

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

**Análise e classificação de séries
temporais não estacionárias utilizando
métodos não-lineares**

por

MARCELO RESENDE THIELO

Dissertação submetida à avaliação, como requisito parcial para
a obtenção do grau de Mestre em
Ciência da Computação

Prof. Dante Augusto Couto Barone
Orientador

Porto Alegre, março de 2000

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Thielo, Marcelo Resende

Análise e classificação de séries temporais não estacionárias utilizando métodos não lineares/ por Marcelo Resende Thielo. – Porto Alegre: PPGC da UFRGS, 2000.

102p:il.

Dissertação(mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2000. Orientador: Barone, Dante Augusto Couto.

1. Séries Temporais 2. Caos 3. Clustering I. Barone, Dante Augusto Couto. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitora: Profª Wrana Panizzi

Pró-Reitor de Pós-Graduação: Prof. Franz. Rainer Semmelmann

Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux

Coordenadora do PPGC: Profª Carla Maria Dal Sasso Freitas

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

Agradecimentos

Agradeço à CAPES pela bolsa concedida, devido à qual a realização deste trabalho se fez possível.

Um grupo de pessoas que teve comparável importância no decorrer destes anos de trabalho duro foram a minha família e meu orientador. Não seria exagero se eu dissesse que considero o professor Dante Barone como um familiar muito próximo. A esta pessoa de mente e espírito privilegiados devo um “muito obrigado” dos grandes.

Quanto a estas pessoas, acho que ‘agradecer’ é muito pouco frente à dívida que tenho com elas. Minha mãe Mariluz, vó Alda, mana Marisol, tia Rosângela e prima Aline sempre mantiveram o pique e seguraram as pontas nos momentos difíceis em que infelizmente não pude estar presente. Muito obrigado, mulheres da minha vida. Vocês são o máximo.

Sem amigos não somos ninguém. Um muito obrigado ao amigo Günther e sua adorável esposa Suzana pela amizade e frutíferas discussões técnicas. Obrigado também ao pessoal que ficou em Rio Grande, em especial aos colegas da física (Ives), por terem mantido os antigos laços mesmo com minhas pouco frequentes viagens à terrinha. Obrigado Maurício e André pelo companheirismo durante esse período.

Ao pessoal /97 (Andre, Mauro, Francine, Sandra, Edson, Evandro, Hércules) com quem fiz minhas primeiras amizades em PoA, ao pessoal /98, e todos aqueles com quem costumo trocar pelo menos um ‘oi’ no corredor, agradeço pelo calor humano.

Para finalizar, agradeço aos professores e funcionários do Instituto de Informática, (especialmente à professora Ana Lúcia Bazzan), do Instituto de Física (em especial aos professores Márcia Barbosa, Leonardo Brunnet e Jason Gallas), ao professor visitante Thomas Schreiber da Universidade de Wuppertal, Alemanha, e todos aqueles com quem tive algum contato nestes últimos anos e que tenham contribuído positivamente de alguma forma para a realização deste trabalho.

In memoriam

Ricardo Juliano

Sumário

<i>Lista de Abreviaturas</i>	7
<i>Lista de Figuras</i>	8
<i>Lista de Tabelas</i>	10
<i>Resumo</i>	11
<i>Abstract</i>	12
1 Introdução	13
1.1 Objetivos.....	14
1.2 Estrutura do texto.....	14
2 Sistemas Lineares	15
2.1 Sistemas lineares em tempo contínuo	16
2.2 Sistemas lineares em tempo discreto.....	17
2.3 Sistemas lineares + estocasticidade	18
3 Sistemas Não-Lineares	20
3.1 Dinâmica de mapas iterados.....	23
3.2 Pontos fixos.....	25
3.3 Pontos n-periódicos	26
3.3.1 Pontos periódicos atrativos e repulsivos	27
3.4 Bacias de atração	28
3.5 Pontos de período três	28
3.5 Caos (alta sensibilidade às condições iniciais).....	29
3.6 Dependência sensitiva às condições iniciais.....	29
3.7 Expoentes de Lyapunov	30
3.8 Dimensões	32
3.9 Estacionariedade.....	35
4 Tratamento de Dados Experimentais	36
4.1 Métodos Clássicos para análise de sinais.....	36
4.2 Modelos estocásticos	36
4.3 Análise não-linear de séries temporais	39
4.3.1 Testes para detecção de não-linearidade	39
4.3.2 Reconstrução do espaço de fases	40
4.3.3 Ajuste dos parâmetros m e τ	41
4.3.4 Função de autocorrelação linear.....	41
4.3.5 Informação mútua	42
4.3.6 Falsos vizinhos.....	44

4.3.7 Filtragem.....	46
4.3.8 Expoentes de Lyapunov	47
4.3.9 O método de Rosenstein.....	48
4.4 Exemplo com uma série real.....	49
5 Estacionariedade em Séries Temporais	56
5.1 Medidas de estacionariedade.....	56
5.2 Clusterização.....	57
5.2.1 Otimização combinatorial.....	58
5.2.2 Algoritmos genéticos (AG).....	58
5.3 Medidas de dissimilaridade	59
5.3.1 Estacionariedade fraca	59
5.3.2 Erros de predições cruzadas	60
5.3.3 Somas de correlações cruzadas normalizadas	61
5.3.4 Distribuições suavizadas	61
6 Experimento	63
6.1 Séries sintéticas	64
6.2 Séries reais (eletroencefalogramas).....	68
7 Conclusão	82
Anexo 1 Gráficos de Resultados.....	85
Bibliografia.....	99

Lista de Abreviaturas

EEG = eletroencefalograma

FNN = false nearest neighbour(s) – falsos vizinhos

AG = Algoritmo Genético

Lista de Figuras

FIGURA 2.1 – Diagrama de primeiro retorno para $a < 1$.	17
FIGURA 2.2 – Diagrama de primeiro retorno para $a = 1$.	18
FIGURA 2.3 – Diagrama de primeiro retorno para $a > 1$.	18
FIGURA 3.1 – Discretização por meio de uma seção de Poincaré.	21
FIGURA 3.2 - Discretização por mapa de tempo τ .	22
FIGURA 3.3 – Diagrama de primeiro retorno	24
FIGURA 3.4 – Trajetória em um mapa de primeiro retorno.	24
FIGURA 3.5 – Diagrama de bifurcações para o mapa logístico	25
FIGURA 3.7- Diagrama da bacia de atração (azul) para o atrator de Hénon (branco). As outras cores estão relacionadas com a quantidade de iterações necessárias para que o valor de x exceda um certo limite positivo k .	28
FIGURA 3.8 – A ordenação de Šarkovskii.	29
FIGURA 3.11 – O Conjunto de Cantor	33
FIGURA 4.1 – Espectro de potências para o sistema de Lorenz, em regime caótico. Mesmo na ausência de estocasticidade, há uma banda contínua em torno do pico central.	39
FIGURA 4.2 – Notação ‘standard’ para equações diferenciais	40
FIGURA 4.3 - Função de autocorrelação para o atrator de Lorenz. O primeiro ponto onde $\sigma < e^{-1}$ é $\tau \approx 7$.	41
FIGURA 4.4 - Atrator de Lorenz reconstruído a partir da coordenada x com $\tau = 7$.	42
FIGURA 4.5 - Informação mútua do atrator de Lorenz. O primeiro mínimo ocorre em $\tau = 3$. O valor escolhido para reconstrução foi $\tau = 2$.	43
FIGURA 4.6 - Lorenz reconstruído a partir da coordenada x com $\tau = 2$	44
FIGURA 4.7 – Percentual de falsos vizinhos (vert) versus dimensão de imersão (horiz) para o sistema de Lorenz.	45
FIGURA 4.8 - Série temporal do laser de NH_3 utilizado na competição de séries temporais do Santa Fe Institute. O eixo vertical representa a amplitude do sinal e o eixo horizontal o tempo.	50
FIGURA 4.9– Detalhe da série temporal do laser de NH_3 . Os eixos são análogos aos da figura 4.8	50
FIGURA 4.10 - Função de autocorrelação para o laser de NH_3 . Na vertical: Correlação σ vs tempo na horizontal.	51
FIGURA 4.11 - Informação mútua para os dados do laser, (I vs τ). O primeiro mínimo utilizado para a escolha do atraso temporal equivale a $\tau = 3$.	51
FIGURA 4.12 – Percentual de falsos vizinhos (vert) versus dimensão de imersão para a série da figura 4.9. Mesmo tendo o gráfico alcançado um valor suficientemente ‘pequeno’ apenas em $m = 6$, assumimos que o sinal está contaminado com ruído de amostragem, de forma que a efetiva dimensão do sistema seja inferior a este valor.	52
FIGURA 4.13 – Gráfico do ‘atrator reconstruído’ tridimensional (x, y, z) com $m = 3$ e $\tau = 3$ a partir do sinal do laser de NH_3	52
FIGURA 4.14 – Gráfico de divergências médias (vertical) para os dados do laser de NH_3 segundo o método de Rosenstein. O valor estimado para o maior expoente de Lyapunov fica em torno de 0.0375.	53
FIGURA 4.15 – Série temporal obtida a partir do eixo z (z vs t) do modelo de Lorenz, para fins de comparação com o laser de NH_3 .	53
FIGURA 4.16 – Detalhe da figura 4.15, eixos análogos.	54
FIGURA 4.17 – Atrator reconstruído a partir do eixo z do sistema de Lorenz.	54
FIGURA 4.18 - Gráfico de divergências médias (vertical) para o sistema de Lorenz segundo o método de Rosenstein. O valor estimado para o maior expoente de Lyapunov fica em torno de 0.125.	55
FIGURA 5.1 Gráficos de recorrência para o mapa de Ikeda estacionário (esq) e com variação contínua de um de seus parâmetros (dir). O significado dos eixos está contido no texto da seção 5.1.	57
FIGURA 5.2 – O algoritmo genético 1+1.	59
FIGURA 5.3 – Operador de mutação modificado	59
FIGURA 6.1 – Modelos geradores das séries numéricas utilizadas para validação. Os sistemas representados por equações diferenciais foram integrados utilizando o método Runge-Kutta de quarta ordem e discretizados por meio de uma seção de Poincaré.	63
FIGURA 6.2 – Séries obtidas a partir de iteradas do sistema de Roessler (coluna da esquerda) e do mapa de Ikeda (coluna da direita).	64

FIGURA 6.3 – Matriz de dissimilaridades para o sistema de Lorenz, sem adição de ruído, utilizando correlações cruzadas.....	66
FIGURA 6.4 – Resultado clusterização para a matriz da figura 6.3, a partir de inicialização ordenada ou aleatória.....	67
FIGURA 6.5 – matriz de dissimilaridade para o sistema de Lorenz sem ruído utilizando previsões cruzadas.....	67
FIGURA 6.6 – Resultado da clusterização para a matriz da figura 6.5, partindo de inicialização sem organização.....	67
FIGURA 6.7 – Resultado da clusterização para a matriz da figura 6.5, inicialização organizada.....	68
FIGURA 6.8 – De cima para baixo, segmentos de 500 pontos dos estágios REM, 1, 2, 3, 4 amostrados nos pontos C_3 - A_2 a uma taxa de 128 Hz.....	68
FIGURA 6.9 - Estágios e ciclos de sono ao longo do tempo. As barras mais escuras indicam sono REM.....	69
FIGURA 6.10 – Posição dos eletrodos para a montagem C_3 - A_2	71
FIGURA 6.11 – Distribuições empíricas de segmentos de 2048 pontos dos estágios 1, 2, 3, 4 e REM.....	72
FIGURA 6.12 – Modelo para as distribuições, onde a , σ , μ e i são parâmetros a serem ajustados.....	72
FIGURA 6.13 – Comparação entre distribuições empírica e teórica.....	72
FIGURA 6.14 – Matriz de dissimilaridade para o início de um ‘ciclo’ de sono do paciente 2.....	73
FIGURA 6.15 - Resultado sobre o segmento da figura anterior (163840) pontos, sendo que 128 pontos=1 segundo. ...	73
FIGURA 6.16 - Matriz de dissimilaridade para um ciclo de sono do paciente 1.....	74
FIGURA 6.17 – Resultado da clusterização para o paciente 1, apresentando uma perceptível continuidade temporal nos clusters encontrados.....	74
FIGURA 6.18 – Matriz de dissimilaridades para o paciente 2.....	75
FIGURA 6.19 – Resultado da clusterização para a matriz da figura 6.11.Note que o número de elementos utilizados foi reduzido em função do ciclo utilizado ter sido mais curto que os demais.	75
FIGURA 6.20 – Matriz de dissimilaridades para um ciclo de sono do paciente 3.....	76
FIGURA 6.22 - Comparação de dois segmentos do estágio 1 com overlap.	77
FIGURA 6.23 - Comparação de dois segmentos disjuntos do estágio 1.....	77
FIGURA 6.24 - Dois segmentos disjuntos estágio 2.....	78
FIGURA 6.25 – Comparação entre segmentos pertencentes aos estágios 1 e 2.....	78
FIGURA 6.26 – Comparação entre segmentos pertencentes aos estágios 2 e 3.....	79
FIGURA 6.27 –Comparação entre segmentos pertencentes aos estágios REM e 1.....	80

Lista de Tabelas

Tabela 2.1 – Tipo de pontos fixos para sistemas lineares	16
Tabela 3.1 – Exemplos de mapas.....	23
Tabela 6.1 – Desempenho das diferentes medidas de dissimilaridade frente aos sinais gerados.....	65

Resumo

Neste trabalho fazemos revisão de alguns dos principais métodos para análise não-linear de séries temporais originadas a partir de sistemas de baixa dimensionalidade com dinâmica predominantemente determinística, dando ênfase ao problema de classificação/clusterização não-supervisionada destas mesmas séries. Várias medidas de dissimilaridade são utilizadas em conjunto com métodos heurísticos baseados em algoritmos estocásticos, para a organização de segmentos de séries temporais não estacionárias em grupos com características em comum, na tentativa de associar a estes alguma característica clínica previamente conhecida. O método é implementado com diferentes medidas de dissimilaridade e um experimento feito com séries temporais sintéticas (obtidas a partir de simulação numérica) com fins de validação e posteriormente aplicado a um problema real, o problema de segmentação de estágios de sono. Os resultados indicam certa promissoriedade do método para aplicação na classificação estágios de sono em eletroencefalogramas.

Palavras-Chave: Séries temporais, caos, clustering

TITLE: "ANALYSIS AND CLASSIFICATION OF NONSTATIONARY TIME SERIES WITH NONLINEAR METHODS"

Abstract

In this work we make a review of some of the main methods available for nonlinear time series analysis for low-dimensional deterministic systems, giving emphasis to the problem of unsupervised classification/clustering of this kind of data. Various dissimilarity measures are used together with heuristic search methods based on stochastic algorithms to organize segments of one (big) nonstationary time series in groups with common characteristics, trying to relate these groups to some known clinical property. The method is implemented with different dissimilarity measures and one experiment made with synthetic (generated by numerical simulations) time series for validation and lately applied to a real problem, the problem of sleep stages segmentation. The results look promising with respect to the applicability of the method to classify sleep stages in electroencephalographic recordings.

Keywords: Time Series, Chaos, Clustering

1 Introdução

Desde os tempos mais remotos o homem vem tentando elaborar modelos que descrevam adequadamente os fenômenos que se observa na natureza. De posse de uma boa descrição de um mecanismo da natureza podemos ser capazes de fazer previsões, controlar, e talvez até *entender* alguns aspectos do sistema observado. Se olharmos para os primórdios da ciência, encontraremos descrições que nos parecerão um tanto quanto bizarras mas que, levando-se em consideração a escassez de ferramental disponível na época, podem ser consideradas verdadeiras obras de arte. Por exemplo, o modelo desenvolvido por Johannes Kepler (que, não por acaso, era pitagórico) para descrever as órbitas dos planetas baseado em esferas e sólidos platônicos aproximava os dados observados com um erro razoavelmente pequeno o que, comparado ao modelo posteriormente desenvolvido por ele mesmo (e em uso até hoje), é uma aproximação bastante impressionante. Kepler elaborou uma descrição geométrica de um sistema físico - todavia, uma descrição que considere aspectos dinâmicos do sistema, ou seja, que forneça informação sobre o que irá acontecer no sistema em um certo tempo futuro dadas as condições atuais, é o que se objetiva normalmente em ciência. As leis dinâmicas a que daremos atenção neste trabalho são aquelas de natureza não-linear.

“As leis fundamentais da física microscópica, como as equações de Newton ou a equação de Schrödinger, ou da física macroscópica como as equações de Navier-Stokes ou a lei de ‘massa de ação’, são inadequadas para entender ou mesmo formular a complexidade induzida pela evolução de sistemas não-lineares. Em contraste atratores, fractais e multifractais, formas normais, expoentes de Lyapunov, entropias, medidas invariantes e funções de correlação são parte do novo vocabulário científico pela moderna ciência não-linear e provêm uma forma pragmática de encarar um desafio frente ao qual as abordagens clássicas falham.”[NIC95]

Neste trabalho serão discutidos rapidamente alguns aspectos teóricos básicos e mais intensivamente aplicações experimentais sobre interpretações não-lineares de sistemas físicos (e biológicos). Iniciaremos com algumas noções teóricas sobre dinâmica linear e processos estocásticos, logo em seguida abordando sistemas não lineares, regime caótico, métodos para evidenciamento experimental de comportamento caótico e caracterização de regimes, finalizando com alguns resultados obtidos a partir da aplicação de tais métodos em algumas séries de dados.

A metodologia adotada para análise de dados supostamente gerados por uma dinâmica caótica determinística basear-se-á nos teoremas de Takens [TAK81], que possuem algumas restrições sobre os dados que precisam ser respeitadas, como a de que estes pertençam a um regime dinâmico estacionário. O problema será abordado com base na proposta feita por Schreiber [SCH97a], e os resultados de um experimento realizado com dados experimentais obtidos a partir de simulações numéricas e leituras de polissonogramas serão interpretados com a finalidade de avaliar a aplicabilidade destes métodos na análise efetiva de dados reais.

1.1 Objetivos

O presente trabalho tem como objetivo geral explorar aspectos teóricos, experimentais e computacionais acerca de sistemas não lineares, concentrando-se na parte de tratamento de séries temporais não-estacionárias. Métodos existentes serão utilizados para análise de dados experimentais, suas vantagens e desvantagens discutidas, e uma diferente abordagem proposta.

Em vista da ainda predominância dos métodos espectrais, adequados apenas para sistemas com dinâmica predominantemente linear estocástica, objetiva-se, neste trabalho, explorar um pouco mais da potencialidade dos métodos baseados na chamada teoria do caos, onde comportamentos complicados podem ser interpretados como sendo gerados por um sistema não linear determinístico de baixa dimensionalidade em regime caótico.

Os conjuntos de dados utilizados serão compostos de séries geradas por computador e polissonogramas, que são dados eletroencefalográficos medidos em múltiplos canais durante o sono de um indivíduo. Os resultados serão interpretados com o auxílio de um médico eletroencefalografista.

1.2 Estrutura do texto

Na primeira metade deste trabalho iremos revisar alguns aspectos teóricos necessários como *background* para quem pretende aplicar algum método de tratamento de sinais. É comum mas extremamente desaconselhável se aplicar um determinado método sem a tentativa de relacionamento entre os fundamentos teóricos deste e a natureza (física, biológica ou mesmo a simples modelagem matemática) do fenômeno em estudo. Tratar sistemas reais como caixas pretas muitas vezes pode induzir conclusões extremamente equivocadas.

Inicialmente serão discutidos alguns aspectos elementares acerca de sistemas lineares e estocasticidade, e na segunda parte abordaremos uma interpretação não-linear para comportamento irregular. Intenciona-se que, após uma leitura desta introdução teórica, o leitor possa compreender mais facilmente os métodos apresentados na segunda parte deste trabalho.

2 Sistemas Lineares

Sistemas dinâmicos lineares são comuns em vários ramos da ciência[HUB95], e estão presentes em grande parte da física, por exemplo. Sistemas lineares autônomos (sem dependência explícita do tempo) pode ter sua evolução no tempo caracterizada em função dos autovalores da matriz de coeficientes. Suas soluções gerais apresentam-se como uma sobreposição de oscilações periódicas (amortecidas ou não). Uma classe de sistemas com comportamento complicado é formada a partir da combinação de dinâmica linear e estocástica, que será discutida na seção 2.3 a seguir. Faremos agora uma breve revisão de alguns aspectos sobre sistemas simples de equações diferenciais e de diferença lineares:

No caso de o modelo em questão considerar o tempo como um continuum, o sistema deve aparecer, de uma forma geral, como:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(t) \quad (2.1)$$

Se o tempo for discretizado, passamos a trabalhar com um mapa de tempo T ou uma seção de Poincaré, e a representação geral se torna:

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) + \mathbf{g}(t) \quad (2.2)$$

onde $\mathbf{f}(\mathbf{x})$ é um vetor de funções lineares que pode ser representada como uma simples matriz de coeficientes \mathbf{A} . Dessa forma:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{g}(t) \quad (2.3)$$

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{g}(t) \quad (2.4)$$

Lembramos que, na maioria das vezes, é necessário que o sistema seja autônomo, ou seja, \mathbf{g} não deve ser uma função explícita de t para que possamos obter soluções gerais representáveis com um conjunto finito de funções elementares. Por exemplo, nenhuma das seguintes equações de aspecto inocente pode ser resolvida por métodos *standard*[HUB95]:

$$\begin{aligned} \dot{x} &= x^2 - t, \\ \dot{x} &= \sin(tx) \\ \dot{x} &= e^{tx} \end{aligned} \quad (2.5)$$

2.1 Sistemas lineares em tempo contínuo

Soluções para um sistema bidimensional

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \quad (2.6)$$

com uma matriz associada do tipo

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (2.7)$$

apresentam-se, de um modo geral, sob a forma

$$\mathbf{x}(t) = \boldsymbol{\xi} e^{\lambda t} \quad (2.8)$$

Onde os autovalores λ_1 e λ_2 são obtidos a partir do cálculo das raízes da equação característica

$$\lambda^2 - (a + d)\lambda + ad - bc = 0 \quad (2.9)$$

Substituindo a equação 2.6 na (2.8) temos

$$\begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.10)$$

Substituindo-se nesta equação os autovalores obtidos anteriormente encontramos os autovetores

$$\boldsymbol{\xi}^{(1)} \text{ e } \boldsymbol{\xi}^{(2)} \quad (2.11)$$

Que, substituídos na equação 2.8 dão as soluções gerais

$$\mathbf{x} = c_1 \boldsymbol{\xi}^{(1)} e^{\lambda_1 t} + c_2 \boldsymbol{\xi}^{(2)} e^{\lambda_2 t} \quad (2.12)$$

A seguir temos uma análise resumida do comportamento das soluções em torno do ponto fixo 0 em função dos autovalores:

TABELA 2.1 – Tipo de pontos fixos para sistemas lineares

$\mathbf{x}' = \mathbf{A}\mathbf{x}$	$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$	$\det \mathbf{A} \neq 0$
Autovalores	Tipo de ponto crítico	Estabilidade
$\lambda_1 > \lambda_2 > 0$	Nódulo impróprio	Instável
$\lambda_1 < \lambda_2 < 0$	Nódulo impróprio	Assintoticamente estável
$\lambda_1 < 0 < \lambda_2$	Ponto de sela	Instável
$\lambda_1 = \lambda_2 > 0$	Nódulo próprio ou impróprio	Instável
$\lambda_1 = \lambda_2 < 0$	Nódulo próprio ou impróprio	Assintoticamente estável
$\lambda_1, \lambda_2 = a \pm ib$:	Ponto espiral	
$a > 0$		Instável
$a < 0$		Assintoticamente estável
$\lambda_1 = ib, \lambda_2 = -ib$	Centro	Estável

Uma discussão mais detalhada sobre problemas de autovalores, estabilidade e equações diferenciais pode ser encontrada, por exemplo, em [BOY98].

2.2 Sistemas lineares em tempo discreto

Nesta seção iremos discutir brevemente o caso unidimensional para mapas lineares do tipo $\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n$, ou seja, $x_{n+1} = ax_n$. A taxa de dissipação do sistema é dada por $1-|a|$. Analisemos os três casos $|a| < 1$, $|a|=1$, $|a|>1$.

Admitamos que uma determinada população na geração $(n+1)$ seja univocamente determinada pela população da geração precedente (n) de acordo com alguma lei do tipo $x_{n+1}=F(x_n)$. O modelo mais simples possível para uma taxa de crescimento populacional é a hipótese Maltusiana de taxa de reprodução constante, isto é, $F(x)=ax$. Para $|a| > 1$ ocorrerá obviamente uma explosão populacional (crescimento exponencial), para $|x_0|> 0$ (que é um ponto fixo repulsivo do mapa). No caso de $|a| < 1$ a tendência da população é diminuir a cada iteração e o ponto 0 é um ponto fixo atrativo do mapa. O parâmetro de controle a mede a taxa de crescimento (ou diminuição) monotônica, e geometricamente representa a inclinação da reta descrita pela função linear $F(x)$.

Primeiro caso: $a < 1$. Obviamente, para qualquer x_n diferente de zero, x_{n+1} será exatamente uma fração equivalente a $(a \times 100)$ % de x_n . Dessa forma, x_{n+1} irá decrescer sempre na mesma razão de x_n , que é a . O n -ésimo valor de x pode ser expressado como

$$x_n = x_0 a^n \quad (2.13)$$

Um diagrama de primeiro retorno para um valor de $a < 1$:

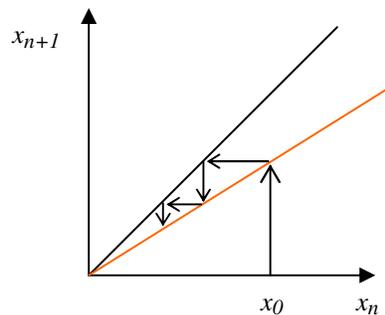
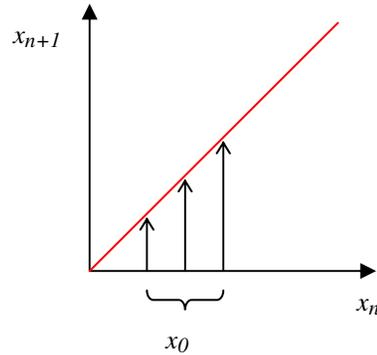
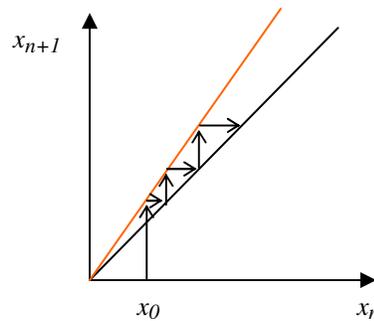


FIGURA 2.1 – Diagrama de primeiro retorno para $a < 1$.

Para $a = 1$, a taxa de dissipação é nula, e temos então um sistema conservativo. Note que não existe um ponto atrativo ou repulsivo no mapa:

FIGURA 2.2 – Diagrama de primeiro retorno para $a=1$.

Para $|a| > 1$, o sistema recebe energia indefinidamente até divergir ($|x| \rightarrow \infty$). Este modelo não possui uma interpretação física, mas está relacionado com os autovalores da matriz jacobiana local de sistemas não-lineares em regime caótico.

FIGURA 2.3 – Diagrama de primeiro retorno para $a>1$.

2.3 Sistemas lineares + estocasticidade

No caso de $g(t)$ não poder ser desprezado, retornamos ao sistema

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(t) \quad (2.14)$$

ou

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) + \mathbf{g}(n) \quad (\text{tempo discreto } n) \quad (2.15)$$

que passa a ser chamado não-homogêneo. Sistemas com uma $g(t)$ não nula ou não constante raramente possuem soluções gerais analiticamente obtíveis sob a forma de um número finito de funções elementares, exceto quando $g(t)$ é formada de combinações lineares de finitos termos com a forma $t^k e^{at}$. Quando analiticamente calculáveis, tais resultados freqüentemente mostram-se tão

complicados de serem interpretados que tornam-se praticamente inúteis. Aproximar $g(t)$ por uma sobreposição de infinitos termos deste mesmo tipo é uma prática comum via transformada de Fourier o que, apesar de não garantir solução analítica geral, pode facilitar a obtenção numérica de $\mathbf{x}(t)$.

Em processos reais, muitas vezes $g(t)$ apresenta-se como uma componente de ruído dinâmico de natureza estocástica. Tais modelos são normalmente estudados utilizando-se métodos espectrais (seção 4.2), tratados nos livros de probabilidade, e não nos aprofundaremos em seus aspectos teóricos aqui.

3 Sistemas Não-Lineares

Um sistema não-linear é um conjunto de equações não-lineares, as quais podem ser algébricas, funcionais, diferenciais ordinárias, diferenciais parciais, integrais ou uma combinação destas, e podem possuir dependência de um ou mais parâmetros. “Sistemas dinâmicos” é um termo agora utilizado como sinônimo de sistema não linear quando as equações não-lineares representam a evolução de uma solução no tempo ou alguma variável similar a tempo. Também podemos interpretar um sistema não linear como um elo de retroalimentação no qual a saída de um elemento não é proporcional à sua entrada[DRA92].

Nesta seção será dada atenção principalmente aos sistemas dinâmicos em tempo discreto ou mapas iterados, sempre tendo em mente que tais mapas estão relacionados com a aplicação de algum método de discretização a um sistema contínuo real, pois estes carregam em sua dinâmica muitas características do sistema original e a interpretação em tempo discreto é a mais adequada para este trabalho, que visa tratamento de sinais experimentais medidos em intervalos de tempo fixos.

Sistemas dinâmicos não-lineares com tempo discreto podem ser representados, de forma similar aos sistemas lineares, como uma equação de diferenças:

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) + \mathbf{g}(t) \quad (3.1)$$

Neste caso, $\mathbf{f}(\mathbf{x})$ é um vetor/matriz de funções não-lineares e não uma simples matriz de coeficientes. Uma análise de autovalores passa a ter uma interpretação local e soluções globais passam a ser possíveis em casos muito raros, mesmo com $\mathbf{g}(t) = 0$.

Dos diferentes métodos de discretização existentes, faremos menção a dois: O mapa de tempo τ e o mapa de Poincaré.

Toda equação diferencial dá origem a um mapa chamado “mapa de tempo um”, definido pelo avanço do fluxo uma unidade de tempo. Se a equação diferencial contém um termo ou termos periódicos no tempo, então o mapa de tempo τ (onde τ é o período) é muito útil. O mapa de tempo τ em um sistema com termos periódicos é também chamado um mapa estroboscópico, visto que estamos olhando efetivamente para uma posição no espaço de fases com uma “iluminação estroboscópica” ajustada com período τ .

Em sistemas autônomos (sem termos dependentes do tempo nas equações), pode também ser possível definir uma seção de Poincaré e reduzir a dimensão do espaço de fase de uma de suas componentes. Aqui a discretização é definida não por meio de um intervalo de tempo fixo, mas por tempos sucessivos quando uma órbita cruza um hiperplano fixo no espaço de fases. Este hiperplano nada mais é do que uma variedade com uma dimensão a menos que o espaço de fases.

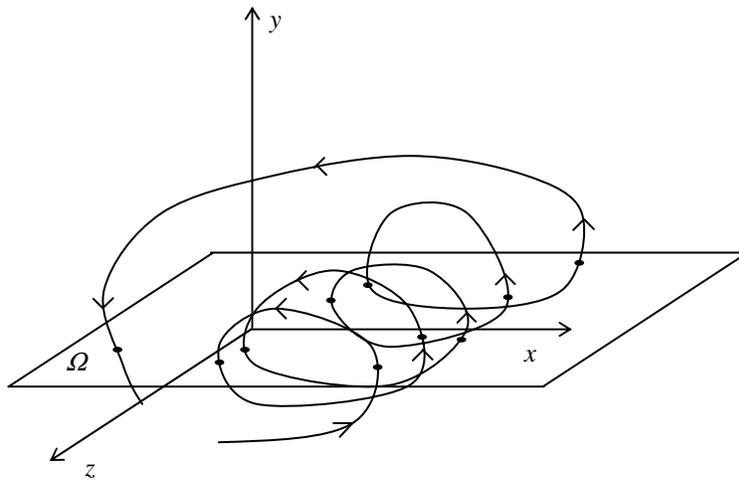


FIGURA 3.1 – Discretização por meio de uma seção de Poincaré

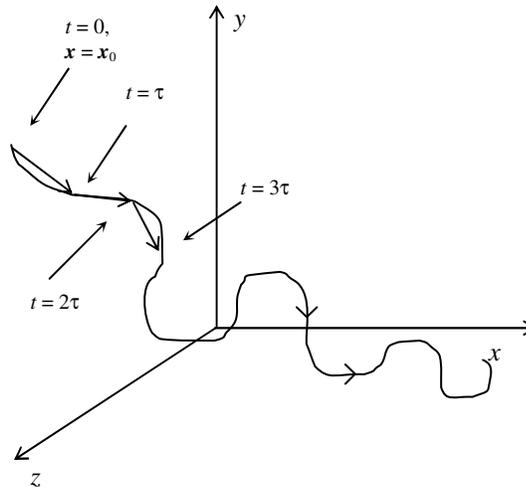


FIGURA 3.2 - Discretização por mapa de tempo τ

É importante que se tenha em mente que nem todo fluxo tem uma seção de Poincaré global, a qual precisaria ser transversa à toda possível órbita. Diferentes escolhas para o plano de secção podem resultar em mapas bastante diferentes.

