2cell(1+serie(i:(i+1)));
    idx=sub2ind(DIM,coor{:});
    count([idx 1])=count([idx 1])+1;
end
[numboxes ~]=size(find(count>0));
totnumboxes=prod(DIM);
bcdim=log(numboxes)/log(1/r);
end
```

### • Função CorrIntOld.m

```
function [d corrint et] = CorrIntOld(vec,k,r)
%CorrIntOld Calcula a Dimensao de Correlacao a Partir de:
%   -uma dada serie temporal (vec)
%   -uma dimensao de imersao (k)
%   -uma distancia de vizinhanca (r)
%Devolve:
%   -d: dimensao de correlacao
%   -corrint: o integral de correlcao
%   -et: tempo de computacao
if nargin==2
    r=.1;
end
if nargin==1
    r=.1;
    k=1;
end
tic;
format longg;
vec=(vec-min(vec))/(max(vec)-min(vec));
countR=0;
for i=1:(length(vec)-k)
    for j=(i+1):(length(vec)-k+1)
        if norm(vec(i:(i+k-1))-vec(j:(j+k-1)),Inf)<=r
            countR=countR+1;
        end
    end
end
countN=(length(vec)-k)*((length(vec)-k)+1);
corrint=2*countR/countN;
d=log(corrint)/log(r);
et=toc;
end
```

### • Função CorrIntNew.m

```
function [alph intcorr et] = CorrIntNew(serie,k,r)
import edu.stanford.covert.util.SparseMat;
%CorrIntNew Calcula a Dimensao de Correlacao a Partir de:
%   -uma dada serie temporal (serie)
%   -uma dada dimensao de imersao (k)
%   -uma distancia de vizinhanca (r)
%Devolve:
%   -alph: dimensao de correlacao
%   -intcorr: o integral de correlcao
%   -et: tempo de computacao
if nargin==2
    r=.1;
end
if nargin==1
    r=.1;
    k=1;
end
tic
```

```

format longg;
N=length(serie);
serie=floor(((serie-min(serie))/(max(serie)-min(serie)))*(1/r));
clear a;clear count;clear DIM;
DIM= repmat(ceil(1/r+1),1,k);
if k==1
    count=zeros(DIM+1,1);
for i=1:(length(serie)-k+1)
    idx=1+serie(i);
    count(idx)=count(idx)+1;
end
pares=sum(count.^2)-sum(count);
numtotalpares=(N-k+1)*(N-k);
intcorr=pares/numtotalpares;
alph=log(intcorr)/log(r);
elseif k>1
a(2,1)=uint32(0);
count=SparseMat(a);
count((max(DIM))^k,1)=1;
for i=1:(length(serie)-k+1)
    coor=num2cell(1+serie(i:(i+k-1)));
    idx=sub2ind(DIM,coor{:});
    count([idx 1])=count([idx 1])+1;
end
pares=collapse(count.^2)-collapse(count);
numtotalpares=(N-k+1)*(N-k);
intcorr=pares/numtotalpares;
alph=log(intcorr)/log(r);
end
et=toc;
end

```

### • Função LyapunovExp.m

```

function [lyap] = LyapunovExp(serie,k)
%LyapunovExp Calcula o Expoente de Lyapunov a Partir de:
%     -uma dada serie temporal (serie)
%     -uma dada dimensao de imersao (k)
%Devolve:
%     -lyap: o expoente de lyapunov
if nargin==1
    k=1;
end
N=length(serie);
M=N-k;
Y=Reconstruct(serie,k);
r=1e-9; % vizinhanca inicial
l=1;
dis=0;
while (length(dis)<50)
    for i=1:k:floor(N/2)
        x0=ones(M,1)*Y(i,:);
        distance=max(abs(Y-x0),[],2);
        for j=1:(length(distance))
            if (distance(j)<r && distance(j)>0 && length(dis)<100)

```

```

        dis(l)=distance(j);
        posinic(l)=i;
        posiprox(l)=j;
        l=l+1;
    end
end
end
r=r*10; % incremento em caso de falta de pontos
if r>max(Y(:))
    error('Serie Demasiado Pequena Para Estimacao')
end
end
lyap=zeros(length(dis),1);
if exist('posiprox','var')
    for p=1:length(dis);
clear dist;
l=1;
        for S=0:k:(min(M-posiprox(p)-1,4*k));
dist(l)=max(abs(Y(posinic(p),:)-Y(posiprox(p)+S,:)),[],2);
l=l+1;
        end
record(p,:)=dist;
    end
S=0:k:(min(M-posiprox(p)-1,4*k));
fit1=polyfit(S'+1,log(mean(record))',1);
lyap=fit1(1);
else
d=Inf;
end
end
end

```

### • Função Analogs.m

```

function [prev1 prevp et] = Analogs(serie,k,ep)
%Analogs Calcula Uma Previsao Para o Proximo Elemento a Partir de:
%     -uma dada serie temporal (serie)
%     -uma dada dimensao de imersao (k)
%     -uma dada vizinhanca (ep)
%Devolve:
%     -prev1: previsao para o proximo elemento da serie temporal
%     -prevp: previsao para o proximo vector reconstruido
%     -et: tempo de computacao
if nargin==2
    ep=.0001;
end
if nargin==1
    k=1;
    ep=.0001;
end
tic;
N=length(serie);
M=N-k;
Y=Reconstruct(serie,k);
miny=min(Y(:));
amp=(max(Y(:))-miny);

```

```

Y=(Y-miny)/amp;
num=0;
den=0;
count=0;
l=1;
xT=ones(M,1)*Y(M,:);
while (length(count)<11) % existe sempre pelo menos 10 analogos
    for i=1:(M-1)
        if (max(abs(xT(i,:)-Y(i,:)))<ep)
            dis=max(max(abs(xT(i,:)-Y(i,:))),1e-014); % dis != 0
            num=num+Y(i+k,:)*(1/dis);
            den=den+1/dis;
            count(l)=1;
            l=l+1;
        end
    end
    ep=ep*10;
end
prevp=(num./den).*amp+miny;
prevl=prevp(1);
et=toc;
end

```

#### • Função Reconstruct.m

```

function Y=Reconstruct(serie,k)
%Reconstruct Calcula Matriz do Espaco de Fases Reconstruido a partir de:
%     -uma dada serie temporal (serie)
%     -uma dada dimensao de imersao (k)
%Devolve:
%     -Y: matriz de vectores de reconstrucao
N=length(serie);
Y=zeros(N-k,k);
for i=1:k
    Y(:,i)=serie((1:(N-k))+(i-1))';
end
end

```