MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2017/07.03.19.54-TDI

ANÁLISE COMPUTACIONAL DE PADRÕES ESTRUTURAIS NÃO-LINEARES A PARTIR DE IMAGENS DIGITAIS COM ESTUDOS DE CASO EM CIÊNCIAS AMBIENTAIS E ESPACIAIS

Vanessa Cristina Oliveira de Souza

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Reinaldo Roberto Rosa, e Arcilan Trevenzoli Assireu, aprovada em 09 de junho de 2017.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34P/3P7RU3B>

INPE
São José dos Campos
2017

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/6921
E-mail: pubtc@inpe.br

COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):
Presidente:
Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)
Membros:
Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)
Dr. André de Castro Milone - Coordenação de Ciências Espaciais e Atmosféricas (CEA)
Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (CTE)
Dr. Evandro Marconi Rocco - Coordenação de Engenharia e Tecnologia Espacial (ETE)
Dr. Hermann Johann Heinrich Kux - Coordenação de Observação da Terra (OBT)
Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CPT)
Silvia Castro Marcelino - Serviço de Informação e Documentação (SID)
BIBLIOTECA DIGITAL:
Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Serviço de Informação e Documentação (SID)
REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:
Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SID)
Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)
EDITORAÇÃO ELETRÔNICA:
Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)
André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2017/07.03.19.54-TDI

ANÁLISE COMPUTACIONAL DE PADRÕES ESTRUTURAIS NÃO-LINEARES A PARTIR DE IMAGENS DIGITAIS COM ESTUDOS DE CASO EM CIÊNCIAS AMBIENTAIS E ESPACIAIS

Vanessa Cristina Oliveira de Souza

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Reinaldo Roberto Rosa, e Arcilan Trevenzoli Assireu, aprovada em 09 de junho de 2017.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34P/3P7RU3B>

INPE
São José dos Campos
2017

Dados Internacionais de Catalogação na Publicação (CIP)

Souza, Vanessa Cristina Oliveira de.

So89a Análise computacional de padrões estruturais não-lineares a partir de imagens digitais com estudos de caso em ciências ambientais e espaciais / Vanessa Cristina Oliveira de Souza. – São José dos Campos : INPE, 2017.

xxii + 166 p. ; (sid.inpe.br/mtc-m21b/2017/07.03.19.54-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2017.

Orientadores : Drs. Reinaldo Roberto Rosa, e Arcilan Trevenzoli Assireu.

1.Análise de flutuação. 2. Destendenciada bi-dimensional. 3. Textura. 4. Paralelização. 5. GPGPU/CUDA. I.Título.

CDU 004.932.2:502/504

Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

Aluno (a): *Vanessa Cristina Oliveira de Souza*

Título: "ANÁLISE COMPUTACIONAL DE PADRÕES ESTRUTURAIS NÃO-LINEARES A PARTIR DE IMAGENS DIGITAIS COM ESTUDOS DE CASO EM CIÊNCIAS AMBIENTAIS E ESPACIAIS"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Computação Aplicada

Dr. Stephan Stephany

Presidente / INPE / SJCampos - SP

Dr. Reinaldo Roberto Rosa

Orientador(a) / INPE / SJCampos - SP

Dr. Arcilan Trevenzoli Assireu

Orientador(a) / UNIFEI / Itajubá - MG

Dr. Lamartine Nogueira Frutuoso Guimarães

Membro da Banca / INPE / SJCampos - SP

Dr. Marcelo Panaro de Moraes Zamith

Convidado(a) / UFRRJ / Nova Iguaçu - RJ

Dr. Maurício José Alves Bolzan

Convidado(a) / UFG / Jataí - GO

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 09 de junho de 2017

*À **Malu**, com todo meu amor*

AGRADECIMENTOS

Agradeço aos meus orientadores, Dr. Reinaldo Roberto Rosa e Dr. Arcilan Trevenzolli Assireu pelos ensinamentos, dedicação a este trabalho e pela gentileza de sempre.

A minha família, em especial a meu marido, Rafael, por compartilhar comigo as angústias e sempre me acalentar.

Aos meus pais, por terem cedido tanto tempo deles para me ajudar nessa empreitada.

Aos meus colegas do Instituto de Matemática e Computação da Universidade Federal de Itajubá, por serem solidários, solícitos e compreensivos, especialmente no último ano.

À Universidade Federal de Itajubá pelo apoio.

À FAPEMIG (Projeto APQ 1575/14) e ao *National Renewable Energy Laboratory* (NREL) pelos dados de vento utilizados neste trabalho.

RESUMO

A Análise de Flutuação Destendenciada (DFA) tem sido amplamente utilizada na verificação de propriedades de escala de séries temporais unidimensionais. Além de revelar a presença ou não de memória na série (persistência), o método DFA possibilita compreender melhor o processo que originou o sinal analisado e a força da correlação. Nesse ponto, o operador do DFA (α) infere também sobre a rugosidade do sinal, isto porque quanto maior a persistência, menor a rugosidade. Quando tal característica é expandida para sinais bi-dimensionais, em especial imagens digitais, a noção de persistência infere também sobre a textura desses sinais. O DFA foi generalizado para operar em sinais bi-dimensionais em 2006 (DFA-2D) e, desde então, diversos estudos vêm sendo feitos, especialmente utilizando o α como um operador textural. Diferente dos operadores texturais comuns que atuam sobre a variação de brilho na imagem, o DFA-2D utiliza o arcabouço teórico da teoria dos fractais e infere a textura a partir da característica de autossimilaridade do sinal. Neste contexto, os objetivos desse trabalho foram ; i) explorar questões controversas ou não tratadas ainda para o DFA-2D na literatura e ii) avançar no estado da arte da técnica bi-dimensional em suas versões mono e multifractal, avaliando a resposta em um conjunto diverso de dados e também melhorando sua performance por meio da paralelização, utilizando GPGPU/CUDA. Sendo assim, essa tese pretendeu contribuir com dois aspectos do DFA-2D criticados na literatura : a dificuldade de interpretação do operador α e o alto custo computacional. Foram estudados conjuntos de dados simulados cuja resposta é conhecida (fBm e fGn), conjuntos clássicos de dados simulados não tratados na literatura com o DFA-2D (como ruídos do tipo $1/f$ e redes de mapas acoplados). A análise também incluiu conjuntos de dados reais, cujas aplicações utilizando o DFA-2D são inéditas, como a classificação morfológica de galáxias e a inferência de rugosidade efetiva para estudos de energia eólica, por meio da classificação de Modelos Digitais de Elevação topográfica (MDE). Para os sinais simulados, o DFA-2D (mono e multifractal) caracterizou de forma satisfatória os sinais. Para os sinais reais, os resultados mostraram que, mesmo quando o operador α não atinge uma boa taxa de classificação, seu uso é muito informativo, caracterizando o sinal, e não apenas classificando-o. Além disso, o DFA-2D aplicado a dados reais apresentou dificuldades e desafios imperceptíveis nos sinais simulados. Por fim, a paralelização mostrou-se eficaz, diminuido consideravelmente o tempo de processamento pelo DFA-2D.

Palavras-chave: Análise de Flutuação Destendenciada bi-dimensional. Textura. Classificação. Paralelização. GPGPU/CUDA.

COMPUTATIONAL ANALYSIS OF NONLINEAR STRUCTURAL PATTERNS FROM DIGITAL IMAGES AT ENVIRONMENTAL AND SPACE SCIENCES

ABSTRACT

The Detrended Fluctuation Analysis (DFA) has been widely used to verify the scaling properties of unidimensional time series. Besides revealing the presence or absence of memory in the series (persistence), the DFA method allows understanding the process that originated the analyzed signal, as well as the strength of the correlation. The DFA operator (α) also infers about the signal roughness, because the larger the persistence, the smaller the roughness. When such feature is expanded to bi-dimensional signals, especially in digital images, the notion of persistence also infers on the texture of these signals. The DFA method was generalized to operate on bi-dimensional signals in 2006 (DFA-2D) and, since then, several studies have been performed using the (α) as a textural operator. Differently from the common textural operators, which focus on the brightness variation of an image, the DFA-2D exploits the theoretic framework from the fractal theory and infers the texture using the auto similarity feature of the signal. In this context, the goals of this work were: i) to explore controversial or not yet treated issues for the DFA-2D in the literature and ii) advance the state of the art in the bi-dimensional technique in its two versions – mono e multifractal – evaluating the response in a diverse set of data and also improving its performance by means of parallelization, using GPGPU/CUDA. Therefore, this thesis provides contributions in two aspects of the DFA-2D which are criticized in the literature: the interpretation difficulty related to the α operator and high computational cost of the method. Simulated sets of data that present a well-known response (fBm and fGn) has been studied, as well as classical sets of simulated data not treated in the literature using the DFA-2D (such as noise data of type $1/f$ and coupled map networks). The analysis also included sets of real data, providing umprecedent applications using the DFA-2D, such as the morphological classification of galaxies and the inference of effective roughness for eolic energy studies, by using the classification of Models of Digital Elevation (MDE). The DFA-2D (mono and multifractal) was able to categorize simulated signals in a satisfactory manner. For real signals, the results show that, even when the α operator does not reach a good classification rate, its usage is very informative, characterizing the signal and not only classifying it. Besides, the DFA-2D applied to real data presented difficulties and challenges imperceptible in simulated signals. Lastly, the parallelization has proven to be effective in order to lower the processing time required by the DFA-2D.

Keywords: Bi-Dimensional Detrended Fluctuation Analysis. Texture. Classification. Parallelization. GPGPU/CUDA.

LISTA DE FIGURAS

Pág.

2.1 Séries fGn e fBm obtidas para os mesmos expoentes de Hurst. As séries foram geradas com 65.536 amostras. 16
2.2 Perfis das séries apresentadas na Figura 2.1, obtidos após a soma acumulada da série (Passo a do algoritmo, equação (2.3). 17
2.3 Exemplo de superfície monofractal (a) e multifractal(b) e os gráficos $q \times h(q)$ e $q \times \tau(q)$ comparativos. 26
2.4 Espectro de singularidades de uma superfície multifractal e uma fractal (h=0.2). Os conceitos definidos por Makowiec et al. (2011) são representados na superfície multifractal. 27

3.1 Amostras de simulações de processos fBm (topo) e fGn (baixo) para diferentes expoentes de Hurst. 36
3.2 Visão tridimensional das superfícies fBm com diferentes expoentes de Hurst, ilustrando a diferença entre a rugosidade das mesmas. a) H=0.1. b) H=0.5. c) H=0.8. 36
3.3 Ruídos típicos. a)Ruído branco gaussiano (γ=0). b) Ruído rosa (γ=1). c) Ruído Vermelho (γ=2). 37
3.4 Amostras de evolução de séries espaço-temporais dos mapas logísticos acoplados, em diferentes instantes de tempo. Em A) mapa logístico. Em B) mapa dyadic bernoulli e em C) mapa SOM. 39
3.5 Esquema proposto por Hubble para classificação das galáxias a partir de sua morfologia visual. FONTE : Adaptado de Wikimedia Commons . . . 40
3.6 Exemplo de galáxia elíptica (A) e espiral (B). Fonte : Galaxy Zoo. 40
3.7 Amostras de imagens de galáxias elíticas (topo) e espirais (abaixo) extraídas do Galaxy Zoo. 42
3.8 Amostra dos modelos digitais de elevação. GDEM (Coluna da esquerda) e SRTM (coluna da direita). 45

4.1 *Ranking* das dez linguagens de programação mais utilizadas em 2016. FONTE : (IEEE, 2016) 49
4.2 Média do valor estimado do Expoente de Hurst com o DFA-2D para amostras de superfícies fBm. $0.05 \leq H < 1$, com incremento de 0.05. . . . 50
4.3 Diferença entre as fórmulas de integração. a) Soma acumulada (*cumsum*). b) Formulação dada pela equação 4.1. 52

4.4 Média do valor estimado do Expoente de Hurst com o DFA-2D para amostras de superfícies fBm, sem realizar a integração do sinal. 56
4.5 Padrões correlacionados utilizados para avaliar o comportamento do DFA-2D. 57
4.6 Padrões do tipo *chessboard* utilizados para avaliar o comportamento do DFA-2D. A intercalação dos valores zero e um variou de 1 pixel (primeira imagem) até 256 (última imagem). 58
4.7 Divisão lógica de unidades de trabalho na GPU. FONTE : (NVIDIA; UIUC, 2015). 61
4.8 Passos do DFA-2D paralelizados com CUDA. 63
4.9 Trecho do código PyCUDA que dispara a função de *kernel* nomeada 'fit', para o cálculo paralelo do DFA-2D. 65
4.10 Comparação do tempo de processamento entre as abordagens sequencial e paralela com o DFA-2D Monofractal. Os tempos são dados em segundos. 67
4.11 Comparação do tempo de processamento entre as abordagens sequencial e paralela com o DFA-2D Multifractal. Os tempos são dados em segundos. 67

5.1 Gráfico *log-log* dos sinais típicos fBm e fGn com Hurst 0.5 e dos ruídos branco, rosa e vermelho. Os dados são médias de quinhentas amostras. As linhas e as equações destacadas referem-se ao fit dos dados. A inclinação dessa reta revela o valor do parâmetro α. 71
5.2 Exemplos de espectro de potência radial para a)Ruído Branco ($\beta = 0$); b) Ruído Rosa ($\beta = 1$); c) Ruído Vermelho ($\beta = 2$); d) fGn (H=0.5); e)fBm (H=0.5). 72
5.3 Análise temporal dos modelos de mapas acoplados com o DFA-2D. Na parte de baixo da figura, as séries com seus respectivos desvios padrão referente às dez amostras. 74
5.4 Série original versus série deslocada em uma unidade no tempo. 75
5.5 Iterações escolhidas para análise de fractalidade dos modelos. 76
5.6 Resultado do MFDFA-2D para as iterações 1, 5, 10, 15, 25, 50, 100, 200 e 300 do mapa Logístico. Em a) Gráfico $q \times h(q)$. Em b) relação $q \times \tau(q)$. Em c) Espectro de singularidade. Os valores apresentados são médias das dez séries estudadas. 77
5.7 Resultado do MFDFA-2D para as iterações 1, 5, 10, 15, 25, 50, 100, 200 e 300 do mapa Dyadic Bernoulli. Em a) Gráfico $q \times h(q)$. Em b) relação $q \times \tau(q)$. Em c) Espectro de singularidade. 78

5.8 Resultado do MFDFA-2D para as iterações 1, 5, 10, 15, 25, 50, 100, 200 e 300 do mapa SOM. Em a) Gráfico $q \times h(q)$. Em b) relação $q \times \tau(q)$. Em c) Espectro de singularidade. Os valores apresentados são médias das dez séries estudadas. 80
5.9 Espectro de singularidade característico para sinais que apresentam monofractalidade para $q < 0$ e $q > 0$. A interpretação física é que a série apresenta um comportamento de escala para $q < 0$ e outro para $q > 0$. . 81
5.10 Relação de h(q) e q média para os mapas estudados. 82
5.11 Média dos valores de $\tau(q)$ para os mapas estudados. 82
5.12 Espectro de singularidades médio para os mapas estudados. 83

6.1 Amostras de imagens com erros que acarretaram em valores espúrios dos descritores de Haralick. 88
6.2 Resultado do MF-DFA-2D para as galáxias elípticas (a) e espirais (b). Na primeira coluna, o gráfico $q \times h(q)$, na coluna do meio, o gráfico $q \times \tau(q)$ e na terceira coluna, o espectro de singularidades. 89
6.3 Modelo de classificação de galáxias SDSS gerado pelo algoritmo J4.8 (α_{Raw} + descritores de Haralick selecionados). 96

7.1 Perfil vertical da velocidade do vento desde a superfície até a altura do vento geostrófico. O comprimento de rugosidade (z_0) é a altura onde a velocidade é nula. FONTE : Adaptado de (MARTINS et al., 2008) 100
7.2 Perfil vertical do vento e a influência da rugosidade da superfície. 101
7.3 Localização espacial da Torre M2 do NREL. 112
7.4 Representação dos quadrantes processados com os MDEs a)GDEM; b)SRTM para região do NREL. 112
7.5 a) Imagem de satélite da região de estudo NREL, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os anos 2010 e 2016, medidos a 2 metros. 114
7.6 Representação do perfil vertical da velocidade média do vento na região do NREL. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura) 115
7.7 Comprimento de Rugosidade Aerodinâmica na região do NREL, calculada pelo perfil vertical dos ventos. 116
7.8 Localização das áreas de Guapé, Camocim e Paracuru. 117
7.9 Localização espacial da torre de Guapé/MG. 118
7.10 Representação dos quadrantes processados com os MDEs a)ASTER/GDEM; b)SRTM para região de Guapé/MG 119

7.11 a) Imagem de satélite da região de estudo de Guapé, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os meses de Maio e Agosto de 2014, medidos a 3m. 120
7.12 Representação do perfil vertical da velocidade média do vento na região de Guapé. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura). 121
7.13 Comprimento de Rugosidade Aerodinâmica na região de Guapé, calculada pelo perfil vertical dos ventos. 122
7.14 Representação dos quadrantes processados com os MDEs a)ASTER/GDEM; b)SRTM para a região de Camocim/CE. 123
7.15 : a) Imagem de satélite da região de Camocim, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os anos 2004 e 2006, medidos a 20.6m. 125
7.16 Representação do perfil vertical da velocidade média do vento na região de Camocim. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura). 125
7.17 Perfil do relevo para os quadrantes 2 e 4 na região de Camocim. 127
7.18 Comprimento de Rugosidade Aerodinâmica na região de Camocim, calculada pelo perfil vertical dos ventos. 128
7.19 Representação dos quadrantes processados com os MDEs a)GDEM; b)SRTM para a região de Paracuru/CE. 129
7.20 a) Imagem de satélite da região de Paracuru, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os anos 2004 e 2006, medidas a 20.6 metros. 130
7.21 Representação do perfil vertical da velocidade média do vento na região de Paracuru. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura). 131
7.22 Comprimento de Rugosidade Aerodinâmica na região de Paracuru, calculada pelo perfil vertical dos ventos. 132

A.1 Exemplo de geração de uma matriz de co-ocorrência. Em a) Imagem quantizada em 3 níveis de cinza - 0, 1 e 2; Em b) Matriz de co-ocorrência da imagem, na direção 0° e distância de 1 pixel; Em c)Matriz de Co-Ocorrência normalizada pela soma dos valores da matriz. 160

LISTA DE TABELAS

Pág.