Mapas originados de amostragem estroboscópica ou seção de Poincaré de um fluxo são necessariamente invertíveis, pois o fluxo possui uma única solução através de um ponto no espaço de fase – a solução é única tanto para a frente ou para trás no tempo. De qualquer modo, mapas não-invertíveis podem ser relevantes para estudo de sistemas descritos por equações diferenciais: mapas de Poincaré são, às vezes, muito bem aproximados por mapas não-invertíveis. Por exemplo, o mapa de Hénon

$$H_{ab} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix} \quad (3.2)$$

com valor pequeno de $|b|$ é próximo ao mapa quadrático.

Veja que qualquer solução por procedimento numérico para um problema diferencial de valor inicial que utiliza passos discretos de tempo na aproximação é efetivamente um mapa (e.g. método de Runge-Kutta de 4ª ordem). Esta não é uma observação trivial, e ajuda a explicar por que um sistema de tempo contínuo que não deve exibir caos pode ter soluções numéricas que o façam.

3.1 Dinâmica de mapas iterados

Como visto anteriormente, um sistema dinâmico pode ser temporalmente discretizado através de uma seção de Poincaré ou um mapa de tempo τ . Tal simplificação permite que se estude alguns aspectos da dinâmica do sistema original de uma forma mais fácil, através da análise das órbitas obtidas a partir de iterações sucessivas dos mapas discretos. Consideremos x_0 como sendo uma dada condição inicial e um mapa $f : \mathfrak{R} \rightarrow \mathfrak{R}$. Obtemos a órbita correspondente fazendo sucessivas composições da função f , da seguinte forma:

$$x_0, f(x_0), f(f(x_0)), f(f(f(x_0))), f(f(f(f(x_0)))) \dots \quad (3.3)$$

assim,

$$x_n = f(x_{n-1}) \quad (3.4)$$

Chamamos $f(x_0)$ a primeira iterada de x_0 para f , $f(f(x_0))$ a segunda iterada de x_0 para f , e assim sucessivamente. Uma notação mais conveniente é utilizar $f^{[1]}$ para a primeira iterada da função, de forma que $f^{[n]}$ denota a n-ésima iterada de f .

Exemplo:

TABELA 3.1 – Exemplos de mapas

$f(x)$	x_0	Órbita de x_0 para f
$f(x)=x^2$	1	1, 1, 1, 1, 1, 1, ...
$f(x)=x^2-1$	-1	-1, 0, -1, 0, -1, 0, -1, 0, ...
$f(x)=ix$	1	1, i , -1, $-i$, 1, i , -1, $-i$, ...
$f(x)=4x+4x^2$	$\frac{1}{4}$	0.3333333..., 0.8888888..., 0.395061..., 0.955951...,...

Um dos principais métodos utilizados para a análise da dinâmica de um mapa(unidimensional) é a análise gráfica das iteradas, ou mapa de primeiro retorno. Consiste basicamente em um gráfico da função em seu domínio, e uma bissetriz diagonal com inclinação de 45 graus.

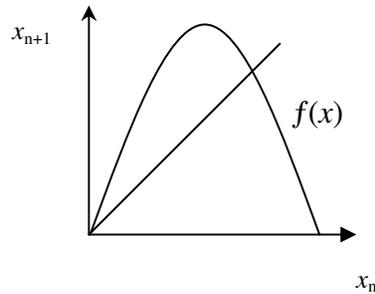


FIGURA 3.3 – Diagrama de primeiro retorno

A partir de uma condição inicial x_0 , procede-se da seguinte maneira: a partir do ponto x_0 , traça-se uma linha perpendicular ao eixo x_n , até a curva da função. Em seguida, une-se o ponto onde a linha toca a curva da função à bissetriz por uma reta paralela ao eixo de x_n , e retorna-se à curva da função através de outra reta perpendicular ao eixo de x_n , e assim sucessivamente, como mostrado na figura 3.4.

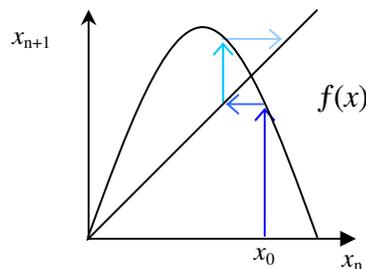


FIGURA 3.4 – Trajetória em um mapa de primeiro retorno

A coleção de coordenadas x_n dos pontos onde as setas atingem a curva da função formam a órbita do mapa f a partir da condição inicial x_0 .

Modelos com componentes não-lineares oferecem uma dinâmica bastante rica, e merecem uma atenção especial. O modelo clássico para estudo de dinâmica não-linear é o mapa logístico, que é um mapa unidimensional unimodal. Mapas unidimensionais unimodais são mapas definidos num intervalo finito e com um único máximo. Mais exatamente, mapas unimodais são mapas continuamente diferenciáveis que mapeiam o intervalo unitário $[0,1]$ nele mesmo, que apresentam um máximo único em $x = 1/2$ e são monotônicos para $0 \leq x \leq 1/2$ e $1/2 < x \leq 1$. Tais mapas apresentam padrões de bifurcação similares quando o parâmetro de controle é variado[FIE94].

O citado mapa logístico $x_{n+1} = \mu x_n(1 - x_n) = F_\mu(x_n)$ apresenta uma dinâmica impressionantemente rica sob a forma de cascatas de bifurcação e caos em função da variação do parâmetro μ . Tal mapa é um modelo (extremamente simplificado) de crescimento populacional, e tem a seguinte interpretação: Seja x_n a quantidade de indivíduos em uma população no instante n . Admite-se que a taxa de reprodução diminua quando a população torna-se grande. Com efeito, o segundo termo ($-\mu x_n^2$), dominante para x grande, é negativo e contribui para a diminuição relativa da população.

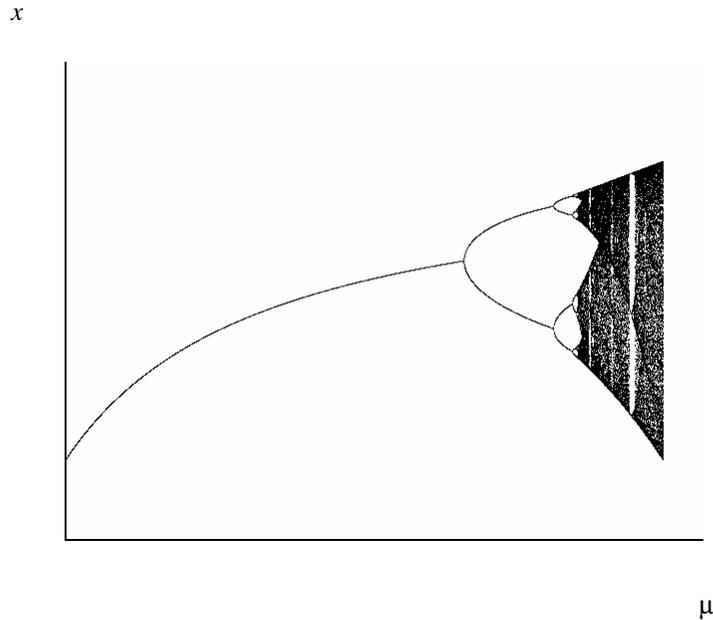


FIGURA 3.5 – Diagrama de bifurcações para o mapa logístico

3.2 Pontos fixos

Um ponto cujas iteradas são o próprio ponto é chamado um ponto fixo. Pontos fixos podem ser atrativos, repulsivos ou pontos de bifurcação (pontos onde existe uma mudança de natureza).

Definição: Esteja p no domínio de f . Então p é um ponto fixo de f se $f(p)=p$.

Graficamente, um ponto p no domínio de f é um ponto fixo de f se e somente se o gráfico de f toca (ou cruza) a linha $y = x$ em (p,p) .

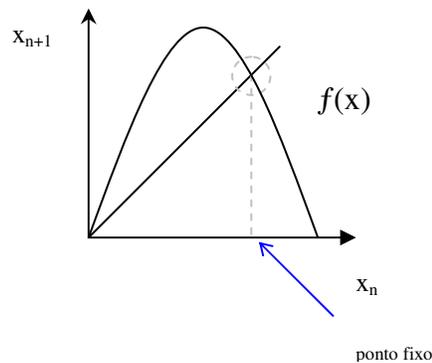


FIGURA 3.6 - Ponto fixo de um mapa

Aplicando a análise gráfica, podemos observar diversos comportamentos para as iteradas de vários pontos. Somos, portanto, levados às seguintes definições

- O ponto p é um ponto fixo atrativo de f quando existir um intervalo $(p - \varepsilon, p + \varepsilon)$ contendo p tal que se x pertence ao intervalo, então $f^{[n]}(x) \rightarrow p$ quando n cresce sem limites. Tais pontos são também chamados de pontos assintoticamente estáveis na literatura.
- O ponto p é um ponto fixo repulsivo de f quando existir um intervalo $(p - \varepsilon, p + \varepsilon)$ contendo p tal que se x pertence ao intervalo e $x \neq p$, então $|f(x) - p| > |x - p|$.

Outros tipos de pontos fixos, que não são nem atrativos nem repulsivos, podem ocorrer.

Critério: Suponha que a função f seja diferenciável num ponto fixo p . Então

- Se $|f'(p)| < 1$, então p é atrativo.
- Se $|f'(p)| > 1$, então p é repulsivo.
- Se $|f'(p)| = 1$, então p pode ser atrativo, repulsivo, ambos (ponto de sela) ou nenhum dos dois casos.

3.3 Pontos n-periódicos

Dizemos que um ponto x_0 tem período n (ou é um ponto n -periódico) se $f^{[n]}(x_0) = x_0$, sendo todos os pontos $x_0, f(x_0), f^{[2]}(x_0), \dots, f^{[n-1]}(x_0)$ distintos. Se x_0 tem período n , então a órbita de x_0 , que é a órbita

$$\{x_0, f(x_0), f^{[2]}(x_0), \dots, f^{[n-1]}(x_0)\}, \quad (3.5)$$

é uma órbita periódica, sendo chamada de um ciclo- n . Por convenção, chamamos ciclos-1 de pontos fixos.

3.3.1 Pontos periódicos atrativos e repulsivos

Suponha que x é um ponto n -periódico de f . Então x é um ponto fixo de $f^{[n]}$ e, portanto, temos um meio natural de definir pontos periódicos atrativos e repulsivos.

Seja x um ponto n -periódico de uma função f . Então x é um ponto n -periódico atrativo se x for um ponto fixo atrativo de $f^{[n]}$; analogamente, x é um ponto n -periódico repulsivo se x é um ponto fixo repulsivo de $f^{[n]}$.

Seja x um ponto n -periódico de f . Se x for atrativo (repulsivo), então cada um dos pontos da seqüência $\{x, f(x), f^{[2]}(x), \dots, f^{[n-1]}(x)\}$ é um ponto n -periódico atrativo (repulsivo), de modo que dizemos ser o ciclo- n $\{x, f(x), f^{[2]}(x), \dots, f^{[n-1]}(x)\}$ é atrativo (repulsivo).

Em particular, se $n = 2$, então o ponto 2-periódico x é atrativo se e somente se existir um intervalo $(x - \epsilon, x + \epsilon)$ tal que, sempre que y pertencer a este intervalo,

$$f^{[2n]}(y) \rightarrow x \text{ e } f^{[2n+1]}(y) \rightarrow f(x) \quad (3.6)$$

à medida que n aumenta sem limites.

Apresentamos anteriormente um critério para discriminar pontos fixos atrativos e pontos fixos repulsivos. Tal critério envolve a derivada da função. Analogamente, existe um critério para ciclos atrativos e repulsivos, envolvendo derivadas. Antes de enunciá-lo, consideremos um resultado preliminar.

Seja $\{x, z\}$ um ciclo-2 de f . Se $f^{[2]}$ for diferenciável em x e em z , então

$$(f^{[2]})'(x) = f'(x)f'(z) = (f^{[2]})'(z). \quad (3.7)$$

Isto é fácil de se provar. Basta usar a regra da cadeia e efetuar a derivada indicada:

$$(f^{[2]})'(x) = (f \circ f)'(x) = [f'(f(x))][f'(x)] = f'(x)f'(z). \quad (3.8)$$

Por simetria, temos também $(f^{[2]})'(z) = f'(x)f'(z)$.

Critério: Seja $\{x, z\}$ um ciclo-2 de f .

- Se $|f'(x)f'(z)| < 1$, então o ciclo-2 é ATRATIVO.
- Se $|f'(x)f'(z)| > 1$, então o ciclo-2 é REPULSIVO.

Se $|f'(x)f'(z)| = 1$, então não se pode concluir nada sobre o ciclo ser ou não atrativo, repulsivo ou nenhum dos dois. Por exemplo, considere $f(x) = 1/x$. Então $f^{[2]}(x) = x$, de modo que $\{x, 1/x\}$ é um ciclo-2 para todo $x \neq 0$. Evidentemente tal ciclo - 2 não é nem atrativo nem repulsivo, pois $|f'(x)f'(z)| = 1$ para todo $x \neq 0$.

Quando $|f'(x)f'(z)| \neq 1$, o critério pode efetivamente nos informar se $\{x, z\}$ é atrativo ou repulsivo.

3.4 Bacias de atração

Se um ponto fixo p de f é atrativo, então todos pontos perto de p são “atraídos” em direção a p , no sentido que suas iteradas convergem para p . O conjunto de todos os pontos cujas iteradas convergem para p é chamado de bacia de atração de p (em inglês, basin of attraction).

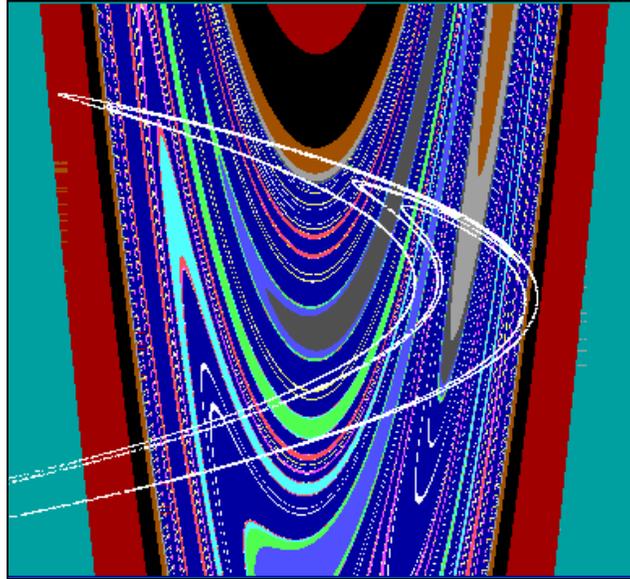


FIGURA 3.7- Diagrama da bacia de atração (azul) para o atrator de Hénon (branco). As outras cores estão relacionadas com a quantidade de iterações necessárias para que o valor de x exceda um certo limite positivo k .

Exemplo. Considere a função $f(x) = x^2$. Determinar a bacia de atração do ponto fixo 0.

Solução: Se $|x| < 1$, então $f^{(n)}(x) = x^{2^n} \rightarrow 0$ à medida que n aumenta sem limite, de modo que tais valores de x pertencem à bacia de 0. Em contraste, se $|x| \geq 1$, então $|f^{(n)}| \geq 1$, de modo que tais valores de x não pertencem à bacia de 0. Portanto, a bacia de 0 consiste no intervalo $(-1,1)$.

Observe que, se p for um ponto fixo repulsivo, sua bacia de atração pode consistir de um único ponto: o próprio p . Um exemplo desta situação é o ponto 0 se $f(x)=2x$. No extremo oposto do espectro, a bacia de atração do ponto fixo 0 para $\text{sen}(x)$ consiste de todos os números reais.

3.5 Pontos de período três

Nesta secção será apresentado um argumento para relacionar a presença de órbitas de período três com órbitas de outros períodos. Tal argumento é baseado na ordenação de Šarkovskii[ŠAR64], que aqui será ilustrada apenas a título de informação. A demonstração do respectivo teorema é excessivamente técnica e foge ao propósito deste trabalho.

A ordenação de Šarkovskii dos inteiros positivos é a seguinte:

$$3 \curvearrowright 5 \curvearrowright 7 \curvearrowright \dots \curvearrowright 2 \cdot 3 \curvearrowright 2 \cdot 5 \curvearrowright 2 \cdot 7 \curvearrowright \dots \curvearrowright 2^2 \cdot 3 \curvearrowright 2^2 \cdot 5 \curvearrowright 2^2 \cdot 7 \curvearrowright \dots \curvearrowright \dots \curvearrowright 2^3 \curvearrowright 2^2 \curvearrowright 2 \curvearrowright 1$$

ímpares inteiros
2·(ímpares inteiros)
2²·(ímpares inteiros)
potências de 2

FIGURA 3.8 – A ordenação de Šarkovskii

O que o teorema de Šarkovskii diz é que, se uma f possui um ciclo- n , possuirá órbitas de todos os ciclos situados à direita desta, de acordo com a ordenação acima. O teorema de Li-Yorke[LIY75] pode ser interpretado como um corolário do teorema de Šarkovskii; o teorema de Li-Yorke prova que, se uma f possui um ciclo-3, então possuirá ciclos com todos outros períodos l , sendo l um inteiro positivo não-nulo. O título do artigo publicado em 1975 é “Período Três Implica Caos”, e está apresentado em uma linguagem relativamente acessível para leitores não especializados na área, se comparado com o trabalho de Šarkovskii.

3.5 Caos (alta sensibilidade às condições iniciais)

Duas são as formas mais utilizadas para descrever o modo pelo qual iteradas de pontos vizinhos separam-se umas das outras: (i) dependência sensível às condições iniciais, e (ii) o expoente de Lyapunov. Estas noções são fundamentais ao entendimento do conceito de caos, que irá também aparecer neste capítulo, e serão abordadas nas subseções 3.6 e 3.7 a seguir.

3.6 Dependência sensível às condições iniciais

Antes de definir “dependência sensível às condições iniciais” adotaremos uma notação que doravante irá facilitar a discussão. Escreveremos $f : A \rightarrow B$ para indicar que o domínio da função f é A e que sua imagem está contida em B . Portanto, $f : J \rightarrow J$ significa que o domínio de f é J e que sua imagem está contida em J .

Definição. Seja J um intervalo e suponha que $f : J \rightarrow J$. Então f tem dependência sensível às condições iniciais em x , ou simplesmente dependência sensível em x se existir um $\varepsilon > 0$ tal que para todo $\delta > 0$, existe um y em J e um inteiro positivo n tais que

$$|x - y| < \delta \quad \text{e} \quad |f^{[n]}(x) - f^{[n]}(y)| > \varepsilon. \quad (3.9)$$

Se f tem dependência sensível nas condições iniciais para todo x em J , dizemos que f tem dependência sensível nas condições iniciais em J , ou que f tem dependência sensível em J , ou que f tem dependência sensível.

As “condições iniciais” na definição referem-se aos pontos dados, ou pontos iniciais, x e y . A definição diz que f tem dependência sensível se arbitrariamente perto de qualquer ponto x no domínio de f existir um ponto y e uma n -ésima iterada que dista da n -ésima iterada de x mais do que uma distância ε . Isto tem um significado prático pois, nesta situação, iteradas de ordem mais

elevada de um valor aproximado de x podem não assemelham-se às verdadeiras iteradas de x . Portanto, cálculos feitos com computadores são muito suscetíveis de fornecerem indícios falsos.

3.7 Expoentes de Lyapunov

Apesar da relativa facilidade de se obter uma noção intuitiva do que é “dependência sensível às condições iniciais”, a determinação na prática se uma dada função tem dependência sensível usualmente não é tão simples como nos casos aqui apresentados. Faz-se necessária a definição de um critério mais formal para a caracterização da dinâmica de um sistema.

Seguiremos o exemplo apresentado em[GAL98], para o mapa tenda:

$$T_{\mu}(x) = \begin{cases} 2\mu x, & 0 \leq x < 1/2 \\ 2\mu(1-x), & 1/2 \leq x \leq 1 \end{cases} \quad (3.10)$$

A separação entre iteradas de pontos próximos é dada por

$$\begin{aligned} |T^{[n]}(x + \varepsilon) - T^{[n]}(x)| &= 2^n \varepsilon \\ |(T^{[n]})'(x)| &= \lim_{\varepsilon \rightarrow 0} \left| \frac{T^{[n]}(x + \varepsilon) - T^{[n]}(x)}{\varepsilon} \right| = 2^n \end{aligned} \quad (3.11)$$

Onde 2 é o valor do módulo da inclinação de cada metade da função e n denota a enésima iterada da função. A seguir generalizamos a idéia acima para outras funções.

Definição. Seja J um intervalo limitado e $f : J \rightarrow J$ continuamente diferenciável em J . Fixe um x em J , e considere $\lambda(x)$ como sendo definida por

$$\lambda(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln |(f^{[n]})'(x)|. \quad (3.12)$$

onde supomos que o limite existe. Neste caso, $\lambda(x)$ é o expoente de Lyapunov de f no ponto x . Se $\lambda(x)$ é independente de x sempre que $\lambda(x)$ estiver definido, então o valor comum de $\lambda(x)$ é denotado por λ e sendo chamado de expoente de Lyapunov da f .

O número $\lambda(x)$ pode ser considerado uma medida da “taxa de perda de informação” das iteradas sucessivas dos pontos perto de x . Se y estiver perto de x , e se as iteradas de x e y permanecerem vizinhas, então $\lambda(x)$ é passível de ser negativo, devido a presença do logaritmo. Em contraste, se as iteradas separam-se umas das outras, então $\lambda(x)$ é positivo. Assim, quanto maior for $\lambda(x)$, maior a perda de informação. O valor de $\lambda(x)$ equivale à taxa de divergência exponencial de órbitas inicialmente vizinhas ($e^{\lambda(x)}$, que para o mapa tenda vale 2). $\lambda(x) \rightarrow 0$ indica que o ponto x é um ponto de bifurcação, o que significa que há uma mudança de estabilidade da órbita de f em x . Outra forma igualmente representativa de se conceituar o expoente de Lyapunov é associando-o à taxa de

deformação de uma hipersfera de condições iniciais em um hiperelipsóide no espaço de fases, com certas direções sofrendo contração e outras sendo distendidas, para escalas de tempo e comprimento suficientemente pequenas.

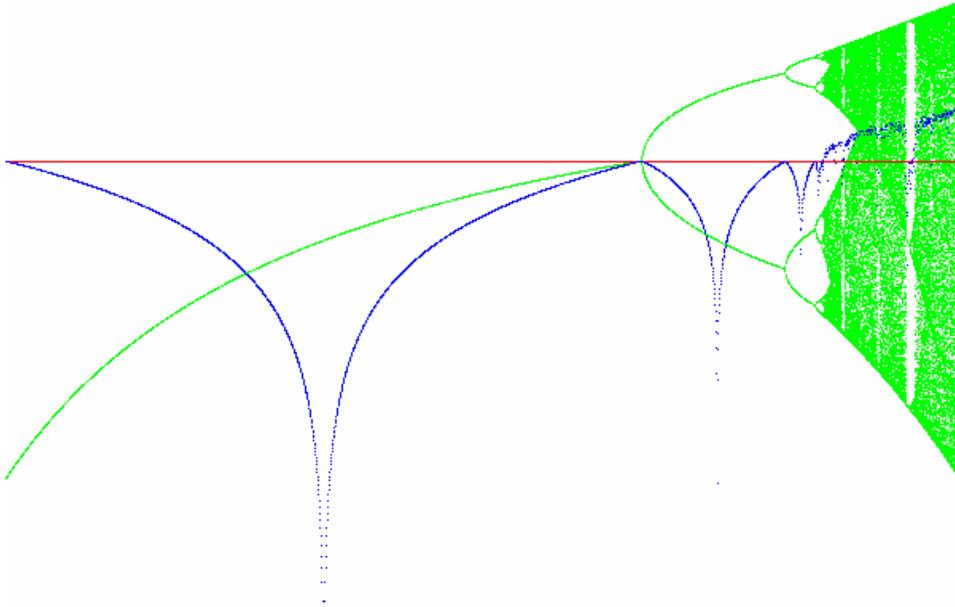


FIGURA 3.9 - Expoentes de Lyapunov (azul) para o mapa logístico (verde)

3.8 Dimensões

O estranho comportamento no tempo de um sistema caótico determinístico tem o seu equivalente na geometria do conjunto no espaço de fase formado pelas trajetórias do sistema, o atrator. Um atrator estranho pode ser espacialmente caracterizado pelos valores de seu espectro de dimensões[OTT94].

$$D_q = \frac{1}{q-1} \lim_{\varepsilon \rightarrow 0} \frac{\ln I(q, \varepsilon)}{\ln \varepsilon} \quad (3.13)$$

onde

$$I(q, \varepsilon) = \sum_{i=1}^{M(\varepsilon)} [\mu(C_i)]^q \quad (3.14)$$

e

$\mu(C_i)$ é a probabilidade do sistema visitar C_i

Considere o mapa logístico parametrizado, como exemplo de um mapa unidimensional unimodal:

$$x_{n+1} = \alpha x_n(1 - x_n) = F_\alpha(x_n) \quad (3.15)$$

É sabido[FIE94] que, a partir de certo valor de α , ocorre uma cascata de bifurcações (via mudança de estabilidade local por duplicação de período das órbitas estáveis ou atrativas) no mapa logístico. Note que este parâmetro α não é o mesmo da equação 3.14.

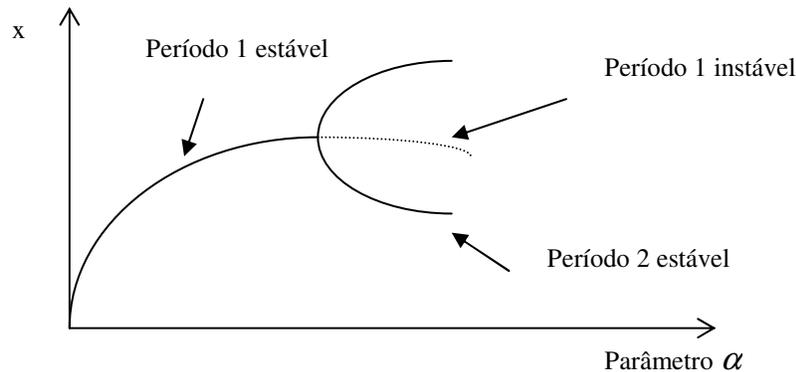


FIGURA 3.10 Bifurcação de forquilha via mudança de estabilidade

No mapa logístico, cada ponto de bifurcação é um ponto de transição onde se tem um dobramento do período da órbita estável até que esta se torne uma órbita de período infinito para $\alpha = 4$, e um crescimento do número de órbitas periódicas instáveis presentes no mapa. Em $\alpha = 4$, o mapa logístico é ‘fortemente caótico’, preenchendo todo o domínio da função com um número infinito de órbitas periódicas instáveis, e uma órbita caótica estável (atrativa) de período infinito com uma dimensão provadamente não-inteira, apresentando uma estrutura do tipo Cantor[FIE94]. O conjunto de Cantor do ‘terço central’ é um dos conjuntos fractais mais simples, e será apresentado aqui como exemplo sem uma discussão mais aprofundada. Imagine um segmento de reta de comprimento ℓ , do qual é extraído o terço central. Dos segmentos restantes extrai-se novamente o terço central.

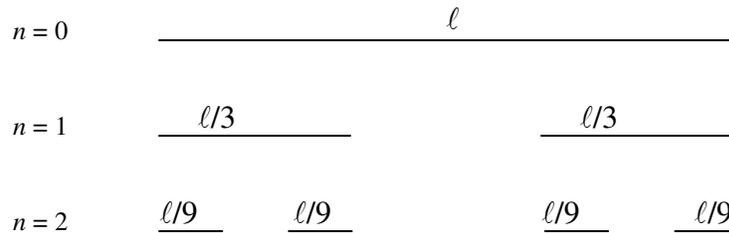


FIGURA 3.11 – O Conjunto de Cantor

O Conjunto de Cantor consiste dos infinitos segmentos de reta de comprimento infinitesimal(pontos) que restarão quando $n \rightarrow \infty$.

Fazendo $q = 0$ na equação 3.13, obtemos facilmente a forma para a equação da dimensão de contagem de caixas:

$$D_0 = \lim_{\varepsilon \rightarrow 0} \frac{\ln M(\varepsilon)}{\ln(1/\varepsilon)} \quad (3.16)$$

Onde $M(\varepsilon)$ é o número mínimo de caixas de lado ε necessárias para cobrir todo o conjunto.

Para o conjunto de Cantor da figura 3.11, tem-se:

$$\begin{aligned} \varepsilon = \ell, M(\varepsilon) &= 1; \\ \varepsilon = \ell/3, M(\varepsilon) &= 2; \\ \varepsilon = \ell/9, M(\varepsilon) &= 4; \end{aligned} \quad (3.17)$$

$$\begin{aligned} \varepsilon = \ell/3^n, M(\varepsilon) &= 2^n \\ \ell &= 1 \end{aligned}$$

$$D_0 = \lim_{\varepsilon \rightarrow 0} \frac{\ln 2^n}{\ln 3^n} = \lim_{\varepsilon \rightarrow 0} \frac{n \ln 2}{n \ln 3} \approx 0.63.$$

Outras dimensões úteis são a dimensão de informação D_1 e a dimensão de correlação D_2 .

Note que não podemos fazer $q = 1$ diretamente. Para calcularmos D_1 precisamos recorrer ao seguinte artifício[FIE94]:

$$\log_2 \sum_i p_i^q = \log_2 \left(\sum_i p_i^q - 1 + 1 \right) \approx \left(\sum_i p_i^q - 1 \right) \log_2 e, \text{ para } q \sim 1 \quad (3.18)$$

considerando

$$\begin{aligned} \sum_i p_i^q &\approx 1, \\ q &\sim 1, \\ \log_2(1+x) &\approx x \log_2 e \\ x &\ll 1 \end{aligned} \quad (3.19)$$

Fazendo $q = 1 + \delta q$, $\delta q \ll 1$, ficamos com

$$\begin{aligned} p_i^q &= p_i^{1+\delta q} = p_i 2^{\delta q \log_2 p_i} \\ &\approx p_i (1 + \delta q \ln 2 \log_2 p_i + \dots) \end{aligned} \quad (3.20)$$

que resulta em

$$\sum_i p_i^q - 1 \approx \sum_i p_i + \delta q \sum_i p_i \ln 2 \log_2 p_i - 1 \approx \delta q \ln 2 \sum_i p_i \log_2 p_i \quad (3.21)$$

admitindo que $\sum_i p_i = 1$,

$$\begin{aligned} \log_2 \sum_i p_i^q &= \log_2 \left(\delta q \ln 2 \sum_i p_i \log_2 p_i + 1 \right) \\ &\approx \delta q \ln 2 \log_2 e \sum_i p_i \log_2 p_i \end{aligned} \quad (3.22)$$

com a aproximação $\log_2(1+x) \approx x \log_2 e$, $x \ll 1$.

Substituindo $\delta q \ln 2 \log_2 e \sum_i p_i \log_2 p_i$ na equação 3.13 ficamos com

$$D_1 = \lim_{\varepsilon \rightarrow 0} \frac{\sum_{i=1}^{N(\varepsilon)} p_i \log p_i}{\log(\frac{1}{\varepsilon})} \quad (3.23)$$

Que é a dimensão de informação. A dimensão de correlação D_2 é obtida com (muito) mais facilidade, basta fazer $q = 2$ na equação 3.16 e obtemos

$$D_2 = \lim_{\varepsilon \rightarrow 0} \frac{\sum_{i=1}^{N(\varepsilon)} \log p_i^2}{\log(\varepsilon)} \quad (3.24)$$

A dimensão de correlação, em particular, é uma grandeza de especial interesse uma vez que caracteriza o quanto o atrator se parece consigo mesmo, de forma análoga à integral de correlação linear.

Grassberger e Procaccia [GRA83] propuseram a fórmula

$$C(\varepsilon) = \frac{1}{N^2} \lim_{N \rightarrow \infty} \sum_{\substack{i,j=1 \\ i \neq j}} \Theta[\varepsilon - |\vec{x}_i - \vec{x}_j|] \quad (3.25)$$

$$D_2 \sim \lim_{\varepsilon \rightarrow 0} \frac{\log C(\varepsilon)}{\log(\varepsilon)}$$

para o cálculo de D_2 . A chamada integral de correlação de Grassberger-Procaccia $C(\varepsilon)$ é uma forma alternativa ao uso de caixas para se obter distribuições empíricas a partir de dados amostrais. Esta medida simplesmente conta o número de vizinhos que um determinado ponto i no espaço vetorial considerado possui.

3.9 Estacionariedade

Os exemplos estudados até aqui abrangeram sistemas onde os graus de liberdade não ativos (parâmetros) se mantinham constantes no tempo. Fenômenos não-lineares nos quais existe variação de um ou mais parâmetros do sistema são chamados não-estacionários e deixa de existir um único atrator no espaço de fases. O estudo deste tipo de sistema pode ser bastante complicado em comparação aos sistemas estacionários.

Uma definição formal de estacionariedade é aquela em que se tem um processo físico no qual todas as probabilidades¹ condicionais existem e se mantém constantes ao longo do tempo. Esta é a definição teórica encontrada nos livros-texto mas, na prática, a finitude do tempo de observação de um processo real (ou simulado) faz emergirem algumas dificuldades com esta interpretação. Ainda não existe uma definição formal de estacionariedade para séries temporais, mas uma interpretação bastante utilizada para sistemas não-lineares determinísticos é considerar-se parâmetros de controle fixos no tempo, bem como ausência de dependência temporal explícita no sistema em observação (ou sejam, t não pode ser um parâmetro nas equações), embora processos com regras fixas também podem apresentar comportamento não-estacionário (e.g. sistemas intermitentes, vôos de Levy [ADA97]). Testes de estacionariedade tentam, muitas vezes utilizando artifícios geométricos, detectar variações nos parâmetros do sistema em estudo.

¹ Para sistemas determinísticos, essas probabilidades são dadas por uma função delta de Dirac $\delta(x' - f(x))$.

4 Tratamento de Dados Experimentais

4.1 Métodos Clássicos para análise de sinais

Neste capítulo iniciaremos a revisão de alguns métodos para tratamento de sinais experimentais abordando, de uma forma bastante resumida, um pouco do paradigma de modelos lineares estocásticos e métodos espectrais para caracterização de sinais, e logo em seguida serão apresentados alguns métodos para a reconstrução de atratores, aos quais será dada a maior ênfase. Tais métodos serão discutidos sempre tendo em mente que sua aplicação será direcionada principalmente a sinais de dinâmica temporal *complicada*.

4.2 Modelos estocásticos

Antes que se conhecesse a existência de processos irregulares oriundos de uma dinâmica puramente determinística, sistemas com comportamento complicado eram tomados como necessariamente possuidores de alguma componente de natureza estocástica. Antes das importantes descobertas teóricas que remontam a Henri Poincaré (final do século XIX), e de se poder utilizar computadores para encontrar soluções particulares de equações diferenciais não-lineares[LOR63], os processos físicos observados (na sua maioria) eram interpretados pelos cientistas como sendo de natureza linear (ou aproximáveis para tal), ou seja, deveriam comportar soluções na forma de oscilações periódicas no tempo. Sistemas com comportamento complicado eram interpretados como possuidores de modos oscilatórios sobrepostos, em conjunto com algum nível de estocasticidade.

Com o surgimento dos computadores, dados experimentais (i.e. séries temporais) puderam então ser processados pelos cientistas com a finalidade de estudar o sistema gerador subjacente. Os modelos predominantes eram os assim chamados “lineares estocásticos” e, nesse contexto, popularizaram-se os modelos ARMA[DIK99]:

$$x_n = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{n-i} + \sum_{j=0}^{M_{MA}} b_j \eta_{n-j}, \quad (4.1)$$

onde M_{AR} e M_{MA} são os números de parâmetros ajustáveis (ordem do processo).

Modelos ARMA (AutoRegressive Moving Averages) são resultado da combinação de modelos AR e MA. Na equação 4.1, x_n é a medida observada, η_n são números randômicos Gaussianos independentes com média zero e variância unitária. É importante notar-se que, para qualquer processo ARMA com entradas Gaussianas, a função de autocorrelação de dois pontos e o espectro de potências são inteiramente determinados pelos coeficientes a_i e b_i e, ainda mais importante, todos os momentos e correlações de maior ordem podem assim ser determinados. Neste sentido, o espectro de potências pode ser utilizado como uma descrição estatística completa do processo.

$$c_v = \frac{1}{\sigma^2} \langle (s_n - \langle s \rangle)(s_{n-v} - \langle s \rangle) \rangle = \frac{\langle s_n - s_{n-v} \rangle - \langle s \rangle^2}{\sigma^2} \quad (4.2)$$

Equação 4.2 - Função de autocorrelação linear

O citado espectro de potências é dado por

$$f(\omega) = \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt \quad (4.3)$$

$$P(\omega) = |f(\omega)|^2 = f(\omega) * f(\omega)$$

Onde $f(\omega) *$ denota o complexo conjugado de $f(\omega)$. Em [OTN72] encontra-se uma abordagem detalhada sobre métodos numéricos para cálculo do espectro de potências contínuo, da transformada rápida de Fourier e vários métodos clássicos para análise de séries temporais.

Modelos ARMA são amplamente utilizados para modelar sistemas com comportamento irregular [OTN72][MAR96][BOX76]. Apesar de lineares, estes modelos não têm um comportamento puramente oscilatório no tempo, ou seja, apresentam oscilações aperiódicas causadas pela presença da componente estocástica. Abaixo temos dois exemplos de modelos lineares estocásticos extremamente simples: O primeiro é um processo autoregressivo distorcido, onde os η_n são números Gaussianos independentes. Tal processo é chamado de processo autoregressivo de ordem um. O segundo exemplo representa um oscilador linear que recebe um golpe a cada intervalo de tempo com uma força de intensidade aleatória segundo uma distribuição Gaussiana.

$$x_n = 0.99 x_{n-1} + \eta_n \quad (4.4)$$

$$\ddot{x} = -\frac{k}{m} x - b\dot{x} + y(t) \quad (4.5)$$

No caso de um número discreto e finito de oscilações superpostas (sem estocasticidade), o comportamento do sistema no tempo poderá ser aparentemente desordenado (veja equação abaixo), mas uma análise a partir do espectro de potências de Fourier, bem como da função de autocorrelação linear pode revelar as frequências formantes e o caráter puramente periódico² do sinal, pois nesse caso o espectro de potências não se apresenta como um continuum, apenas contendo os picos correspondentes às frequências formantes do sistema e seus respectivos harmônicos.

$$Y(x) = 3\sin(x) + 2\sin(2x) + 4\sin(13x) + \sin(17x) \quad (4.6)$$

² Ou quasi-periódico, com a ressalva de que neste caso o espectro de potências apresentará picos em $\omega=\omega_i$ e um background ruidoso[26].

Uma das grandes limitações do método espectral (e também dos métodos não-lineares) para sistemas reais é o fato de que estes dificilmente apresentam uma dinâmica estacionária linear bem-comportada. A utilização de janelas no tempo é uma técnica frequentemente utilizada para compensar este problema, em vista das ondas 'grandes' seno e cosseno não possuírem localização temporal.

4.3 Análise não-linear de séries temporais

Sistemas regidos por regras dinâmicas não-lineares são altamente susceptíveis de apresentarem comportamento caótico, dependendo do condicionamento dos seus parâmetros. Um espectro de potências obtido a partir de um sinal deste tipo apresenta-se como um *continuum*, mesmo sem nenhuma componente estocástica. Métodos de transformada de Fourier não caracterizam adequadamente sistemas cujo comportamento advém de uma dinâmica não-linear determinística³, o que faz necessária a utilização de uma abordagem qualitativamente diferente, levando em consideração propriedades específicas típicas a sistemas desta natureza.

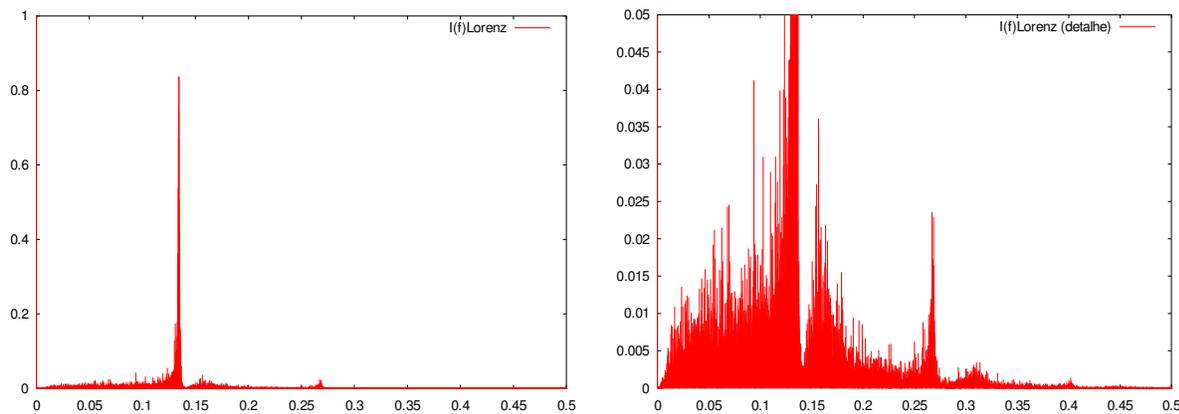


FIGURA 4.1 – Espectro de potências para o sistema de Lorenz, em regime caótico. Mesmo na ausência de estocasticidade, há uma banda contínua em torno do pico central.

A seguir serão revisados alguns dos métodos utilizados para reconstrução de atratores, determinação da dimensão do espaço de imersão, cálculo de grandezas como dimensões e expoentes, e finalmente será abordado o problema de clusterização, com alguns resultados obtidos em simulações e experimentos reais.

4.3.1 Testes para detecção de não-linearidade

Uma série de testes para detectar a presença de não linearidade no sinal (não necessariamente dinâmica) baseia-se em uma técnica bastante simples chamada ‘surrogate data’ [THE92]: primeiro formula-se a hipótese nula de que o sinal seja gerado por uma dinâmica linear estocástica, e a seguir gera-se novos conjuntos de dados com as mesmas propriedades lineares do sinal original (espectro de potências, função de autocorrelação). Calcula-se alguma quantidade não-linear a partir de ambos e estuda-se a discrepância entre os resultados. O percentual de erros fornece uma medida de quanto a série original desvia-se da linearidade suposta. Neste trabalho não empregaremos tal técnica em vista das próprias medidas envolvidas no processo de clusterização funcionarem, de

³ Mesmo sistemas não-lineares com comportamento puramente periódico que possuam picos pronunciados no espectro de potências sempre apresentarão uma banda contínua em torno destes.

certa forma, como um quantificador de não-linearidade, mas uma descrição completa do método pode ser encontrada em [OTT94].

4.3.2 Reconstrução do espaço de fases

Packard e colaboradores, em um artigo seminal publicado em 1980[PAC80], propuseram um método bastante simples para reconstrução de atratores a partir de medidas de um único observável, baseado em sucessivas derivadas deste observável. A aplicação de sucessivas derivadas não é nenhuma novidade se lembrarmos da assim chamada notação standard para equações diferenciais, onde se compacta todo o sistema em uma única equação e os demais eixos do sistema são determinados por sucessivas derivadas desta [HUB95].

$$\mathbf{x}' = \begin{bmatrix} x' \\ x'' \\ x''' \\ \vdots \\ x^{(n)} \end{bmatrix} = \begin{bmatrix} x'_0 \\ x'_1 \\ x'_2 \\ \vdots \\ x'_{n-1} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ f(t, x_0, x_1, \dots, x_{n-1}) \end{bmatrix} \equiv \mathbf{f}(t, \mathbf{x})$$

FIGURA 4.2 – Notação ‘standard’ para equações diferenciais

Além de, com seu artigo, ter dado o que muitos consideram como sendo o ‘pontapé inicial’ na pesquisa em análise não-linear de séries temporais (mais especificamente, reconstrução de atratores), Packard sugeriu, neste mesmo artigo, que atrasos temporais pudessem ser utilizados para reconstruir cada componente do atrator em questão⁴ (a taxa de amostragem finita e freqüentemente não grande o suficiente faz com que o cálculo de sucessivas derivadas através de métodos como diferenças finitas faça com que erros numéricos sejam amplificados a cada operação). Em 1981, Floris Takens publica um artigo [TAK81] que é provavelmente um dos mais importantes na área de ciência não-linear experimental. Neste artigo (de um nível matemático, digamos, ‘divertido’), Takens prova que qualquer transformação difeomórfica que preserve rank é um embedding (mergulho topológico) [PEC92].

Em termos mais simples, Takens provou, entre outras coisas, que a utilização de vetores de estado reconstruídos $\vec{\xi}_i = \{x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)\}$ a partir de, por exemplo, vizinhos temporais, é garantidamente um embedding para dimensões do espaço de imersão $m \geq 2D + 1$, onde D é a dimensão ou ordem do sistema real. Sauer *et al* [SAU91], em um artigo posterior, mostram que esse limite pode ser substituído por $m > 2D_0$, sendo D_0 a dimensão de box-counting do atrator contido em D . A simplicidade de utilização deste método fez com que fosse rapidamente difundido e largamente utilizado nas mais diversas áreas.

⁴ Na verdade, Packard fez essa sugestão devido a uma comunicação privada de David Ruelle, que já possuía trabalhos em comum autoria com Floris Takens.

4.3.3 Ajuste dos parâmetros m e τ

Na construção dos vetores de estado $\bar{\xi}_t = \{x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (m-1)\tau)\}$, os parâmetros m (dimensão do espaço de imersão) e τ (atraso temporal entre componentes) precisam ser determinados. Segundo Takens, τ pode a princípio assumir qualquer valor para uma série de comprimento infinito. Obviamente, séries temporais experimentais possuem um comprimento (bastante) limitado para que se possa seguir tal diretriz, e tal limitação impõe algumas restrições à escolha de τ . Para um valor de τ muito pequeno, os pontos do atrator podem ficar concentrados ao longo da diagonal $x(t) = x(t + \tau) = \dots = x(t + (m-1)\tau)$. Caso um valor muito grande seja atribuído a τ os vetores reconstruídos podem ficar ‘espalhados’ pelo volume do atrator de forma muito pouco correlacionada. Nesse sentido, a literatura apresenta alguns métodos para a determinação empírica de τ . São eles:

4.3.4 Função de autocorrelação linear

Um dos critérios mais simples para a escolha de τ e talvez o mais difundido[FIE94] encontrado na literatura é baseado na função de autocorrelação de dois pontos já citada anteriormente (equação 4.2). Um valor aceitável para τ deve, teoricamente, equivaler ao primeiro instante em que a função de autocorrelação cruza o zero. Esta abordagem garante que $x(t)$ e $x(t + \tau)$ serão linearmente decorrelacionados. Posteriormente, resultados empíricos indicam que um valor mais ‘adequado’ para τ corresponde ao ponto em que a função de autocorrelação cai abaixo de $1/e$.

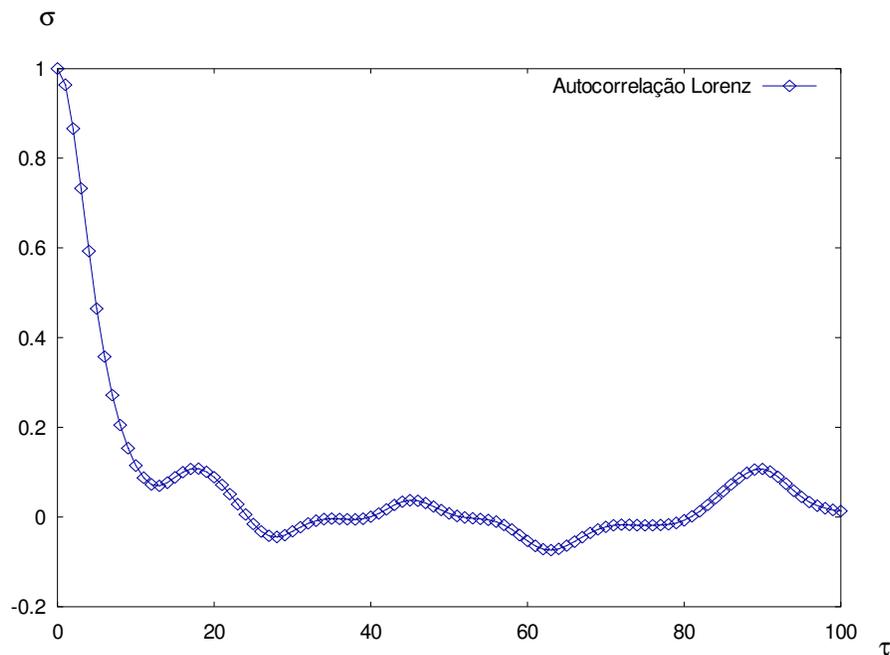


FIGURA 4.3 - Função de autocorrelação para o atrator de Lorenz. O primeiro ponto onde $\sigma < e^{-1}$ é $\tau \approx 7$.

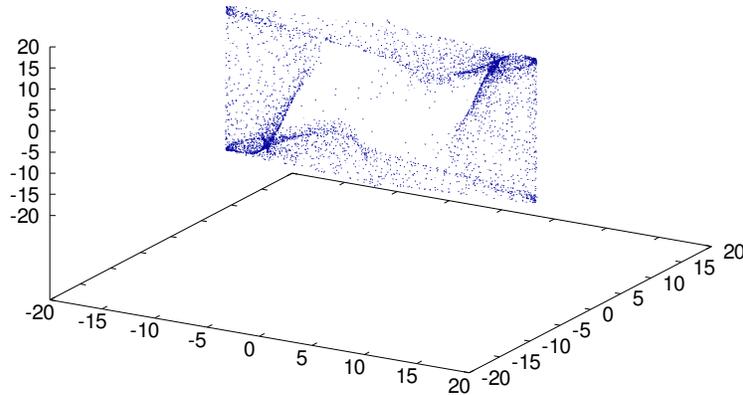


FIGURA 4.4 - Atrator de Lorenz reconstruído a partir da coordenada x com $\tau = 7$.

Esta abordagem apresenta limitações uma vez que apenas correlações lineares são minimizadas. A seguir discutiremos uma medida que, embora de maior complexidade, provê um critério mais refinado para a escolha de τ .

4.3.5 Informação mútua

Uma característica interessante de sistemas caóticos é que estes podem ser interpretados como geradores de informação, no sentido da teoria de informação de Claude Shannon[SHA49]. Com base nesta idéia que Fraser *et al*[FRA86] propuseram um método para a escolha do atraso temporal τ baseado na medida de informação mútua (eq. 4.7a).

$$I(S, Q) = \int P_{sq}(s, q) \log \left[\frac{P_{sq}(s, q)}{P_s(s)P_q(q)} \right] ds dq \quad (4.7a)$$

Considerando um sistema genérico acoplado (S, Q) , onde $P_{sq}(s, q)$ é a probabilidade dos eventos s e q acontecerem 'simultaneamente', $P_s(s)$ é a probabilidade do sistema encontrar-se na partição s e $P_q(q)$ é a probabilidade equivalente para o estado q .

Assumido um espaço discreto (i.e. distribuições empíricas), $[s, q] = [x(t), x(t+\tau)]$ mais uma simples manipulação algébrica obtemos:

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln p_{ij}(\tau) - 2 \sum_i p_i \ln p_i \quad (4.7b)$$

Sobre o intervalo explorado pelos dados, criamos um histograma para a distribuição de probabilidade dos dados. p_i denota a probabilidade⁵ que o sinal visitará o i -ésimo intervalo do histograma, e p_{ij} a probabilidade que uma transição do tempo t para o tempo $t + \tau$ ocorra do intervalo i para o intervalo j . A base escolhida para o logaritmo não é relevante neste caso.

Segundo Fraser e Swinney, R.S. Shaw sugeriu em comunicação privada que um valor adequado para τ equivaleria ao primeiro mínimo de $I(\tau)$, que marca o quanto $x(t+\tau)$ adiciona máxima informação ao nosso conhecimento de $x(t)$ minimizando não só as correlações lineares mas também as não-lineares. Note que $I(\tau) \rightarrow 0$ à medida que $\tau \rightarrow \infty$.

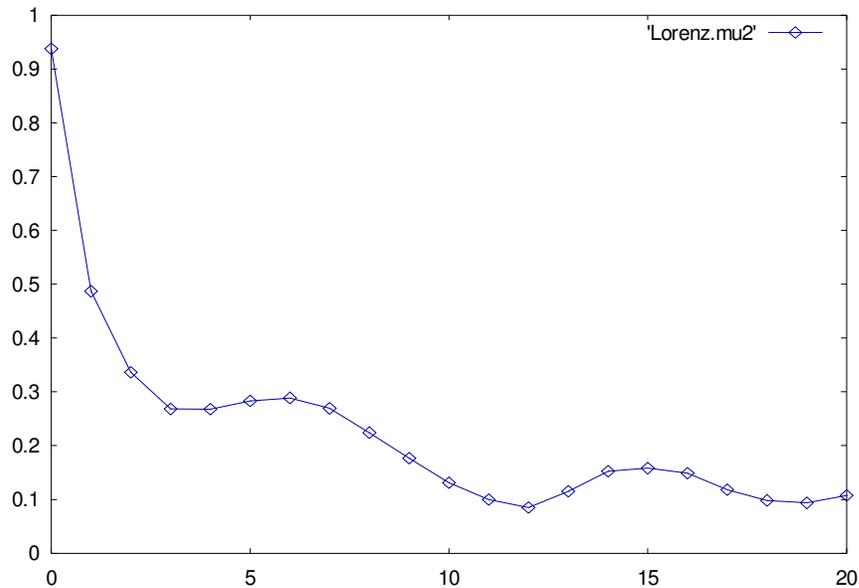


FIGURA 4.5 - Informação mútua do atrator de Lorenz. O primeiro mínimo ocorre em $\tau=3$. O valor escolhido para reconstrução foi $\tau=2$.

Pode-se encontrar uma discussão sobre outros métodos para cálculo do atraso temporal, por exemplo, em [FIE94].

⁵ Esta probabilidade é exatamente a medida invariante $\rho(x_i)$ para o vetor de estado x_i .

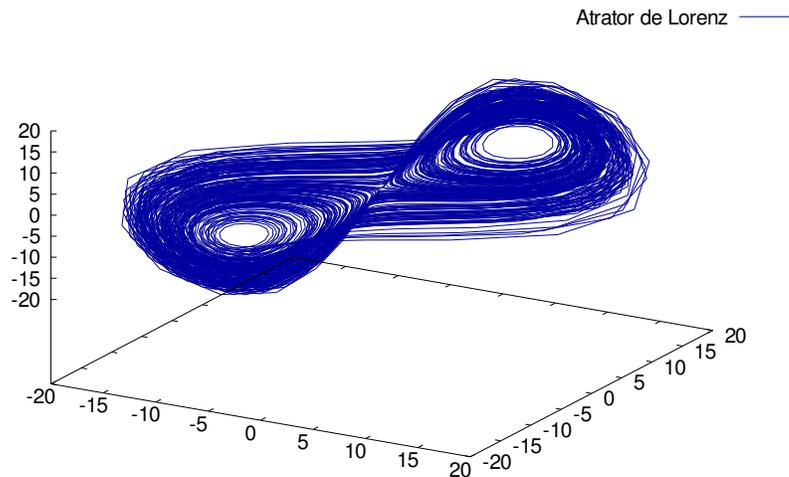


FIGURA 4.6 - Lorenz reconstruído a partir da coordenada x com $\tau = 2$

4.3.6 Falsos vizinhos

Os limites para dimensões sugeridos por Takens e Sauer *et al* tem pouco valor prático visto que raramente o valores para D ou D_0 são conhecidos *a priori*. Além disso, estes valores são limites mínimos para que se possa ter um embedding garantido, partindo de um ponto de vista matematicamente rigoroso, mesmo sendo possível obter-se um embedding “correto” (embora não matematicamente garantido) com uma dimensão de imersão m abaixo destes limites. Um dos métodos mais largamente implementados para este fim é o chamado método dos falsos vizinhos[KEN92]. O algoritmo é bastante simples e elegante, e consiste no seguinte:

- ① Iniciar a reconstrução com um valor mínimo para a dimensão de imersão d (normalmente 1) e o valor pré-calculado para τ .
- ② Encontrar o vizinho mais próximo de cada ponto do conjunto reconstruído imerso no espaço d .(excluindo vizinhos temporais para o caso de fluxos)
- ③ Aumentar a dimensão de imersão d e medir o percentual de falsos vizinhos usando como critério se

$$\frac{|s(k + dT) - s^{NN}(k + dT)|}{R_d(k)} > \theta \quad (\text{onde } \theta \text{ é um threshold } \sim 15) \quad (4.8)$$

④ Um segundo critério deve ser considerado para conjuntos de dados finitos: se o vizinho mais próximo de um ponto $\mathbf{y}(n)$ não estiver próximo deste $[R_{d(n)} \approx R_A]$ e for um falso vizinho, então a distância $R_{d+1}(n)$ resultante da adição da $(d+1)$ -ésima componente aos vetores de dados será $R_{d+1} \approx 2R_A$. Isto é, mesmo vizinhos mais próximos que estejam distantes serão esticados para as extremidades do atrator quando forem desdobrados um do outro, se forem falsos vizinhos. O segundo critério é então:

$$\frac{R_{d+1}(n)}{R_A} > A_{tol} \quad (4.9)$$

como medida de R_A escolhemos o valor

$$R_A^2 = \frac{1}{N} \sum_{n=1}^N [x(n) - \bar{x}]^2, \text{ onde} \quad (4.10)$$

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x(n). \quad (4.11)$$

Quando o percentual de falsos vizinhos cair abaixo de um valor de tolerância, faz-se $m=d$.

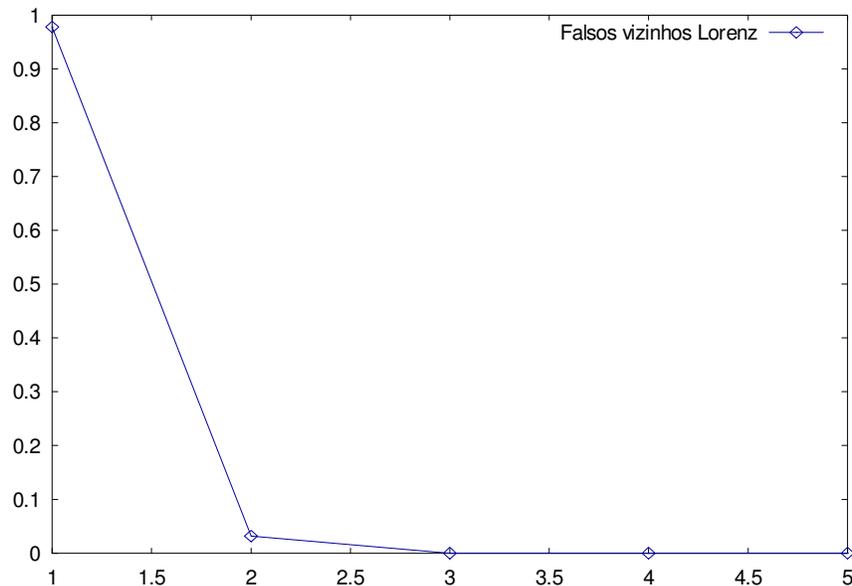


FIGURA 4.7 – Percentual de falsos vizinhos (vert) versus dimensão de imersão (horiz) para o sistema de Lorenz.

Uma discussão mais completa do método pode ser encontrada no artigo original, em [26,27]. Em [OTT94] tem-se uma reimpressão do artigo original.

É importante que se tenha claro que m é a dimensão do espaço de imersão, onde o atrator está mergulhado globalmente, e pode ser bastante diferente da dimensão local do atrator.

4.3.7 Filtragem

Em [KOS88], os autores introduzem um método bastante simples para filtragem de ruído em uma série temporal, aproximando a dinâmica local por um mapa linear

$$x_{i+1} = f(x_i) \approx \mathbf{A}_i x_i + b_i \equiv L(x_i) \quad (4.12)$$

para uma matriz \mathbf{A} e um vetor b_i . (A matriz \mathbf{A}_i é o jacobiano de f em x_i). O ajuste dos parâmetros pode ser feito utilizando-se um procedimento de mínimos quadrados.

Seja $\{x_j\}_{j=1}^n$ um conjunto de pontos em uma epsilon-esfera em torno do i -ésimo ponto de referência, e $y_i = f(x_j)$ denote a imagem observada de x_j . A k -ésima linha $a_i(k)$ de \mathbf{A}_i e k -ésima componente $b_i(k)$ de b_i são dadas pela solução por mínimos quadrados da equação

$$y_j(k) = b_i(k) + (a_i(k) | x_j) \quad (4.13)$$

onde $y_j(k)$ é a k -ésima componente de y_j e $(|)$ denota o produto escalar.

No segundo estágio do método, utiliza-se o mapa linear L para corrigir erros nas trajetórias observadas, da seguinte forma: Dada uma “janela de pontos consecutivos”, $\{x_i, x_{i+1}, \dots, x_{i+p}\}$ na trajetória observada, encontramos uma coleção de pontos $\{\hat{x}_i, \hat{x}_{i+1}, \dots, \hat{x}_{i+p}\}$ mais próximos dos observados que também satisfaçam os correspondentes mapas lineares. Mais precisamente, a nova trajetória $\{\hat{x}_{i+k}\}_{k=0}^p$ minimiza a soma dos quadrados:

$$\sum_{k=0}^p \|\hat{x}_{i+k} - \mathbf{L}(\hat{x}_{i+k-1})\|^2 + \|\hat{x}_{i+k} - x_{i+k}\|^2 + \|\hat{x}_{i+k+1} - \mathbf{L}(\hat{x}_{i+k})\|^2 \quad (4.14)$$

Outro algoritmo também simples para redução de ruído conhecido substitui a coordenada central de cada vetor de embedding pela média local desta coordenada. Isto leva a uma aproximação localmente constante da dinâmica e está baseado na suposição de que a dinâmica é contínua. O algoritmo é descrito em [SCH93], tendo sido uma abordagem similar proposta anteriormente, em [PIK86]. Em um sistema instável, por exemplo caótico, é essencial não substituir-se a primeira e última coordenadas dos vetores de embedding por médias locais. Devido à instabilidade, erros iniciais nestas coordenadas são amplificados ao invés de suprimidos.

Este método de redução de ruído é implementado de uma forma bastante fácil. Primeiro um embedding precisa ser escolhido. Exceto por dados extremamente superamostrados, é vantajoso se escolher um pequeno atraso temporal. A dimensão de embedding m deve ser escolhida pouco acima da requerida pelos teoremas de embeddings (devido à influência do próprio ruído que pode mascarar a dinâmica em uma dimensão m). Então para cada vetor de embedding $\{s_n\}$, uma vizinhança $\mathcal{U}_\varepsilon^{(n)}$ é formada no espaço de fases contendo todos os pontos tal que $\|s_n - s_n'\| < \varepsilon$.