2.1 Relação entre a dimensão fractal, expoente de Hurst e seu significado na rugosidade de superfícies. 11
2.2 Valores do expoente de escala oriundo do DFA e suas implicações fenológicas. 14
2.3 Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm unidimensionais apresentados na figura 2.1 17

4.1 Resultado do DFA-2D aplicado a superfícies fBm com diferentes expoentes de Hurst. 51
4.2 Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais. Os valores são médias de trinta repetições. 52
4.3 Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, utilizando a formulação de soma acumulada apresentada por Gu e Zhou (2006). Os valores são médias de trinta repetições. 53
4.4 Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, para diferentes estratégias sobre a remoção ou não da média. Os valores são médias de trinta repetições. 54
4.5 Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, utilizando a integração da matriz antes da sub-divisão. Os valores são médias de trinta repetições. 54
4.6 Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, utilizando crescimento logarítmico das escalas s. Os valores são médias de trinta repetições. 55
4.7 Resultado do DFA-2D aplicado a superfícies fBm com diferentes expoentes de Hurst, sem a integração do sinal. 56
4.8 Valores de α e Moran para os padrões correlacionados da figura 4.5. . . . 58
4.9 Valores de α e Moran para os padrões '*chessboard*' da figura 4.6, com diferentes janelas de intercalação (em pixels). 58
4.10 Características da placa GeForce GTX 460. 62
4.11 Comparação de tempo de processamento entre as implementações sequenciais em Python e a paralela em PyCuda para o DFA-2D Monofractal. Os tempos são dados em segundos. Média de 50 repetições. 66

4.12 Comparação de tempo de processamento entre as implementações sequenciais em Python e a paralela em PyCuda para o DFA-2D Multifractal. Os tempos são dados em segundos. Média de 50 repetições. 66

5.1 Resumo dos dados utilizados nos testes com ruídos da família $1/f$ 70
5.2 Valores de α (DFA-2D) e β (PSD-2D) para ruídos $1/f$, comparados com os valores para dados fGn e fBm com expoente de Hurst 0.5. A média e o desvio padrão são relativos a quinhentas amostras de cada modelo. . . 70
5.3 Resumo dos dados utilizados nos testes com as redes de mapas acoplados. 73
5.4 Estatísticas básicas de α para os três modelos de mapas acoplados. . . . 75
5.5 Valores numéricos para α, range da função de flutuação e amplitude do espectro de singularidades para as iterações analisadas no mapa Logístico. 78
5.6 Valores numéricos para α, range da função de flutuação e amplitude do espectro de singularidades para as iterações analisadas no mapa Dyadic Bernoulli. Os valores apresentados são médias das dez séries estudadas. . 79
5.7 Valores numéricos para α, range da função de flutuação e amplitude do espectro de singularidades para as iterações analisadas no mapa SOM. . 80

6.1 Valores de $\alpha_{h(q=2)}$ para as galáxias elípticas e espirais após a remoção dos dados espúrios. 88
6.2 Valores do expoente de Hurst ($H_{q=2}$) e do $Hmax(f(\alpha)^{-} = 2)$ obtidos pelo MF-DFA2D aplicado às imagens de galáxias elípticas e espirais. O processamento foi feito com dados integrados e sem integração (Raw). 90
6.3 Principais indicadores multifractais obtidos com o MF-DFA2D. 91
6.4 Resultado dos descritores texturais de Haralick para as galáxias elípticas e espirais 93
6.5 Matriz de confusão da classificação de galáxias elípticas e espirais com o algoritmo de árvore de decisão J48. 95

7.1 Valores aproximados de comprimento de rugosidade da superfície para vários tipos de terreno. FONTE : Adaptado de (MANWELL et al., 2002) . 107
7.2 Comprimento de rugosidade de tipos de superfícies homogêneas (z_0). FONTE : Adaptado de (WIERNGA, 1993) 108
7.3 Valores numéricos de rugosidade efetiva para os quatro quadrantes do NREL, utilizando os MDEs ASTER/GDEM e SRTM. 113
7.4 Quantidade de amostras e respectivos percentuais por quadrantes na região do NREL, medidas a 2 metros. 114
7.5 Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas, para a região do NREL. 116

7.6 Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região do NREL. 117
7.7 Valores numéricos de rugosidade efetiva para os quatro quadrantes de Guapé, utilizando os MDEs GDEM e SRTM. 119
7.8 Quantidade de amostras e respectivos percentuais por quadrantes na região de Guapé, medidos a 3 metros. 120
7.9 Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas, para a região de Guapé. 121
7.10 Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região de Guapé. 122
7.11 Valores numéricos de rugosidade efetiva para os quatro quadrantes de Camocim, utilizando o MDE GDEM e SRTM. 123
7.12 Quantidade de amostras e respectivos percentuais por quadrantes na região de Camocim, medidas a 20.6 metros. 124
7.13 Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas na região de Camocim. 126
7.14 Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região de Camocim. 128
7.15 Valores numéricos de rugosidade efetiva para os quatro quadrantes de Paracuru, utilizando o MDE GDEM e SRTM. 130
7.16 Quantidade de amostras e respectivos percentuais por quadrantes na região de Paracuru, medidas a 20.6 metros. 130
7.17 Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas na região de Paracuru. 131
7.18 Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região de Paracuru. 132
7.19 Matriz de correlação entre os valores obtidos com os descritores texturais de Haralick para as áreas estudadas, e os parâmetros de rugosidade aerodinâmica e efetiva, considerando o MDE de melhor perfomance com a análise do DFA-2D. As células destacadas indicam correlação maior que 90%. 134

SUMÁRIO

Pág.

1 INTRODUÇÃO 1

2 ANÁLISE DE FLUTUAÇÃO DESTENDENCIADA (DFA) E SEUS PRINCIPAIS FUNDAMENTOS TEÓRICOS 5

2.1 Modelos fractais e sua relação com a rugosidade das superfícies 5

2.2 Descrição da rugosidade das superfícies por meio de modelos fractais . . 7

2.3 Expoente de Hurst em superfícies bi-dimensionais 10

2.4 Análise de Flutuação Destendenciada Unidimensional 12

2.4.1 Algoritmo do DFA unidimensional 12

2.4.2 Vantagens e Desvantagens 18

2.4.3 Comparação com outros métodos 19

2.5 Análise de Flutuação Destendenciada bi-dimensional (DFA-2D) 20

2.5.1 DFA-2D Multifractal 22

2.5.2 Complexidade Computacional 27

2.6 DFA e Textura em Imagens Digitais 28

3 CONJUNTOS DE DADOS ANALISADOS 33

3.1 Sinais bi-dimensionais típicos 33

3.1.1 Fractais 33

3.1.2 Ruídos $\frac{1}{f}$ 36

3.1.3 Redes de Mapas Acoplados - CML 37

3.2 Sinais oriundos da ciência espacial 39

3.3 Sinais oriundos da ciência ambiental 42

4 IMPLEMENTAÇÃO DO DFA-2D E DISCUSSÕES RELEVANTES SOBRE A TÉCNICA 47

4.1 Implementação do DFA-2D e do MFDFA-2D 48

4.2 Investigações acerca do algoritmo DFA-2D 51

4.2.1 DFA-2D e a autocorrelação espacial 57

4.3 DFA-2D Paralelo 59

4.3.1 Arquitetura Utilizada 62

4.3.2 Modelagem 63

4.3.3 Ganho de performance com o DFA-2D Paralelo 65

5 RESPOSTA DO DFA-2D À SINAIS BI-DIMENSIONAIS TÍPICOS . . . 69
5.1 Ruídos $\frac{1}{f}$ . . . 69
5.2 Redes de Mapas Acoplados - CML . . . 73
5.3 Considerações Finais . . . 83

6 ANÁLISE E CLASSIFICAÇÃO DE GALÁXIAS SDSS UTILIZANDO DFA-2D MULTIFRACTAL . . . 85
6.1 Métodos de Classificação de Galáxias . . . 85
6.2 DFA-2D Multifractal na classificação de galáxias SDSS . . . 87
6.2.1 Comparação do MFDFA-2D com outros descritores texturais . . . 92
6.3 Considerações Finais . . . 97

7 ANÁLISE DA RUGOSIDADE DE RELEVO UTILIZANDO DFA-2D . . . 99
7.1 Vento e Superfície . . . 99
7.2 Comprimento de Rugosidade . . . 104
7.2.1 Comprimento de Rugosidade Aerodinâmica . . . 105
7.2.2 Comprimento de Rugosidade Efetiva . . . 106
7.3 Comprimento de Rugosidade Efetiva por meio do DFA-2D em MDEs globais . . . 110
7.3.1 Área de Estudo NREL . . . 111
7.3.2 Área de Estudo de Guapé . . . 118
7.3.3 Área de Estudo Camocim . . . 123
7.3.4 Área de Estudo Paracuru . . . 129
7.4 Análise comparativa utilizando os descritores texturais de Haralick . . . 133
7.5 Considerações Finais . . . 135

8 CONCLUSÕES . . . 137

REFERÊNCIAS BIBLIOGRÁFICAS . . . 139

APÊNDICE A - TÉCNICAS COMPLEMENTARES . . . 159
A.1 Matriz de co-ocorrência dos níveis de cinza e descritores texturais . . . 159
A.1.1 Implementação . . . 163
A.2 Índice de Moran . . . 164
A.3 Função de Densidade Espectral de Potências . . . 165
A.4 Árvore de Decisão - Algoritmo J4.8 . . . 165

1 INTRODUÇÃO

Muitos processos na natureza são complexos para serem descritos por modelos simplesmente determinísticos, sendo necessário considerar modelos estocásticos para efetivamente descrevê-los. A tentativa de se construir um esquema teórico geral para descrever sistemas complexos na física deu origem a novas abordagens e conceitos como leis de potência, autossimilaridade, fractais, criticalidade auto-organizada, grupos de renormalização e teoria de campos desordenados, que passaram a fazer parte da física estatística contemporânea (GLERIA et al., 2004; ROSA et al., 2007; PARISI, 1992). Uma das propriedades marcantes de tais sistemas é a presença de leis de escala ou leis de potência, que descrevem quase tudo na natureza, às vezes até sistemas desordenados.

Segundo Kantelhardt (2009), caracterizar e compreender um sistema complexo não é uma tarefa fácil, já que não se pode dividi-lo em subsistemas sem alterar suas propriedades dinâmicas. Para tanto, séries temporais e espaço-temporais podem ser utilizadas, já que elas refletem o estado de um sistema numa representação reduzida de dimensionalidade. Ainda, segundo Kavasseri e Nagarajan (2004), diversas técnicas foram desenvolvidas com o intuito de descrever os mecanismos geradores de tais fenômenos e, principalmente, tentar prevê-los. No entanto, a maioria das técnicas está voltada para tratar fenômenos estacionários. Além disso, séries oriundas de processos complexos tendem a ser curtas (contendo, por exemplo, eventos extremos), o que também diminui a eficiência dos métodos estatísticos tradicionais.

Desde 1994, quando Peng e colaboradores desenvolveram a técnica de *Análise de Flutuação Destendenciada*[1] (DFA) (PENG et al., 1994), a mesma tem sido utilizada de forma recorrente nos estudos que buscam compreender o regime de escala que rege sistemas e processos das mais diferentes áreas além da física. Mesmo contendo algumas deficiências técnicas (HENEGHAN; MCDARBY, 2000; CHEN et al., 2002; BRYCE; SPRAGUE, 2012; HÖLL; KANTZ, 2015b), o DFA tem a vantagem de atacar os dois problemas acima citados : tendência e séries curtas. O método é, na verdade, um aferidor do expoente de Hurst (HURST, 1951), otimizado para análise de sinais não estacionários. O objetivo do DFA é, portanto, quantificar, caso exista, correlação de longo alcance (memória) na série temporal, caracterizando-a como persistente ou anti-persistente e revelando sua intensidade.

Em 2002, Kantelhardt et al. (2002) generalizaram o DFA para caracterizar sinais

[1] Do inglês : *Detrended Fluctuation Analysis*. Nesta tese, o termo *Detrended* está sendo traduzido como Destendenciada.

multifractais. Mais recentemente, Gu e Zhou (2006) expandiram o DFA e o DFA multifractal (MF-DFA) para operar não apenas sobre séries temporais unidimensionais, mas também sobre quaisquer outras dimensões, em especial, sinais bi-dimensionais (DFA-2D e MF-DFA2D), que incluem a observação de padrões a partir de imagens digitais. Além disso, diversas outras generalizações tem sido propostas ao longo dos últimos anos (XIONG; SHANG, 2017).

A noção de persistência quando levada para padrões estruturais bi-dimensionais, pode revelar características relacionadas a flutuação da rugosidade [2] dessas superfícies. Isso porque quanto mais persistente for o sinal, mais constante é a série, e vice-versa (PENG et al., 1994). Dessa forma, o DFA-2D e o MF-DFA2D têm sido aplicados especialmente nessas condições, considerando imagens digitais que revelam padrões espaciais produzidos através de uma dinâmica subjacente. E, apesar de serem ainda pouco explorados, têm sido utilizados, com sucesso, em diferentes aplicações.

A principal motivação dessa tese é, portanto, estudar e aprimorar os algoritmos DFA para aplicações inéditas em dados espaciais e ambientais bi-dimensionais. Partimos do pressuposto que há ganhos em utilizar o DFA-2D como operador textural paralelizado, visto que, além de caracterizar a rugosidade da superfície (textura), o DFA-2D ainda caracteriza a origem do sinal, podendo revelar sua classe de universalidade e, sobretudo, melhorando o entendimento do processo físico subjacente.

Neste sentido, o trabalho desenvolvido nessa tese tem dois grandes objetivos:

a) Levantar, discutir e experimentalmente provar algumas questões do ponto de vista técnico, ou de execução do DFA para aplicações bi-dimensionais;

b) Evoluir o estado da arte em termos de aplicação deste operador para fornecer uma nova métrica para classificação automática de padrões.

A motivação para o primeiro objetivo surge devido à divergências encontradas na literatura sobre a metodologia do DFA-2D, e também por práticas comumente aplicadas ao DFA-1D que não foram comunicadas para o DFA-2D. Para cumprir esse objetivo, foram realizados testes empíricos com dados fBm (*Fractional Brownian Motion*) e fGn (*Fractional Gaussian Noise*), dos quais se conhece o resultado esperado. Toda essa discussão é apresentada no Capítulo 4.

[2] Rugosidade aqui é usada para traduzir termos já consolidados na física como *roughness* e *coarse grainy*

Ainda no Capítulo 4, uma implementação paralela utilizando GPU e CUDA é apresentada, uma vez que alguns autores discutem a performance do DFA-2D mono e multifractal para imagens grandes, ou dentro do contexto de *Big Data* e *Machine Learning*. A implementação para GPGPU-CUDA apresentada nesta tese reduziu significativamente o tempo de processamanto deste operador.

A motivação para o segundo objetivo previamente mencionado, parte da necessidade de se conhecer o comportamento do DFA-2D monofractal e multifractal em conjuntos de dados que não são sabidamente fBm ou fGn. Algo que alguns autores alegam é que o parâmetro oriundo da aplicação da técnica, denominado parâmetro α, não é de fácil interpretação. Nesse sentido, essa tese visa contribuir na melhoria da interpretação dos valores de α e isso foi feito de duas formas: conceitualmente e experimentalmente.

Conjuntos de dados de padrões conhecidos, como fBm, fGn, os ruídos da família $\frac{1}{f}$ e dados caóticos, como as redes de mapas acoplados (CML) foram submetidos ao processamento com o DFA-2D mono e multifractal. O conhecimento e arcabouço conceitual obtidos nessa etapa favoreceram a compreensão e discussão sobre o processamento de dados reais com o DFA-2D. Essa discussão e resultados são apresentados no Capítulo 5.

No âmbito mais experiental, duas diferentes áreas do conhecimento foram escolhidas para os estudos de caso com dados reais : as ciências espaciais (morfometria de galáxias em grandes arquivos de dados) e ambientais (modelos digitais de elevação topográfica). A escolha foi em função delas apresentarem problemas cuja solução poderia passar por uma análise de rugosidade da superfície, e também por já serem objetos de estudo dos orientadores dessa tese.

No campo das ciências espaciais, a classificação de galáxias é um dos maiores desafios dos astrônomos e, portanto, um importante objeto de estudo, pois ela pode fornecer relevantes informações acerca dos processos físicos que ocorrem nas galáxias e também sobre a origem e evolução do universo (ELFATTAH et al., 2014; THANKI et al., 2009). Isso porque galáxias não são objetos estáticos, elas evoluem interagindo com poeira, gás e outras galáxias (AU et al., 2006), obedecendo uma teoria evolutiva que relaciona a origem de galáxias espirias a partir de galáxias elípticas. No entanto, a natureza multivariada das galáxias, em arquivos contendo mais de um bilhão de objetos, torna a classificação difícil e pouco acurada (LEKSHMI et al., 2003).