O raio de vizinhança ε deve ser escolhido grande o suficiente no sentido de cobrir a amplitude de erro, mas ainda menor do que um típico raio de curvatura. Geralmente estas condições não podem ser satisfeitas simultaneamente, neste caso deve-se repetir o processo com

diferentes condições e cuidadosamente avaliar os resultados. Se o nível de ruído é substancialmente menor do que o típico raio de curvatura, vizinhanças de raio em torno de 2-3 vezes o nível de ruído fornecem os melhores resultados com dados artificiais. Para cada vetor de embedding $s_n=(s_{n-(m-1)}, \dots, s_n)$ (o tempo de delay deve ser ajustado para a unidade), uma coordenada média corrigida $\hat{s}_{n-m/2}$ é calculada fazendo-se a média sobre a vizinhança $\mathbf{U}_\varepsilon^{(n)}$:

$$\hat{s}_{n-m/2} = \frac{1}{|\mathbf{U}_\varepsilon^{(n)}|} \sum_{s_n \in \mathbf{U}_\varepsilon^{(n)}} s_{n-m/2}. \quad (4.15)$$

Após uma varredura completa por toda a série temporal, todas as medidas s_n serão substituídas pelos valores corrigidos \hat{s}_n . Naturalmente, para o primeiro e últimos $(m-1)/2$ pontos (se m for ímpar), nenhuma correção será feita. A correção média pode ser tomada como um novo raio de vizinhança para a próxima iteração. Note que a vizinhança de cada ponto contém pelo menos o próprio ponto. Se este for o único membro, a equação de médias acima 4.15 é simplesmente a medida não corrigida e nenhuma mudança é feita. Assim pode-se seguramente executar múltiplas iterações com valores decrescentes de ε até que nenhuma alteração seja feita.

Uma técnica mais recentes e com resultados bastante interessantes para a filtragem emprega projeções de uma elipsóide de pontos em submanifolds de dimensão inferior via decomposição de valores singulares⁶, ou eixos principais, a partir da matriz de covariância da região do elipsóide. Achata-se então a elipsóide no plano perpendicular ao menor valor dos eixos principais, o qual deve existir principalmente devido à contribuição do ruído no sinal. O manifold acaba sofrendo alguma distorção, mas em contrapartida o ganho de informação com a remoção do ruído compensa tal problema.

4.3.8 Expoentes de Lyapunov

Uma medida de grande importância no estudo experimental de sistemas com dinâmica não-linear é a dos conhecidos expoentes de Lyapunov. Tais expoentes medem a divergência exponencial de órbitas inicialmente vizinhas, indicando a taxa de perda de informação com o tempo, de forma que podem ser utilizados (com muito cuidado) para identificar o comportamento de um sistema determinístico como sendo caótico ou não. Dimensões caracterizam a estrutura espacial dos pontos no espaço de fases; expoentes de Lyapunov caracterizam a evolução dinâmica do sistema no tempo [GER98]

Diversos artigos abordam o cálculo dos expoentes de Lyapunov a partir de séries temporais[32,33,34,35,36].

Iremos apresentar inicialmente, por razões históricas, o método descrito por [WOL85]. O algoritmo de Wolf, baseado na reconstrução de Takens, mede a taxa de crescimento da distância entre pontos

⁶ SVD, ou singular value decomposition é um instrumento bastante conhecido da álgebra linear e consiste basicamente em encontrar os valores singulares da matriz \mathbf{A} por meio do cálculo da raiz de cada um dos autovalores do produto $\mathbf{A}^T \mathbf{A}$.

vizinhos dentro de uma hipersfera de raio ε sobre uma trajetória fiducial do atrator. Uma descrição sucinta é encontrada em [FIE94], e aqui segue:

“Considere-se uma trajetória fiducial no atrator reconstruído. Seja essa trajetória descrita pela seqüência de pontos $y(t_0), y(t_1), y(t_2)\dots$. Seja $z_0(t_0)$ o vizinho no atrator reconstruído mais próximo de $y(t_0)$ e L_0 a distância entre $y(t_0)$ e $z_0(t_0)$, isto é, $L_0 = |y(t_0) - z_0(t_0)| < \varepsilon$, ou seja, $z_0(t_0)$ está dentro da hipersfera de raio ε centrada em $y(t_0)$. Acompanha-se então a evolução temporal de z_0 e y_0 até que num instante t_1 a distância entre esses pontos, L'_0 , exceda ε . Nesse momento substitui-se z_0 por um novo vizinho, mais próximo de $y(t_1)$, que esteja na direção do segmento L'_0 e tal que $L_1 = |y(t_1) - z_1(t_1)| < \varepsilon$. Um tal processo prossegue até que todos os pontos $y(t_i)$ tenham sido percorridos. O maior expoente de Lyapunov positivo é obtido como a média de $\log_2 L'_i/L_i$ ao longo da trajetória fiducial, isto é,

$$\lambda_1 = \frac{1}{t_M - t_0} \sum_{i=0}^{M-1} \log_2 \frac{L'_i}{L_i}, \quad (4.16)$$

onde M é o número total de vezes que se escolheu um novo vizinho próximo à trajetória fiducial.

No limite de um número infinito de pontos na série temporal e na ausência de ruído é sempre possível proceder à escolha de um novo vizinho que esteja na direção do segmento L'_{i-1} . Entretanto, na prática, isso não é possível, já que se dispõe de uma série finita. Nesse caso busca-se um novo ponto z_i mais próximo de $y(t_i)$ que esteja dentro de um cone de altura ε cujo eixo de simetria coincida com o segmento L'_{i-1} e com o ângulo de abertura $\theta = \pi/9$. Se nenhum ponto desse cone for localizado aumenta-se o ângulo θ até encontrar-se um vizinho. Caso isso não seja possível, toma-se o vizinho mais próximo de $y(t_i)$, independentemente do ângulo θ e do valor de ε .”

4.3.9 O método de Rosenstein

Rosenstein[ROS93] propôs uma abordagem alternativa para calcular o maior expoente de Lyapunov a partir de uma série temporal. Esta abordagem tem-se difundido bastante por possuir robustez frente à presença de ruído, característica que o algoritmo de Wolf não possui e portanto faz com que este torne-se muito instável. O algoritmo proposto por Rosenstein para o cálculo do maior expoente de Lyapunov consiste basicamente nos seguintes passos:

- ① Reconstrói-se os vetores de estado a partir dos dados $\{x_i\}$.
- ② Para cada ponto i no espaço reconstruído, seleciona-se o vizinho mais próximo j que esteja fora de um limite temporal maior do que o período médio da série (estimado a partir da transformada de Fourier da série), de forma que $|i - j| > \text{período médio}$.
- ③ Acompanha-se a evolução temporal das distâncias entre estes pontos até um limite de tempo Δt , acumulando o valor dos logaritmos das distâncias e calculando a média destes valores (dividindo por Δt)

④ Repete-se o procedimento de (1) a (3) para diferentes pares de pontos plotando o logaritmo da média destas divergências em função do valor atual de Δt .

⑤ Continua-se o procedimento até que Δt alcance um limite pré-estabelecido. Então ajusta-se uma reta ao conjunto de pontos resultantes – a inclinação desta reta deve aproximar o valor do maior expoente de Lyapunov λ_1 .

4.4 Exemplo com uma série real

Com a finalidade de demonstrar a aplicação dos métodos a uma série experimental, escolhemos uma série conhecida ‘limpa’, estacionária e de baixa dimensionalidade, obtida de um experimento de laboratório bem controlado com um laser de NH_3 , para a qual calculamos o atraso temporal τ a partir da função de autocorrelação e informação mútua, a dimensão do espaço de imersão m pelo método dos falsos vizinhos e o maior expoente de Lyapunov com o algoritmo de Rosenstein. As rotinas utilizadas fazem parte do pacote TISEAN[HEG99].

Esta série temporal, com o comprimento de 25000 pontos, foi fornecida ao Santa Fe Institute[WEI91] por Udo Huebner, Phys.-Techn. Bundesanstalt, Braunschweig, Alemanha, e coletado inicialmente por N. B. Abraham e C. O. Weiss. Os dados foram obtidos de um Laser (Far-Infrared-Laser) em um estado caótico; segue abaixo a descrição pelo Dr. Huebner:

“As medidas foram feitas em um laser de 81.5-mícrons 14NH_3 cw (FIR), bombeado opticamente pela linha P(13) de um laser N_2O via a transição vibracional aQ(8,7) NH_3 . Os dados de intensidade foram registrados por um osciloscópio LeCroy. Nenhum processamento posterior aconteceu. A razão experimental sinal/ruído foi em torno de 300 o que significa algo levemente sob a incerteza de meio bit na conversão AD.” [WEI91]

As pulsações do laser seguem mais ou menos o modelo teórico de Lorenz para um sistema de dois níveis.

Para informações mais detalhadas sobre este conjunto de dados, vide [HUE89a],[HUE89b],[HUE89c].

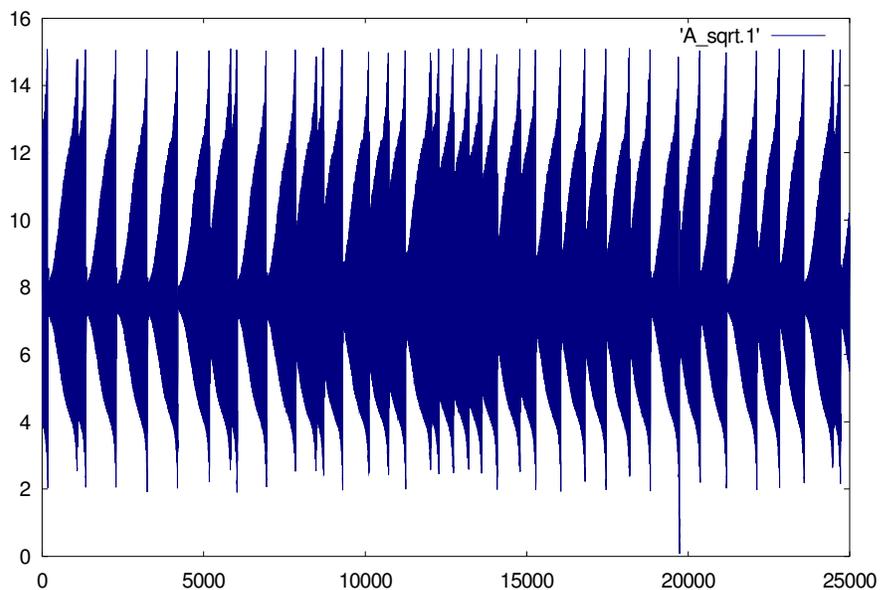


FIGURA 4.8 - Série temporal do laser de NH_3 utilizado na competição de séries temporais do Santa Fe Institute. O eixo vertical representa a amplitude do sinal e o eixo horizontal o tempo.

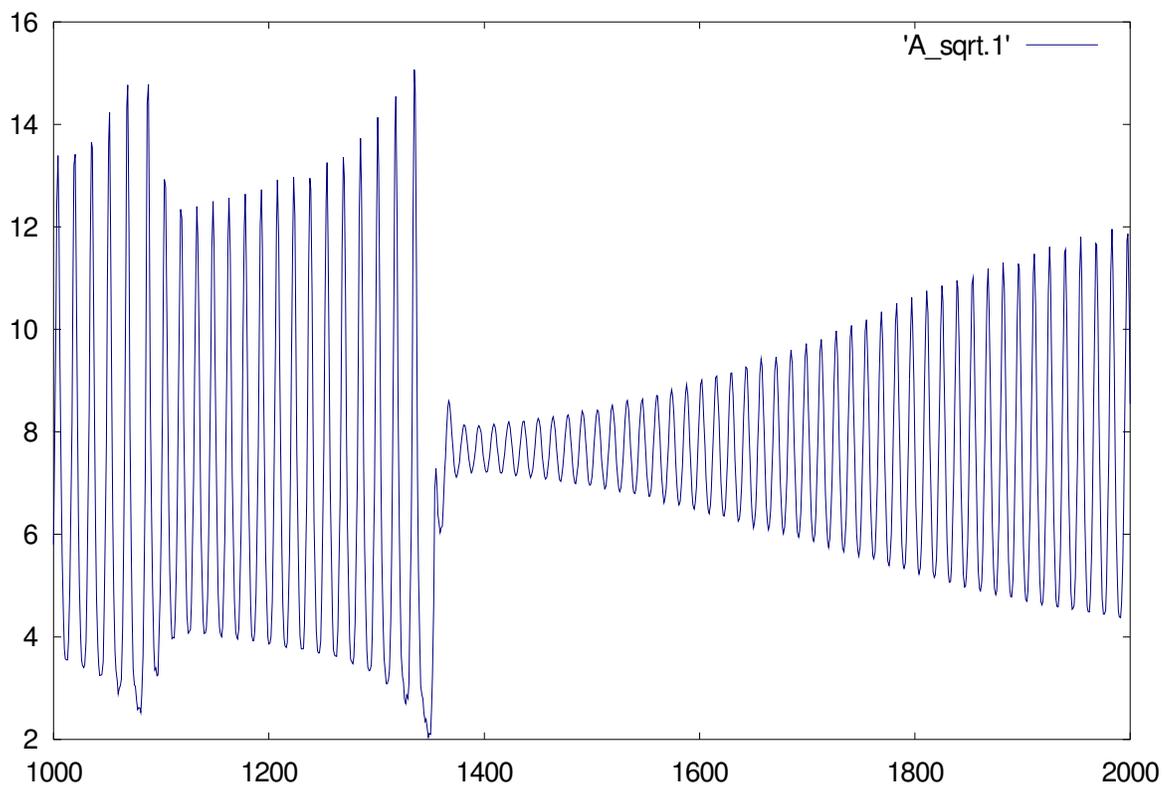


FIGURA 4.9- Detalhe da série temporal do laser de NH_3 . Os eixos são análogos aos da figura 4.8

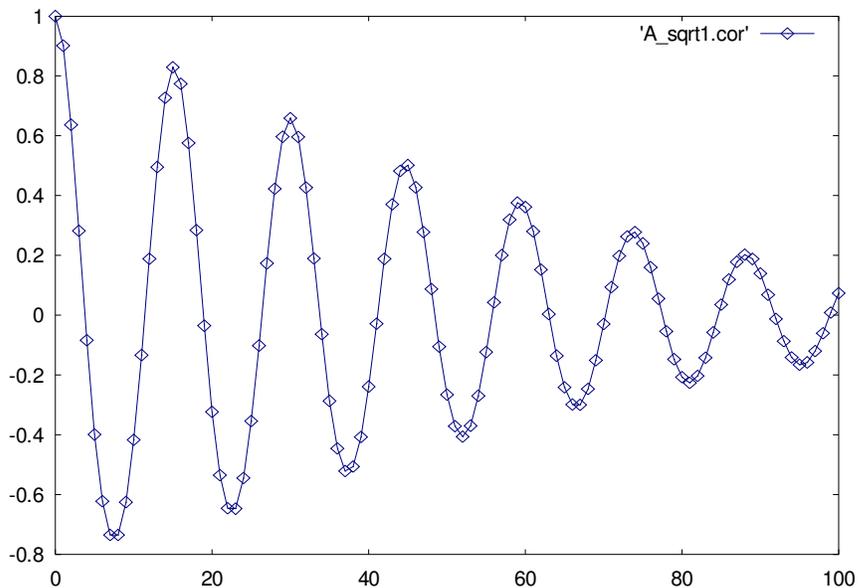


FIGURA 4.10 - Função de autocorrelação para o laser de NH_3 . Na vertical: Correlação σ vs tempo na horizontal.

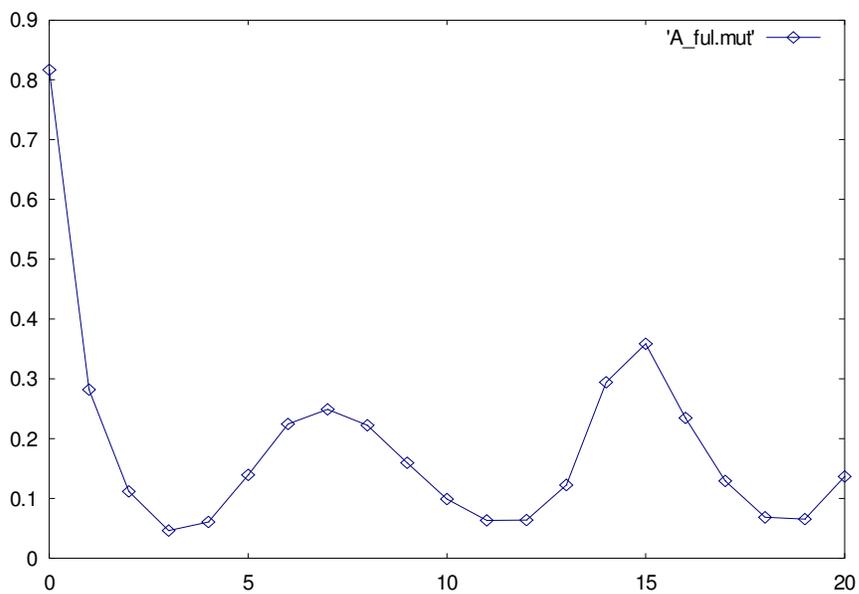


FIGURA 4.11 - Informação mútua para os dados do laser, (I vs τ). O primeiro mínimo utilizado para a escolha do atraso temporal equivale a $\tau=3$.

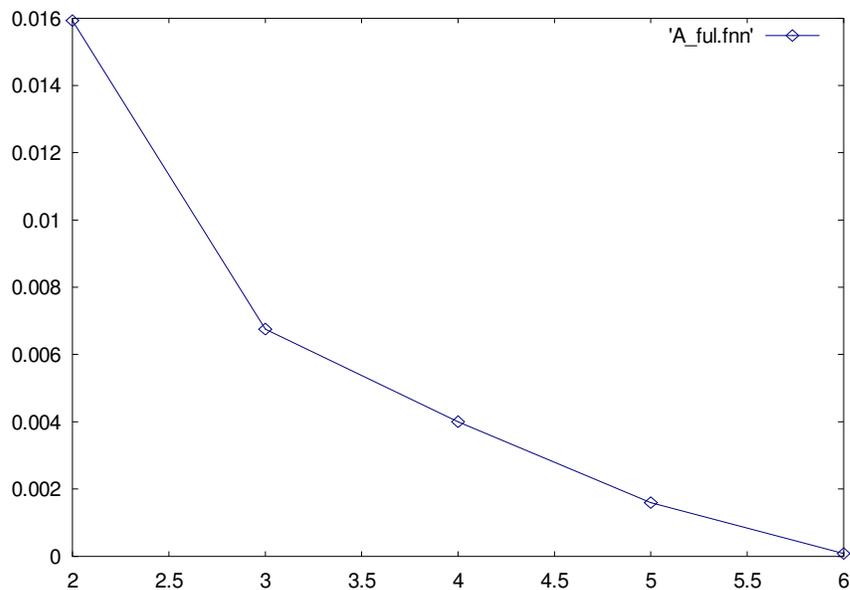


FIGURA 4.12 – Percentual de falsos vizinhos (vert) versus dimensão de imersão para a série da figura 4.9. Mesmo tendo o gráfico alcançado um valor suficientemente ‘pequeno’ apenas em $m=6$, assumimos que o sinal está contaminado com ruído de amostragem, de forma que a efetiva dimensão do sistema seja inferior a este valor.

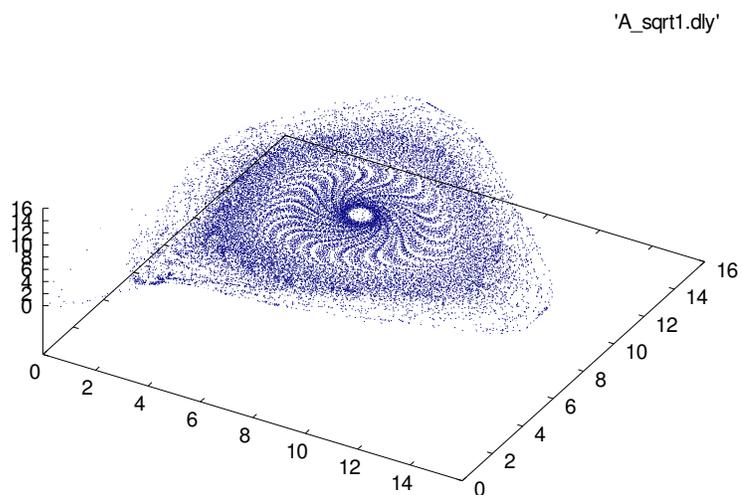


FIGURA 4.13 – Gráfico do ‘atrator reconstruído’ tridimensional (x,y,z) com $m=3$ e $\tau=3$ a partir do sinal do laser de NH_3

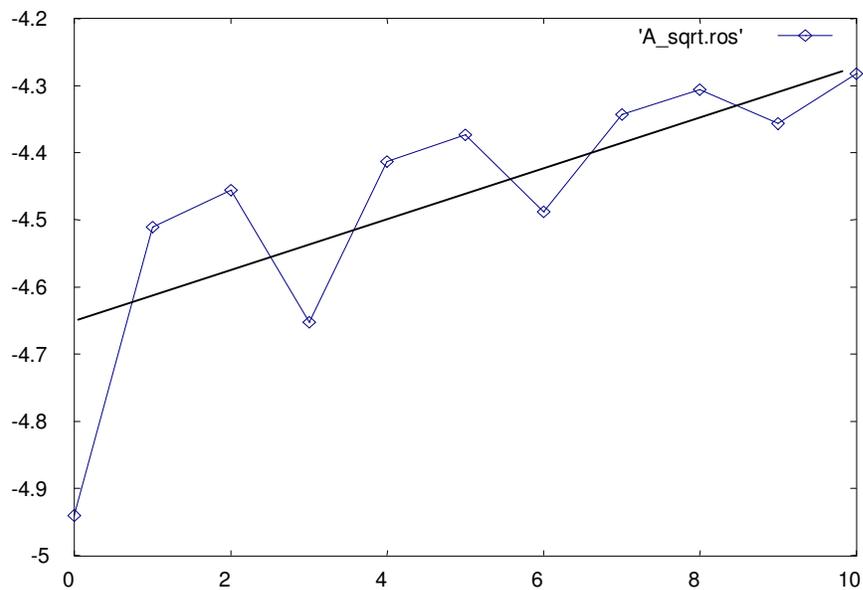


FIGURA 4.14 – Gráfico de divergências médias (vertical) para os dados do laser de NH_3 segundo o método de Rosenstein. O valor estimado para o maior expoente de Lyapunov fica em torno de 0.0375.

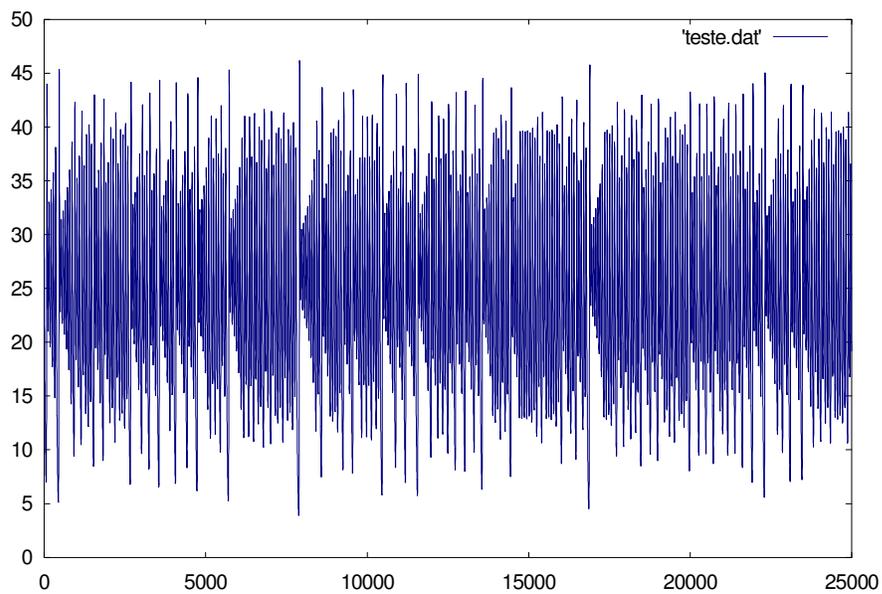


FIGURA 4.15 – Série temporal obtida a partir do eixo z (z vs t) do modelo de Lorenz, para fins de comparação com o laser de NH_3 .

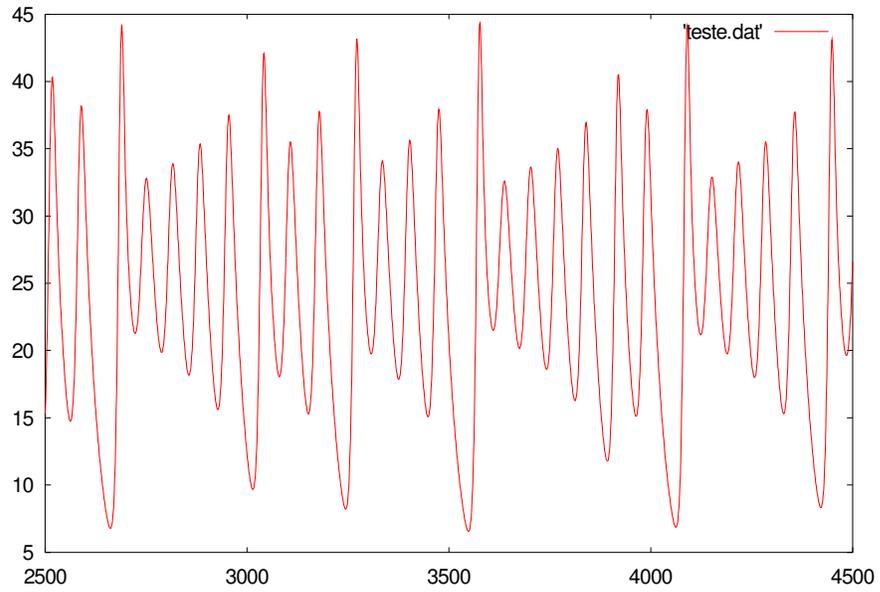


FIGURA 4.16 – Detalhe da figura 4.15, eixos análogos.

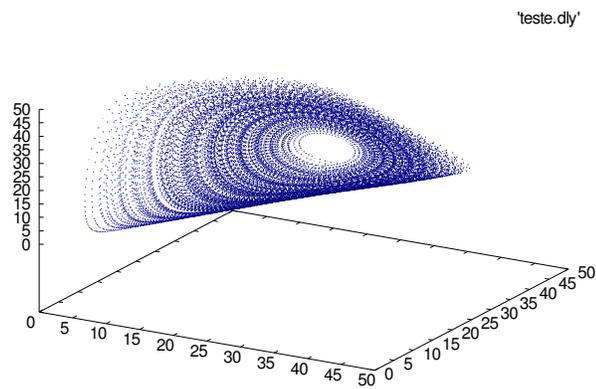


FIGURA 4.17 – Atrator reconstruído a partir do eixo z do sistema de Lorenz

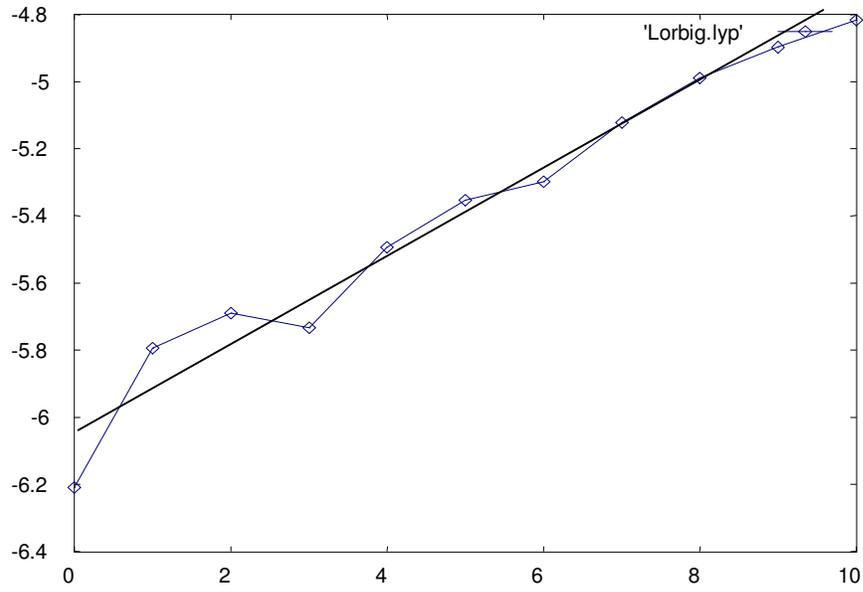


FIGURA 4.18 - Gráfico de divergências médias (vertical) para o sistema de Lorenz segundo o método de Rosenstein. O valor estimado para o maior expoente de Lyapunov fica em torno de 0.125.

Pode-se observar alguma semelhança entre os gráficos das figuras 4.13 e 4.17, bem como entre os das figuras 4.9 e 4.16. Uma possível explicação para as diferenças entre os valores calculados para o maior expoente de Lyapunov dos sinais experimentais e simulações, é que o modelo de Lorenz representa o sistema verdadeiro (laser) de uma forma aproximada apenas, de forma que o número de oscilações “regulares” descritas pelo laser antes do *breakdown* que reinjeta a órbita no interior do atrator é bem maior no sistema real do que no modelo matemático. É possível que esta “maior regularidade” seja o fator que leva a abaixar o valor do maior expoente de Lyapunov do sinal. Uma discussão mais aprofundada sobre as características tanto do sistema de Lorenz ou do laser de NH_3 foge ao escopo deste trabalho.

5 Estacionariedade em Séries Temporais

Os métodos para reconstrução de atratores, não diferentemente de grande parte dos métodos estatísticos convencionais, possuem a restrição de que o sinal em estudo seja estacionário. Em muitos casos, sinais globalmente não-estacionários apresentam janelas de estacionariedade localizadas no tempo; no caso de sistemas caóticos com dinâmica determinística, podemos interpretar essa mudança de estacionariedade como correspondendo a uma variação de um ou mais dos parâmetros de controle do sistema (também chamados de graus de liberdade não-ativos) e assim podemos, então, buscar intervalos de tempo na série temporal nos quais estes parâmetros se mantenham (aproximadamente) constantes, ou seja, variem tão pouco de forma que esta mudança não afete a dinâmica do sistema em observação de modo a impossibilitar o emprego dos métodos de reconstrução. Uma das técnicas disponíveis para tratamento de sinais não-estacionários consiste na tarefa de classificação destes segmentos estacionários em grupos identificando, desta forma, possíveis padrões para as flutuações dos parâmetros do sistema. A seguir, serão discutidas algumas medidas para detecção de não-estacionariedade em uma série temporal e logo em seguida será tratado o problema de clusterização aplicado a séries temporais.

5.1 Medidas de estacionariedade

Antes de aplicar algum método para tratamento de sinal não-estacionário, é interessante que possamos saber se o dito sinal é realmente não-estacionário. O procedimento mais comum consiste em se dividir o conjunto de dados em duas partes e calcular algum parâmetro estatístico de cada um dos segmentos, na esperança de que a diferença entre estes parâmetros reflita o caráter não-estacionário do sinal. Diversos trabalhos propõem métodos que argumentam ser mais adequados para detecção de não-estacionariedade em séries temporais. Dentre os diversos métodos existentes iremos dar uma breve revisão naqueles propostos por Kennel[KEN98] e Eckmann[ECK87].

O método proposto por Kennel para detecção de não-estacionariedade baseia-se na idéia que, em uma dinâmica não estacionária, pontos próximos no espaço são também próximos no tempo (excluídos vizinhos de um tempo de correlação pré-determinado). O método não exige que se divida a série em segmentos, pois baseia-se em histogramas de uma quantidade $D = |T(\mathbf{x}^m) - T(\mathbf{x})|$, a qual denota a distância temporal entre um ponto \mathbf{x} e seu vizinho mais próximo \mathbf{x}^m . Para dados não-estacionários, espera-se que os valores de D apresentem valores notavelmente menores do que para dados estacionários. É essencial para a aplicação correta do método que se excluam vizinhos temporais de um tempo de correlação W , que pode ser determinado como algo em torno de 3 a 5 vezes o valor do primeiro mínimo de informação mútua. Igualmente importante mas menos óbvia é a exclusão de pares de trajetórias com correlações seriais, ou seja, pontos vizinhos que quando iterados continuam sendo vizinhos, de forma que não produzem nova informação. Mais detalhes sobre o método podem ser encontrados em [KEN98].

A proposta de Eckman *et al* é bastante popular, e consiste no emprego do que os autores chamaram de gráficos de recorrência (em inglês, recurrence plots), para determinar se um sinal é estacionário ou não. O método é bastante simples e baseia-se na construção de gráficos

bidimensionais de dimensão $N \times N$ onde cada elemento pode estar ligado ou não dependendo de o ponto x_i estar dentro de uma esfera de raio ε centrada no ponto x_j no espaço reconstruído. Para sinais estacionários, os pontos devem se apresentar distribuídos de maneira aproximadamente uniforme no *lattice*; para dados não-estacionários notar-se-á a presença de regiões mais densas e mais rarefeitas, formando padrões freqüentemente interessantes, cuja interpretação quantitativa ainda não se encontra formulada. Uma revisão bastante rica sobre o método encontra-se em [CAS97].