A classificação tradicional de galáxias é feita visualmente por observadores treinados.

No entanto, no contexto *Big Data* e com o advento de telescópios modernos, uma quantidade enorme de dados de objetos celestiais tem sido capturada (cerca de 1 TB/h), tornando indispensável o estudo de metodologias que sejam capazes de classificar automaticamente as galáxias com base em sua morfologia. Nesta tese, o DFA-2D multifractal foi testado para esse fim. A literatura já apresentava o uso de operadores fractais para essa tarefa, porém não incluindo o uso do DFA-2D. Esse estudo inédito é apresentado no Capítulo 6.

Já no campo das ciências ambientais, o problema tratado foi a quantificação do coeficiente de rugosidade aerodiâmica (z_0), utilizado especialmente em modelos teóricos para extrapolação de dados de vento. Esses modelos são muito úteis nos estudos de energia eólica. A instalação de um parque, ou mesmo de uma torre anemométrica depende do estudo da intensidade e direção do vento na região, que em geral é realizado em altitudes menores do que a da torre. Sendo assim, o perfil do vento nas altitudes alcançadas pelas torres eólicas necessita dos modelos teóricos, como o modelo logaritmico, que é dependente do coeficiente de rugosidade aerodinâmica.

Este coeficiente é de difícil determinação, uma vez que é influenciado por diversas variáveis. Em locais onde o relevo é uniforme (plano), a estimação desse parâmetro é feita com base no uso e cobertura do solo da região. No entanto, em locais onde o relevo é acidentado, esse tipo de estimação perde o sentido, uma vez que a rugosidade do relevo passa a ser mais importante do que a rugosidade da cobertura do solo. É justamente esse o ponto atacado nessa tese. Por meio de imagens oriundas de Modelos Digitais de Elevação (MDE) de baixa resolução, verificou-se se a utilidade do DFA-2D de aferir a rugosidade do terreno no entorno de torres anemométricas e ainda se há uma relação direta entre a rugosidade do terreno e a rugosidade aerodinâmica. Esse estudo está descrito no Capítulo 7.

Além dos capítulos mencionados acima, os capítulos 2 e 3 e o apêndice , foram incorporados com as seguintes finalidades:

- Capítulo 2: Apresenta o DFA em suas versões unidimensional, bidimensional e multifractal, além dos principais conceitos necessários para compreender seu funcionamento.
- Capítulo 3: Neste capítulo é feita uma descrição dos dados utilizados nessa tese, tanto os simulados, quanto os reais.
- Apêndice : apresenta as técnicas complementares utilizadas para fins de comparação com o DFA-2D.

2 ANÁLISE DE FLUTUAÇÃO DESTENDENCIADA (DFA) E SEUS PRINCIPAIS FUNDAMENTOS TEÓRICOS

Este capítulo apresenta a Análise de Flutuação Destendenciada (DFA) em suas múltiplas versões : unidimensional, bi-dimensional e multifractal. O começo do capítulo aborda os conceitos que formam o arcabouço teórico do DFA, oriundos da teoria dos fractais. Posteriormente, o algoritmo do DFA unidimensional é apresentado, visto que o DFA bi-dimensional é uma generalização do unidimensional e, sendo assim, as vantagens, desvantages e críticas ao método unidimensional são também aplicadas a sua versão bi-dimensional. O capítulo destaca ainda os principais conceitos envolvidos na aplicação de técnicas fractais como operadores texturais em imagens digitais.

2.1 Modelos fractais e sua relação com a rugosidade das superfícies

Apesar de não haver uma definição rigorosa, um fractal pode ser definido como uma forma geométrica complexa que pode ser subdividida em partes onde cada parte é a cópia reduzida do todo. Os fractais são reconhecidos por compartilharem algumas características, tais como (MENDONÇA et al., 2007): a) autossimilaridade; b) lei de potência; c) propriedades estatísticas anômalas[1]; d) dimensão fractal, em geral não inteira; e) lei ou algoritmo de construção. Ainda, segundo Santos (2013), a fractalidade de um objeto ou fenômeno pode ser entendida como um conjunto de regras sistemáticas que governam as propriedades deste objeto (ou fenômeno) ao longo do tempo e espaço. A identificação desse conjunto de regras passa pela investigação de características, tais como a autossimilaridade e a dimensão fractal.

A autossimilaridade, ou autossemelhança, é uma das principais características dos fractais. Autossimilaridade significa que cada parte menor do objeto ou processo fractal se assemelha ao todo. Isso implica que as partes de um sistema, em diferentes escalas espaciais ou temporais, são semelhantes ao sistema como um todo. A dimensão de autossimilaridade (dimensão fractal) é uma quantidade fracionária que representa o grau de ocupação da estrutura no espaço que a contém e o nível de irregularidade de um fractal. As dimensões fractais procuram, portanto, mensurar quão densamente um fractal ocupa o espaço métrico no qual está imerso (GLERIA et al., 2004; KANTELHARDT, 2009; MENDONÇA et al., 2007).

Sendo assim, fractais autossimilares são estruturas que podem ser reescalonadas

[1] As propriedades estatísticas anômalas de objetos e processos fractais caracterizam-se pela inexistência de média e variância.

isotropicamente. Ou seja, as mudanças de escala em qualquer direção são as mesmas ou estão afetadas por um mesmo fator. Neste caso, a dimensão fractal, D, é única. No entanto, pode acontecer certa anisotropia na estrutura do objeto fractal, gerando os chamados fractais auto-afins. Neste caso, para cada direção, é possível encontrar diferentes propriedades de escalonamento e dimensão fractal. Nesses processos, em escalas menores, há flutuações em torno de uma configuração média e o coeficiente de autocorrelação não é exato.

No caso de fractais auto-afins, a autossimilaridade é estatística e está ligada ao fato de uma parte do perfil ser estatisticamente semelhante ao todo. A autossimilaridade estatística envolve a busca de propriedades estatísticas que se mantêm inalteradas diante de mudanças de escala, tais como expoente de Hurst e a dimensão fractal, e é caracterizada por leis de escala (SANTOS, 2013).

Segundo Cruz (2000), existem leis de escala que são básicas e demonstram que sistemas aparentemente diferentes podem apresentar comportamentos comuns. A caracterização de sistemas através de expoentes globais leva à definição de classes de universalidade. Ou seja, sistemas pertencentes à mesma classe de universalidade podem ser descritos por leis de formação semelhantes. A lei de escala é uma relação polinomial que exibe a propriedade de invariância de escala na forma:

$$f(x) = A * x^{\alpha} \tag{2.1}$$

onde A é chamada de constante de normalização; α é o expoente de escala e x é uma variável do sistema físico estudada.

Sendo assim, é possível encontrar propriedades de escala para fenômenos complexos através das flutuações existentes nas suas séries temporais e espaço-temporais, e visualizar o nível de periodicidade do sistema (MARINHO et al., 2015). Essas flutuações podem ser descritas como sendo: a) ruído branco não correlacionado; b) correlações de curto alcance; ou c) correlações ou memória de longo alcance com lei de potência (escala).

Batista (2006) afirma que não existe uma definição totalmente formal sobre memória longa em séries temporais, e que normalmente essas séries apresentam persistência nas autocorrelações amostrais, isto é, dependência significativa entre observações separadas por um longo período de tempo. Na prática, a memória de longo alcance significa que o efeito de um evento ocorrido em um instante pode ser detectado depois de muitos intervalos de tempo.

Witt e Malamud (2013) explicam que a persistência na série temporal é caracterizada quando, em média, valores grandes tendem surgir depois de valores grandes e pequenos valores tendem a surgir depois de valores pequenos. Em contraste, a anti-persistência é caracterizada quando valores grandes tendem a surgir depois de valores pequenos, e pequenos tendem a surgir depois de valores grandes. Tanto a persistência, quanto a anti-persistência, pode ter uma força que varia de fraca a muito forte.

De forma geral, correlações de longo alcance estão intimamente relacionadas à propriedade de auto-similaridade e são comumente observadas na natureza, economia, hidrologia e redes de comunicação. Diversos métodos foram desenvolvidos para detecção de memória (correlações de longo alcance) em domínio temporal, sendo o principal deles, o expoente de Hurst (HURST, 1951), que será detalhado na seção 2.3. O uso desses métodos assume que os processos apresentam comportamento fractal e que podem ser modelados pelo Ruído Gaussiano Fracionário (fGn) ou pelo Movimento Browniano Fracionário (fBm).

Movimento Browniano Fracionário é uma generalização do Movimento Browniano Ordinário que possui média igual a zero e auto-similaridade estatística dos incrementos. O fBm, diferente do movimento browniano, possui incrementos estacionários, porém dependentes e, consequentemente, memória longa.

O Ruído Gaussiano Fracionário é um sinal estacionário, caracterizado por sua variância e expoente de Hurst.

Maiores detalhes sobre fractais e seus modelos (Ruído Gaussiano Fracionário (fGn) e Movimento Browniano Fracionário (fBm) poderão ser vistos na seção 3.1.1.

No caso de superfícies bi-dimensionais, a auto-similaridade e a persistência podem ser associadas com a rugosidade da superfície. Nesse caso, uma superfície que exibe um comportamento de escala persistente tende a ser mais lisa do que aquela que possui comportamento anti-persistente. Sendo assim, modelos fractais têm sido utilizados satisfatoriamente para descrever a rugosidade de superfícies, como será visto na próxima seção.

2.2 Descrição da rugosidade das superfícies por meio de modelos fractais

Existe uma diversidade de metodologias e algoritmos dedicados à análise e extração de medidas de textura e rugosidade de imagens digitais. Isso porque há grande interesse em se conhecer a natureza da rugosidade presente nas diversas superfícies

reais. Zuñiga et al. (2014) agrupam esses métodos em quatro classes:

- Estruturais : onde as texturas são descritas por meio de primitivas[2];
- Estatísticos : onde as texturas são caracterizadas por medidas de distribuição não determinísticas;
- Baseados em modelos : onde as texturas são descritas por modelos físicos e matemáticos. Exemplos incluem modelos fractais, autoregressivos, randômicos, etc .
- Espectrais : onde as análises são feitas no domínio da frequência

A análise de textura por fractais se enquadra, portanto, na classe baseada em modelos e torna-se uma ferramenta relevante, já que muitos objetos na natureza exibem padrões randômicos, porém persistentes que resultam na aparência de textura do tipo nuvem [3] (*cloud-like*). Neste caso, elas precisam de modelos capazes de capturar a característica auto-afim.

Ohanian e Dubes (1992) citam três motivos que tornam modelos fractais interessantes nos estudos de classificação, segmentação e compressão de imagens : 1) os modelos gerados computacionalmente são bastante realistas e oferecem informação essencial sobre a textura; 2) os objetos fractais produzem imagens fractais com a mesma dimensão fractal e, 3) forma compacta de representação da textura e a propriedade multi-escala. Ainda segundo Xie (2008), funções fractais representam um bom modelo para descrever a rugosidade típica de superfícies naturais.

Em um trabalho referência na área, Schmittbuhl et al. (1995) classificam os métodos que mensuram expoentes auto-afins em dois grupos. No primeiro grupo, encontram-se os métodos clássicos, que buscam encontrar a dimensão fractal dos objetos. Dentre esses, podem ser citados o *box-counting* e o método da divisão. O outro grupo seria de métodos que buscam calcular o expoente de auto-afinidade propriamente dito. Os métodos largura de banda variável, probabilidade de retorno e espectro de potência

[2]Primitivas de textura, ou *texton*, são elementos que se repetem numa área da imagem. Esses elementos podem ser descritos como cor, altura, largura, orientação, entre outros. Dessa forma, a abordagem estrutural pode utiizar, por exemplo, paralelismo de linhas e seus espaçamentos na caracterização e diferenciação de objetos de textura (FILHO, 2012).

[3]Kaplan e Kuo (1995)definem essas texturas como aquelas onde há o decaimento, mas a função de correlação é significante. Exemplos incluem imagens de mamografia, modelos digitais de elevação, imagens de fogo, nuvens e fumaça. Essas texturas podem ser representadas por modelos fBm.

são alguns citados como sendo integrantes desse segundo grupo. Schmittbuhl et al. (1995)detalham todos os métodos e avaliam seus erros de acurácia.

Neste cenário, a dimensão fractal tem sido aplicada na análise de textura, segmentação, mensuração de forma e classificação de imagens e análise gráfica (LI et al., 2009; OHANIAN; DUBES, 1992). Em geral, a dimensão fractal de uma superfície lisa é próxima de 2, enquanto que a dimensão fractal de uma superfície extremamente rugosa é próxima de 3. Valores próximos a 2.5 indicam a presença de fenômenos estocásticos.

Segundo Myint (2003), existe uma relação entre a rugosidade, caracterizada pela diferença no brilho da imagem em pequenos intervalos espaciais, com a escala. Zuñiga et al. (2014) corroboram, afirmando que mesmo pequenas mudanças entre os padrões de textura produzem mudanças significativas na assinatura espectral (intensidade do brilho) da imagem.

Srinivasan e Shobha (2008) fizeram uma análise de diferentes metodologias utilizadas para realizar análise de textura em imagens digitais 2D. Dentre elas, aquelas baseadas em fractais. Os autores afirmam que a descrição da textura de uma imagem é altamente dependente da escala e que para diminuir a sensibilidade da escala, a textura precisa ser descrita em diferentes resoluções e, posteriormente, deve-se escolher uma que melhor a descreva. Neste ponto há uma relação direta com o fato dos objetos reais não serem verdadeiramente fractais. Sendo assim, não se pode esperar autossimilaridade em todas as escalas, como acontece com objetos fractais matematicamente construídos, e a análise deve ser feita em uma gama limitada de escalas sujeitas a considerações físicas ou biológicas (GONZALES-BARRON; BUTLER, 2008).

Sarkar e Chaudhuri (1992)também fizeram um estudo comparando diferentes técnicas fractais na análise de textura de imagens digitais.

Pantic et al. (2016) utilizaram a técnica *box-counting* para avaliar diferenças entre células renais sadias e lesionadas. Os autores descobriram que a dimensão fractal diminui em tecidos lesionados. Ainda na área biomédica, Borowska et al. (2015) utilizaram cinco diferentes métodos baseados na teoria de fractais para avaliar tecidos no processo de cicatrização após perda óssea. Os autores concluíram que também há diminuição da dimensão fractal em tecidos com perda óssea. E, assim como Myint (2003), concluíram que métodos diferentes levam a valores de dimensão fractal diferentes também.

Martins et al. (2003) aplicaram o método dos Prismas Triangulares, proposto por Clarke (1986), para separar a textura de três diferentes usos da terra em uma aerofoto digitalizada da cidade de Porto Alegre: área urbana, água e vegetação. Os autores realizaram uma transformação na imagem espectral de modo a convertê-la em uma 'banda fractal'. Usando janela móvel de 9x9, os autores conseguiram uma separabilidade de 98.62% resultado cerca de 15% superior ao da classificação da imagem espectral.

Ibanez (2012) utilizou a dimensão fractal de autossimilaridade para caracterizar a rugosidade do Modelo Digital de Elevação (MDE) oriundo a missão SRTM[4] .

Apesar dos diversos trabalhos de pesquisa na área, há uma discussão na literatura sobre o uso de modelos fractais na análise de imagens digitais. Isso porque tais objetos não podem, de fato, serem considerados fractais. Outro aspecto importante é que os operadores fractais revelam uma média das propriedades texturais da superfície e podem evidenciar uma mesma dimensão fractal para objetos que apresentam texturas visualmente diferentes (MYINT, 2003; XIE, 2008).

No entanto, modelos fractais e aferições de dimensão fractal têm sido utilizados como medidas de complexidade da imagem, e não na caracterização desses objetos como fractais, ou na estimativa fiel de sua dimensão fractal. Neste caso, normalmente considera-se que as superfícies reais são aproximadamente auto-similares e não estacionárias.

Como citado anteriormente, os operadores fractais buscam aferir a dimensão fractal ou o expoente de autossimilaridade (Hurst). A próxima seção descreve o expoente de Hurst, sua relação com a dimensão fractal e como aferidor de textura em imagens digitais.

2.3 Expoente de Hurst em superfícies bi-dimensionais

Estudando as inundações anuais do rio Nilo, no Egito, o engenheiro Harold Edwin Hurst aferiu empiricamente um método estatístico, nomeado R/S ou análise de reescalonamento. O R/S é amplamente utilizado na caracterização de estruturas relacionadas a fenômenos naturais e econômicos dependentes do tempo, podendo ainda classificar séries temporais e distingui-las quanto a sua randomicidade. Detalhes e formulação matemática do expoente de Hurst podem ser obtidos em (GNEITING; SCHLATHER, 2001; HURST, 1951).

[4] *Shuttle Radar Topography Mission* - `www2.jpl.nasa.gov/srtm/`

Considerando superfícies bi-dimensionais irregulares, complexas e auto-afins, o expoente de Hurst controla a rugosidade visual da textura em todas as escalas. Diversos algoritmos de classificação de imagens bi-dimensionais têm sido desenvolvidos utilizando esse *framework* teórico. Carbone (2007) cita alguns métodos propostos para esse fim e propõe um algoritmo para estimar o expoente de Hurst em fractais n-dimensionais.

Gneiting e Schlather (2001) esclarecem a relação entre o expoente de Hurst e a dimensão fractal. Tais autores afirmam que, à princípio, dimensão fractal e expoente de Hurst são medidas independentes, já que a dimensão fractal é uma propriedade local da superfície e o expoente de Hurst é uma propriedade global da série e/ou superfície, relacionada à persistência (memória de longo alcance). No entanto, para processos auto-afins, a propriedade local é refletida na global, da seguinte forma (GNEITING; SCHLATHER, 2001):

$$D + H = n + 1 \tag{2.2}$$

onde,

D é a dimensão fractal; H é o expoente de Hurst;n é a quantidade de dimensões do espaço.

Dessa forma, no caso de imagens bi-dimensionais, $n = 2$ e, consequentemente, $D \equiv 3 - H$ (FORTIN et al., 1992).

A tabela 2.1 resume os relacionamentos e propriedades acima citados. No caso, quanto mais próximo dos limites inferiores e superiores, mais forte a correlação.

Tabela 2.1 - Relação entre a dimensão fractal, expoente de Hurst e seu significado na rugosidade de superfícies.

Tipo de Superfície	**Dimensão Fractal**	**Expoente de Hurst**	**Correlação de Longo Alcance**
Lisas e Homongêneas	[2 a 2.5)	(0.5 a 1]	Persistente
Estocásticas	2.5	0.5	Sem memória
Rugosas	(2.5 a 3]	[0 a 0.5)	Anti-Persistente

Atualmente, o cálculo do expoente de Hurst pode ser feito no domínio do tempo por métodos como R/S, Desvio Quadrático Médio (RMS) e Análise de Flutuação Destendenciada (DFA); ou no domínio da frequência, com métodos como periodograma, *Power Spectrum* e *Wavelets*. Lacasa et al. (2009) introduziram ainda um método di-

ferente, baseado na teoria dos grafos. Uma boa revisão dos métodos unidimensionais pode ser obtida em Rea et al. (2009) e Mielniczuk e Wojdyłło (2007).

Alvarez-ramirez et al. (2006) reforçam o conceito de que a noção de persistência provê informações sobre textura em diferentes escalas. No entanto, os autores chamam atenção para o fato de que tendências nos dados (ou não-estacionaridade) podem enviesar o expoente de Hurst. Nesse sentido, a nálise de Flutuação Destendenciada (DFA) tem sido utilizada justamente por ser um método destendenciado.

A próxima seção aborda o DFA unidimensional, precursor do DFA-2D.

2.4 Análise de Flutuação Destendenciada Unidimensional

O DFA foi originalmente definido por Peng et al. (1994) para análise de sequências gênicas e, desde então, tem se tornado uma técnica robusta para quantificar correlações em séries temporais não estacionárias, utilizada em diversas aplicações. O DFA é um aperfeiçoamento do método denominado *Fluctuation Analysis* - FA e uma versão modificada do desvio médio quadrático, aplicado sobre um *random walk*.

Peng et al. (1995) explicam que, do ponto de vista prático, caso seja possível separar o ruído descorrelacionado das flutuações geradas pelo sistema dinâmico gerador da série temporal, então, apenas as flutuações deverão apresentar correlação de longo-alcance.

2.4.1 Algoritmo do DFA unidimensional

O algoritmo para calcular o DFA consiste basicamente de seis passos (KAVASSERI; NAGARAJAN, 2004; PENG et al., 1994):

a) A série temporal X_k com k = 1, ...,N é integrada na chamada função perfil, dada por:

$$Y(k) = \sum_{k=1}^{N} (X_k - \bar{X}) \tag{2.3}$$

onde $\bar{X}$ é a média da série temporal.

b) A série (ou perfil) $Y(k)$ é dividida em n intervalos não sobrepostos de tamanho s, onde $n_s = \lfloor N/s \rfloor$.

c) Para cada intervalo n_s, a tendência local (t) é calculada por meio do ajuste local de mínimos quadrados e calcula-se a série "destendenciada"através da

diferença entre o perfil $Y(k)$ e o polinômio $P_t(k)$.

$$\tilde{Y}_s(k) = Y(k) - P_t(k) \tag{2.4}$$

Polinômios de primeiro grau removem tendências lineares da função perfill $Y(k)$ e são normalmente, os mais utilizados. Porém, podem ser utilizados polinômios de segundo, terceiro, quarto, quinto e até sexto grau, que removerão tendências de ordens maiores (FREITAS et al., 2009).

d) Calcula-se a variância para cada segmento $v = 1, 2, ..., n_s$, obtendo-se a função de flutuação local $F^2(v, s)$:

$$F^2(v, s) = \frac{1}{s} \sum_{i=1}^{s} {Y_s}^2[(v-1)s + i] \tag{2.5}$$

e) Calcula-se a função de flutuação global ($F(s)$) como a média das variâncias de cada intervalo:

$$F(s) = \left[\frac{1}{s} \sum_{v=1}^{n_s} F^2(v, s))\right]^{\frac{1}{2}} \tag{2.6}$$

f) Para determinar a relação entre a função de flutuação $F(s)$ e a escala s, a computação da função de flutuação gobal $F(s)$ é repetida sobre várias escalas temporais. O expoente de correlação α é calculado por regressão linear da curva de variação de um gráfico *log-log*.

Caso a série analisada seja auto-similar, a função $F(s)$ apresentará comportamento de lei de potência do tipo $F(s) \approx s^\alpha$, onde α é o expoente de correlação de longo alcance.

Em geral, para cada grau do polinômio utilizado no cálculo do DFA, usa-se a nomenclatura α_1 (polinômio linear), α_2 (polinômio quadrático), e assim sucessivamente, para os polinômios de maior grau.

Na referência original (PENG et al., 1994), as escalas s variam linearmente de 4 até $\frac{n}{4}$, sendo n o tamanho da série. No entanto, testes empíricos levaram a fórmulas logarítmicas para o crescimento de s, sem afetar a performance do método e trazendo vantagens computacionais. Fornari (2016) faz um estudo da complexidade algorítmica do DFA-1D mono e multifractal. Para o caso monofractal, utilizando crescimento logarítmico de escalas, têm-se uma complexidade assintótica da ordem de $O(nlog(n)b^2)$, onde b representa o grau do polinômio utilizado no destendencia-

mento da série.

Segundo Souza et al. (2015), em geral, a interpretação física do problema estudado deve considerar o comportamento do expoente de correlação, já que o mesmo quantifica a natureza das correlações da série temporal, conforme tabela 2.2.

Tabela 2.2 - Valores do expoente de escala oriundo do DFA e suas implicações fenológicas.

Valores de α	**Caracterização**	**Consequência**
$0 \leq \alpha < 0.5$	Correlações de longo alcance anti-persistentes	Valores altos são normalmente seguidos por valores baixos e vice-versa
$\alpha \equiv 0.5$	Ruído branco	Sinal sem nenhuma correlação
$0.5 < \alpha < 1$	Correlações de longo alcance persistentes	Valores altos (baixos) são normalmente seguidos por valores altos (baixos)
$\alpha = 1$ α =1.5	Correspondem a ruído $\frac{1}{f}$,(*pink noise*) e *Random Walk*	

Alpatov et al. (2013) explicam que se todos os elementos da série são independentes (como em um modelo Gaussiano), caracteriza-se um caso de '*random walk*' discreto (1D). Por outro lado, a probabilidade de incrementos positivos e negativos dependerá do grau de correlação entre os elementos vizinhos da série; e o domínio positivo ou negativo dessa correlação indica o efeito de memória. Segundo Kantelhardt (2009), quanto maior o valor de α, mais forte é a correlação entre as amostras.

Algumas considerações são relevantes sobre o método do DFA e pouco discutidas pelos autores que o usam apenas como um índice discriminatório. O primeiro ponto é que, em seu artigo original, Peng et al. (1994) não descrevem o passo 1 do algoritmo acima citado, relativo à integração da série. Isso porque os autores trabalhavam com séries de nucleotídeos. Na ocasião, cada sequência de nucleotídeos era representada como uma cadeia de caracteres de purinas e pirimidinas e, para ser possível lidar com valores numéricos, a série foi codificada como um '*DNA-Walk*', de forma que, para cada pirimidina encontrada na posição i da cadeia de caracteres, dava-se um passo para frente ($u(i) = +1$) na série codificada; e, para cada purina encontrada, dava-se um passo para trás ($u(i) = -1$). Sendo assim, o processo gerava uma superfície fractal, cuja altitude $y(n)$ é o deslocamento líquido depois de n passos (BULDYREV et al., 1993):

$$y(n) = \sum_{i=1}^{n} u(i) \tag{2.7}$$

ou seja, o perfil identificado no passo 1 do algoritmo.

A remoção da média da série é justificada por Eke et al. (2000), que afirmam que isso melhora a estimativa para sinais fBm, mas não têm efeito nos sinais fGn. Já Kantelhardt et al. (2002) esclarecem que a remoção da média não faz diferença no DFA-1D, visto que a mesma é novamente calculada no passo e do algoritmo (equação (2.6)). De fato, experimentos empíricos realizados nesse trabalho mostraram que a remoção ou não da média levam aos mesmos valores de α, tanto para sinais fGn, quanto para fBm.

Outro aspecto bastante relevante é citado por Delignieres et al. (2006). Os autores frisam que o expoente de Hurst pode ser calculado para séries fBm e fGn. No entanto, devido às diferentes características dos modelos, métodos diferentes precisam ser utilizados. Sendo assim, é necessário identificar primeiro o tipo de modelo. Porém, o método de Densidade Espectral de Potência (PSD) e o DFA são capazes de distinguir entre séries fBm e fGn.

No DFA fica claro que a integração da série no primeiro passo do algoritmo contribui para essa distinção. Ao integrar um sinal fGn ele é convertido em fBm. Caso o sinal já seja um fBm, ele passa a ser um fBm acumulado. Nesse caso, o parâmetro α, na verdade, varia entre 0 e 2, de forma que (DELIGNIERES et al., 2006; EKE et al., 2000):

- Se o sinal for estacionário, i.e., fGn, $0 \leq \alpha \leq 1$ e $\hat{H} \equiv \alpha$
- Se o sinal for não estacionário, i.e., fBm, $1 \leq \alpha \leq 2$ e $\hat{H} \equiv \alpha - 1$

Sendo $\hat{H}$ o expoente de Hurst aferido pelo DFA.

Os conceitos acima são ilustrados na figura 2.1 e tabela 2.3. Na figura 2.1, a coluna da esquerda apresenta sinais fGn e na direita, sinais fBm com o mesmo valor de Hurst. Os sinais não são derivados um do outro, são amostras independentes. Na tabela 2.3 são apresentados os valores de α obtidos por meio do DFA-1D nas séries apresentadas na figura 2.1. Percebe-se, portanto, que os valores e interpretações contidos na tabela 2.2 são, na verdade, para o Hurst esperado (ou aferido), $\hat{H}$ e não para o α diretamente.

A segunda linha da figura 2.1, onde H=0.5 representa, portanto, o ruído branco (fGn) e o *random walk* (fBm), que nada mais é do que a integração de um ruído branco. Tais sinais são descorrelacionados, isto é, sem memória. A última linha da

figura 2.1, onde H=1, representa o ruído rosa (*pink noise* - $1/f$), muito importante em diversas áreas, em especial as biomédicas.

A interpretabilidade do DFA é citada por alguns autores como difícil de ser feita. Sendo assim, compreender essas nuances traz ganhos para a análise de resultados do DFA.

Figura 2.1 - Séries fGn e fBm obtidas para os mesmos expoentes de Hurst. As séries foram geradas com 65.536 amostras.

A figura 2.2 mostra ainda as séries da figura 2.1 após o passo 1 do algoritmo do DFA, que consiste em integrar e remover a média da série de entrada. Neste momento, os sinais fGn são convertidos em fBm e o fBm torna-se um fBm acumulado e o valor de α é acrescido de um para esse último caso.

Tabela 2.3 - Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm unidimensionais apresentados na figura 2.1

Expoente de Hurst	**fGn**		**fBm**	
	α	$\hat{H}$	α	$\hat{H}$
0.2	0.20373	0.20	1.1309	0.14
0.5	0.50161	0.51	1.5204	0.52
0.9	0.88061	0.89	1.8441	0.84
1	0.99092	1.00	2.0051	1.01

Figura 2.2 - Perfis das séries apresentadas na Figura 2.1, obtidos após a soma acumulada da série (Passo a do algoritmo, equação (2.3).

2.4.2 Vantagens e Desvantagens

A principal vantagem do método DFA reside no fato de ele conseguir detectar e eliminar tendências de diferentes ordens e com correlações polinomiais. Kantelhardt et al. (2001) afirmam que é fundamental distinguir as tendências das flutuações de longo alcance intrínsecas nos dados. Segundo esses autores, tendências são causadas por fatores externos, como, por exemplo, o efeito estufa e as variações sazonais da temperatura e apresentam comportamento suave, monotônico e lentamente oscilante. Tendências fortes podem conduzir a uma falsa detecção de correlação de longo alcance. Nesse caso, o DFA (e demais métodos destendenciados) tira vantagem por eliminar sistematicamente as tendências de diferentes ordens. Ainda segundo Alvarez-ramirez et al. (2006), reduzindo os efeitos das tendências no dado, o DFA possibilita uma estimação mais precisa do expoente de escala e posições de '*crossover points*', os quais evidenciam uma mudança no regime da série analisada.

Outra vantagem é que enquanto a maioria dos métodos analisa a série temporal diretamente, o DFA, assim como as demais técnicas baseadas em *random walk*, integra a série a priori. A integração reduz o nível de ruído oriundo das medições imperfeitas, resultando em uma função de flutuação mais estável que àquela obtida a partir da série sem esse tratamento (ALVAREZ-RAMIREZ et al., 2006; BASHAN et al., 2008; KANTELHARDT et al., 2001).

Peng et al. (1994) resumem essas vantagens, afirmando que o DFA permite diferenciar duas propriedades que são difíceis de serem obtidas com outros métodos : o valor de α, que mensura a força da correlação; e a indicação dos pontos de *crossover*, que indica diferentes regimes que a série pode ter.

Finalmente, segundo Gu e Zhou (2006), a simplicidade na implementação do DFA fez com que esse método tenha se tornado um dos mais importantes na investigação das propriedades fractais e multifractais de séries temporais.

Uma das primeiras desvantagens apontadas no uso do DFA é sua dificuldade de interpretação. Além disso, o DFA não é dito específico, ou seja, aponta falso-positivo. Um algoritmo é dito sensível quando detecta a correlação de longo alcance quando ela realmente existe. Por outro lado, o algoritmo é dito específico, quando rejeita a correlação quando ela não existe. Um algoritmo ótimo é sensível e específico, quando não revela falso-positivo e nem falso-negativo. Maraun et al. (2004) estudaram sistematicamente o DFA e concluíram que o algoritmo não é específico.

Chen et al. (2002) estudaram a resposta do DFA à três tipos de não-estacionaridades inseridas em séries temporais estacionárias e correlacionadas. Segundo esses autores, as tendências inseridas levaram ao aparecimento de pontos de *crossover* no comportamento de escala dos dados. Dessa forma, os autores relatam ser aconselhável realizar pré-processamentos nos dados brutos antes da análise, de forma a minimizar os efeitos da não-estacionaridades intrínsecas dos dados.

Höll e Kantz (2015a) e Höll e Kantz (2015b) afirmam que, apesar da sua popularidade, não existem provas matemáticas da corretude do DFA, nem para processos de memória curta, nem memória longa. Segundo os autores, a correlação entre a função de flutuação e a função de autocorrelação oriunda do DFA só pode ser derivada para séries temporais estacionárias. Esses trabalhos apresentam estudos analíticos que visam correlacionar a função de autocorrelação e a de flutuação.

2.4.3 Comparação com outros métodos

Heneghan e McDarby (2000) questionam investigações acerca do DFA onde o parâmetro α é dito não poder ter sido obtido de outra maneira, ou que o DFA é uma medida inerentemente superior às demais técnicas existentes. Dessa forma, ou autores investigaram o cálculo do expoente de Hurst (H) usando densidade espectral de potência (*Power spectral density*) e DFA e concluíram que ambas as medidas originam valores altamente semelhantes de H, e que, portanto, não devem ser tratadas como medidas independentes das propriedades de memória longa de um sinal estocástico.

Bryce e Sprague (2012) examinaram criticamente a estrutura algorítmica do DFA, a fim de determinar em quais condições pode-se esperar que a técnica realmente mitigue os efeitos da não-estacionaridade. Concluíram que o DFA não é tão robusto para séries tendenciadas e argumentam que o FA supera o DFA. Isso porque o DFA apresenta uma curvatura significativa no gráfico *log-log* para as janelas de tempo pequenas, enquanto o FA é linear. Segundo os autores, devido a essa curvatura espúria resultante, permanecem alguns problemas na interpretação dos resultados.

Shao et al. (2012) compararam o FA, DFA e duas variações chamadas B*ackward Detrending Moving Average* (BDMA) e *Centred Detrending Moving Average* (CDMA) para caracterização da correlação de longo-alcance em séries temporais sintéticas. O FA teve a pior performance, enquanto o DFA e o CDMA obtiveram os melhores resultados. Os autores rebatem os resultados encontrados por Bryce e Sprague (2012) acima citados, argumentando que o FA só é superior ao DFA em casos muito

específicos (H=0.3) e que, sendo assim, não se pode generalizar os resultados. Os autores concluíram ainda que o tamanho da série não afeta significativamente os resultados, exceto para séries pequenas (500 e 2000 elementos), as quais tiveram maior flutuação.

Mais recentemente, Witt e Malamud (2013) compararam diversas técnicas para quantificar persistência em séries temporais, incluindo técnicas no domínio do tempo (R/S, semivariograma e DFA), e no domínio da frequência (técnicas de densidade espectral). Por meio de comparações estatísticas, os autores concluíram que:

- o R/S não é recomendado devido aos grandes erros sistemáticos. De fato, Kavasseri e Nagarajan (2004) também concluíram que as diferentes formas de calcular o expoente de Hurst apresentam valores muito distintos entre si, o que indica potenciais dificuldades para aplicar e interpretar a técnica em dados obtidos de processos físicos.
- a análise de semivariograma apresenta grandes erros randômicos para sinais auto-afins, cuja força de persistência varie entre 1.2 e 2.8;
- o DFA é adequado para séries temporais com distribuição de probabilidade de cauda curta e cuja força de persistência seja maior que 0.
- as técnicas espectrais obtiveram melhor performance entre as avaliadas.