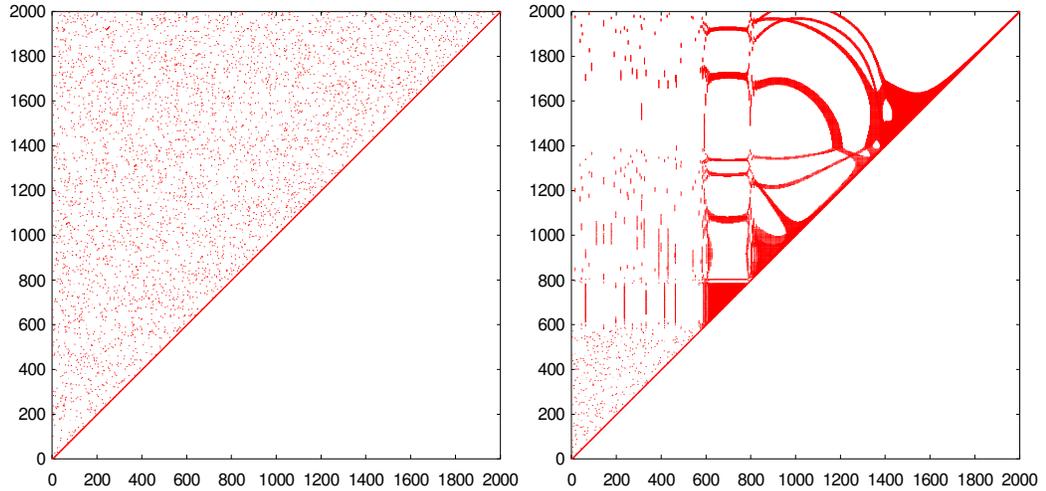


FIGURA 5.1 Gráficos de recorrência para o mapa de Ikeda estacionário (esq) e com variação contínua de um de seus parâmetros (dir). O significado dos eixos está contido no texto da seção 5.1.

5.2 Clusterização

Técnicas para agrupar N elementos similares em K grupos são chamadas técnicas de *clusterização*. Podem ser encontradas na literatura implementações destas técnicas sob as mais diferentes modalidades – supervisionadas, não-supervisionadas, utilizando redes neurais artificiais [AMI89], etc. Nosso ponto central aqui será a clusterização não-supervisionada por pares baseada em uma função custo que representa uma soma das dissimilaridades médias internas de cada cluster, tendo a forma[BUH94]:

$$\varepsilon_K^{pc}(\mathbf{M}) = \sum_{v=1}^K \frac{1}{2p_v N} \sum_{k=1}^N \sum_{t=1}^N M_{kv} M_{tv} D_{kt} \quad (5.1)$$

com $p_v = \sum_{k=1}^N \frac{M_{kv}}{N}$

Sendo K o número de clusters pré-determinado, N o número total de elementos a serem classificados, $\mathbf{D} \in \mathfrak{R}^{N \times N}$ a matriz de proximidade (ou dissimilaridade) entre cada elemento, e $\mathbf{M} \in$

$\{0,1\}^{N \times K}$ é a matriz de associação que será manipulada utilizando um método estocástico com fins de minimizar a função custo $\varepsilon(\mathbf{M})$. O problema encaixa-se, então, em uma classe de problemas chamados de problemas de otimização combinatorial.

5.2.1 Otimização combinatorial

Grande parte dos métodos utilizados para a tarefa de otimização combinatorial envolvem algoritmos estocásticos, mais especificamente particularizações de cadeias de Markov [FOG00] como, por exemplo, simulated annealing e algoritmos genéticos.

Simulated annealing é um método bastante simples e recente para otimização estocástica. Baseia-se no processo físico chamado recozimento (annealing) utilizado para se minimizar a energia interna de um sólido, que consiste em aumentar a temperatura do banho térmico em que o sólido se encontra até um valor máximo quando o sólido muda para a fase líquida, fazendo com que diminua a intensidade das interações intermoleculares. Em seguida resfria-se o sólido lentamente de forma que este atinja seu estado fundamental com energia interna mínima.

Em 1953, Metropolis, usando técnicas do tipo Monte Carlo, introduziu um algoritmo simples que simula o processo físico de recozimento. A função a ser minimizada desempenha o papel da energia, sendo cada novo estado aceito ou não de acordo com uma regra probabilística [PRE97]. Em [ROS90] os autores propõem a remoção da componente estocástica do algoritmo, chamando-o de “deterministic annealing”.

5.2.2 Algoritmos genéticos (AG)

Segundo [FOG00], o método para evolução simulada conhecido como algoritmos genéticos surgiu pelo menos em três ocasiões independentes ao longo de aproximadamente uma década, e foi inicialmente proposto por Fraser, Bremermann e em publicações de estudantes de John Holland na Universidade de Michigan (vide referências em [FOG00]). Alguns textos conhecidos abordando o assunto são [50,51,52].

Essencialmente, a técnica incorpora uma população de indivíduos codificados como “cromossomos” em uma representação normalmente binária, propaga cópias destes indivíduos baseado em um critério de avaliação externo, e gera novos indivíduos para cada nova geração aplicando mutação e recombinando diferentes membros da população.

Algoritmos genéticos são largamente utilizados na resolução de problemas de otimização combinatorial, principalmente quando a função custo é do tipo multimodal [TAN95].

O método selecionado para manipular a matriz \mathbf{M} baseia-se em um modelo de algoritmos genéticos consistindo de uma variação daquele conhecido como algoritmo 1+1 (um pai gera um filho por mutação – o melhor sobrevive) [BÄC96], que é um dos modelos de AG mais simples conhecidos. O algoritmo 1+1 é descrito na figura 5.2.

```

t := 0;
inicializa P(0) := {ā(0)} ∈ I onde I = {0,1}^l;
avaliar P(0) : {Φ(ā(0))} onde Φ = f;
enquanto (t(P(t)) ≠ verdade) fazer
    mutar : ā'(t) := m_{pm}(ā(t));
    avaliar P'(t) := {ā'(t)} :
        {Φ(ā'(t))} onde Φ = f;
    selecionar : P(t+1) := s_{(1+1)}(P(t) ∪ P'(t));
    t := t + 1;
fim enquanto

```

FIGURA 5.2 – O algoritmo genético 1+1

Com o objetivo de minimizar a tendência de que a solução fique presa em um mínimo local induzida pelas mutações locais normalmente utilizadas, foi feita uma pequena alteração no algoritmo introduzindo a possibilidade de que a solução dê um salto com uma probabilidade $p_s=1-p_m$. Dessa forma, o operador de mutação $m_{\{pm\}}(\cdot)$ toma a seguinte descrita na figura 5.3.

```

m_{pm}(ā(t)) :=
{
p_m := random[0,1];
se p_m < p_{randomize} fazer ā'(t) = random(t);
se p_m > p_{randomize} fazer ā'(t) = mut_local(ā(t));
}

```

FIGURA 5.3 – Operador de mutação modificado

A inclusão de mais indivíduos na população deve produzir uma melhoria no desempenho do algoritmo, apesar de causar algum aumento no custo computacional. Neste trabalho foi utilizada a versão clássica do 1+1, com um pai e um filho, pela simplicidade de implementação e desempenho satisfatório na tarefa proposta.

5.3 Medidas de dissimilaridade

5.3.1 Estacionariedade fraca

Em se tratando de análise clássica de sinais, uma medida de estacionariedade muito usada é a chamada medida de estacionariedade fraca, baseada na média e na variância do sinal. Para

processos lineares estocásticos com ruídos puramente gaussianos, esta estatística fornece uma medida razoável de (não) estacionariedade uma vez que as estatísticas baseadas em média, variância e função de autocorrelação carregam total informação sobre processos deste tipo.

Embora somente estimativas de média e função de autocorrelação ou uma análise espectral possam caracterizar o comportamento dinâmico do sistema (linear) de uma forma *completa*, uma medida de estacionariedade (ou similaridade) do tipo RMS pode nos dizer se o sistema não é estacionário, embora um teste de não-estacionariedade com um resultado negativo não possa ser interpretado como indicativo de um sistema em regime estacionário. Comparações entre conjuntos de dados originados de sistemas com dinâmica não-linear baseadas em estatísticas lineares não caracterizam adequadamente a estacionariedade deste tipo de dado [KAN97].

5.3.2 Erros de predições cruzadas

Em [SCH97b], o autor argumenta que estatísticas comparativas para sistemas não-lineares baseadas em medidas de valor único acarretam uma considerável perda de informação. De fato, pode-se utilizar expoentes de Lyapunov, dimensões ou outras medidas para agrupar segmentos similares de uma série temporal por meio de histogramas, embora tal método de classificação pode fornecer resultados bastante enganosos. Por exemplo, é possível se variar um dos parâmetros do mapa do padeiro generalizado sem que se altere o maior expoente de Lyapunov [FAR83]. Uma abordagem conveniente é comparar-se as séries entre si, gerando uma medida de similaridade a posteriori, ao invés de comparar os escalares gerados a priori. Neste sentido, em [SCH97b] é proposto o uso de erros de predições cruzadas para a comparação de séries temporais geradas a partir de sistemas não-lineares em regime caótico, alegando que estas medidas possuem robustez quando se compara sinais “ruidosos”.

A dissimilaridade γ_{ij} entre os elementos i e j é calculada a partir do erro médio de predição quando utilizamos uma série temporal como base de dados para prever o futuro de uma outra. Isto é feito da seguinte forma:

① - Monta-se os vetores de embedding $\vec{x}_k = (x_{k-(m-1)\tau}, x_{k-(m-1)\tau}, \dots, x_k)$

② - Para cada ponto \vec{x}_n pertencente à ε -vizinhança de \vec{x}_k (cuja última componente é x_k), computamos o valor médio para os pontos futuros x_{n+l} (última componentes do vetor \vec{x}_{n+l})

③ - Calculamos a diferença entre x_{k+l} e o valor médio obtido anteriormente, acumulando no elemento γ_{nk} sob a forma de um erro médio quadrado.

Segundo o autor, esta medida apresenta bastante robustez quando se compara sinais caóticos na presença de ruído. Tal medida ainda deve servir como um quantificador de ruído para dados não-lineares determinísticos.

5.3.3 Somas de correlações cruzadas normalizadas

A soma de correlação cruzada generalizada

$$C_{XY}(\varepsilon) = \frac{1}{|X||Y|} \sum_{\bar{x} \in X} \sum_{\bar{y} \in Y} \Theta(\varepsilon - |\bar{x} - \bar{y}|) \quad (5.2)$$

introduzida em [KAN94a] é uma variação da soma de autocorrelação de Grassberger-Procaccia (eq 3.25) como uma medida comparativa entre dois conjuntos. Uma possível medida de dissimilaridade entre dois conjuntos pode ser a própria C_{XY} ou $C_{XY} / \sqrt{C_{XX} C_{YY}}$. Se estes dois conjuntos formam atratores similares, então em escalas grandes as duas integrais de correlação e a correlação cruzada possuirão valores comparáveis. A diferença entre dois atratores irá mostrar-se em escalas menores, onde as correlações cruzadas caem rapidamente a zero.

Note que a integral de correlação de Grassberger-Procaccia prevê que, para fluxos contínuos, deva-se excluir da soma vizinhos temporais. Assim, quando comparamos uma série com ela mesma (no caso de um elemento da diagonal da matriz de dissimilaridades) devemos excluir $|i - j| < \tau$ onde τ é um tempo de correlação que deve valer no mínimo 1 para o caso de mapas.

Apesar de possuir uma fundamentação teórica bastante consistente, as somas de correlação apresentam uma grande instabilidade quando aplicadas a dados ruidosos. Em [ECK92] encontramos argumentos rigorosos provando que este método de estimativa possui uma limitação natural não fornecendo valores além de $2 \log_{10} N$ sendo N o número de pontos para o conjunto de dados disponível. Quando esta borda é saturada não se pode concluir que exista uma dinâmica de baixa dimensionalidade presente.

5.3.4 Distribuições suavizadas

Press et al, no conhecido Numerical Recipes [PRE97] sugerem o uso do teste χ^2 para a comparação de duas distribuições empíricas. No entanto, o uso ingênuo desta abordagem pode levar a resultados enganosos. A medida invariante $\rho(\mathbf{x})$ pode sofrer uma pequena translação de forma que as caixas preenchidas da nova e da antiga distribuição não possuam mais nenhuma superposição, embora estruturalmente o sistema possa ter sofrido alterações praticamente desprezíveis. Ainda pior, se utilizarmos o teste χ^2 para comparar distribuições $\rho_1(\mathbf{x})$, $\rho_2(\mathbf{x})$ onde se tem sobreposição de oscilações periódicas e comportamento estocástico, é bem provável que a distância entre medidas $\rho_1(\mathbf{x})$ e $\rho_2(\mathbf{x})$ não forneça a dissimilaridade corretamente.

Em [DIK96] os autores introduzem uma medida de distância entre distribuições com o conceito de *smoothed distributions*, fazendo uma convolução da distribuição medida com uma gaussiana (eqs. 5.3 e 5.4).

$$\rho'_k(\vec{r}) = \int d\vec{s} \rho_k(\vec{s}) \kappa(\vec{r}, \vec{s}) \quad \text{para } k \in \{1, 2\}, \quad (5.3)$$

$$\text{onde } \kappa(\vec{r}, \vec{s}) = (\sqrt{2\pi d})^{-m} e^{-|\vec{r}-\vec{s}|^2/(2d^2)} \quad (5.4)$$

com $||$ denotando a norma Euclideana em \mathfrak{R}^m e $d > 0$ sendo um comprimento de escala fixo, ou largura de banda.

A quantidade

$$Q = (2d\sqrt{\pi})^m \int d\vec{r} [\rho'_1(\vec{r}) - \rho'_2(\vec{r})]^2 \quad (5.5)$$

define a distância entre as distribuições de probabilidade ρ'_1 e ρ'_2 , para qualquer $d > 0$.

Esta abordagem apresenta-se particularmente útil quando tratamos com um conjunto de dados composto, por exemplo, por sobreposição de oscilações periódicas puras e movimento estocástico. No mesmo artigo, os autores mostram que tal medida de dissimilaridade só assume o valor zero quando as distribuições ρ_1 e ρ_2 são realmente idênticas.

Medidas baseadas em probabilidades tornam-se interessantes à medida que se perde determinismo. Estatísticas baseadas em distribuições espaciais podem quantificar características geométricas de objetos no espaço de estados como flutuações da medida invariante do sistema em estudo, seja ela oriunda de uma dinâmica determinística ou processo estocástico, com a desvantagem de não conter informação sobre as regras dinâmicas do fenômeno observado. Medidas de dissimilaridade utilizando probabilidades de transição, apesar de carregarem informação sobre a dinâmica do sistema (a definição de estacionariedade baseia-se nestas probabilidades), exigem um número muito maior (exponencialmente) de pontos amostrais e a construção de histogramas usualmente requer uma grande quantidade de registros de alta precisão.

6 Experimento

Com a finalidade de verificar o desempenho dos métodos vistos nos capítulos 4 e 5 para o tratamento de um problema real, realizou-se um experimento envolvendo diferentes tipos de dados irregulares, obtidos de simulações numéricas e sinais experimentais reais.

Este experimento foi dividido em duas partes: em uma etapa preliminar, com o objetivo de validar a consistência dos métodos para dados irregulares compostos de diferentes quantidades de caos e estocasticidade, foram submetidos à *clusterização* dados artificialmente com estas características. Foram geradas séries temporais compostas de segmentos estacionários previamente conhecidos, para que se pudesse verificar se os elementos eram agrupados de forma correta.

Em seguida, partiu-se para o estudo de dados reais consistindo de leituras eletroencefalográficas de superfície. Tendo em vista que os métodos lineares apresentam algum sucesso na separação de estágios do sono, foram aplicadas diferentes medidas não-lineares com o objetivo de, pelo menos, discriminar algum dos estágios dos demais. Embora os métodos espectrais (lineares) mostrem um desempenho razoável na tarefa de identificação dos estágios de sono, estes não possuem concordância com a física envolvida, uma vez que não se espera que o cérebro seja um mecanismo linear[AMI89]. A princípio, medidas não-lineares devem conseguir extrair mais profundamente a riqueza do sistema em questão. O problema de segmentação automática de estágios de sono ainda é um problema em aberto sem uma solução ótima, o que estimula o desenvolvimento de novos métodos.

As séries numéricas utilizadas foram geradas, por meio de métodos iterativos, a partir de órbitas dos sistemas de Ikeda, Rössler e Lorenz (descartando-se um certo período de transiente), e misturadas com diferentes quantidades de ruído estocástico. Os sistemas de Roessler e Lorenz foram discretizados utilizando uma seção de Poincaré, para que se obtivessem mapas bidimensionais em tempo discreto. Todos os algoritmos foram implementados utilizando os compiladores Gnu C e DJGPP (versão do Gnu C para MS-DOS), devido a sua compatibilidade com o padrão ANSI para a linguagem C, gerando executáveis via linha de comando. Esta opção garante independência de plataforma e fácil portabilidade, permitindo que usuários de praticamente qualquer sistema operacional atual possam utilizá-los, caso haja interesse.

$$\begin{aligned}
 \text{Roessler} & \left\{ \begin{array}{l} \dot{x} = -(y + z) \\ \dot{y} = x + \frac{1}{5}y \\ \dot{z} = \frac{1}{5} + z(x - \mu) \end{array} \right. \\
 \text{Lorenz} & \left\{ \begin{array}{l} \dot{x} = -\sigma(x - y) \\ \dot{y} = -xz + \rho x - y \\ \dot{z} = xy - bz \end{array} \right. \\
 \text{Ikeda} & f_{\mu}(z) = a + bze^{ik - [i\mu/(1+|z|^2)]}
 \end{aligned}$$

FIGURA 6.1 – Modelos geradores das séries numéricas utilizadas para validação. Os sistemas representados por equações diferenciais foram integrados utilizando o método Runge-Kutta de quarta ordem e discretizados por meio de uma seção de Poincaré.

6.1 Séries sintéticas

Na figura 6.2 vemos gráficos de algumas das séries utilizadas para a validação de cada uma das medidas de dissimilaridade e do minimizador, ainda sem a adição de ruído. Todas as medidas não-lineares (erros de predições cruzadas, somas de correlação, distribuições) apresentaram resultados excelentes e praticamente 100% de sucesso na clusterização de dados ‘limpos’ gerados por simulações numéricas de sistemas como Lorenz, Ikeda e Roessler, apresentando alguns problemas (já previstos) devido ao condicionamento de alguns parâmetros, como o raio ϵ da hipersfera utilizada nas predições cruzadas ou na soma de correlações. Após ‘calibrados’ estes parâmetros, os algoritmos passaram a fornecer resultados coerentes com o esperado, como pode ser visto no anexo 1. As séries utilizadas consistiram de 80000 pontos, sendo que o sinal gerado para a segunda metade diferia do primeiro por uma pequena variação de algum dos parâmetros do sistema (vide figura 6.1).

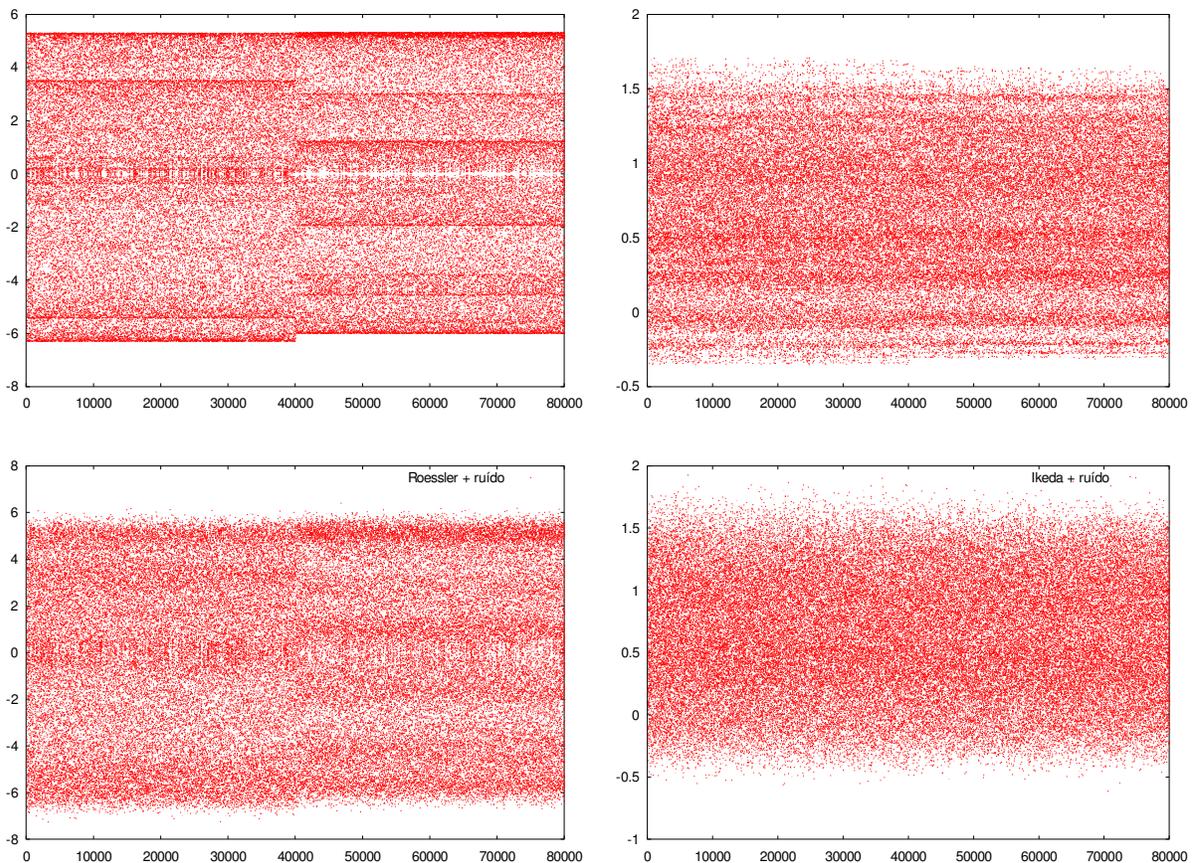


FIGURA 6.2 – Séries obtidas a partir de iteradas do sistema de Roessler (coluna da esquerda) e do mapa de Ikeda (coluna da direita).

Proporcionalmente à quantidade de ruído estocástico inserido (foram utilizadas duas amplitudes absolutas de ruído gaussiano, 0.1 e 0.3), foi notada uma degeneração na qualidade dos resultados de cada medida, de acordo com a tabela 6.1. Os rótulos dados para o desempenho de cada medida representam, aproximadamente: Excelente \approx 0-10% de erros; Muito bom \approx 20-30% ; Bom \approx 30 – 40%; Regular \approx 50 - 60%; Ruim \approx 60 - 70% ; Péssimo \geq 70%.

TABELA 6.1 – Desempenho das diferentes medidas de dissimilaridade frente aos sinais gerados.

γ_{ij}	Somas de correlação cruzadas	Erros de predição cruzados	Distribuições ‘suavizadas’	Média e Variância
Tipo de dados				
Séries geradas numericamente	Muito Bom	Excelente	Excelente	Ruim
Séries geradas + ruído estocástico	Ruim	Muito bom	Muito bom / Bom	Ruim

De acordo com os resultados obtidos, percebe-se que a medida de dissimilaridade baseada em erros de predições cruzadas, apesar de fundamentar-se na existência de uma regra dinâmica determinística e de baixa dimensionalidade no mecanismo gerador dos dados, apresenta uma considerável robustez frente à presença de ruído estocástico, principalmente quando comparamos seu desempenho com o das somas de correlação, o que concorda com os resultados obtidos em [SCH97a]. Tal robustez sugere que se dê preferência a esta medida quando se quer comparar sistemas onde predomina um regime caótico determinístico, com uma pequena quantidade de ruído presente. Predições também são úteis quando se deseja quantificar o nível de ruído do sinal, comparando-se séries limpas e ruidosas, de forma análoga a um processo de *surrogate data*. Neste caso, os erros de predição fazem o papel da quantidade não-linear mencionada na seção 4.3.1 deste texto.

Diferenças entre média e variância, apesar de serem largamente utilizadas como medidas de estacionariedade, não caracterizam corretamente flutuações na estacionariedade de sistemas não-lineares e, mesmo para sistemas estocástico-lineares, não são capazes de caracterizar completamente a dinâmica do processo (e por isso esta medida é chamada de estacionariedade fraca), tarefa que é desempenhada pela média em conjunto com a função de autocorrelação do sinal ou pelo espectro de potências, ou seja, estas medidas refletem as condições de todos os parâmetros da equação 4.1. Por estas razões, dissimilaridades baseadas em média e variância não se apresentam adequadas dentro do contexto deste trabalho.

No que diz respeito à *clusterização* propriamente dita, é importante ressaltar que o processo de minimização da equação 5.1 tende a agrupar elementos menos dissimilares e pode, dependendo da forma da matriz de dissimilaridades, reunir elementos de ‘diferentes classes’ em um mesmo cluster, em vista destas serem muito similares entre si quando comparadas aos demais elementos, pertencentes a outras classes.

Com a finalidade de evitar que elementos de diferentes grupos altamente similares entre si em relação aos demais sejam classificados de forma indesejada (i.e. devido a flutuações na matriz de dissimilaridades, a classificação desejada corresponde a um mínimo local com um custo muito próximo ao do mínimo global de $\epsilon(\mathbf{M})$), pode-se dar preferência para *clusters* previamente organizados de uma forma semelhante à da configuração esperada (próximos, de acordo com a mesma métrica utilizada na função custo, à solução considerada ótima). Um artifício conveniente de efeito análogo ao proposto em [SCH97a] onde o autor sugere a adição de um termo de *bias* à função custo, favorecendo clusters de tamanhos iguais, é inicializar-se \mathbf{M} com segmentos inteiros agrupados em cada cluster. Experimentos realizados indicaram que o método se mostrou eficiente, mesmo quando ainda existiam alguns elementos fora do cluster correto.

Uma explicação para a ocorrência de tal fenômeno é que, devido às características discriminativas específicas a cada medida de dissimilaridade, estas podem gerar hipersuperfícies de erro com mínimos locais correspondentes a diversos padrões mais ou menos aceitáveis. Espera-se que a qualidade destes resultados aumente, em média, à medida que decresce $\epsilon(\mathbf{M})$, para que a minimização seja útil. Dessa forma, faz-se interessante a múltipla inicialização do algoritmo de busca com diferentes condições iniciais, e a comparação dos padrões obtidos. No caso dos padrões desejados não corresponderem ao mínimo absoluto da função custo, os valores de $\epsilon(\mathbf{M})$ para estes devem se encontrar na vizinhança do valor mínimo.

A múltipla inicialização (com diferentes condições iniciais), utilizada neste trabalho, é uma alternativa ao uso de populações maiores de soluções, com a vantagem de que cada solução obtida não é “contaminada” com os genes do(s) melhor(es) indivíduos, o que pode facilmente acontecer em algoritmos genéticos com múltiplos indivíduos.

Nas figuras 6.3 e 6.5 temos exemplos de matrizes de dissimilaridade obtidas a partir do sistema de Lorenz utilizando somas de correlação e erros de predições cruzadas, respectivamente. Devido às já mencionadas flutuações na matriz de dissimilaridades, pode ocorrer (como neste caso) que o resultado desejado corresponda a outro mínimo que não seja o global de $\epsilon(\mathbf{M})$ [SCH97b]. A existência de outros atratores locais na superfície de custo sugere que se experimente diferentes condições iniciais, pois diferentes padrões de agrupamento para os elementos podem ser identificados. No primeiro caso a inicialização de \mathbf{M} foi totalmente aleatória, No segundo \mathbf{M} foi inicializada com valores próximos do resultado esperado para se evitar o mínimo global, pelo menos por algumas iterações.

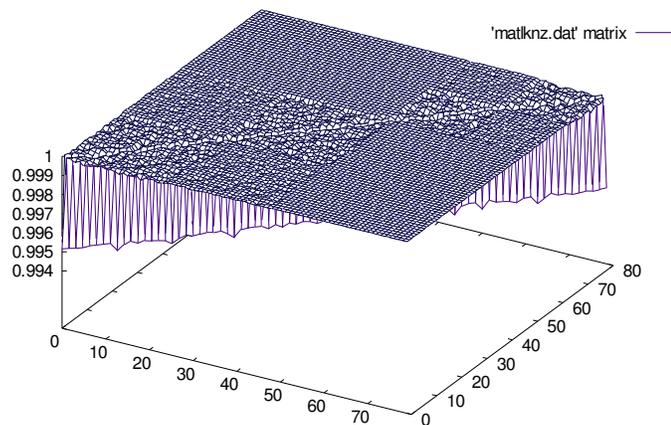


FIGURA 6.3 – Matriz de dissimilaridades para o sistema de Lorenz, sem adição de ruído, utilizando correlações cruzadas.

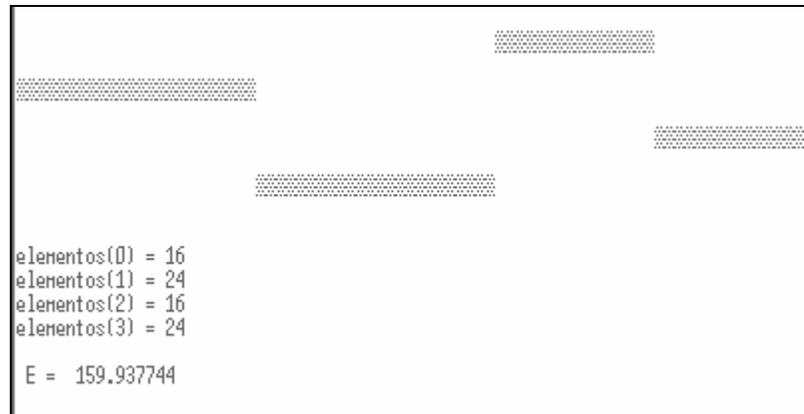


FIGURA 6.4 – Resultado clusterização para a matriz da figura 6.3, a partir de inicialização ordenada ou aleatória.

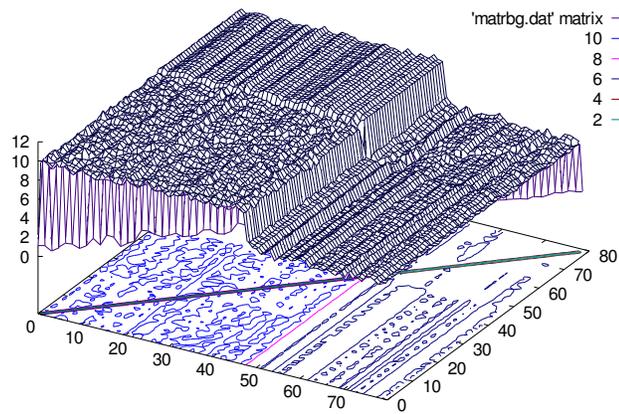


FIGURA 6.5 – matriz de dissimilaridade para o sistema de Lorenz sem ruído utilizando predições cruzadas

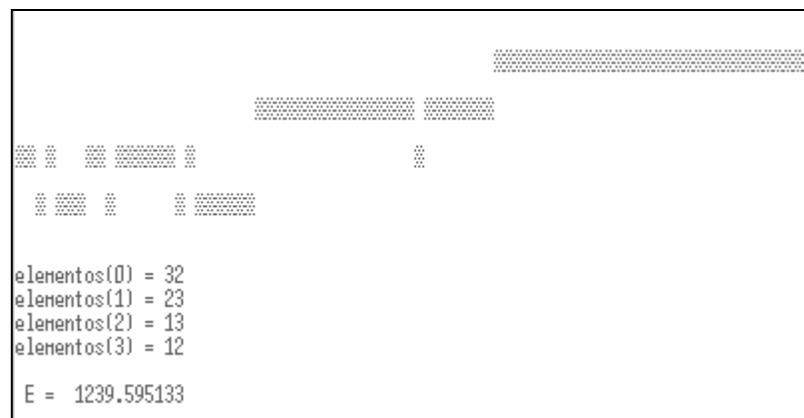


FIGURA 6.6 – Resultado da clusterização para a matriz da figura 6.5, partindo de inicialização sem organização

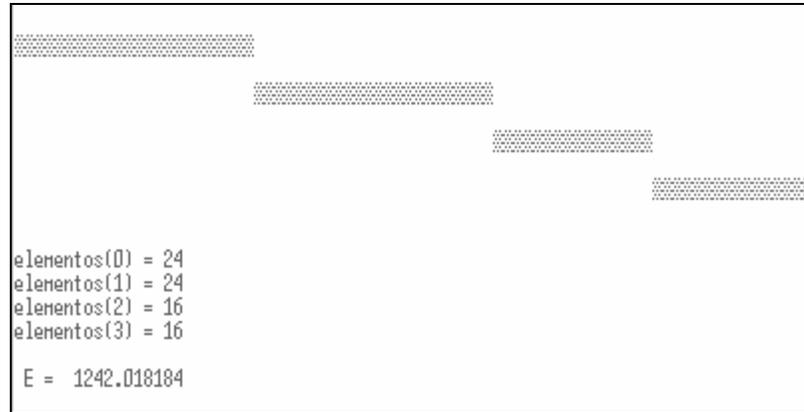


FIGURA 6.7 – Resultado da clusterização para a matriz da figura 6.5, inicialização organizada

Em todos os casos o algoritmo genético se mostrou como sendo a componente menos crítica do processo, convergindo rapidamente para o valor correspondendo ao padrão conhecido, em menos de dois minutos para 80 elementos, em um Pentium 233MHz. Resultados como o apresentado na figura 6.6 mostraram-se muito raros, surgindo apenas quando as flutuações na matriz de dissimilaridade eram muito intensas (veja o degrau da figura 6.5, que não existe na figura 6.6).