A próxima seção apresenta a generalização do DFA unidimensional para operar em dados bi-dimensionais. A versão apresentada e utilizada neste trabalho foi definida por Gu e Zhou (2006).

2.5 Análise de Flutuação Destendenciada bi-dimensional (DFA-2D)

Gu e Zhou (2006) adaptaram o DFA e o DFA Multifractal (MDFA) para operar sobre duas ou mais dimensões. Neste trabalho de referência, os autores testaram o DFA-2D com superfícies Brownianas fracionadas (fBm) e imagens reais, sendo uma da superfície de Marte e uma típica imagem microscópica digitalizada (SEM) de dióxido de carbono. Os autores concluíram que o DFA-2D é capaz de capturar a natureza auto-similar de superfícies Brownianas fracionadas e que estima com precisão o índice de Hurst.

O DFA-2D é uma adaptação do método DFA para séries temporais unidimensionais (2.4.1) e pode ser acompanhado no algoritmo 1 (GU; ZHOU, 2006). Basicamente, para

cada escala s analisada, a imagem é dividida em sub-imagens, que são integradas e destendenciadas. A variância do resíduo obtido pela diferença entre a superfície integrada e a ajustada pelo plano é calculada como a função de flutuação local. A média das variâncias locais é computada com a função de flutuação global para cada escala. No final do processo, o expoente de escala α é calculado como a inclinação da reta no gráfico *log-log*.

Assim como na versão unidimensional, o destendenciamento da superfície pode ser feito em diferentes ordens. Gu e Zhou (2006) sugerem cinco equações diferentes, a saber:

$$\tilde{y}_{v,w}(i,j) = ai + bj + c \tag{2.8}$$

$$\tilde{y}_{v,w}(i,j) = ai^2 + bj^2 + c \tag{2.9}$$

$$\tilde{y}_{v,w}(i,j) = aij + bi + cj + d \tag{2.10}$$

$$\tilde{y}_{v,w}(i,j) = ai^2 + bj^2 + ci + dj + e \tag{2.11}$$

$$\tilde{y}_{v,w}(i,j) = ai^2 + bj^2 + cij + di + ej + f \tag{2.12}$$

Onde $1 \leq i, j \leq s$ e a, b, c, d, e e f são parâmetros livres estimados.

O expoente α é considerado uma generalização do expoente de Hurst e, consequentemente, sua interpretação também. Zhou et al. (2013) discutem a relação entre os expoentes α e H para o DFA-2D e concluem que se o sinal é estacionário, então $H \approx \alpha$, como na versão unidimensional. Já para superfícies não estacionários, $H = \alpha - 2$.

Dessa forma, para o DFA-2D, sinais estacionários (fGn) retornam valores de α entre 0 e 1, enquanto sinais não estacionários (fBm), retornam valores entre 2 e 3. Valores de α entre 1 e 2 indicam que o sinal não é auto-afim. O entendimento do significado do valor do parâmetro α ajuda a compreender não apenas a força da correlação do sinal analisado, mas também a origem do mesmo.

Algorithm 1 Algoritmo DFA-2D

procedure DFA-2D(M) ▷ M é uma superfície auto-similar de tamanho $(R \times C)$

  for $(s = 6; s \leq min(R, C)/4)$ **do**

    Divida a superfície M em sub-matrizes $M_s(k)$ de tamanho $(s \times s)$, onde $1 \leq k \leq (ml * mc)$ ▷ $ml = \left\lfloor \frac{R}{s} \right\rfloor$ e $mc = \left\lfloor \frac{C}{s} \right\rfloor$

    for cada sub-matriz $M_s(k)$ **do**

      Obtenha o perfil $Y_s(k)$ a partir da integração da superfície $M_s(k)$: $Y_s(k) = \sum_{k_1=1}^{i} \sum_{k_2=1}^{j} M_s(k)(k_1, k_2)$

      Obtenha a matriz destendenciada $z_s(k)$ como a diferença entre $Y_s(k)$ e a superfície $\tilde{Y}_s(k)$ ajustada por um plano (mínimos quadrados).

      Compute a função de flutuação local: $F_s^2(k) = \frac{1}{s^2} \sum_{i=1}^{s} \sum_{j=1}^{s} (z_s(k))^2$

    end for

    Compute a função de flutuação global para s : $F_2(s) = \left(\frac{1}{ml*mc} \sum_{i=1}^{ml} \sum_{j=1}^{mc} F_s^2(k) \right)^{\frac{1}{2}}$

  end for

  Calcule o expoente α obtido a partir da relação entre $F(s) \sim s^{\alpha}$, como a inclinação da reta no gráfico log-log $F_2(s)$ and s.

  retorne α

end procedure

2.5.1 DFA-2D Multifractal

O formalismo multifractal (OLSEN, 1995) busca avaliar estatisticamente propriedades de escala de medidas singulares que surgem em várias situações físicas. Ele é considerado uma extensão de ordem mais alta da análise monofractal (CZARNECKI; GRECH, 2010; MUZY et al., 1993).

A principal diferença no método multifractal é que, ao invés de analisar apenas um expoente de escala, a função de flutuação é avaliada sob diversas ordens. Czarnecki e Grech (2010) explicam que a relação $F(s) \approx s^H$ pressupõe que o sistema apresenta a mesma propriedade de escala (regime) em todas as escalas. Assim, H é constante. No entanto, em alguns sistemas, as propriedades fractais diferem para diferentes conjuntos de escala. Nesse caso, a estrutura fractal é descrita por múltiplas regras de escala.

É importante ressaltar a diferença entre *crossover point* e multifractalidade. Como

já mencionado, os *crossovers* revelam diferentes expoentes de escala para diferentes escalas (regimes). Já a característica multifractal revela que, para escalas diferentes, o processo possui diferentes leis de potência. Numa visão bi-dimensional, um ponto de *crossover* indica que, para diferentes escalas espaciais, há diferentes leis de escala. Por exemplo, as escalas espaciais de 6x6 a 20x20, são regidas por um expoente de Hurst 0.3. Acima da escala 20x20, o sistema é regido pelo expoente de Hurst 0.5. Já a multifractalidade indica que há diferentes leis de escala para uma mesma ou diversas escalas espaciais. Nesse caso, a análise é sobre o conjunto de valores possíveis de H.

Na presença de multifractalidade, Czarnecki e Grech (2010) afirmam que o DFA padrão (monofractal) revela a regra de escala principal (aquela que possui as maiores ou mais frequente flutuações). No entanto, devido às características do sinal, esse valor pode ser distorcido e, nesse caso, é necessário utilizar técnicas que permitam amplificar flutuações, como o DFA-Multifractal, definido por (KANTELHARDT et al., 2002).

Makowiec et al. (2011) resumem o formalismo multifractal. Flutuações são estimadas em diferentes escalas (n) e quantificadas por momentos q, formando a chamada função de partição $P(q, n)$. A propriedade de escala da função de partição, isto é, $P(q, n) \approx n^{\tau(q)}$ determina o expoente da função $\tau(q)$, o qual leva, via transformada de Legendre, ao espectro multifractal $(h, D(h))$. Isso é exatamente o que o DFA multifractal faz.

Algorithm 2 Algoritmo DFA-2D Multifractal

1: **procedure** DFA-2D(M) ▷ M é uma superfície auto-similar de tamanho $(R \times C)$

2: **for** $(s = 6; s \leq min(R, C)/4)$ **do**

3: Divida a superfície M em sub-matrizes $M_s(k)$ de tamanho $(s \times s)$, onde $1 \leq k \leq (ml * mc)$ ▷ $ml = \left\lfloor \frac{R}{s} \right\rfloor$ e $mc = \left\lfloor \frac{C}{s} \right\rfloor$

4: **for** cada sub-matriz $M_s(k)$ **do**

5: $y_{v,w}(i,j) \leftarrow \sum_{k_1=1}^{i} \sum_{k_2=1}^{j} M_s(k)(k_1, k_2)$

6: $z_s(i) \leftarrow Y_s(k) - \tilde{Y}_s(k)$

7: $F_2(s) \leftarrow \left(\frac{1}{ml*mc} \sum_{i=1}^{ml} \sum_{j=1}^{mc} F_s^2(k) \right)^{\frac{1}{2}}$

8: **end for**

9: Compute a função de flutuação global para $F_q(s)$ para as diferentes ordens q como :

10: **for** $(q = qMin; q = qMax)$ **do**

11: **if** $q \neq 0$ **then**

12: $F^2_{s(q)}(i) = \left(\frac{1}{(ml*mc)^2} \sum_{i=1}^{ml} \sum_{j=1}^{mc} \left(z_s(k)^{\frac{q}{2}} \right) \right)^{\frac{1}{q}}$

13: **else**

14: $F^2_{s(0)}(i) \equiv \frac{1}{(ml*mc)^2} \sum_{i=1}^{s} \sum_{j=1}^{s} \ln \left(z_s(k) \right)^{\frac{q}{2}}$

15: **end if**

16: **end for**

17: **end for**

18: Calcule o expoente $h(q)$ obtido a partir da relação entre $F(s) \sim s^{h(q)}$, como a inclinação da reta no gráfico log-log entre $F_q(s)$ and s.

19: **Retorne** $h(q)$ para cada q

20: **end procedure**

O algoritmo do DFA-2D multifractal (algoritmo 2) apresenta os mesmos passos ilustrados no algoritmo 1 entre as linhas 1 e 11. A diferença do DFA multifractal é que diversos momentos, indexados pela variável q, são analisados na função de flutuação global e, consequentemente, diversos expoentes de escala ($h(q)$) são computados. Um para cada momento analisado. A variável q varia de um número real negativo a um número real positivo. Quando $q = 2$, $h(q) \equiv \alpha$. Para sinais monofractais, $h(q)$ independe de q e a função $h(q)$ é representada por uma reta. Caso contrário, o sinal apresenta característica multifractal.

Outra forma de caracterizar a multifractalidade do dado é por meio do expoente de massa $\tau(q)$, obtido pela fórmula da equação 2.13, onde D_f é a dimensão fractal do dado analisado (GU; ZHOU, 2006; KANTELHARDT et al., 2002; WANG et al., 2015a; WANG et al., 2014):

$$\tau(q) = q * h(q) - D_f \tag{2.13}$$

Em geral, para séries unidimensionais, $D_f = 1$ e, para superfícies bi-dimensionais, $D_f = 2$.

Outro conjunto de variáveis importante é o $\alpha - f(\alpha)$, definido via transformada de Legendre como:

$$\alpha = h(q) + qh'(q) \tag{2.14}$$

$$f(\alpha) = q[\alpha - h(q)] + 1 \tag{2.15}$$

onde α é o expoente de Lipschitz-Hölder, que caracteriza as singularidades de uma série temporal. O espectro de singularidade $f(\alpha)$ descreve o conteúdo de singularidades do dado analisado. A intensidade da multifractalidade da série (ω) pode ser caracterizada pela diferença entre os valores máximo e mínimo de α, de forma que $\omega = \alpha_{max} - \alpha_{min}$.

Portanto, a caracterização do padrão multifractal depende de dois conjuntos de parâmetros : o global, composto pela dimensão de correlação generalizada $D(q)$ e o expoente de massa $\tau(q)$; e o local, composto pelo expoente de singularidade α e a dimensão fractal $f(\alpha)$. E, nesse caso, a transformada de Legendre pode ser usada para conectar e transformar parâmetros globais com locais e vice versa.

A figura 2.3 apresenta um exemplo de superfície monofractal gerada com expoente de Hurst=0.2 (a) e multifractal gerada com o modelo p-model(b), e o resultado do processamento das mesmas com o MF-DFA2D, por meio dos gráficos $q \times h(q)$ e $q \times \tau(q)$. Percebe-se pelo gráfico $q \times h(q)$ da superfície multifractal que $h(q)$ assume uma diversidade de valores ($1.4438 \leq h(q) \leq 2.2396$), numa função monotônica decrescente. Já para a superfície monofractal, há pouca variação de $h(q)$ ($1.9487 \leq h(q) \leq 2.1504$). No gráfico $q \times \tau(q)$, a superfície monofractal gera uma reta ascendente, enquanto a multifractal apresenta um desvio para valores positivos de q.

A figura 2.4 mostra o espectro de singularidades para a superfície fractal e a multifractal. Percebe-se que para a superfície multifractal, o espectro é amplo ($\omega = 1.0629$), envolvendo diversos expoentes de singularidades (α). Já a superfí-

Figura 2.3 - Exemplo de superfície monofractal (a) e multifractal(b) e os gráficos $q \times h(q)$ e $q \times \tau(q)$ comparativos.

cie monofractal apresenta um espectro curto ($\omega = 0.3941$).

Conforme explicam Makowiec et al. (2011), o espectro multifractal resume as informações mais importantes do sinal:

- h_{max}
  Singularidade que aponta no máximo do espectro.
  Refere-se à singularidade mais provável
- Expoente de Hurst (h)
  Expoente de auto-similaridade : $\tau(2)$ ou $H(2)$
- Eventos Raros (h_l)
  Obtido quando $q \to \infty$.
  Refere-se ao lado esquerdo do espectro de singularidades
- Parte Suave (h_r)
  Obtido quando $q \to -\infty$.
  Refere-se ao lado direito do espectro de singularidades
- Amplitude do espectro de singularidades (ω)

Esses conceitos são ilustrados na figura 2.4, que apresenta os espectros de singularidade das superfícies mono e multifractal da figura 2.3.

Figura 2.4 - Espectro de singularidades de uma superfície multifractal e uma fractal (h=0.2). Os conceitos definidos por Makowiec et al. (2011) são representados na superfície multifractal.

Na figura 2.4 é possível verificar ainda uma 'torção' no espectro de singularidade da superfície monofractal. Essa torção deve-se a desvios da monofractalidade, visiveis no gráfico $q \times h(q)$ da figura 2.3, que tornam a função $h(q)$ não-monotônica. Ao aplicar a transformada de Legendre em funções não-monotônicas, obtêm-se essa forma incomum no espectro de singularidades. O efeito não-monotônico pode acontecer quando se lida com números muito pequenos e/ou o sinal apresenta não-estacionaridades. Nesses casos, um estudo mais detalhado torna-se necessário (CZARNECKI; GRECH, 2010; MAKOWIEC et al., 2011).

2.5.2 Complexidade Computacional

A complexidade computacional pode ser definida como uma medida de custo para execução de um algoritmo em função de uma dada instância. Ela pode ser medida em função do tempo de processamento ou do espaço utilizado. No caso do DFA-2D, a medida que interessa é aquela relativa ao tempo de processamento.

Nesta análise, será considerada uma matriz quadrada e um polinômio de grau 1 ($\tilde{y}_{v,w}(i,j) = ai + bj + c$). Além disso, consideraremos que, mesmo que a escala s não seja múltipla de m, o algoritmo não irá recomeçar pelos demais cantos.

Como discutido para o DFA-1D por Fornari (2016), o que influencia o tempo de processamento do algoritmo é a quantidade de escalas. No caso bi-dimensional, para cada escala s, são processadas $\left\lfloor \frac{m}{s} \right\rfloor^2$ sub-matrizes. Para cada sub-matriz calcula-se os seguintes custos:

- sua soma acumulada $\rightarrow s^2$
- o destendenciamento por mínimos quadrados $\rightarrow s^3$[5]
- cálculo da função de flutuação local $\rightarrow s^2$

E ainda, para cada escala, calcula-se a função de flutuação global, ao custo de $\left\lfloor \frac{m}{s} \right\rfloor^2$.

Considerando que s varia entre 6 e $m/4$, então uma função aproximada da complexidade do DFA-2D seria:

$$\sum_{s=6}^{\frac{m}{4}} \left(\left\lfloor \frac{m}{s} \right\rfloor^2 * (2s^2 + s^3) \right) + \sum_{s=6}^{\frac{m}{4}} \left\lfloor \frac{m}{s} \right\rfloor^2 \tag{2.16}$$

Escolhendo $s = m$, teríamos que realizar todos os passos do algoritmo para a matriz inteira $(m \times m)$, acarretando complexidade assintótica cúbica $\theta(s^3)$. A relação $s = m/4$, também acarreta em um custo cúbico.

O DFA-2D multifractal tem a mesma complexidade assintótica que o DFA-2D. No entanto, o custo para o cálculo da função global deve ser multiplicado pela quantidade de momentos analisados q :

$$\sum_{s=6}^{\frac{m}{4}} \left(\left\lfloor \frac{m}{s} \right\rfloor^2 * (2s^2 + s^3) \right) + q * \sum_{s=6}^{\frac{m}{4}} \left\lfloor \frac{m}{s} \right\rfloor^2 \tag{2.17}$$

2.6 DFA e Textura em Imagens Digitais

Segundo Alvarez-ramirez et al. (2006), como o DFA é comumente utilizado para detectar correlações de longo alcance em séries não estacionárias, ele é uma técnica promissora na análise de rugosidade em imagens, já que, de certa forma, esse atributo está relacionado às propriedades de correlação e à organização dos objetos dentro da imagem. De fato, Peng et al. (1995) afirmam que o parâmetro α pode ser um indicador de rugosidade da série temporal, sendo que, quanto maior o valor de α, mais suave é a série temporal. Afirmação que pode ser comprovada pelas séries apresentadas na figura 2.1.

[5] Considerando o uso de fatoração LU para o cálculo.

Especialmente no caso do DFA, existem duas abordagens na literatura: aquelas em que o DFA é calculado em diferentes linhas e colunas da imagem e, posteriormente, calcula-se uma média; e a abordagem onde o DFA-1D foi de fato generalizado para trabalhar com mais de 2 dimensões.

Como exemplos da primeira abordagem, os trabalhos de Alvarez-ramirez et al. (2006), EINSTEIN et al. (2012) e Freitas et al. (2009) adaptaram o uso do DFA-1D em matrizes 2D para diferentes análises em imagens digitais.

Alvarez-ramirez et al. (2006) aplicaram o DFA para estudar características de rugosidade de imagens naturais. Os autores selecionaram linhas e colunas em diversas direções de quatro diferentes imagens (grama, emulsão microscópica de água-óleo-água, superfície de Marte e granulação solar) e concluíram que o DFA pode ser usado para caracterizar a textura nessas diferentes imagens. Os autores discutem o fato de que um único expoente de escala não é suficiente para representar a rugosidade de uma imagem, já que orientações diferentes da imagem podem apresentar diferentes comportamentos em termos texturais.

Freitas et al. (2009) utilizaram o DFA para avaliar a anisotropia de imagens de radar em plataformas de petróleo localizadas na região da bacia de Campos, RJ. Mantendo a plataforma de petróleo fixa no centro, a matriz 2D foi transformada em segmentos radiais. Posteriormente, calculou-se o DFA-1D de cada segmento e, ao final, os valores de α foram visualizados em coordenadas polares, para diferentes raios e direções.

EINSTEIN et al. (2012) aplicaram o DFA em imagens de células cancerígenas (câncer de mama) benignas e malignas, com o intuito de melhorar os modelos de classificação da doença. O DFA foi utilizado para caracterizar a rugosidade nos núcleos das células de 41 pacientes, dos quais 22 possuíam tumores malignos e 19 benignos. Núcleos de células benignas apresentam maior variabilidade espacial (textura) do que as malignas. Os autores aplicaram o DFA a cada linha e coluna da imagem e, posteriormente, calcularam a média dos expoentes de escala encontrados (α_{avg}). A diferença de α_{avg} entre as células malignas e benignas não foi estatisticamente suficiente para caracterizar a doença. Os autores usaram então o desvio padrão e concluíram que nas células malignas existe menos variabilidade na textura.

No caso da segunda abordagem, diversas pesquisas científicas têm sido feitas na direção de generalizar as técnicas de análise de flutuação para dados n-dimensionais (ALPATOV et al., 2013; GU; ZHOU, 2006; YEH et al., 2012). A mais comum delas é a

elaborada por Gu e Zhou (2006), que foi apresentada seção 2.5.

Os trabalhos apresentados abaixo utilizaram o DFA-2D tal qual proposto por Gu e Zhou (2006) em aplicações com imagens para diferentes fins. Percebe-se que muitos estudos utilizam o DFA-2D, como já dito, como um parâmetro discriminatório de diferentes estágios no tempo de algum processo. No entanto, o α está sempre diretamente relacionado com a textura ou rugosidade da superfície analisada e apresenta resultados satisfatórios quando utilizado dessa forma.

Barrera et al. (2010) aplicaram o DFA-2D e o MDFA-2D em imagens de superfícies de revestimento preto de molibdênio submetidas a diferentes tempos de exposição à substâncias químicas. O objetivo dos autores era estudar as características fractais do revestimento e relacioná-las às propriedades óticas dos mesmos.

Em trabalho semelhante, Velazquez-Camilo et al. (2010) utilizaram a análise de fractal, por meio do DFA-2D, para estudar padrões em imagens de cristalização de chorume, como na cana-de-açúcar, submetidas a diferentes tempos de cristalização. Os autores tiveram melhores resultados com imagens binarizadas do que com imagens em nível de cinza.

Yadav et al. (2012) aplicaram o DFA-2D multifractal na caracterização da superfície morfológica de películas com diferentes espessuras. Os autores investigavam como a geometria dessas superfícies afeta as propriedades físicas do sistema em questão e concluíram que o expoente de rugosidade diminui com a espessura, além de que tais superfícies apresentam comportamento multifractal.

Freitas (2012) utilizou o DFA-2D para detectar mudanças de uso e cobertura da terra em séries espaço-temporais. Foram utilizadas imagens Modis(EVI) de tamanho 128x128, entre os anos 2000 e 2010, para caracterizar áreas com grandes mudanças na vegetação. O autor afirma que o DFA-2D é capaz de indicar áreas que a vegetação possui comportamento espaço-temporal similar e que o desafio passa pela correta interpretação do parâmetro α e sua correlação com as demais variáveis.

Vargas-Olmos et al. (2015) utilizaram o DFA-2D na determinação de propriedades de escala em imagens criptografadas. Foram utilizadas 13 imagens de 512x512 pixels e 5 imagens de 1024x1024 pixels. Os autores concluíram que imagens criptografadas apresentam uma propriedade de escala característica, possibilitando a identificação desse tipo de imagem; e sugerem que o expoente de escala α pode ser apropriado para medir a qualidade de criptografia das imagens.

Wang et al. (2015b) aplicaram o MFDFA-2D para detectar espécies de plantas a partir de imagens. Segundo os autores, o método foi eficiente para o objetivo proposto.

Nie et al. (2015) investigaram padrões urbanos de superfícies impermeáveis em Shanghai, China, utilizando imagens de satélite. Para tanto, utilizaram índice de Moran para medir a correlação espacial das áreas, DFA-2D, MFDFA-2D e quatro imagens do satélite Landsat, nos anos de 1997, 2002, 2007 e 2010. Os autores concluíram que as superfícies impermeáveis apresentam comportamento de escala monofractal, forte correlação de longo alcance e alta correlação espacial.

Os trabalhos acima citados confirmam que o DFA-2D e o MFDFA-2D são bons aferidores de textura em diferentes tipos de imagens digitais, com diferentes resoluções e formas de obtenção.

3 CONJUNTOS DE DADOS ANALISADOS

Este capítulo apresenta os dados utilizados neste trabalho, os quais foram separados em três grupos : sinais bi-dimensionais típicos, imagens oriundas da ciência espacial e imagens oriundas da ciência ambiental. Os sinais típicos foram incluídos com dois objetivos principais: provar a corretude do método implementado e verificar o comportamento dos mesmos no espaço 2D. As imagens oriundas das ciências espacial e ambiental são aplicações reais que serão estudadas com o DFA-2D e MFDFA-2D. No caso da ciência espacial, os dados são imagens digitais de galáxias elípticas e espirais. Os dados ambientais são modelos digitais de elevação topográfica (MDE). A ideia foi verificar se o DFA-2D é capaz de caracterizar e classificar corretamente esses sinais por meio de suas respostas fractais e, consequentemente, sua rugosidade.

3.1 Sinais bi-dimensionais típicos

Nessa sessão estão incluídos os dados fractais fBm e fGn, os ruídos do tipo $1/f$ e uma amostra de dados caóticos, representados por redes de mapas acoplados (CML).

3.1.1 Fractais

Há dois tipos de sinais fractais : Ruído Gaussiano Fracionário (fGn) e o Movimento Browniano Fracionário (fBm). O movimento browniano fracionário foi definido por Mandelbrot e Van Ness (1968) e é usado para modelar séries temporais fractais não estacionárias, sendo uma generalização do movimento Browniano (ordinário).

Movimento Browniano Ordinário é o nome de um processo estocástico contínuo com incrementos Gaussianos independentes. Possui média igual a zero e auto-similaridade estatística dos incrementos.

A definição clássica de fBm por Mandelbrot é dada como : Dado $B(t)$ como um movimento Browniano, e H um parâmetro que satisfaça a condição $0 < H < 1$, fBm com expoente H é uma média móvel dos incrementos de $B(t)$ (MANDELBROT; Van Ness, 1968). Dessa forma, a propriedade de independência dos incrementos que faz do movimento browniano tão útil para modelar processos, não é válida para o fBm. Matematicamente, o fBm é caracterizado pela seguinte lei de potência (DELIGNIERES et al., 2006):

$$\left\langle \Delta x^2 \right\rangle \propto \Delta t^2 H \tag{3.1}$$

ou seja, o deslocamento quadrático esperado (Δx) é uma função de potência do

intervalo de tempo observado (Δt). O fBm ainda apresenta as seguintes propriedades (LUNDAHL et al., 1986):

- Os incrementos têm média zero
- Os incrementos são estacionários
- Os incrementos são autossimilares e, por esse motivo, o fBm é classificado como um processo auto-afim.
- A variância (var) entre os incrementos é dada por : $var[B(t_2; H) - B(t_1; H] = V_H(t_2 - t_1)^2 H$, onde V_H é uma constante que depende de H.

Quando $H = 0.5$, o fBm é um movimento Browniano ordinário (sem correlação). Apesar de tornar a análise dos dados mais difícil, Mandelbrot e Van Ness (1968) lembram que estudos empíricos normalmente sugerem uma forte interdependência entre amostras distantes, tornando imprescindível o uso de fBm em tais modelagens.

O ruído gaussiano fracionário pode ser obtido por diferenças sucessivas entre elementos da série fBm. Neste caso, o sinal é estacionário e caracterizado por sua variância e expoente de Hurst. Sendo assim, os modelos apresentam características bastante específicas. fBm é não estacionário com variância dependente do tempo. Já o fGn é estacionário, com média e variância constantes. fBm e fGn são interconvertíveis, ou seja (DELIGNIERES et al., 2006; EKE et al., 2000):

$$fBm(t) = \sum_{i=1}^{t} fGn(x_i) \tag{3.2}$$

$$fGn(t) = \lim_{\xi \to 0} \frac{fBm(t+\xi) - fBm(t)}{\xi} \tag{3.3}$$

Cada fBm tem seu respectivo fGn, ambos caracterizados pelo mesmo expoente de autossimilaridade H. O fBm pode ser expandido para outras dimensões. Em especial, no caso 2D, e, idêntico ao fBm unidimensional, o 2D-fBm é definido como um processo estocástico randômico, não estacionário, com propriedades autossimilares. A função de correlação e densidade espectral de potência seguem leis de potência e a natureza fractal é caracterizada pelo expoente de Hurst (FORTIN et al., 1992; HENEGHAN et al., 1996; KAPLAN; KUO, 1996). Desta forma, as propriedades do 2D-fBm $B_H(t_x, t_y)$ podem ser sumarizadas como (HENEGHAN et al., 1996; KAPLAN; KUO, 1996; MCGAUGHEY; AITKEN, 2002):

- $B_H(0,0) = 0$

- Os incrementos $B_H(t_x+r_x, t_y+r_y)-B_H(t_x, t_y)$ possuem distribuição Gaussiana com média zero, são estacionários e isotrópicos
- A variância dos incrementos $B_H(t_x+r_x, t_y+r_y)-B_H(t_x, t_y)$ depende apenas da distância entre os pontos x e y

Computacionalmente, sinais 2D-fBm podem ser gerados a partir de diversos métodos, como wavelet (HENEGHAN et al., 1996) , transformadas de Fourier incrementais (KAPLAN; KUO, 1996) ou à partir de processos gaussianos (MCGAUGHEY; AITKEN, 2002).

O modelo 2D-fGn também pode ser computacionalmente gerado por diferentes métodos e, como os modelos são reversíveis, é possível obter o fGn a partir da derivada parcial de segunda ordem mista de um fBm(B_H) da seguinte forma (MCGAUGHEY; AITKEN, 2002):

$$\frac{\partial^2}{\partial x \partial y} B_H(x,y) = [B_H(x,y) - B_H(x-1,y)] - [B_H(x,y-1) - B_H(x-1,y-1)] \quad (3.4)$$

onde $B_H(x,-1) = 0$ e $B_H(-1,y) = 0$[1]

Em Qian et al. (1998) há uma rica discussão matemática sobre o modelo 2D-fBm e Penttinen e Virtamo (2004) apresentam e discutem o modelo 2D-fGn.

A figura 3.1 apresenta uma amostra dos dados fBm e fGn bi-dimensionais para diferentes expoentes de Hurst.

Em uma visão tri-dimensional (figura 3.2), é possível perceber a noção de rugosidade da superfície, que é um dos objetos de estudo principais dessa tese. Em termos de textura, a figura 3.2 também ilustra o já mencionado acima, que quanto menor o expoente de Hurst, maior a rugosidade da imagem. Verifica-se também que as superfícies tendem a aumentar sua dependência espacial à medida que o expoente de Hurst aumenta.

Com o intuito de seguir o padrão de testes definido por Gu e Zhou (2006), as simulações de processo fBm foram gerados em Matlab com o *software* Fraclab (INRIA, 2016). Foram geradas matrizes de tamanho 500x500 e com expoentes de Hurst variando de 0.05 até 1, com intervalo de 0.05. Para cada valor de expoente de Hurst foram geradas quinhentas simulações diferentes de fractal.

[1] Considerando que os índices da matriz em questão variem de 0 a (n-1).

Figura 3.1 - Amostras de simulações de processos fBm (topo) e fGn (baixo) para diferentes expoentes de Hurst.

Figura 3.2 - Visão tridimensional das superfícies fBm com diferentes expoentes de Hurst, ilustrando a diferença entre a rugosidade das mesmas. a) H=0.1. b) H=0.5. c) H=0.8.

Os dados fGn foram gerados a partir do fBm utilizando a equação 3.4.

3.1.2 Ruídos $\frac{1}{f}$

Voss (1992) afirma que diversos sistemas físicos apresentam correlações do tipo lei de potência no espaço (fractal) ou no tempo ($1/f^{\gamma}$ noise). Gilden et al. (1995) definem $1/f$ *noise* como uma forma de flutuação temporal que tem intensidade de potência inversamente proporcional a frequência. Os ruídos do tipo $1/f$ variam sua força de correlação entre o ruído branco ($1/f^0$) e o ruído vermelho [2] ($1/f^2$). Entre esses dois extremos há diversos outros ruídos, incluindo o intermediário ruído rosa ($1/f^1$), que contém influências iguais de todos os intervalos de escala de tempo em todo o espectro (HALLEY; KUNIN, 1999).

[2] Também chamado ruído marrom, ou browniano

Ao contrário do ruído branco que não tem memória, os processos processos regidos pelo ruído vermelho são dominados por sua história recente. Ambos são não estacionários. O ruído rosa compartilha propriedades dos ruídos branco e vermelho, é considerado quase-estacionário e, por essas características intermediárias, é o modelo mais utilizado para modelar eventos reais.

A figura 3.3 apresenta uma amostra dos ruídos branco, rosa e vermelho.

Figura 3.3 - Ruídos típicos. a)Ruído branco gaussiano (γ=0). b) Ruído rosa (γ=1). c) Ruído Vermelho (γ=2).

Os dados referentes aos ruídos branco (*white gaussian noise*), rosa e vermelho utilizados nesse trabalho foram gerados com código em Matlab, distribuído por Yearsley (2016), que é baseado na metodologia descrita por Lennon (2000), utilizando a transformada inversa de Fourier. Para cada tipo de ruído foram realizadas quinhentas simulações, com matrizes de tamanho 500×500.

3.1.3 Redes de Mapas Acoplados - CML

Os mapas acoplados (do inglês, *Coupled Map Lattice* - CML) são sistemas dinâmicos estendidos que representam, de forma simplificada, o caos espaço-temporal. Kaneko (1992) define um CML como um sistema dinâmico com tempo e espaço discretos e uma variável de estado contínua. Apesar de simples, os mapas acoplados podem exibir diferentes comportamentos, variando desde uma estrutura espaço-temporal periódica à intermitência e ao caos. Estudos com redes deste tipo não se limitam ao campo de caos espaço-temporal, e têm ganhado relevância em áreas como biologia, matemática, engenharia e física.

Grigoriev (1999) explica algumas razões para esse crescimento no uso de CML. Segundo o autor, devido a estrutura espaço-temporal explícita, o comportamento do CML é mais rica do que o dos sistemas caóticos unidimensionais. Além disso, os

CMLs são altamente simétricos, o que é uma propriedade importante nas aplicações práticas.

No modelo CML, os mapas são acoplados formando uma malha uniforme em que cada elemento conecta-se a seus elementos vizinhos. Nesse processo dois atributos são importantes : a função de mapeamento (*map*) e a função de acoplamento.

Neste trabalho foram utilizados três modelos de mapas : logístico, doubling e SOM.

O mapa logístico é uma equação de evolução não linear que atua continuamente sobre uma variável x ($0 \leq x \leq 1$). A evolução é definida no tempo discreto por interações em um mapeamento polinomial de segundo grau, dado por:

$$x_{n+1} = r * x_n(1 - x_n) \tag{3.5}$$

onde o parâmetro r ($0 \leq r \leq 4$) direciona a evolução temporal e define o comportamento da série. $r = 4$ é o valor comumente utilizado para se obter comportamento caótico.

O mapa doubling também é conhecido como Dyadic Bernoulli e é o mapa caótico mais simples, cuja regra de evolução temporal segue a equação polinomial de primeiro grau, dado por (HASEGAWA; SAPHIR, 1992):

$$x_{n+1} = 2x_n mod 1 \tag{3.6}$$

O mapa SOM (abreviação de 'So-Ose-Mori') é um mapa linear-por-partes, cuja regra de evolução temporal é dada por (CAVALCANTE, 2003; MIYAZAKI; HIGUCHI, 2013):

$$x(n+1) = \begin{cases} ax(n) + 0.2 & (0 \leq x \leq \gamma) \\ \alpha^{-1}(x(n) - 0.8) + 1 & (\gamma \leq x \leq 0.8) \\ -\beta^{-1}(x(n) - 1) & (0.8 \leq x \leq 1) \end{cases} \tag{3.7}$$

Os valores comuns para α, γ e β são 0.6, 0.2 e $(0.8/1 + \alpha)$, respectivamente. Neste caso, segundo Cavalcante (2003), o modelo apresenta um comportamento crítico que não é lei de potência, como nos mapas polinomiais, mas logarítmico.