6.2 Séries reais (eletroencefalogramas)

Eletroencefalogramas não-invasivos (medidos sobre o escalpo) são um recurso largamente utilizado para estudar ciclos e estágios de sono (um ciclo é composto dos estágios 1, 2, 3, 4 e REM), bem como para diagnosticar patologias como epilepsia ou coma. Consistem na colocação de dois eletrodos sobre o escalpo do paciente, um chamado eletrodo ativo que é posicionado sobre a região que se pretende medir e um eletrodo denominado indiferente que é colocado a alguma distância do primeiro. Dessa maneira, a diferença de potencial entre os eletrodos é registrada periodicamente sob a forma de uma série temporal, possuindo amplitudes que variam em torno de 20-100 μ V e ‘frequências’ que oscilam entre 1-30Hz. Na figura 6.8 vemos segmentos de alguns dos estágios que compõem um ciclo de sono.

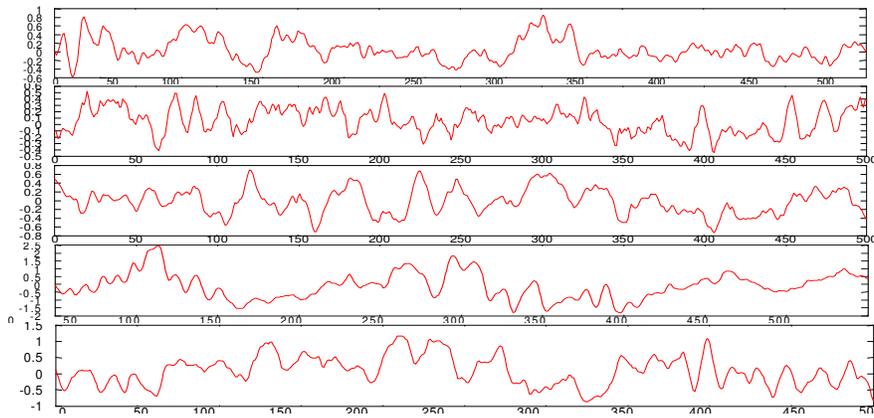


FIGURA 6.8 – De cima para baixo, segmentos de 500 pontos dos estágios REM, 1, 2, 3, 4 amostrados nos pontos C₃-A₂ a uma taxa de 128 Hz.

É comum a associação dos estágios 1 e REM com sono menos profundo até o estágio 4 como sono mais profundo. De acordo com a distribuição dos estágios no sinal, o médico especialista classifica os segmentos do eletroencefalograma em ciclos (vide figura 6.9), podendo assim, a partir do estudo da quantidade, duração e estrutura interna destes ciclos, diagnosticar anomalias e indicar tratamentos. Por exemplo, a maioria dos indivíduos depressivos apresentam um padrão de sono característico, com período pré-REM em torno de 25-35 minutos, ao contrário dos 80-90 minutos de pacientes normais[KAN91].

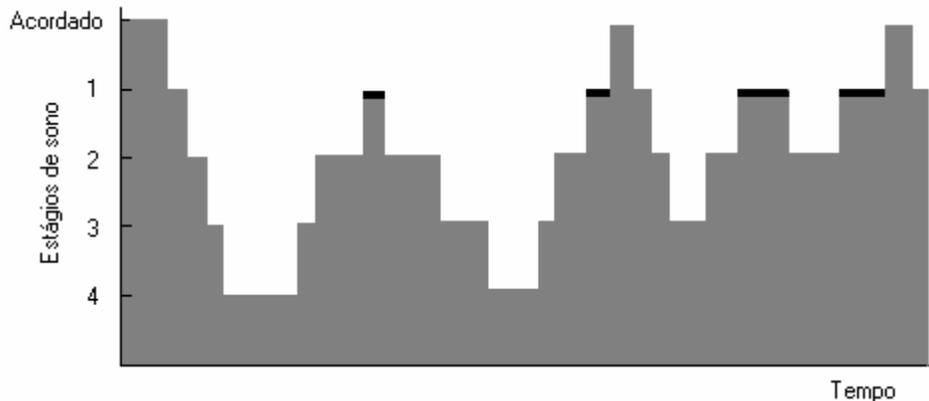


FIGURA 6.9 - Estágios e ciclos de sono ao longo do tempo. As barras mais escuras indicam sono REM.

Usualmente, esta classificação é feita a olho nu pelo próprio médico, observando o sinal e identificando o padrão de cada estágio em função da quantidade e distribuição de cada um dos assim chamados grafo-elementos presentes. Estes grafo-elementos são oscilações de determinadas frequências, com características próprias, localizados no tempo. Alguns exemplos são ondas alfa, ondas vertex, complexos K e fusos (spindles)[KAN91].

Outro tipo de elemento presente (este indesejado) em eletroencefalogramas são denominados artefatos, que consistem em diversas formas de ruído injetado no sinal por alguma causa externa, como descolamento de um eletrodo, acoplamento das canetas (no caso do material impresso) ou algum outro problema com respeito ao equipamento.

Com a finalidade de facilitar o trabalho do médico especialista, acelerando a obtenção de resultados e barateando o custo do processo, faz-se interessante o desenvolvimento de métodos automáticos para auxiliar o profissional na segmentação de eletroencefalogramas. O grande desafio reside no fato de que ainda não há disponível um critério quantitativo ótimo para a segmentação destes sinais.

Uma série de trabalhos recentes abordam seqüências de medidas eletroencefalográficas como possuindo uma componente dinâmica não-linear determinística observável (e.g. [60,61]), utilizando-se dos métodos de reconstrução de atratores[2,20] para tentar identificar, a partir de sinais de EEGs, possíveis atratores (caóticos ou periódicos) correspondentes a uma dinâmica determinística de baixa dimensionalidade. Estes atratores poderiam estar relacionados a estágios de sono normal ou crises epiléticas. O fato de eletroencefalogramas obtidos durante um ataque epilético serem consideravelmente mais regulares do que aqueles durante o sono normal suscitou,

durante muito tempo, a idéia de que pudesse existir uma cascata de bifurcações inversa precedendo o ataque. Em [HER95], os autores utilizam, com sucesso, uma abordagem não-linear para classificar segmentos de eletroencefalogramas contendo ritmos teta (4-7.5 Hz) monomórficos e polimórficos, de indivíduos normais.

Em contrapartida, outros trabalhos igualmente recentes [SCH99] questionam a presença significativa de uma dinâmica (caótica) determinística de baixa dimensionalidade em medidas eletroencefalográficas (mesmo nas intracraniais, onde o número de neurônios sensoriados é muito menor do que em medidas não-invasivas) para a utilização dos métodos de espaço de estados, enfatizando o cuidado que deve ser tomado ao adotar-se uma determinada técnica para tratar um certo sinal e como interpretar os resultados obtidos.

De fato, seria muito intrigante caso fosse comprovadamente possível se assumir que em todo e qualquer EEG exista uma dinâmica determinística de baixa dimensionalidade predominante, em vista da quantidade enorme de elementos (neurônios) envolvidos na geração destes sinais, além da forte atenuação do crânio e tecidos no caso de EEGs de superfície.

Com o intuito de estudar um pouco mais a fundo as razões desta diversidade de opiniões sobre a natureza dos eletroencefalogramas e aplicar os métodos estudados para tal problema, buscando desenvolver uma abordagem correta e eficiente, utilizou-se algumas das técnicas não-lineares vistas anteriormente para a tentativa de detecção de padrões em EEGs de superfície de pacientes normais durante períodos de sono. O problema escolhido foi a segmentação de cada um dos estágios de sono, de um subgrupo ou de pelo menos um estágio em relação aos demais. O estímulo para se estudar os resultados da utilização de uma abordagem não-linear para este tipo de sinal uma vez que as abordagens lineares já apresentam algum resultado (não ótimo) é que, por não se esperar que o cérebro seja um mecanismo com uma dinâmica inerentemente linear, parece claro que os métodos espectrais não sejam capazes de capturar toda a riqueza presente devido à não-linearidade presente neste órgão. Sabendo-se que os EEGs são sinais não-estacionários oriundos de processos biofísicos não-lineares, faz-se desaconselhável o uso de medidas comparativas baseadas na chamada estacionariedade fraca [KAN97].

Os eletroencefalogramas utilizados consistiram de registros multicanais de diferenças de potencial entre dois eletrodos, amostrados a uma taxa de 128Hz. Os dados foram gentilmente cedidos por voluntários normais no laboratório de eletroencefalografia do HCPA/POA e disponibilizados sob a forma de arquivos ASCII contendo números decimais de ponto fixo com precisão de 8 bits.

Cada segmento selecionado de sinal foi subdividido em janelas de comprimento igual a uma época (16 segundos) e estas, por sua vez, submetidas ao algoritmo. Foi selecionado o canal 1 (montagem C_3-A_2) [KAN91] devido a este ser um dos mais utilizados para inspeção visual, sendo assim mais fácil comparar os resultados obtidos pelo método com o observado pelo médico. No total, foram analisados sinais de três indivíduos, o que é uma amostra relativamente pequena, mas a idéia central deste trabalho não é desenvolver uma técnica única (o que pode ser impossível) para a análise de um determinado tipo de sinal, e sim fazer um estudo *qualitativo* referente ao desempenho de uma abordagem mais plausível fisicamente, de acordo um paradigma não-linear (estocástico ou determinístico), a um sinal irregular real.

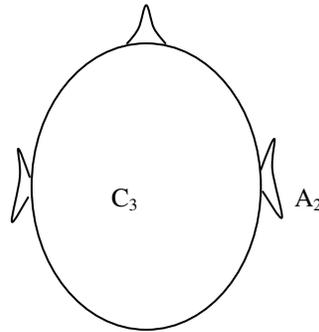


FIGURA 6.10 – Posição dos eletrodos para a montagem C_3 - A_2 .

O número de clusters utilizado foi igual ao número de (macro)segmentos a serem classificados mais um, em vista da possibilidade de existência de artefatos e inomogeneidades nos sinais. Dessa forma, pode-se isolar elementos muito dissimilares do resto (desde que sejam similares entre si). Como sugerido anteriormente, foi utilizada múltipla inicialização, partindo de condições iniciais aleatórias e organizadas similarmente ao resultado esperado.

Com vista nos resultados obtidos anteriormente com os modelos (anexo 1), aplicou-se as diversas medidas de dissimilaridade para clusterização de séries eletroencefalográficas, esperando que, caso existisse um atrator caótico no sinal do EEG, os erros de predições cruzadas, forneceriam a melhor medida comparativa, pois como já era de se esperar as medidas baseadas em correlações cruzadas apresentariam alta instabilidade frente à menor presença de ruído.

Surpreendentemente, os erros de predições cruzadas não apresentaram resultados melhores do que as somas de correlação que, por sua vez, resultaram em elementos distribuídos com muito pouca adjacência em um mesmo cluster ou uma concentração de quase todos os elementos em apenas um dos clusters. Os resultados em ambos os casos foram tão irregulares se comparados com as distribuições suavizadas que não iremos nos deter discutindo-os. Tais resultados sugerem que, caso exista uma dinâmica caótica determinística de baixa dimensionalidade neste sinal, a mesma encontra-se mergulhada em uma quantidade tão grande de ruído que medidas baseadas em reconstrução de atratores baseadas em regras puramente determinísticas inevitavelmente falham.

A partir da análise de alguns histogramas de EEGs, observou-se que a distribuição de densidade de probabilidade do processo poderia ser modelada aproximadamente como sobreposição de uma gaussiana com distribuições de diferentes oscilações periódicas, o que possui certa plausibilidade com o processo físico subjacente, uma vez que os grafo-elementos destacam-se como oscilações aparentemente regulares sobre um fundo ruidoso.

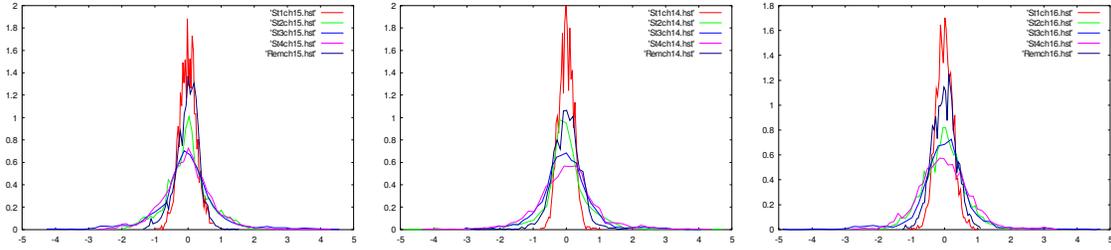


FIGURA 6.11 – Distribuições empíricas de segmentos de 2048 pontos dos estágios 1, 2, 3, 4 e REM.

$$g(x) = \begin{cases} \frac{1}{\sqrt{a-x^2}}, & -1 < x < 1 \\ 0, & x \notin]-1,1[\end{cases}$$

$$\rho(x) \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \sum_{i=1}^N g_i(x+x_c)$$

FIGURA 6.12 – Modelo para as distribuições, onde a , σ , μ e i são parâmetros a serem ajustados.

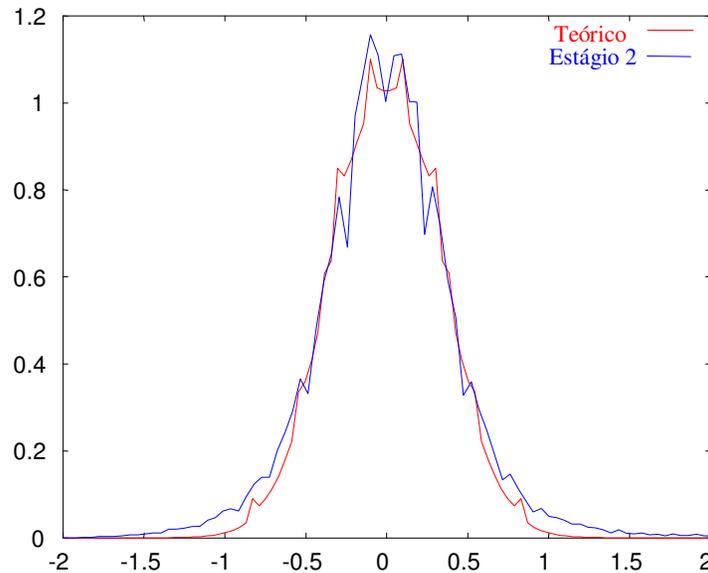


FIGURA 6.13 – Comparação entre distribuições empírica e teórica.

Apesar da atraente possibilidade de se fazer um ajuste dos parâmetros do modelo da figura 6.12 de forma a “fitar” a curva teórica aos dados experimentais e, a partir de diferenças nos parâmetros se gerar uma medida de distância entre duas distribuições, o que desejamos na prática é comparar duas distribuições empíricas – comparar vetores de parâmetros, neste caso, iria apenas diminuir a acuracidade da medida, além de que o processo de ajuste seria consideravelmente mais custoso do que a comparação direta adotada; a este custo seria ainda adicionado o da etapa de suavização, uma vez que esta não poderia ser descartada justamente devido à presença de picos correspondentes a

oscilações periódicas. Alguns exemplos de resultados obtidos podem ser observados nas figuras 6.14 a 6.21.

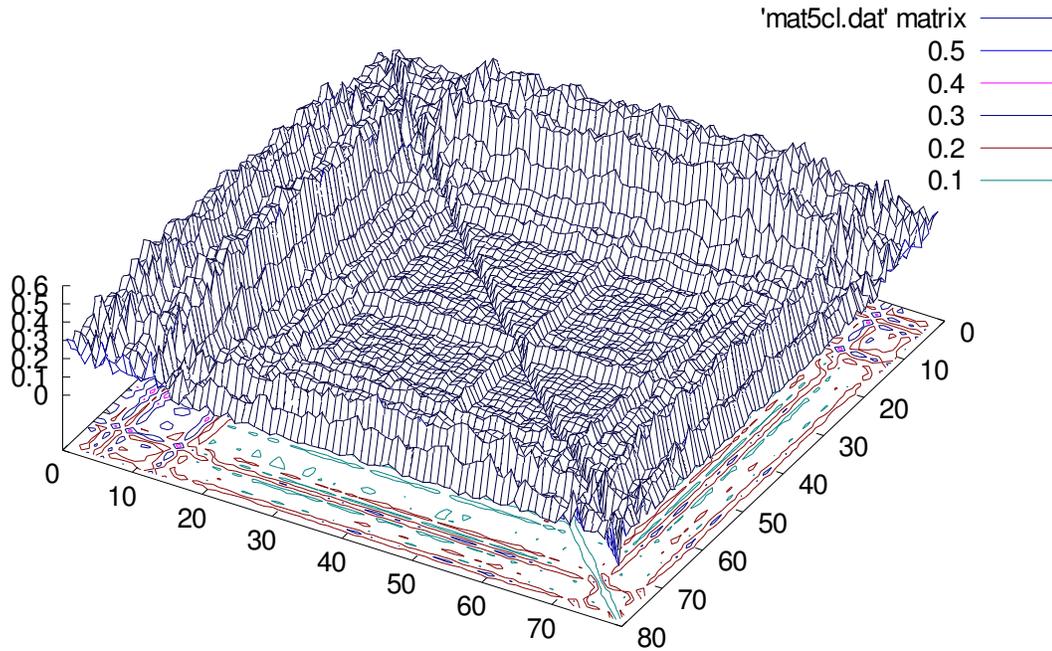


FIGURA 6.14 – Matriz de dissimilaridade para o início de um ‘ciclo’ de sono do paciente 2.



FIGURA 6.15 - Resultado sobre o segmento da figura anterior (163840) pontos, sendo que 128 pontos=1 segundo.

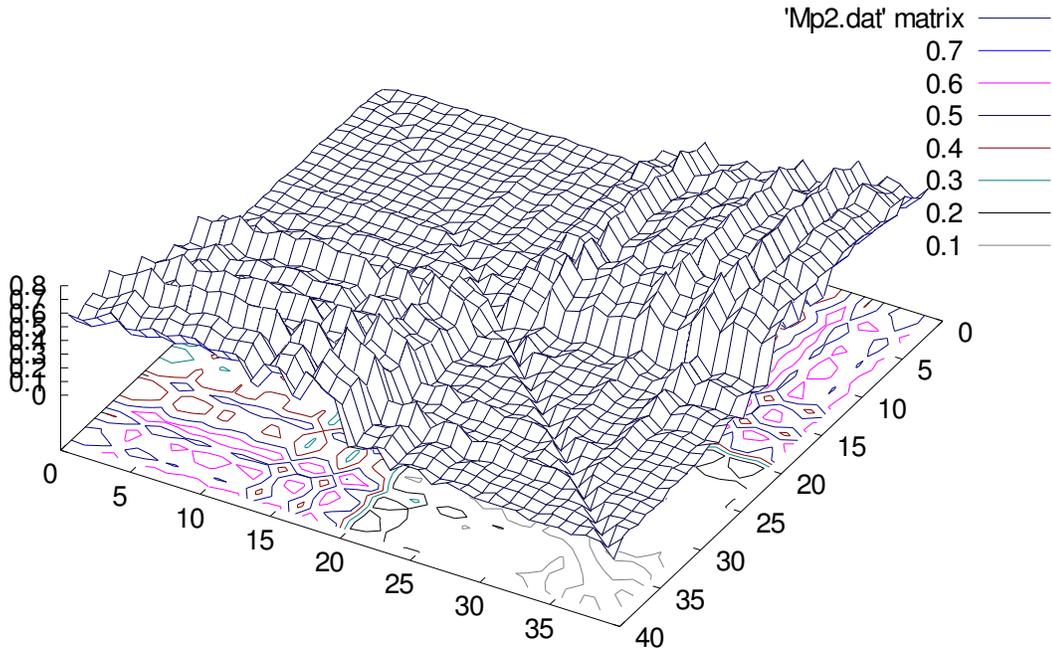


FIGURA 6.16 - Matriz de dissimilaridade para um ciclo de sono do paciente 1

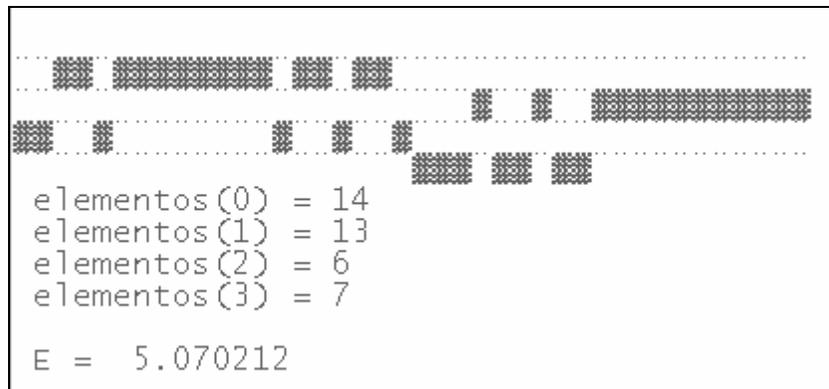


FIGURA 6.17 – Resultado da clusterização para o paciente 1, apresentando uma perceptível continuidade temporal nos clusters encontrados.

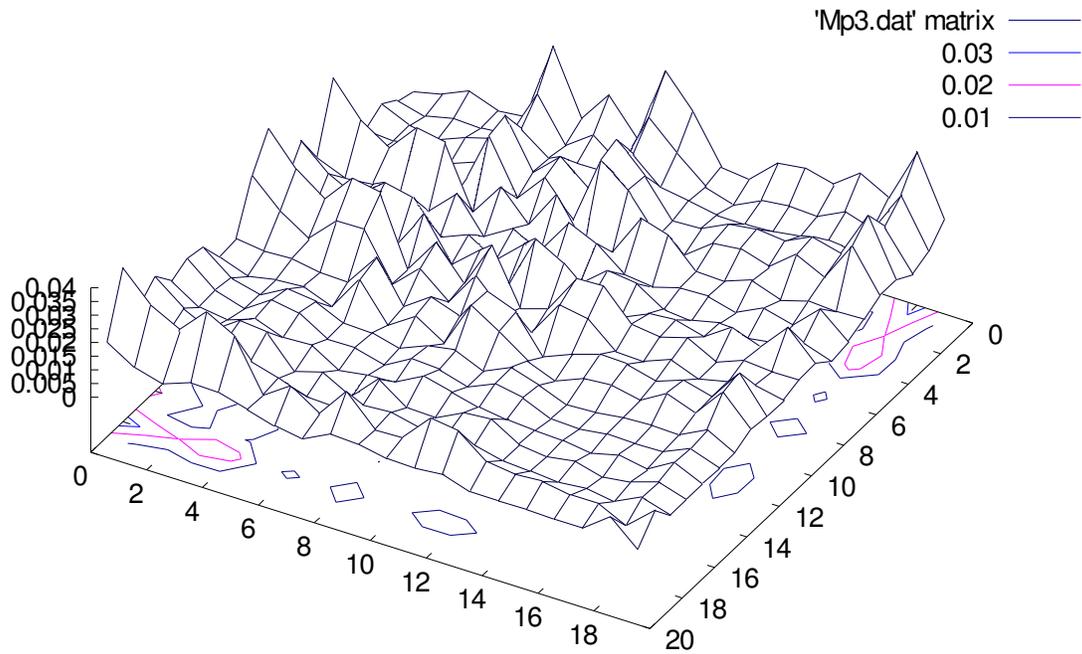


FIGURA 6.18 – Matriz de dissimilaridades para o paciente 2

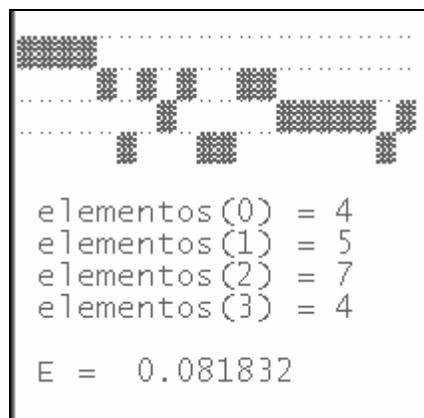


FIGURA 6.19 – Resultado da clusterização para a matriz da figura 6.11. Note que o número de elementos utilizados foi reduzido em função do ciclo utilizado ter sido mais curto que os demais.

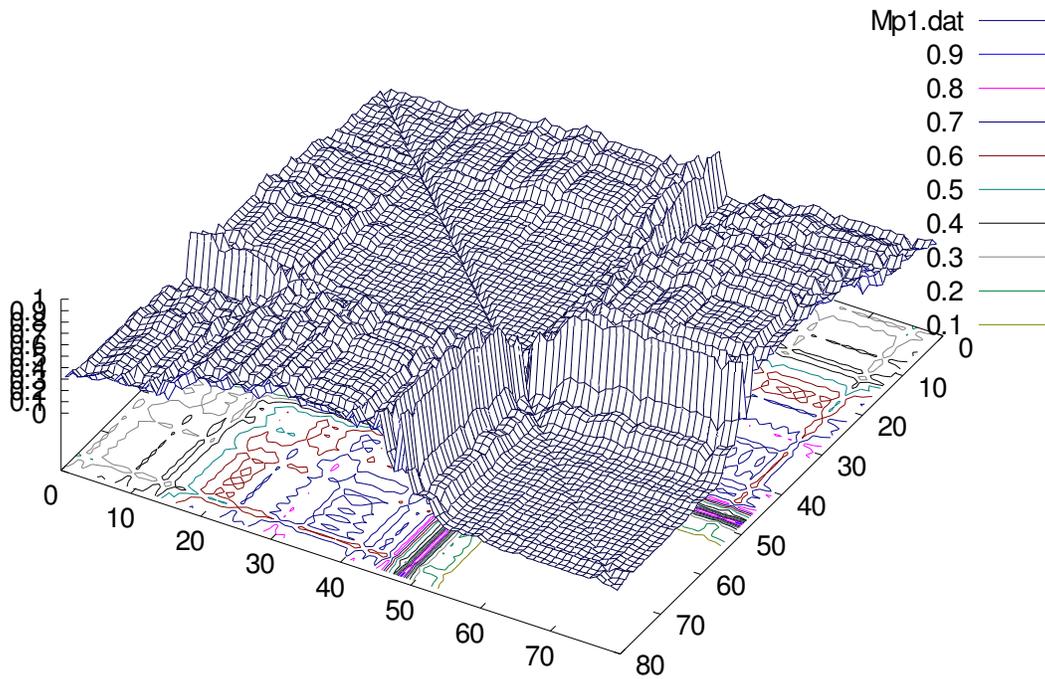


FIGURA 6.20 – Matriz de dissimilaridades para um ciclo de sono do paciente 3.

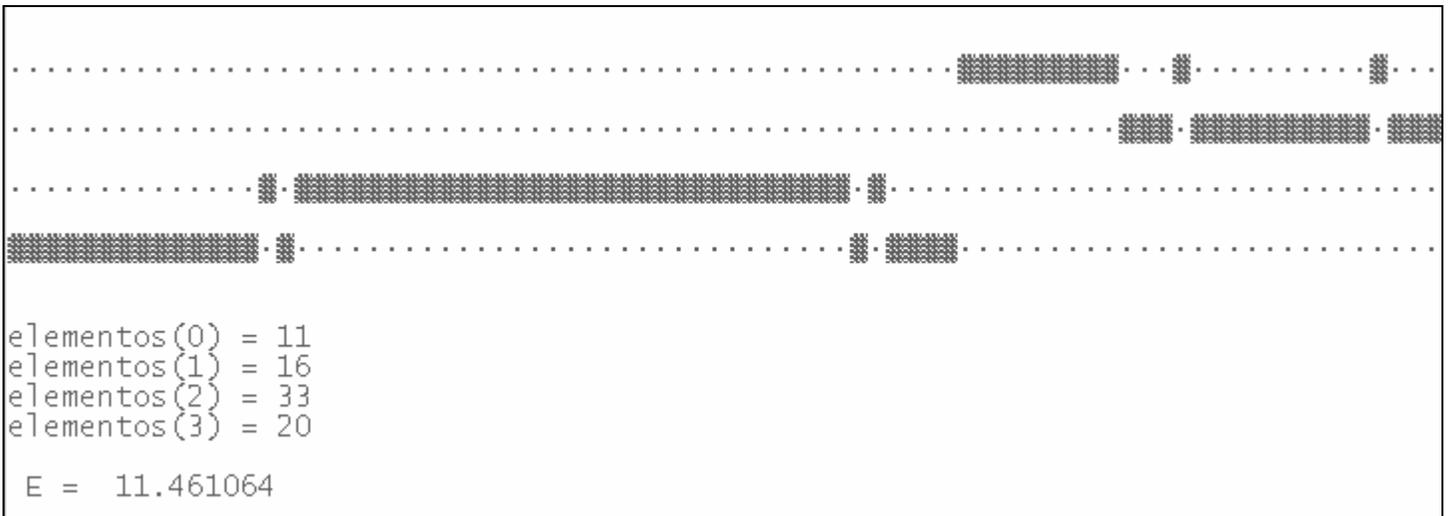


FIGURA 6.21 - Resultado da clusterização para o paciente 3. Neste caso haviam mais pontos disponíveis no ciclo, de forma que se pode utilizar um número maior de janelas.

As figuras de 6.22 até 6.27 representam alguns dos casos onde se aplicou o método para diferentes segmentos de sinal para um ciclo com estágios já conhecidos, do paciente 3. Por economia de espaço, à esquerda encontra-se a matriz de dissimilaridade entre os elementos e à direita o resultado do processo de clusterização.

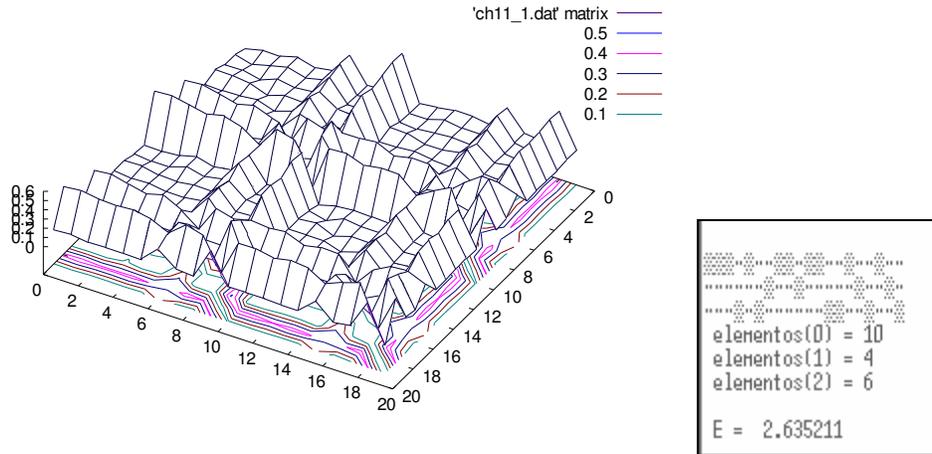


FIGURA 6.22 - Comparação de dois segmentos do estágio 1 com overlap.