A figura 3.4 mostra algumas iterações para os três modelos analisados.

Figura 3.4 - Amostras de evolução de séries espaço-temporais dos mapas logísticos acoplados, em diferentes instantes de tempo. Em A) mapa logístico. Em B) mapa dyadic bernoulli e em C) mapa SOM.

As séries de mapas acoplados foram geradas utilizando o código de Sautter (2017), escrito em Python. O código gera séries espaço-temporais nos modelos logístico, dyadic bernoulli e SOM, com a quantidade de iterações definida pelo usuário. Na geração dos mapas acoplados é utilizada a vizinhança de Von Neumann. A matriz inicial pode ser gaussiana ou bessel. Para cada modelo de mapa foram geradas dez séries, com 300 iterações cada e com tamanho 500x500. Para ter diversidade, o código precisou ser modificado, de forma que o desvio padrão da matriz inicial (escolhida como gaussiana) fosse randômico.

3.2 Sinais oriundos da ciência espacial

Uma galáxia pode ser definida como um aglomerado de estrelas, poeiras e gás, que aparece isolada no espaço. Nela, todas as estrelas (e demais constituintes) são mantidas juntas pela gravidade e, cada constituinte emite luz sobre diferentes comprimentos de onda e com diferentes intensidades. No Universo existem galáxias de diversas formas, tamanhos e idades. Normalmente elas são categorizadas de acordo com sua forma aparente, referida como sua morfologia visual.

Em 1936, Hubble (HUBBLE, 1936) propôs uma classificação de galáxias baseada em sua morfologia. O esquema proposto divide as galáxias em dois grandes grupos: elípticas e espirais, que são ainda classificadas em sub-grupos. Um terceiro grupo é

formado por galáxias irregulares (figura 3.5).

Figura 3.5 - Esquema proposto por Hubble para classificação das galáxias a partir de sua morfologia visual. FONTE : Adaptado de Wikimedia Commons

As galáxias espirais apresentam aglomerados de gás, bolhas e super bolhas em diferentes escalas. Já as elípticas, em alta resolução, apresentam muitos *clusters* globulares (THANKI et al., 2009). As galáxias elípticas tem forma de um elipsoide e uma superfície mais lisa. Já as espirais apresentam curvaturas na forma elipsoide e estruturas espirais em torno delas (Figura 3.6) (AU et al., 2006).

Figura 3.6 - Exemplo de galáxia elíptica (A) e espiral (B). Fonte : Galaxy Zoo.

Os dados de galáxias elípticas e espirais utilizados nesse trabalho foram extraídos do

Galaxy Zoo[3], projeto colaborativo de classificação de imagens de galáxias, incluído no conceito de ciência cidadã. O termo ciência cidadã serve para descrever, de forma geral, projetos com finalidades científicas que usam não cientistas como voluntários.

O Galaxy Zoo foi lançado em 2007 na Universidade de Oxford, Reino Unido. O objetivo do projeto era que o voluntário, ao observar uma imagem coletada pelo consórcio internacional *Sloan Digital Sky Survey*, decidisse se a imagem continha alguma galáxia e, se afirmativo, qual era a sua forma (elíptica ou espiral). Se fosse espiral, o voluntário ainda poderia informar o sentido da sua rotação.

Os resultados do Galaxy Zoo foram surpreendentes. Em três semanas, 80 mil usuários se cadastraram e realizaram mais de 10 milhões de tarefas de classificação de imagens de galáxias. Ao final de um ano havia 150 mil usuários, que classificaram mais de 50 milhões de imagens. Em média, cada galáxia foi classificada mais de 50 vezes.

Nesse trabalho os dados foram extraídos do Galaxy Zoo 1, onde as imagens são classificadas apenas em elípticas ou espirais. Do conjunto disponível, foram selecionadas as galáxias maiores, seguindo o seguinte critério:

$$\pi * R_{50} * \left(\frac{B}{A}\right) \geq 40 * \pi * FWHM \tag{3.8}$$

Onde:

- R_{50} é o raio Petrosian;
- A and B são os eixos da elipse que contém a isofota;
- $FWHM$ é a largura a meia altura da PSF

Esse critério retorna cerca de 60.000 (sessenta mil) galáxias. Dessas, mil galáxias de cada tipo (elíptica e espiral) foram selecionadas utilizando um segundo critério de maior razão de eixo : $\frac{B}{A} \approx 1$.

As isofotas passaram ainda por um pré-processamento, onde foram recortadas em tamanhos de 180x180 pixels e tiveram as galáxias centralizadas[4].

A figura 3.7 apresenta amostras dos tipos de galáxias elíticas (topo) e espirais (baixo) após o pré-processamento realizado.

[3]https://data.galaxyzoo.org/

[4]Os dados foram processados por Rubens Sautter (rubens.sautter@gmail.com) e gentilmente cedidos para esse trabalho de tese.

Figura 3.7 - Amostras de imagens de galáxias elíticas (topo) e espirais (abaixo) extraídas do Galaxy Zoo.

3.3 Sinais oriundos da ciência ambiental

Um modelo digital de elevação topográfica (MDE) é uma representação matemática contínua da distribuição espacial das variações de altitude numa área. Atualmente, com a popularização dos produtos de sensoriamento remoto, são cada vez mais comuns os modelos digitais de elevação orbitais.

Em um MDE orbital, os valores de altitude reportados não referem-se necessariamente ao nível do solo, já que os sinais são refletidos por alvos mais altos, tais como torres ou dossel das árvores. Os modelos cuja altura refere-se ao nível do solo são chamados Modelo Digital de Terreno (MDT) ou Modelo Numérico de Terreno (MNT) e podem ser obtidos, dentre outras técnicas, por levantamentos realizados com GPS ou com laser (GROHMANN, 2008).

Dois são os produtos de MDE orbitais e globais mais conhecidos e populares na atualidade : o SRTM[5] e o ASTER GDEM[6] .

O MDE SRTM - *Shuttle Radar Topography Mission* é fruto de uma missão espacial ocorrida em Fevereiro de 2000, que mapeou a Terra entre 56 °S e 60 °N utilizando interferometria por sensoriamento remoto ativo (radar) na faixa de microondas. No

[5] http://www2.jpl.nasa.gov/srtm/
[6] http://gdem.ersdac.jspacesystems.or.jp/

site Earth Explorer (USGS, 2016), que é a fonte dos dados utilizados nesse trabalho, há três produtos SRTM disponíveis para *download*:

- SRTM *Non-Void filled* : produto SRTM tratado para delinear e nivelar corpos d'água, melhor definição de linhas costeiras, remover picos e poços e preencher pequenos vazios. A resolução para os EUA é de 1 arco-segundo (aproximadamente 30 metros) e, para fora dos EUA é de 3 arco-segundos (aproximadamente 90 metros).
- SRTM *Void Filled* : áreas de dados perdidos ou vazios na coleção *non-void filled* foram preenchidas utilizando algoritmos de interpolação. A resolução para os EUA é de 1 arco-segundo e para fora dos EUA é de 3 arco-segundos.
- SRTM 1 Arc-Second : cobertura mundial da coleção *Void Filled* com a resolução de 1 arco-segundo. No entanto, algumas cenas podem conter vazios, ou falta de dados.

O Modelo Digital de Elevação Global (*Global Digital Elevation Model* - GDEM) (USGS, 2017) oferece produtos construídos a partir de pares estereoscópicos de imagens oriundas da plataforma EOS AM-1 com o instrumento ASTER (*Advanced Spacebone Thermal Emission and Reflection Radiometer*), sensor VNIR. Para obtenção dos dados de elevação da superfície o sensor opera no modo estereoscópico com as bandas 3N (nadir) e 3B (*backward*) correspondentes a faixa do infravermelho próximo (0.78 - 0.86 μm). Este produto apresenta uma resolução espacial de 30m (RODRIGUES et al., 2010). A obtenção dos dados se deu entre os anos de 2000 e 2008 e mapeou a Terra entre 83 °N e 83 °S. ASTER GDEM é um produto da NASA (*National Aeronautics and Space Administration*) e do METI (*Japan's Ministry of Economy, Trade, and Industry*). O produto utilizado nesse trabalho é o ASTER GDEM *Version* 2, lançado em outubro de 2011 e que foi melhorado com dados obtidos despois de setembro de 2008.

Landau e Guimarães (2011) compararam os produtos SRTM (90 metros), Topodata[7] e ASTER na região do Rio de Janeiro/Brasil e concluíram que o ASTER, apesar de melhor resolução espacial, tende a subestimar altitudes de locais mais elevados, enquanto o SRTM apresenta limitações para áreas pequenas. A subestimação das cotas de altitude pelo ASTER também foi comunicada por Gómez-Gutiérrez et al.

[7] O projeto TOPODATA melhorou a resolução das informações do SRTM de 90 para 30 metros no Brasil. `http://www.dsr.inpe.br/topodata/`

(2011), que compararam os MDEs na Espanha e concluíram que o uso dos modelos globais deve ser limitado à descrição, visualização e representação do relevo.

Hayakawa et al. (2008) compararam os modelos GDEM e SRTM-3 (90 metros) na região oeste do Japão e concluíram que o GDEM oferece representações mais suaves e mais realistas de planícies, vales, declives íngremes e cumes de montanha, enquanto que o SRTM tende a superestimar a elevação do vale e subestimar a elevação do cume. Segundo esses autores, o GDEM tem melhor representação topográfica do que o SRTM.

Neste trabalho foram escolhidas quatro áreas de estudo em função da presença de torres anemométricas, cujas medidas de vento em diferentes alturas estão disponíveis. São elas : NREL, Guapé, Paracuru e Camocim.

Em todas as áreas foram obtidos os MDEs do Aster/GDEM e do SRTM, ambos com resolução de 30m. As imagens, em formato geotiff, foram baixadas do catálogo disponível no EarthExplorer (USGS, 2016), no conjunto de dados '*Digital Elevation*' e os produtos *'Aster Global DEM'* e '*SRTM/SRTM 1 Arc-Second Global*'. A pesquisa no site do Earth Explorer foi feita pelas coordenadas das torres anemométricas.

A figura 3.8 apresenta os MDEs GDEM e SRTM das áreas estudadas. O SRTM da região de Paracuru apresenta falta de dados.

Figura 3.8 - Amostra dos modelos digitais de elevação. GDEM (Coluna da esquerda) e SRTM (coluna da direita).

4 IMPLEMENTAÇÃO DO DFA-2D E DISCUSSÕES RELEVANTES SOBRE A TÉCNICA

Neste capítulo a implementação do DFA-2D em python é descrita e provada sua corretude por meio da análise de modelos fractais fBm e fGn. Além disso, o mesmo conjunto de dados será utilizado em algumas discussões acerca da técnica. Ao contrário de outros trabalhos relacionados, o objetivo não é discutir se o DFA-2D é ou não afetado por tendências no sinal, ou pelo tamanho da série. O objetivo aqui foi averiguar e validar questões acerca do método em si. Especialmente aquelas onde há divergência na literatura ou características que são válidas para o DFA-1D e não foram comunicadas para o DFA-2D.

As seguintes questões foram levantadas e respondidas nesse capítulo:

- A formulação matemática de integração apresentada por Gu e Zhou (2006) é adequada?
- Faz diferença integrar o sinal antes ou depois da subdivisão da matriz?
- É necessário remover a média da matriz, ou das sub-matrizes no momento da integração?
- Se o sinal é sabidamente fBm, há ganhos se o mesmo não for integrado?
- Como é a resposta do DFA-2D para a prática de crescimento logarítmico de escala utilizado no DFA-1D?
- O DFA-2D é capaz de distinguir diferentes padrões de autocorrelação espacial?

Para tanto, os dados fractais (fBm e fGn) foram utilizados, uma vez que se conhece exatamente o comportamento esperado da técnica nesses casos. Além disso, os fractais foram utilizados na verificação de corretude dos métodos implementados e nas avaliações de performance do algoritmo. Foram realizados testes empíricos com três diferentes expoentes de Hurst : 0.2, 0.5 e 0.9, em um conjunto de trinta fractais fBm e trinta fGn, para cada valor de expoente. Cada matriz tinha as dimensões de 500 $\times$ 500.

A última questão levantada vem porque, como um dos interesses principais do DFA-2D é na avaliação de textura de imagens, há uma necessidade de compreender seu

comportamento em relação a autocorrelação espacial. Para tanto, foram gerados padrões espaciais sintéticos e o resultado do DFA-2D foi comparado ao já bem consolidado índice de correlação espacial de Moran (MORAN, 1950).

O capítulo apresenta também a implementação paralela do DFA-2D em GPU, assim como os testes de performance da abordagem utilizada na parelelização.

4.1 Implementação do DFA-2D e do MFDFA-2D

A implementação do DFA-2D em suas versões monofractal e multifractal foi feita utilizando a linguagem Python e, especialmente, as bibliotecas Numpy e Scipy (SCIPY.ORG, 2016). Python é uma linguagem de alto nível, baseada em *scripts*, interpretada, multi-plataforma, fortemente tipada, multiparadigma (envolvendo orientação a objetos, programação imperativa e programação funcional), de propósito geral e livre (licença GPL). Além disso, fornece estruturas de dados de alto nível, como listas e dicionários, tipagem e vinculação dinâmica de variáveis, gerenciamento automático de memória e sintaxe concisa. Além da biblioteca padrão, Python tem diversos módulos e *frameworks* que podem ser integrados, facilitando e agilizando o desenvolvimento de soluções. De fato, quando criada, em 1991, por Guido van Rossum, o objetivo era prover um ambiente de programação que fosse rápido e legível.

Python tem sido utilizada em diversas áreas, como web, desktop, sistemas embarcados, aprendizado de máquina, animações, sistemas 3D, e também em computação científica e análise de dados. A linguagem cresceu muito em popularidade nos últimos anos. Retrato disso é o ranking das linguagens programação feito pela IEEE, em que, no ano de 2016, Python superou a linguagem C++ e passou ocupar o terceiro lugar (Figura 4.1) (IEEE, 2016). O *IEEE Spectrum* usa diversas métricas para medir o uso atual de uma linguagem.

Langtangen (2008) faz uma revisão sobre porque linguagens de *script* são úteis na computação científica e, dentre os motivos, cita o fato de que, em geral, cientistas preferem ter ambientes completos, onde é possível, com poucas linhas de código, realizar o pré-processamento, processamento e visualização dos dados. Cai et al. (2005) expandem esse conceito, afirmando que a investigação de problemas científicos requer *softwares* que ofereçam interface com usuário, gerenciamento de entrada e saída de dados, computação numérica, análise de dados, visualização, manipulação de arquivos e diretórios e geração de relatórios.

A linguagem Python oferece essa possibilidade dentro de um ambiente gratuito,

Language Rank	Types
1. C	
2. Java	
3. Python	
4. C++	
5. R	
6. C#	
7. PHP	
8. JavaScript	
9. Ruby	
10. Go	

Figura 4.1 - *Ranking* das dez linguagens de programação mais utilizadas em 2016. FONTE : (IEEE, 2016)

com uma comunidade ativa, diversos fóruns para compartilhar ideias e tirar dúvidas e, principalmente, com performance. Cai et al. (2005) compararam a performance de Python com as linguagens C e Fortran e concluíram que os *scripts* escritos em Python podem ser tão eficientes quanto os programas desenvolvidos em baixo nível, como C e Fortran.

Especificamente para a computação científica, a biblioteca padrão do Python não é adequada, especialmente devido à execução lenta de *loops* aninhados e a falta de estruturas de dados eficientes para tratar grandes quantidades de dados. Sendo assim, é necessário fazer uso da biblioteca NumPy (*Numerical Python*), especialmente desenvolvida e otimizada para computação científica. Essa biblioteca permite trabalhar com matrizes e *arrays* n-dimensionais de forma bastante natural, além de oferecer diversas funções matemáticas úteis em álgebra linear, transformada de Fourier, números randômicos, entre outros. A programação com NumPy deve seguir o paradigma de vetorização[1], evitando o uso de *loops* aninhados, assim como no Matlab. Para programadores que vêm de linguagens procedurais como C e C++, talvez esse seja o maior desafio em se trabalhar com NumPy, que é pensar algoritmicamente de forma vetorial e não procedural.

O NumPy está contido no *framework* SciPy, que também inclui outras bibliotecas importantes para a computação científica, como o Matplotlib e o Pandas (SCIPY.ORG,

[1]Capacidade de escrever o código condensado para aplicar uma ação a cada elemento de uma matriz com uma única linha de código, por meio de operadores vetorizados fornecidos pela linguagem. Tais operadores realizam as tarefas de forma mais rápida do que os códigos utilizando estruturas de repetição, como *for*.

2016).

O NumPy não oferece funções diretas para o cálculo dos mínimos quadrados em matrizes bi-dimensionais. Sendo assim, para realizar o destendenciamento de cada sub-matriz, utilizou-se fatoração LU[2], a partir do modelo plano dado por $y_{v,w}(i,j) = ai + bj + c$.

No desenvolvimento desse trabalho, foi utilizado o compilador Python versão 3.4, numpy 1.11 e scipy 0.17. Os códigos em python do DFA-2D mono e multifractal estão disponíveis em `https://github.com/vanessavcos/DFA-2D`.