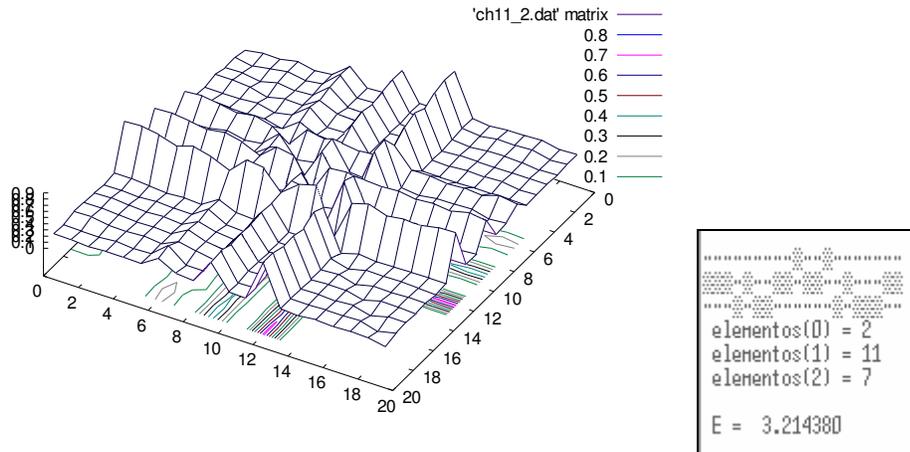


FIGURA 6.23 - Comparação de dois segmentos disjuntos do estágio 1

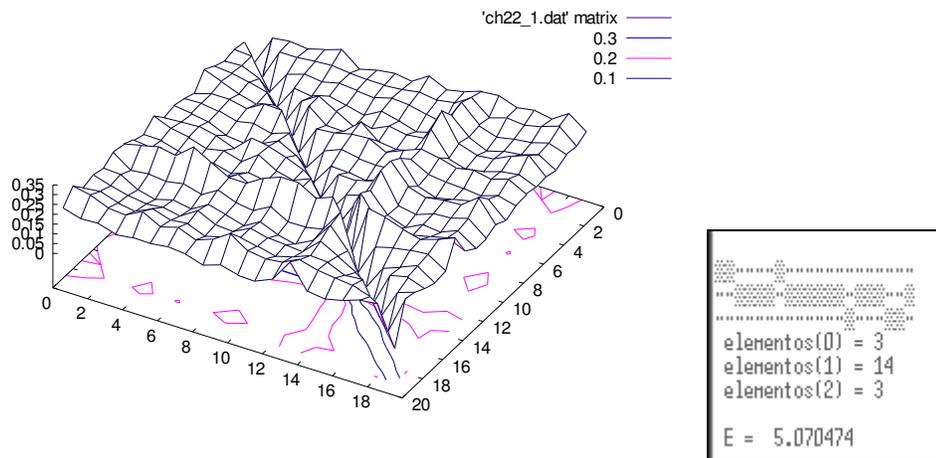


FIGURA 6.24 - Dois segmentos disjuntos estágio 2

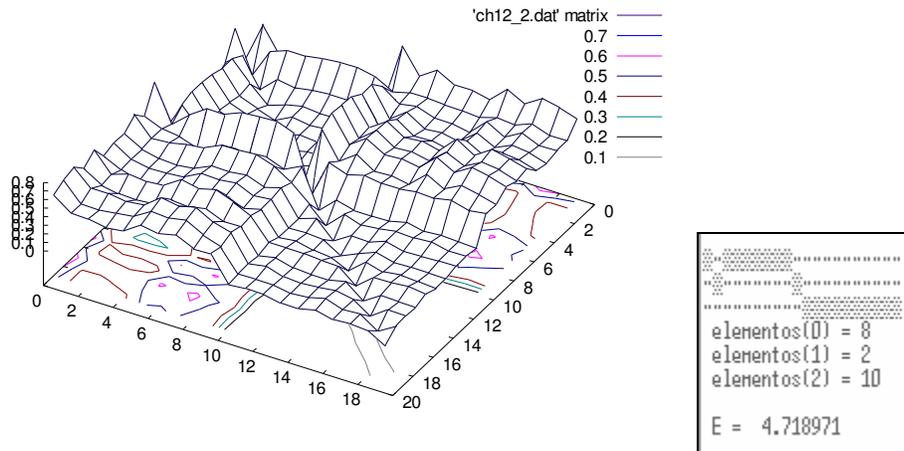
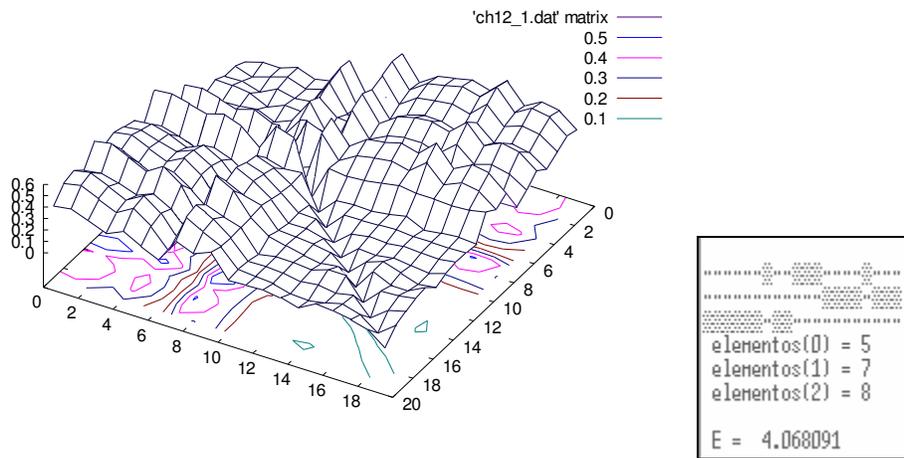


FIGURA 6.25 – Comparação entre segmentos pertencentes aos estágios 1 e 2

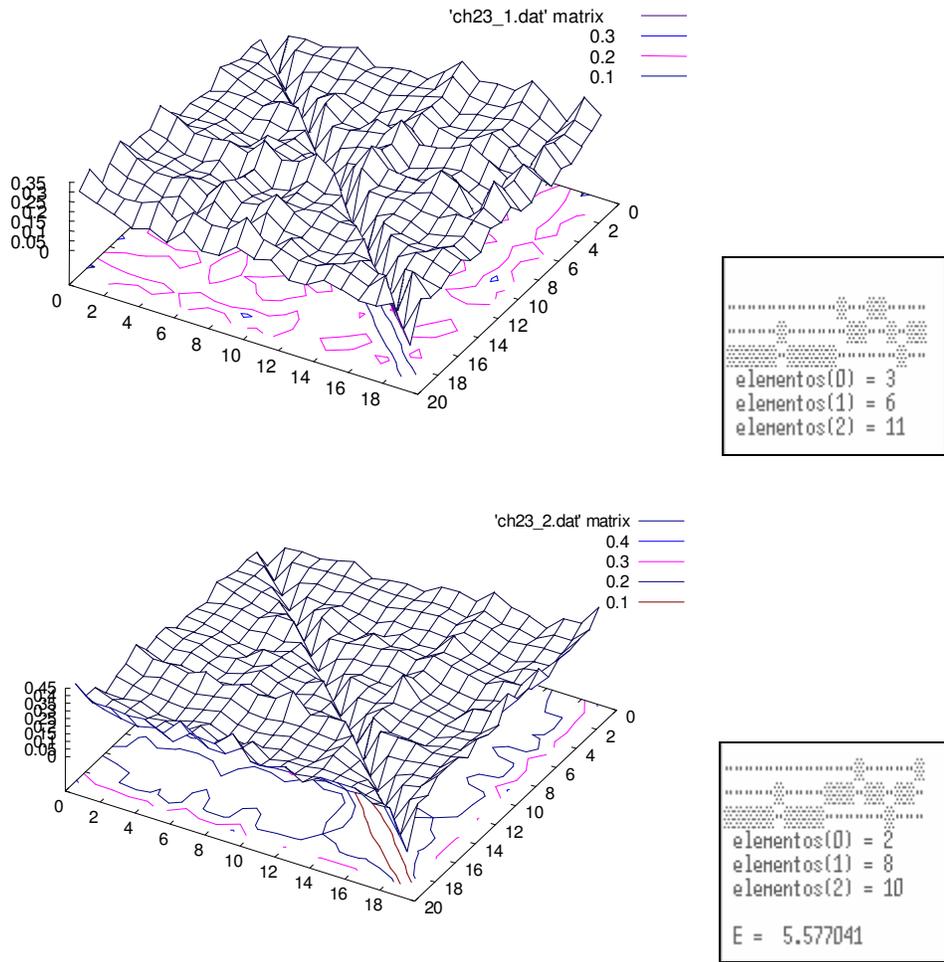


FIGURA 6.26 – Comparação entre segmentos pertencentes aos estágios 2 e 3

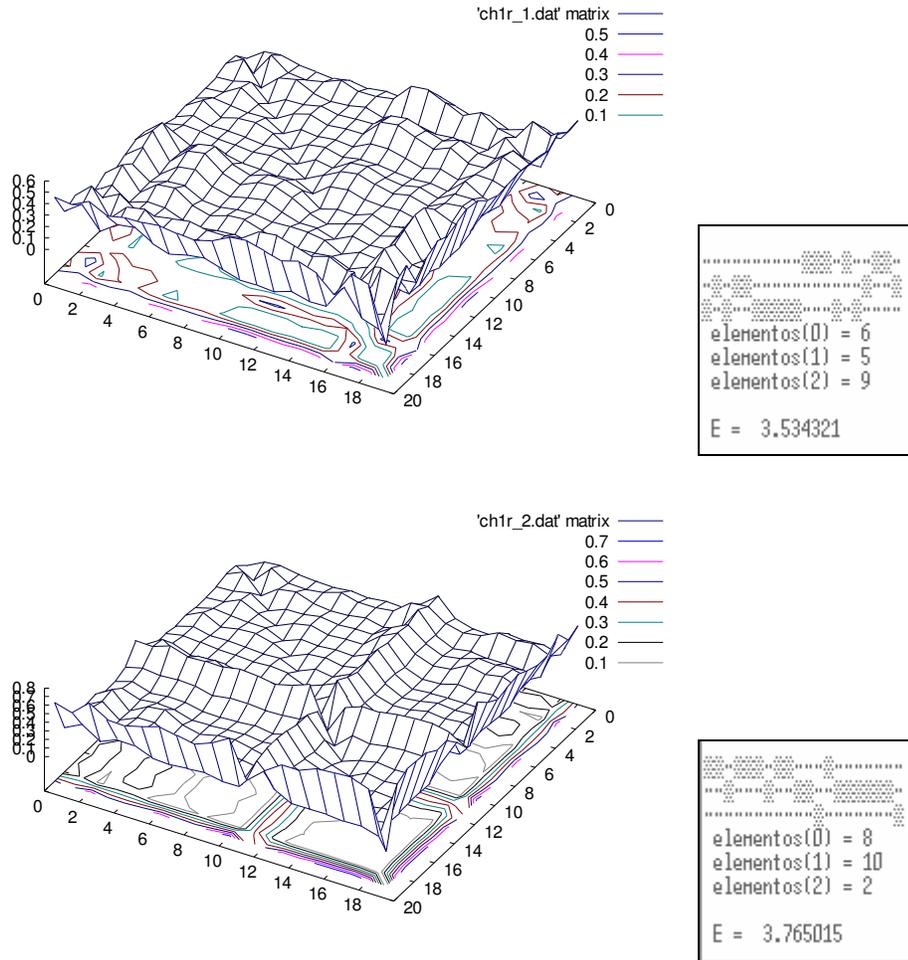


FIGURA 6.27 –Comparação entre segmentos pertencentes aos estágios REM e 1

Como se pode concluir a partir da observação destas figuras, as medidas baseadas em comparações de distribuições suavizadas da distribuição de densidade de probabilidade $\rho(x)$ em apenas uma dimensão já apresentaram resultados de qualidade bastante superior, detectando padrões de continuidade no sinal. Foi escolhido estimar a medida de probabilidade invariante em uma única dimensão por motivos de simplificação, como diminuição do tempo de processamento e maximização dos pontos amostrais por caixa, embora não haja nenhum empecilho fundamental para se experimentar dimensões maiores. Uma limitação para a utilização de espaços de maior dimensionalidade a ser frisada é o crescimento exponencial do número de caixas para construção dos histogramas e do tempo de CPU necessário para processar dados deste espaço.

Uma possível explicação para a melhoria nos resultados obtidos com a abordagem probabilística é que, apesar de possuírem alguma sensibilidade aos parâmetros (e.g. tamanho das caixas), medidas invariantes são instrumentos comuns no estudo de sistemas determinísticos e estocásticos. Uma flutuação nas características de um processo estocástico ou determinístico que se reflita em uma mudança observável na medida invariante do processo (um sinal de não-estacionariedade) deverá ser detectada por uma estatística comparativa deste tipo, seja este processo linear ou não-linear. Ou seja, mesmo que o processo seja predominantemente estocástico, espera-se que, na maioria dos casos, comparações entre distribuições de densidade de probabilidade tenham sensibilidade para detectar mudanças que outras medidas não são capazes de quantificar. Pelo fato da medida invariante ser uma densidade de mais baixa ordem, não carregando informação sobre a dinâmica do sistema, tomamos a liberdade de batizar esta medida como “medida de estacionariedade fraca para sistemas com não-linearidades”, por analogia àquela baseada em média e variância para sistemas lineares.

Em seguida, tentou-se verificar se os clusters encontrados possuíam relação com alguma variável clínica já conhecida. Em muitos casos, os clusters encontrados coincidiram com os estágios do sono. Os únicos estágios que o algoritmo não foi capaz de discernir de uma forma satisfatória foram os estágios 3 e 4, devido às distribuições $\rho(x)$ destes estágios serem muito similares. A informação obtida junto ao especialista é de que, durante a análise clínica, muitas vezes se consideram estes estágios como sendo um único, pois freqüentemente não há vantagem em separá-los. Ainda assim, é interessante que se dê continuidade ao processo de análise de significância dos clusters devido ao pequeno número de amostras disponíveis. Uma perspectiva otimista é que talvez ainda seja possível extrair informações clinicamente relevantes que ainda não identificamos.

Em certas ocasiões notou-se a ocorrência de um fenômeno interessante: um dos segmentos encontrava-se inteiro em um cluster e o outro particionava-se de forma mais ou menos uniforme em outros dois clusters, sem que elementos de um segmento aparecessem no cluster do outro. Mesmo tendo-se perdido continuidade temporal em um dos clusters, pode-se dizer que, apesar da separação não ter ocorrido no sentido estrito do termo, é perceptível, visualmente, que existe uma diferença qualitativa entre os dois segmentos, visível pela forma dos três clusters formados.

7 Conclusão

Apesar dos sinais de EEGs apresentarem um comportamento bastante irregular, e não esperarmos que sejam descritos por modelos lineares estocásticos (uma vez que o cérebro não deve ser um mecanismo linear), é arriscado acreditar-se que tão complexo mecanismo possa ser descrito por um sistema caótico determinístico de baixa dimensionalidade *simplesmente*. Apesar de serem conhecidos alguns trabalhos que tratam, com sucesso, medidas eletroencefalográficas como sinais caóticos determinísticos[60,61], os resultados obtidos neste trabalho indicam que tal abordagem não pode ser generalizada para toda e qualquer medida, principalmente quando se trata daquelas feitas sobre a superfície do crânio, onde o número de neurônios que contribuem na leitura é muito maior do que em medidas intracraniais. Este tipo de limitação faz com que resultados obtidos a partir de experimentos com eletroencefalogramas, baseados em técnicas de reconstrução de atratores de baixa dimensão não sejam diretamente generalizáveis, o que limita o valor dos mesmos.

O presente trabalho se propôs a explorar possíveis aspectos não-lineares presentes em eletroencefalogramas de pacientes normais durante períodos de sono também normal, e os resultados obtidos com os experimentos sugerem que, em vista da enorme quantidade de neurônios(unidades não-lineares) sensoriados simultaneamente, faz sentido tratar o sistema como uma mistura de (baixo) determinismo não-linear com um forte ruído estocástico de fundo. Para sistemas com um número muito grande de graus de liberdade, é ponto pacífico que se deva adotar uma abordagem probabilística. Neste contexto, métodos específicos para sinais não-lineares determinísticos (métodos de reconstrução de atratores) ou para sinais lineares estocásticos puros (métodos espectrais) não devem apresentar um desempenho *ótimo*, apesar dos métodos espectrais serem corriqueiramente utilizados para a análise deste tipo de sinal (uma vez que a estocasticidade é mais forte que o caos determinístico) com considerável sucesso[WES98]. Qualquer bom resultado obtido com métodos que não possuam concordância com a física envolvida no problema, sem devida justificativa, ou seja, “métodos de caixa preta”, têm seu valor científico comprometido. Evitar tais procedimentos foi uma das diretrizes deste trabalho.

No artigo de Hernández *et al*[HER95] os autores utilizam, com sucesso, comparações baseadas em modelos preditivos não-lineares para comparar ritmos teta monomórficos e polimórficos, argumentando que a análise espectral não é capaz de diferenciá-los. A aparente vantagem da abordagem não-linear adotada por estes autores provavelmente deve seu sucesso a outro fator aleatório senão a presença de dinâmica determinística forte, caos, atratores ou bifurcações nestes sinais. É clara, a olho nu, a presença de componentes de alta frequência e baixa amplitude nos ritmos teta polimórficos que não são vistos nos ritmos teta monomórficos. Por si só, esta componente pode ser a responsável pela diferenciação dos dois ritmos via previsões não-lineares. Não-linearidades que comprometam o desempenho da análise espectral não são evidência de dinâmica não-linear, pois podem ser introduzidas pelo processo de observação[THE92].

É conveniente se frisar que, a despeito da idéia de necessidade de adoção incondicional de um procedimento científico rigoroso ao extremo em qualquer situação, exigindo que cada detalhe de um método seja cirurgicamente especificado e justificado, dificultando e até impossibilitando

uma grande quantidade de experimentos de serem realizados, acredita-se que não há problemas sérios em se utilizar, empiricamente, algum método conhecido para tratar um problema ainda obscuro tendo em vista fins tecnológicos, dependendo obviamente do nível de compatibilidade entre o método e o problema. O risco maior surge quando se começa a interpretar tais resultados e, com base nestes, inferir (às vezes afirmar) sobre características do problema, ou sobre a natureza física do sistema estudado. Em [SCH99], cita-se um trabalho que reporta a dinâmica pico-onda de um ataque epilético como sendo caótica, mas que é mais tarde questionado por outro autor.

Neste trabalho experimentou-se diferentes medidas comparativas para sinais oriundos de sistemas não-lineares biológicos e simulações numéricas em busca de uma medida mais adequada fisicamente suficientemente eficiente para a tarefa de segmentação de séries eletroencefalográficas não-estacionárias. Três medidas de dissimilaridade foram utilizadas, respectivamente baseadas em predições mútuas[SCH97b], na integral de correlação cruzada de Kantz[KAN94b] e em diferenças entre distribuições suavizadas segundo Diks[DIK96]. Como a distribuição dos elementos nos clusters utilizando predições e somas de correlação não possuía coerência com os estágios de sono previamente classificados buscou-se, em conjunto com um médico especialista, encontrar alguma relação dos elementos com alguma variável clínica, mas esta relação não foi encontrada. Em contrapartida, a medida de dissimilaridade baseada na densidade de probabilidade $\rho(\mathbf{x})$ aplicada para sinais determinísticos com diversos níveis de ruído e para sinais eletroencefalográficos foi capaz de identificar diferenças significativas entre segmentos do sinal, o que leva a concluir que, dadas as características fortemente estocásticas do EEG, estas medidas apresentaram um desempenho melhor do que aquelas puramente baseadas em aspectos caóticos determinísticos, por levarem em consideração possíveis combinações dos aspectos probabilísticos (medidas invariantes), comuns tanto a sistemas caóticos quanto estocásticos. Resultados experimentais positivos com a utilização de estimativas de $\rho(\mathbf{x})$ para a detecção de não-estacionariedade em um sinal irregular, apesar de limitados por adotarem uma grande simplificação devido a não utilização de probabilidades de transição de nenhuma ordem, apresentam a vantagem de minimizarem a quantidade de dados exigidos e o custo computacional da comparação. Mesmo comparada com as medidas de dissimilaridade que não apresentaram bons resultados, as estimativas e comparações entre as distribuições apresentaram um custo consideravelmente menor (mesmo em uma máquina considerada rápida atualmente como um Pentium III 450MHz, seria impossível, por exemplo, utilizar predições e somas de correlação em tempo real, o que pode ser feito utilizando distribuições). A simplicidade do algoritmo genético 1+1, por utilizar uma população composta de apenas um indivíduo, também contribuiu para a rapidez na convergência dos resultados. Uma possibilidade interessante é a paralelização do algoritmo de clusterização, tendo em vista que algoritmos genéticos são naturalmente paralelizáveis na maioria dos casos.

Dessa forma, com máquinas de desempenho equivalente ao daquelas disponíveis no mercado atualmente, torna-se viável, a custos acessíveis, o desenvolvimento de software para a segmentação de eletroencefalogramas em tempo real o que, além de economizar o tempo extra necessário para análise, abre a possibilidade de, realizados experimentos com os devidos conjuntos de dados, se verificar a aplicabilidade do método para a predição de possíveis ataques epiléticos.

De acordo com os resultados experimentais, o percentual de sucesso obtido na aplicação do método para segmentação de diferentes estágios de sono sugere que este possa vir a ser utilizado

como ferramenta auxiliar na tarefa de análise de medidas eletroencefalográficas. Refinamentos no método tornar-se-ão disponíveis à medida em que mais testes possam vir a serem feitos.

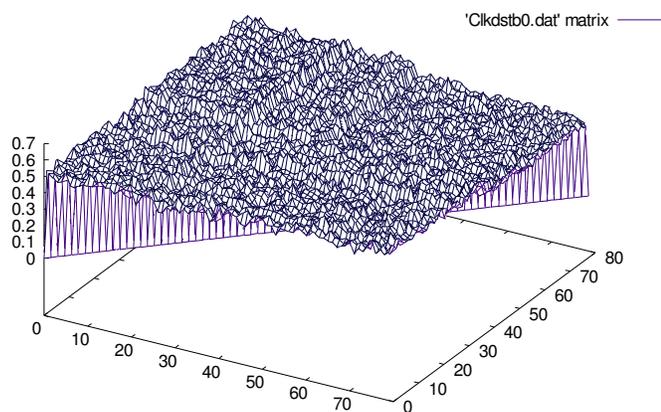
Em um trabalho futuro, devido às estatísticas que apresentaram os melhores resultados para os EEGs serem baseadas em estimativas da medida invariante, possuindo a desvantagem de não carregarem informação sobre a dinâmica do processo, pretende-se estudar a viabilidade de uma medida de similaridade e sincronização baseada em estimativas de distribuições de probabilidades de transição suavizadas, utilizando o método aqui abordado. No contexto de eletroencefalogramas medidos em canais simultâneos, esta proposta se torna bastante interessante, uma vez que tais distribuições de mais alta ordem são necessárias na detecção de eventuais aumentos de sincronização entre diferentes regiões do cérebro, que podem ser o fator causador de doenças como a epilepsia.

Anexo 1 – Gráficos de Resultados

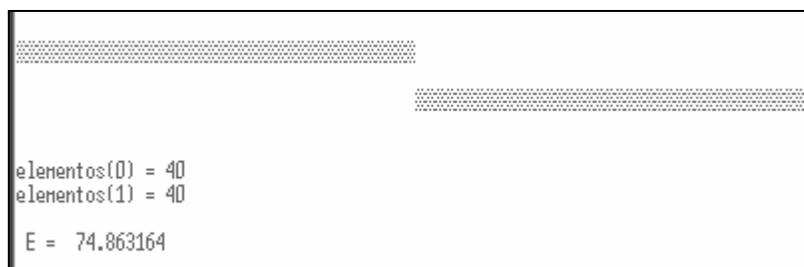
Abaixo seguem alguns exemplos de resultados obtidos com séries geradas a partir de simulação numérica. Na parte superior das figuras aparecem as matrizes de dissimilaridade, e logo abaixo o resultado da clusterização. As linhas denotam os clusters e as colunas os elementos:

Distribuições suavizadas

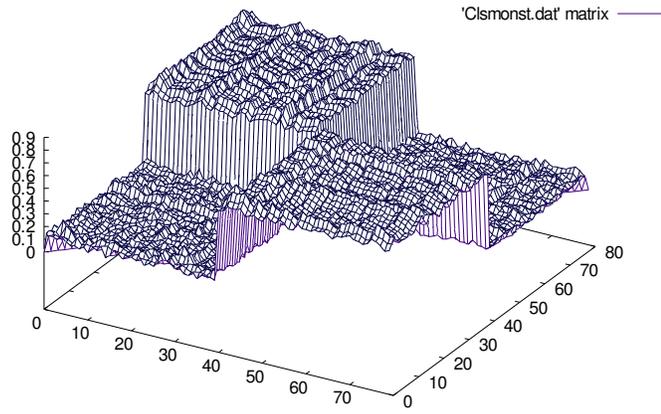
Sinais sem ruído



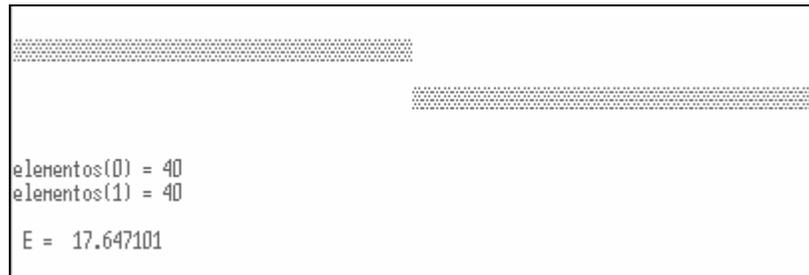
Matriz de dissimilaridade mapa de Ikeda



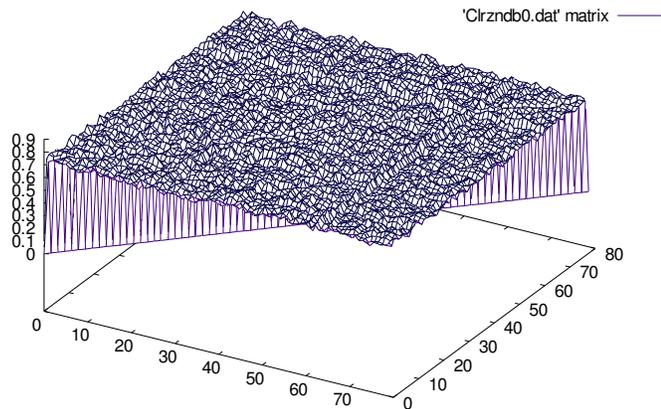
Resultado clusterização



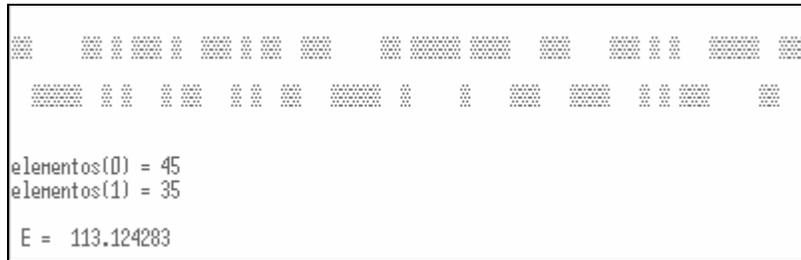
Matriz de dissimilaridade para o sistema de Roessler



Resultado Clusterização



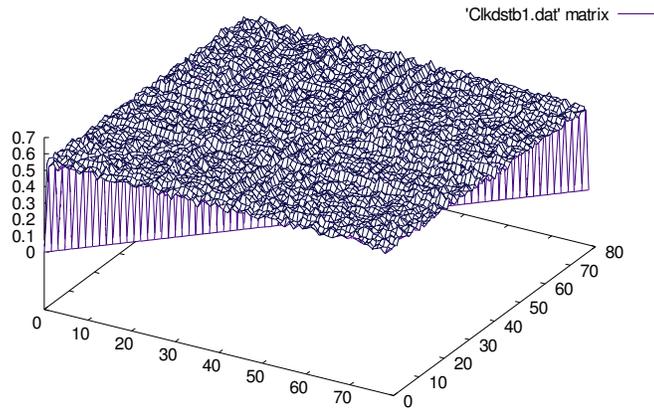
Matriz de dissimilaridade para o sistema de Lorenz



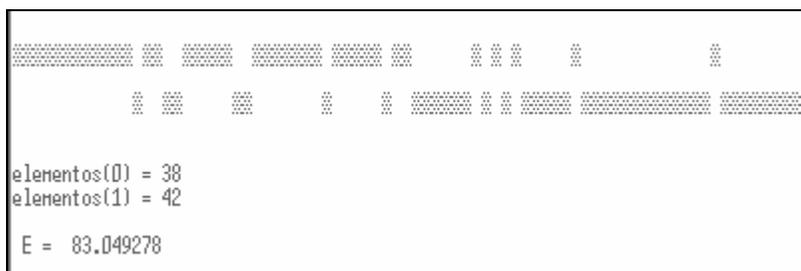
Resultado da clusterização

Note que o resultado para a clusterização do sistema de Lorenz reflete a pouca variação dos elementos da matriz de dissimilaridade, uma vez que o parâmetro escolhido sofreu uma flutuação muito pequena (~ 0.01).

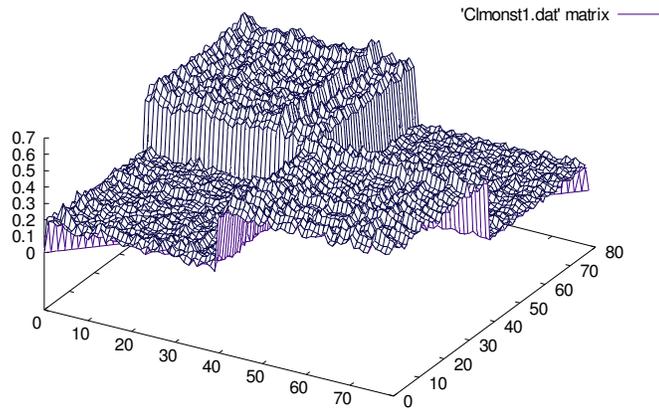
Sinais com ruído gaussiano de amplitude absoluta = 0.1



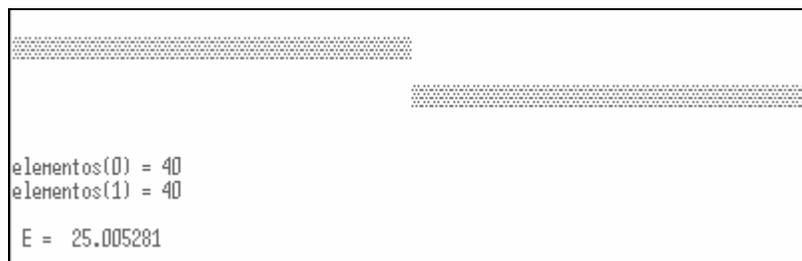
Matriz de dissimilaridade para o mapa de Ikeda



Resultado Clusterização

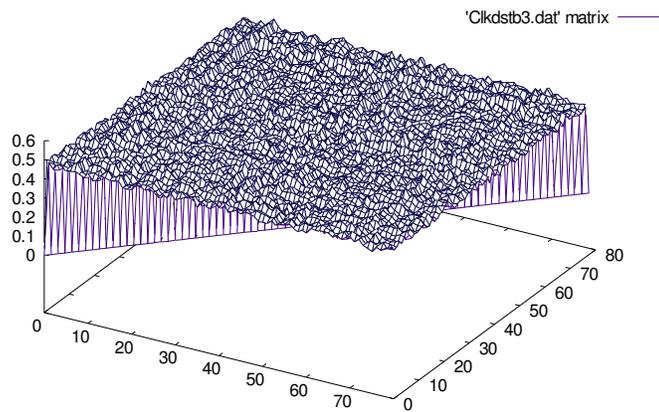


Matriz de dissimilaridade para o sistema de Roessler

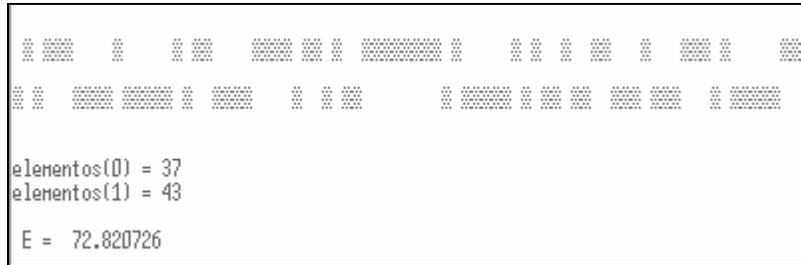


Resultado clusterização

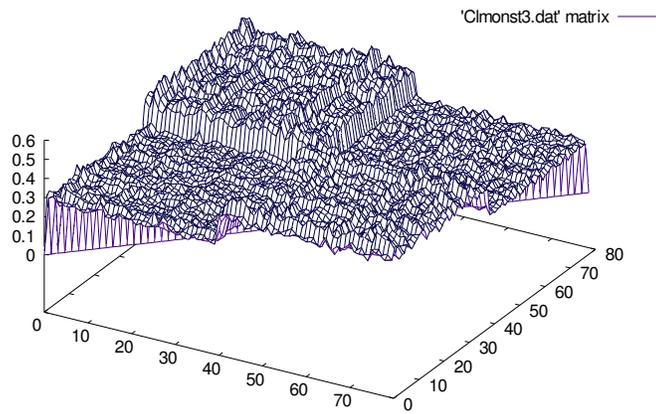
Séries com ruído gaussiano de amplitude absoluta = 0.3



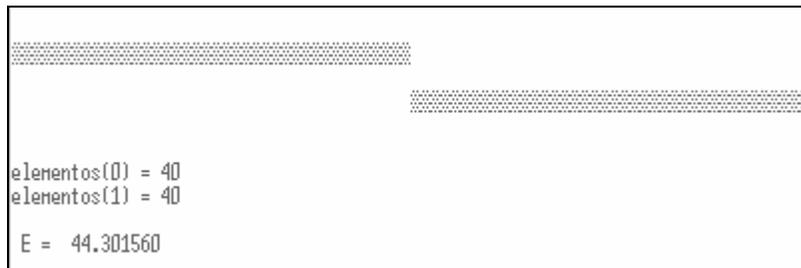
Matriz de dissimilaridade para o sistema de Ikeda



Resultado clusterização

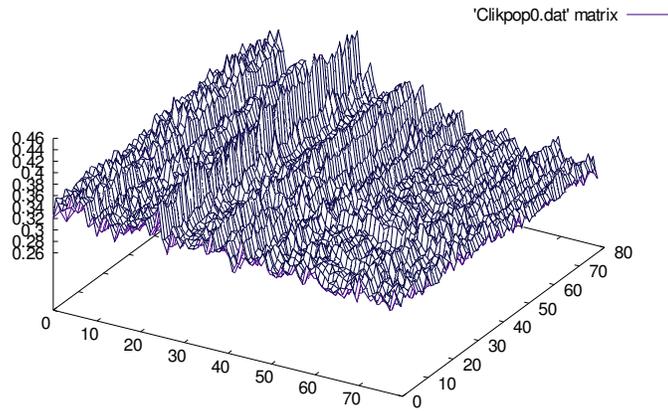


Matriz de dissimilaridade para o sistema de Roessler

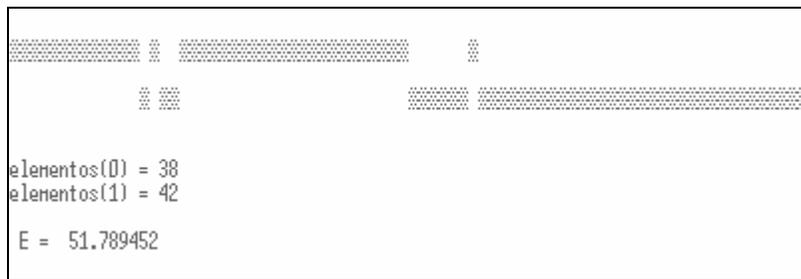


Resultado da clusterização

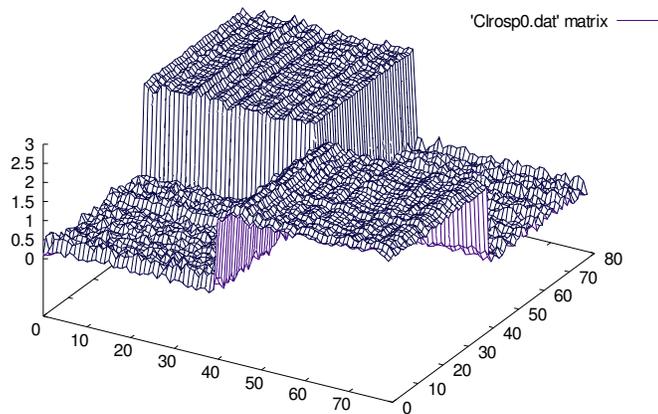
Medida de dissimilaridade : Erros de predições cruzadas



Mapa de Ikeda sem ruído



Resultado clusterização



Matriz de dissimilaridade para o sistema de Roessler sem ruído

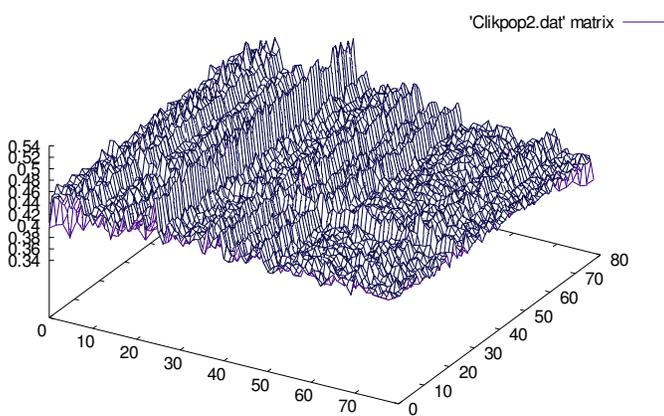
```

.....................................................................
.....................................................................
elementos(0) = 40
elementos(1) = 40
E = 46.665643

```

Resultado clusterização

Sinais com adição de ruído gaussiano de amplitude = 0.1



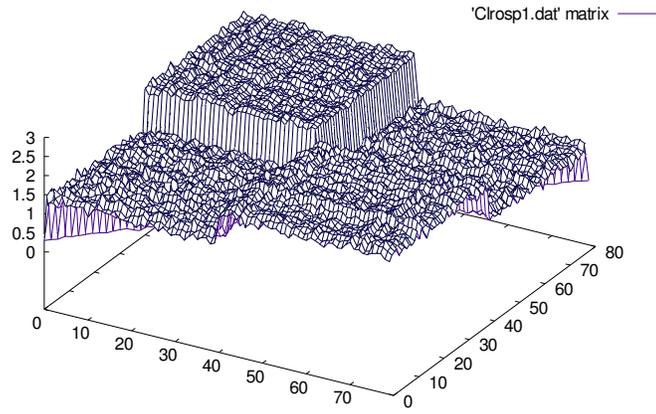
Matriz de dissimilaridade para o mapa de Ikeda

```

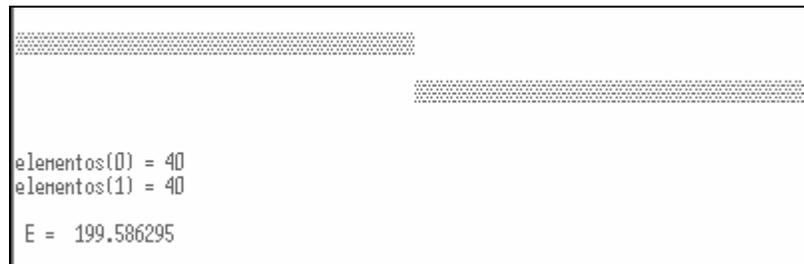
.....................................................................
.....................................................................
elementos(0) = 34
elementos(1) = 46
E = 68.134066

```

Resultado clusterização

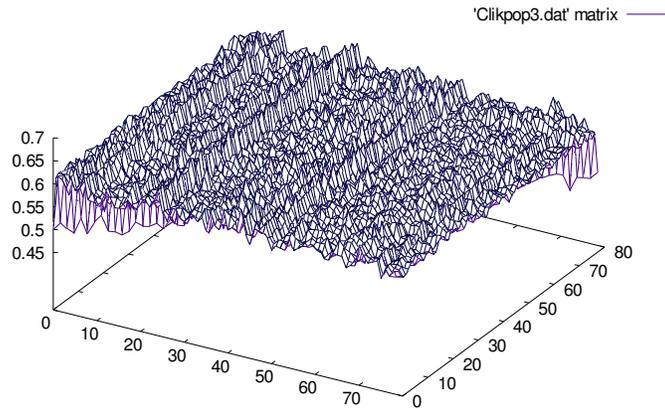


Matriz de dissimilaridade para o sistema de Roessler

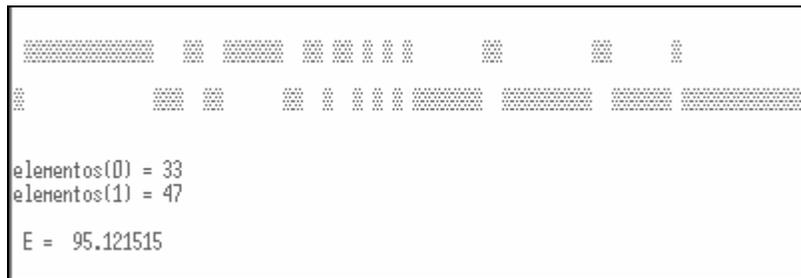


Resultado clusterização

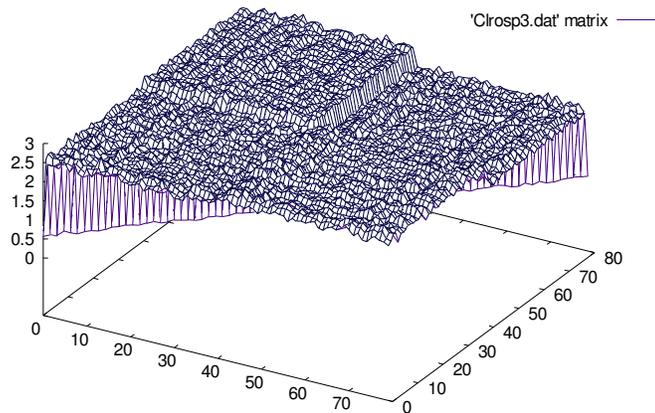
Sinais com adição de ruído gaussiano com amplitude = 0.3



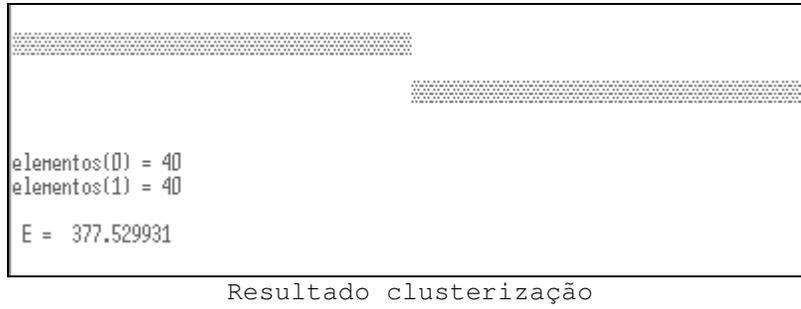
Matriz de dissimilaridade para o mapa de Ikeda



Resultado clusterização

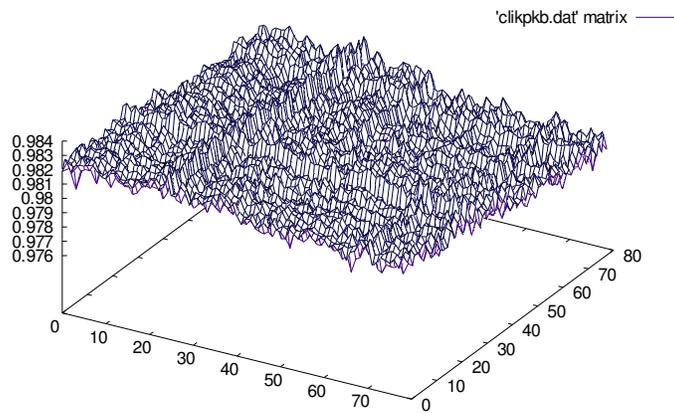


Matriz de dissimilaridade para o sistema de Roessler



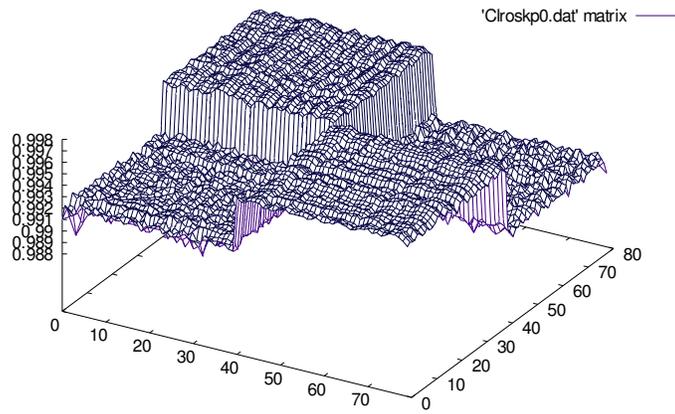
Medida de dissimilaridade: Somas de correlação cruzadas

Sinais sem adição de ruído

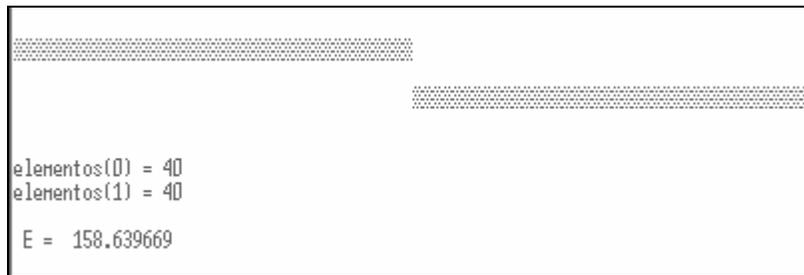


Matriz de dissimilaridade para o mapa de Ikeda



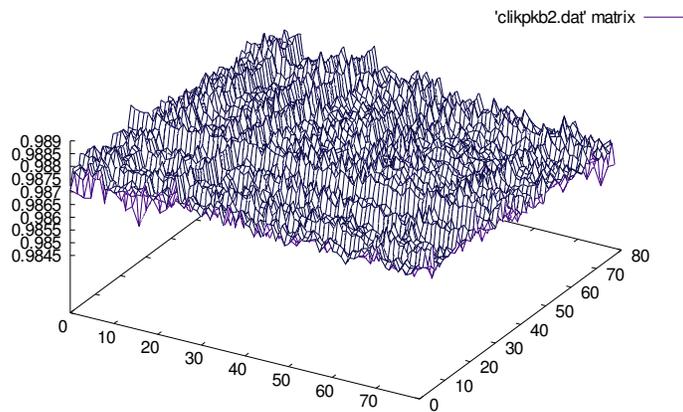


Matriz de dissimilaridade para o sistema de Roessler

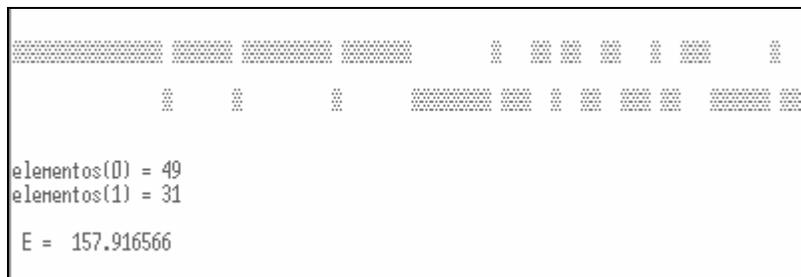


Resultado clusterização

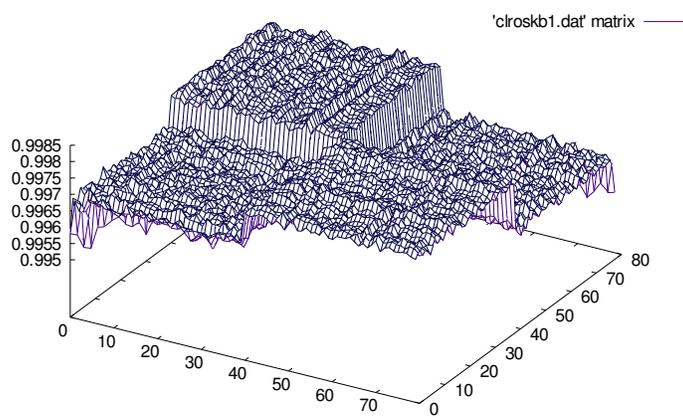
Sinais com adição de ruído gaussiano com amplitude absoluta de 0.1



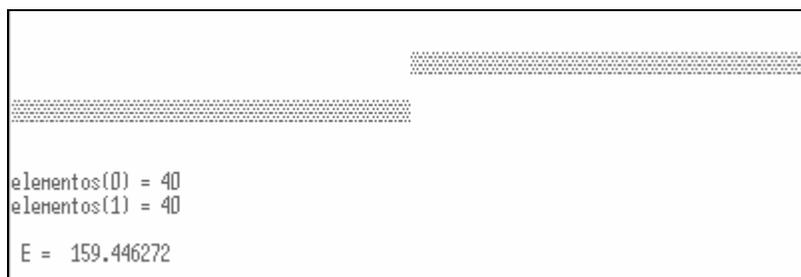
Matriz de dissimilaridade para o mapa de Ikeda



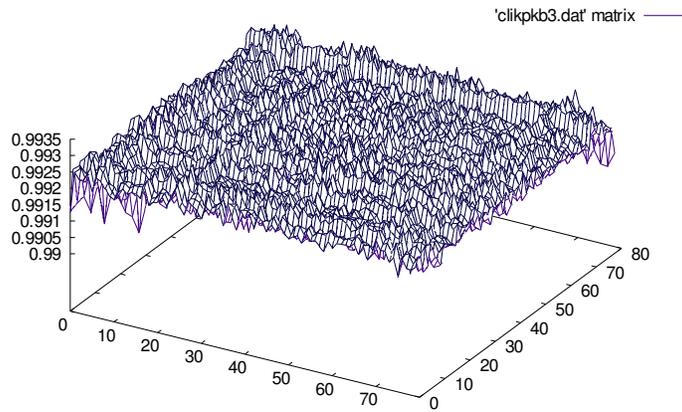
Resultado clusterização



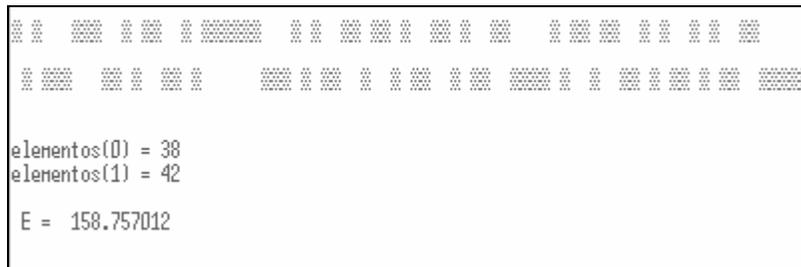
Matriz de dissimilaridade para o sistema de Roessler



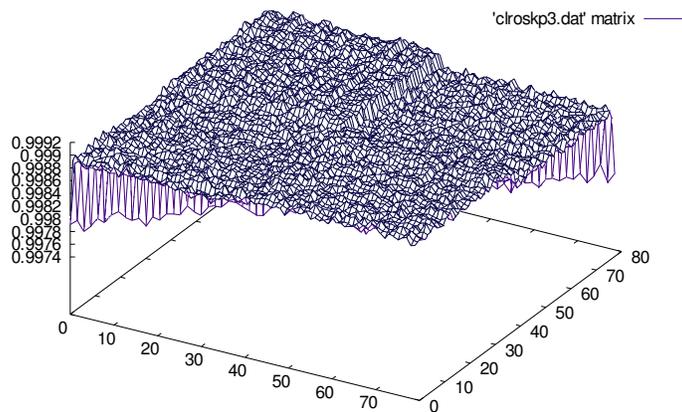
Sinais com adição de ruído gaussiano com amplitude absoluta 0.3



Matriz de dissimilaridade para o mapa de Ikeda



Resultado Clusterização



Matriz de dissimilaridade para o sistema de Rössler

```
.....  
.....  
elementos(0) = 40  
elementos(1) = 40  
E = 159.806297
```

Resultado clusterização

Bibliografia

- [ABA96] ABARBANEL, H.D.I. **Analysis of Observed Chaotic Data**. New York: Springer, 1996.
- [ADA97] ADAMI, C. **Introduction to Artificial Life**. New York: Springer, 1997.
- [AMI89] AMIT, D. **Modelling Brain Function**. Cambridge: Cambridge University Press, 1989.
- [BAB85] BABLOYANTZ, A.; SALAZAR, J.M.; NICOLIS, C. Evidence of chaotic dynamics of brain activity during the sleep cycle. **Phys. Lett. A**, Amsterdam, v.111, n.3, p. 152-156, 1985.
- [BÄC96] BÄCK, T. **Evolutionary Algorithms in Theory and Practice**. Oxford: Oxford Univ. Press, 1996.
- [BOX76] BOX, G.E.P.; JENKINS, G.M. **Time Series Analysis**. San Francisco: Holden-Day, 1976.
- [BOY98] BOYCE, W. ; DIPRIMA, R. **Equações Diferenciais Elementares e Problemas de Valores de Contorno**. São Paulo: LTC ,1998.
- [BRY90] BRYANT, P.; BROWN, R.; ABARBANEL, H. Lyapunov exponents from observed time series. **Phys. Rev. Lett.**, Woodbury, v. 65, p.1523, 1990.
- [BUH94] BUHMANN, J.M.; HOFMANN, T. A Maximum entropy approach to Pair wise data Clustering. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 12., 1994, Jerusalem, Israel. **Proceedings...** Los Alamitos: IEEE Computer Society Press, 1994.
- [CAS97] CASDAGLI, M. Recurrence plots revisited. **Physica D**, Amsterdam, v.108, p.12, 1997.
- [DAV91] DAVIS, L. (Ed.). **Handbook of Genetic Algorithms**. New York: Van Nostrand Reinhold, 1991.
- [DIK96] DIKS, C. Detecting differences between delay vector distributions. **Phys. Rev. E**, Woodbury, v. 53, p. 2169, 1996.
- [DIK99] DIKS, C. **Nonlinear Time Series Analysis: Methods and Applications**. Singapore; River Edge, NJ: World Scientific, 1999.
- [DRA92] DRAZIN, E. **Nonlinear Systems**. Cambridge: Cambridge University Press, 1992.
- [ECK86] ECKMANN, J.P. et al. Liapunov exponents from time series. **Phys. Rev. A**, New York, v.34, p.4971, 1986.

- [ECK87] ECKMANN, J.P.; OLIFFSON KAMPHORST, S.; RUELLE, D. Recurrence plots of dynamical systems. **Europhys. Lett.**, Paris, v. 4, p.973, 1987.
- [ECK92] ECKMANN, J.-P.; RUELLE, D. Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. **Physica D**, Amsterdam, v.56, p.185, 1992.
- [FAR83] FARMER, J.D.; OTT, E.; YORKE, J.A. The Dimension of Chaotic Attractors. **Physica D**, Amsterdam, v.7, p.153, 1983.
- [FIE94] FIEDLER-FERRARA, N. ; CINTRA DO PRADO, C. **Caos: uma introdução**. São Paulo: E. Blücher , 1994.
- [FOG00] FOGEL, D. B. **Evolutionary Computation**. New York: IEEE, 2000.
- [FRA86] FRASER, A. M.; SWINNEY, H.L. Independent coordinates for strange attractors from mutual information. **Phys. Rev. A**, New York, v. 33, p.1134, 1986.
- [GAL98] GALLAS, J. Notas de aula. 1998. Curso Caos em Sistemas Dissipativos Discretos. IF-UFRGS, Porto Alegre.
- [GER98] GERHARDT, G.J.L. **O caos em abordagens experimentais**. [S.l.: s.n.], 1998.
- [GOL89] GOLDBERG, D.E. **Genetic Algorithms in Search, Optimization and Machine Learning**. Reading, MA. Addison-Wesley, 1989.
- [GRA83] GRASSBERGER, P.; PROCACCIA, I. Measuring the Strangeness of Strange Attractors. **Physica D**, Amsterdam, v. 9, p.189, 1983.
- [HEG99] HEGGER R.; KANTZ, H.; SCHREIBER, T. Practical Implementation of nonlinear time series methods: The TISEAN package. Artigo aceito por CHAOS, 1999. Disponível por WWW em <http://www.mpi-pks-dresden.mpg.de/~tisean>. Dez. 2000.
- [HER95] HERNÁNDEZ, J.L. et al. R. Measuring the dissimilarity between EEG recordings through a non-linear dynamical system approach. **Int. J. Bio-Med. Comput.**, Limerick, v.38, p.121, 1995.
- [HUB95] HUBBARD, J.H. ; WEST, B.H. **Differential Equations, a Dynamical Systems Approach**. Berlin: Springer-Verlag, 1995. 2v.
- [HUE89a] HUEBNER, U.; ABRAHAM, N. B.; WEISS, C. O. Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH₃ laser. **Phys. Rev. A**, New York, v. 40, p. 6354, 1989.
- [HUE89b] HUEBNER ,U. On problems encountered with dimension calculations. In: **Measures of Complexity and Chaos**. New York: Plenum Press, 1989. p. 133.

- [HUE89c] HUEBNER ,U. Comparison of Lorenz-like laser behavior with the Lorenz model. In: CONFERENCE COHERENCE AND QUANTUM OPTICS, 6., 1989. **Proceedings...** New York: Plenum Press, 1989. p. 517.
- [KAN91] KANDEL E.R.; SCHWARTZ, J.H.; JESSEL, T.M. (Ed.). **Principles of Neural Science**. Englewood Cliffs: Prentice Hall, 1991.
- [KAN94a] KANTZ, H. A robust method to estimate the maximal Lyapunov exponent of a time series. **Phys. Rev. Lett.**, Woodbury, v.56, p.889, 1994.
- [KAN94b] KANTZ, H. Quantifying the closeness of fractal measures. **Phys. Rev E**, New York, v. 49, p.5091, 1994.
- [KAN97] KANTZ, H.; SCHREIBER, T. **Nonlinear Time Series Analysis**. Cambridge: Cambridge University Press, 1997.
- [KEN92] KENNEL, M.; BROWN, R.; ABARBANEL, H. Determining embedding dimension for phase-space reconstruction using a geometrical construction. **Phys. Rev. A**, New York, v. 45, p.3404-3411, 1992.
- [KEN98] KENNEL, M.B. Statistical test for dynamical nonstationarity in observed time-series data. Disponível por WWW <http://xyz.lanl.gov>. (nov. 1998).
- [KOS88] KOSTELICH, E.J.; YORKE, J.A., Noise reduction in dynamical systems. **Phys. Rev. A**, New York, v. 38, p.1649, 1988.
- [LIY75] LI, T.Y. ; YORKE J. A. Period Three Implies Chaos. **Amer. Math. Monthly**, [S.I.], v. 82, p. 459-467, 1975.
- [LOR63] LORENZ, E. N. Deterministic Nonperiodic Flow. **J. Atmospheric Sciences**, Boston, v. 20, p. 130-131, 1963.
- [MAR96] MARQUES, I. da S.; BARONE D.A.C. **Algoritmos Genéticos em Previsão de Séries Temporais** : trabalho individual. Porto Alegre: CPGCC da UFRGS, 1996.
- [MIC92] MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs**. New York: Springer-Verlag, 1992.
- [NIC95] NICOLIS, G. **Introduction to Nonlinear Science**. Cambridge: Cambridge University Press, 1995.
- [OTN72] OTNES, R. K.; ENOCHSON, L. **Digital Time Series Analysis**. New York: Wiley, 1972.
- [OTT94] OTT , E.; SAUER,T.; YORKE J. A. **Coping with Chaos** : Analysis of chaotic data and

- exploitation of chaotic systems. New York: John Wiley & Sons, 1994.
- [PAC80] PACKARD, N. et al. Geometry from a time series. **Phys. Rev. Lett.**, Woodbury, v. 45, p. 712, 1980.
- [PAR92] PARLITZ, U. Identification of true and spurious Lyapunov exponents from time series. **Int. J. Bif. Chaos**, [S.l.], v. 2, p. 155, 1992.
- [PEC92] PECORA, L.M. ; CARROL, T.L.; HEAGY, J.F. Statistics for mathematical properties of maps between time series embeddings. **Phys. Rev. E**, Woodbury, v. 52, p. 3420, 1992.
- [PIK86] PIKOVSKY, A. Discrete-time dynamic noise filtering. **Sov. J. Commun. Technol. Electron.**, Silver Spring, MD, v. 31, p. 81, 1986.
- [PRE97] PRESS, W.H. et al. **Numerical Recipes: The Art of Scientific Computing**. Cambridge: Cambridge University Press, 1997.
- [ROS90] ROSE, K.; GUREWITZ, E.; FOX, G. Statistical mechanics and phase transitions in clustering. **Physical Review Letters**, Woodbury, v. 65, n.8, p. 945-948, 1990.
- [ROS93] ROSENSTEIN, M.T.; COLLINS, J.J.; DELUCA, C.J. A practical method for calculating largest Lyapunov exponents from small data sets. **Physica D**, Amsterdam, v. 65, p.117, 1993.
- [ŠAR64] ŠARKOVSKII, A.N. Co-existence of Cycles of a Continuous Mapping of a Line into Itself. **Ukrainian Math. J.**, [S.l.], v. 16, p. 61-71, 1964.
- [SAU91] SAUER, T.; YORKE, J.A.; CASDAGLI, M.. Embedology. **J.Stat. Phys.**, [S.l.], v. 65, p. 579, 1991.
- [SCH93] SCHREIBER, T. Extremely Simple Nonlinear Noise Reduction Method. **Phys. Rev. E**, New York, v. 47, p. 2401, 1993.
- [SCH97a] SCHREIBER, T. Classification of Time Series Data with Nonlinear Similarity Measures. **Phys. Rev. Lett.**, Woodbury, v. 79, p. 1475, 1997.
- [SCH97b] SCHREIBER, T. Detecting and analysing nonstationarity in a time series with nonlinear cross predictions. **Phys. Ver. Lett.**, Woodbury, v.78, p.843, 1997.
- [SCH99] SCHREIBER, T. Is nonlinearity evident in time series of brain electrical activity? In: ["Chaos in Brain?"](http://www.mpipks-dresden.mpg.de/~schreibe/myrefs/brain.ps.gz), 2000. **Proceedings...** Disponível por WWW em <http://www.mpipks-dresden.mpg.de/~schreibe/myrefs/brain.ps.gz> (jan.2000).
- [SHA49] SHANNON, C. E. ; WEAVER, W. **The Mathematical theory of information**. Urbana, IL: University of Illinois Press, 1949.

- [TAK81] TAKENS, F. Detecting strange attractors in turbulence. In: **Dynamical Systems and Tubulance**. Berlin: Springer, 1981. p.366-381. (Lecture Notes in Mathematics, v.898).
- [TAN95] TANOMARU, J. Motivação, Fundamentos e Aplicações de Algoritmos Genéticos. In: CONGRESSO BRASILEIRO DE REDES NEURAIIS, 3., 1995. **Anais...** Curitiba: Copel, 1995.
- [THE92] THEILER, J. et al. Testing for nonlinearity in time series: The method of surrogate data. **Physica D**, Amsterdam, v. 58, p.77, 1992.
- [WEI91] WEIGEND, A. Competição de Santa Fe de Séries Temporais. Disponível por WWW em <http://www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html>
- [WES98] WEST, M.; PRADO, R.; KRYSTAL, A. **Evaluation and Comparison of EEG Traces: Latent Structure in Nonstationary Time Series**. Disponível por WWW em <http://www.amstat.org/publications/jasa/west.pdf> (ago.1999).
- [WOL85] WOLF, A. et al. Determining Lyapunov Exponents from a Time Series. **Physica D**, Amsterdam, v.16, p.285, 1985.