O gráfico da figura 4.2 apresenta o resultado da reprodução do experimento de Gu e Zhou (2006) para o DFA-2D monofractal aplicado às simulações de fBm. A barra de erros mostra o desvio padrão das 500 amostras de cada valor de H. Assim como relatado por Gu e Zhou (2006), o desvio padrão aumenta com o aumento de H. Os valores numéricos podem ser vistos na tabela 4.1. Nessa implementação a escala s variou entre 4 e $500/4 = 125$. Nessa implementação também não houve recomeço do cálculo pelos demais cantos da matriz, quando a mesma não era múltipla da escala s.

Figura 4.2 - Média do valor estimado do Expoente de Hurst com o DFA-2D para amostras de superfícies fBm. $0.05 \leq H < 1$, com incremento de 0.05.

[2]Método da álgebra linear para solução de sistemas lineares.

Tabela 4.1 - Resultado do DFA-2D aplicado a superfícies fBm com diferentes expoentes de Hurst.

	H	
$\hat{H}$	**Média**	**Desvio Padrão**
0.05	-0.019	0.049206766
0.10	0.067	0.050815274
0.15	0.126	0.055048521
0.20	0.175	0.056346921
0.25	0.230	0.055964043
0.30	0.285	0.059591404
0.35	0.334	0.059998687
0.40	0.382	0.064528759
0.45	0.428	0.061990103
0.50	0.480	0.067590915
0.55	0.524	0.066471521
0.60	0.576	0.072649855
0.65	0.624	0.073777266
0.70	0.674	0.08136611
0.75	0.722	0.084127308
0.80	0.759	0.082903514
0.85	0.797	0.084368217
0.90	0.837	0.083488529
0.95	0.890	0.082179447

A figura 4.2 e a tabela tabela 4.1 provam a corretude do método implementado.

Os testes empíricos foram realizados em uma máquina desktop, com processador Core i7 950, 3,07 GHZ, 8GB de memória RAM, rodando Linux Centos 7.2.15.11, totalmente dedicada. Assim, em termos de processamento, cada matriz levou, em média, 11 segundos para ser computada com o DFA-2D monofractal. O código do DFA-2D multifractal é mais lento, levando cerca de 19 segundos para ser computado.

4.2 Investigações acerca do algoritmo DFA-2D

O objetivo dessa seção será discutir algumas divergências presentes na literatura sobre o DFA-2D. Todas as discussões foram validadas empiricamente. Para tanto, foram utilizadas trinta simulações de sinais fBm e seus respectivos fGn. Na tabela 4.2 são apresentados os valores do α e do expoente de Hurst esperado ($\hat{H}$) que servirão para validar ou invalidar os demais resultados encontrados.

A primeira discussão importante sobre o DFA-2D é sua fórmula de integração apre-

Tabela 4.2 - Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bidimensionais. Os valores são médias de trinta repetições.

	fGn		**fBm**	
Expoente de Hurst	α	$\hat{H}$	α	$\hat{H}$
0.2	0.1321 ± 0.0065	0.13	2.1596 ± 0.0474	0.16
0.5	0.4100 ± 0.0231	0.41	2.5069 ± 0.0691	0.51
0.9	0.8411 ± 0.0371	0.84	2.8451 ± 0.0913	0.85

sentada por Gu e Zhou (2006). O DFA-2D faz a soma acumulada dos valores da matriz e isso é consenso na literatura. No entanto, a formulação apresentada no artigo acumula os valores de forma diferente. A figura 4.3 ilustra uma sub-matriz $X_{v,w}$ e mostra as células que serão somadas para obter o valor da posição destacada em vermelho no perfil $Y_{v,w}$. A figura 4.3a, exemplifica a soma acumulada e a 4.3b, a formulação indicada no trabalho de referência e repetida na equação 4.1[3].

$$Y_{v,w}(i,j) = \sum_{k_1=1}^{i} \sum_{k_2=1}^{j} X_{v,w}(k_1, k_2) \tag{4.1}$$

Figura 4.3 - Diferença entre as fórmulas de integração. a) Soma acumulada (*cumsum*). b) Formulação dada pela equação 4.1.

Testes empíricos, sumarizados na tabela 4.3, utilizando a formulação 4.1 mostraram que a mesma gera resultados errôneos para sinais fBm, mas aproximou melhor os valores para fGn. Esse resultado indica que a formulação utilizada acerta ao transformar um fGn em um fBm, mas erra ao transformar um fBm em um fBm acumulado.

[3]Considerando que os índices da matriz iniciam em 1.

Outro fator desfavorável é o tempo de processamento que, nessa abordagem, aumenta cerca de três vezes.

Tabela 4.3 - Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bidimensionais, utilizando a formulação de soma acumulada apresentada por Gu e Zhou (2006). Os valores são médias de trinta repetições.

	fGn		**fBm**	
Expoente de Hurst	α	$\hat{H}$	α	$\hat{H}$
0.2	0.1753 ± 0.0051	0.18	1.9554 ± 0.0334	-
0.5	0.4688 ± 0.0151	0.47	1.9772 ± 0.0282	-
0.9	0.8834 ± 0.0265	0.88	1.9713 ± 0.0280	-

Ainda na geração do perfil $Y_{v,w}$, há uma divergência na literatura sobre a necessidade de se remover ou não a média antes da integração. Gu e Zhou (2006) não o fazem, mas Vargas-Olmos et al. (2015) sugerem que seja feita a remoção da média da sub-matriz antes da integração (equação 4.2).

$$Y_{v,w}(i,j) = \sum_{k_1=1}^{i} \sum_{k_2=1}^{j} X_{v,w}(k_1,k_2) - \bar{X}_{v,w} \tag{4.2}$$

onde $\bar{X}_{v,w}$ é a média da sub-matriz $X_{v,w}$.

Os testes empíricos mostraram que não remover a média, remover a média de toda imagem (antes da divisão da matriz), e remover a média das sub-matrizes levam ao mesmo valor de α (tabela 4.4). No entanto, o tempo de processamento aumenta nas abordagens que removem a média da imagem. Comparando os resultados da tabela 4.4 com os da tabela 4.2, conclui-se ser desnecessário remover a média durante a geração do perfil. Kantelhardt et al. (2002) explicam que para o DFA-1D também não existe a necessidade de remover a média, já que a mesma é feita em passos posteriores do algoritmo.

Gu e Zhou (2006) afirmam que há duas formas de generalizar o DFA-1D para o DFA-2D : dividir a imagem em sub-imagens e depois integrar, ou integrar a imagem toda e depois dividi-la. Os autores afirmam que para o DFA-2D monofractal não faz diferença, mas para a versão multifractal, a integração antes da divisão leva a valores errôneos para fBm. A tabela 4.5 apresenta os valores de α para a abordagem de integração à priori. Verifica-se que os resultados não são coincidentes com aqueles apresentados na tabela 4.2 e que mesmo para o DFA-2D monofractal, a integração

Tabela 4.4 - Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, para diferentes estratégias sobre a remoção ou não da média. Os valores são médias de trinta repetições.

		fGn		**fBm**	
Estratégia	**Expoente de Hurst**	α	$\hat{H}$	α	$\hat{H}$
Removendo a média da matriz antes da divisão	0.2	0.1321 + 0.0065	0.13	2.1596 + 0.0474	0.16
	0.5	0.4100 + 0.0231	0.41	2.5069 + 0.0691	0.51
	0.9	0.8411 + 0.0371	0.84	2.8451 + 0.0913	0.85
Removendo a média da sub-matriz	0.2	0.1321 + 0.0065	0.13	2.1596 + 0.0471	0.16
	0.5	0.4100 + 0.0231	0.41	2.5069 + 0.0691	0.51
	0.9	0.8411 + 0.0371	0.84	2.8457 + 0.0913	0.85

antes da divisão da matriz leva a valores errôneos de α, tanto para modelos fBm, quanto fGn.

Tabela 4.5 - Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, utilizando a integração da matriz antes da sub-divisão. Os valores são médias de trinta repetições.

	fGn		**fBm**	
Expoente de Hurst	α	$\hat{H}$	α	$\hat{H}$
0.2	0.1051 ± 0.0223	0.11	1.5959 ± 0.0952	-
0.5	0.3319 ± 0.0574	0.33	1.8534 ± 0.0783	-
0.9	0.3386 ± 0.1708	0.34	1.9677 ± 0.0629	-

Uma preocupação relevante com o DFA-2D é com relação ao tempo de processamento. Freitas (2012) e Yeh et al. (2012) destacaram em seus trabalhos que o tempo de processamento do DFA-2D pode inviabilizar sua utilização. Códigos estáveis e utilizados por diversos pesquisadores para o DFA-1D, como o disponível no fórum do projeto Physionet (GOLDBERGER et al., 2000; MIETUS et al., 2005), não utilizam o crescimento linear da variável s, que controla o tamanho das escalas e, com menos escalas, há a redução da complexidade assintótica do algoritmo (FORNARI, 2016). A estratégia mencionada usa a equação exponencial 4.3 no incremento do parâmetro s. Dessa forma, o processamento alcança mais rápido o limite superior de $\frac{n}{4}$, preservando especialmente as escalas menores.

$$s = \left\lfloor s * 2^{\frac{1}{8}} \right\rceil \tag{4.3}$$

onde $6 \leq s \leq \frac{n}{4}$, sendo n o tamanho da série.

A tabela 4.6 apresenta os resultados do DFA-2D utilizando o crescimento exponen-

cial de escalas conforme equação 4.3. Verifica-se que para a amostra utilizada, o crescimento exponencial de escala obteve resultado muito próximo dos apresentados na tabela 4.2, com a vantagem de diminuir o processamento e, consequentemente, o tempo.

Tabela 4.6 - Valores de α e expoente de Hurst esperado $\hat{H}$ para os sinais fGn e fBm bi-dimensionais, utilizando crescimento logarítmico das escalas s. Os valores são médias de trinta repetições.

	fGn		**fBm**	
Expoente de Hurst	α	$\hat{H}$	α	$\hat{H}$
0.2	0.1401 ± 0.0065	0.14	2.1559 ± 0.0350	0.16
0.5	0.4070 ± 0.0183	0.41	2.5023 ± 0.0570	0.50
0.9	0.8332 ± 0.0249	0.83	2.8520 ± 0.0817	0.85

A formulação exponencial reduziu em 1/4 a quantidade de escalas avaliadas, passando de 120 para 30. Com isso houve uma redução de cerca de 50% no tempo de processamento.

Apesar de aproximar muito bem os resultados e reduzir o tempo de processamento de forma significativa, o uso do crescimento exponencial no DFA-2D deve ser visto com precaução, uma vez que a quantidade de escalas já é limitada pelas dimensões das imagens, que são menores do que as das séries unidimensionais.

Por fim, é sabido que a geração do perfil $Y_{v,w}$ serve para transformar um possível sinal fGn em um fBm, sobre o qual o DFA foi desenvolvido para atuar. Neste caso, sinais fBm são transformados em um fBm acumulado e tem o valor de α adicionado de 1 para o caso unidimensional e de 2 para o caso bi-dimensional. Testes empíricos apresentados na figura 4.4 e tabela 4.7 revelam que, se o dado é sabidamente fBm, não integrar o sinal diminui consideravelmente a barra de desvio e, consequentemente, o erro de aproximação do expoente de Hurst. Esse experimento é idêntico ao apresentado na figura 4.2, porém sem integrar o sinal antes de calcular o expoente α.

A importância desse resultado é que, ao saber que a superfície possui comportamento de escala modelado por fBm, é possível melhorar a aproximação do expoente de Hurst (e consequente dimensão fractal), realizando o processamento sem a integração. É importante ressaltar que esses resultados só são válidos para fBm. Não integrar um sinal fGn, por exemplo, leva a resultados errôneos por parte do DFA.

Figura 4.4 - Média do valor estimado do Expoente de Hurst com o DFA-2D para amostras de superfícies fBm, sem realizar a integração do sinal.

Tabela 4.7 - Resultado do DFA-2D aplicado a superfícies fBm com diferentes expoentes de Hurst, sem a integração do sinal.

	H	
$\hat{H}$	**Média**	**Desvio Padrão**
0.05	0.050	0
0.10	0.100	0
0.15	0.150	0.01
0.20	0.200	0.01
0.25	0.250	0.01
0.30	0.300	0.01
0.35	0.350	0.01
0.40	0.400	0.01
0.45	0.450	0.02
0.50	0.500	0.02
0.55	0.550	0.02
0.60	0.600	0.02
0.65	0.650	0.02
0.70	0.600	0.02
0.75	0.750	0.02
0.80	0.800	0.02
0.85	0.850	0.03
0.90	0.900	0.03
0.95	0.950	0.03

4.2.1 DFA-2D e a autocorrelação espacial

A dependência espacial é a primeira lei da geografia que prega que "*todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes*". A dependência espacial está presente em todas as direções e fica mais fraca à medida em que aumenta a dispersão na localização dos dados. Nesse sentido, os eventos apresentam uma relação que depende da distância (CÂMARA et al., 2004).

A autocorrelação espacial é a expressão computacional da dependência espacial. Indica que a medida de correlação sobre a mesma variável, medida em locais distintos no espaço.

A referência na literatura para medir a autocorrelação espacial é o índice de Moran (MORAN, 1950), cujos detalhes podem ser vistos no Apêndice A.2. O índice de Moran pode receber valores entre -1 (anti-correlacionado) e 1 (correlacionado). O zero representa superfícies totalmente descorrelacionadas.

A autocorrelação espacial pode ser considerada uma medida de rugosidade. Foram avaliados os padrões apresentados nas figuras 4.5 e 4.6. Na figura 4.5, padrões altamente correlacionados e, na figura 4.6, o padrão tabuleiro (*chessboard*), com zeros e uns intercalados. O primeiro padrão da figura 4.6 tem um pixel intercalado, isto é, um pixel com zero e outro com um. O padrão foi evoluindo com intervalos de pixels de 2, 4, 8, 16, 32, 64, 128 e 256. A imagem tem tamanho 500 × 500. A ideia era reproduzir superfícies anti-correlacionadas, cuja rugosidade é alta. As matrizes foram geradas em Python.

Figura 4.5 - Padrões correlacionados utilizados para avaliar o comportamento do DFA-2D.

A tabela 4.8 apresenta os resultados para o DFA-2D e o índice de Moran para os padrões apresentados na figura 4.5. E a tabela 4.9 apresenta os resultados para os

Figura 4.6 - Padrões do tipo *chessboard* utilizados para avaliar o comportamento do DFA-2D. A intercalação dos valores zero e um variou de 1 pixel (primeira imagem) até 256 (última imagem).

padrões da figura 4.6.

Tabela 4.8 - Valores de α e Moran para os padrões correlacionados da figura 4.5.

	P1	**P2**
α	3.4025	2.9521
Moran	0.9980	0.9980

Tabela 4.9 - Valores de α e Moran para os padrões '*chessboard*' da figura 4.6, com diferentes janelas de intercalação (em pixels).

	1	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**
α	0	0.0016	0.0232	0.6425	1.2423	1.8416	2.4708	2.8859	3.3331
Moran	-1	0.0027	0.5033	0.7536	0.8787	0.9413	0.9726	0.9883	0.9961

Nas tabelas 4.8 e 4.9 verifica-se que o objetivo de simular superfícies correlacionadas e anti-correlacionadas foi atingido, uma vez que o Moran apresentou valores próximos de 1 para os padrões da figura 4.5 e -1 para o primeiro padrão da figura 4.6. Os valores apresentados mostram ainda que o DFA-2D é capaz de aferir sobre a dependência espacial das superfícies. Na tabela 4.9, para superfícies correlacionadas, o α recebeu valores próximos de 3, que revelam superfícies extremamente persistentes. É importante notar que o Moran gerou o mesmo valor para os dois diferentes padrões, enquanto o DFA-2D mostrou que eles eram diferentes.

A tabela 4.9 confirma o resultado. Para o primeiro padrão analisado, o DFA-2D revelou valor zero, ou seja, máxima anti-correlação (-1 para Moran). O valor de α cresceu à medida que a janela de intercalação também cresceu, como o Moran. O último padrão, de janela 256×256 revelou valor próximo de 3, indicando novamente correlação, ou persistência, quando o Moran indicou valor próximo de 1.

Valores de α maiores que 3 acontecem porque os dados simulados atingiram características estacionárias, para o qual o DFA não opera e, consequentemente, gera valores estranhos. No entanto, a conclusão importante é que o DFA-2D foi capaz de identificar os diferentes padrões espaciais.

A próxima seção apresenta a implementação paralela em GPU do DFA-2D e MFDFA-2D.

4.3 DFA-2D Paralelo

Como já mencionado anteriormente, o DFA-2D tem um alto custo computacional e alguns autores já chamaram atenção para o tempo de processamento do método. No entanto, o algoritmo é altamente paralelizável, não pelo método em si, mas porque o mesmo processo é realizado em todas as sub-matrizes de forma independente. Essas duas constatações nortearam a ideia de paralelizar o DFA-2D.

Nos últimos anos, o uso das Unidades de Processamento Gráfico (GPU) para propósitos gerais (GPGPU), juntamente com arquiteturas e diferentes modelos de programação, tais como a NVIDIA CUDA tem ganhado espaço e permitido que, mesmo com pouca experiência, seja possível desenvolver programas paralelos utilizando extensões de linguagens conhecidas, como C (JIMÉNEZ; Ruiz de Miras, 2012).

Dois conceitos são importantes na programação com GPU : paralelismo de dados e independência. Esses atributos permitem que a mesma tarefa de computação seja aplicada a fluxos de dados diferentes, os quais dependem muito pouco uns dos outros.

Isto é, permitem a execução simultânea de uma mesma instrução em diferentes porções de um conjunto de dados.

O paralelismo de dados é uma técnica de programação que divide uma grande quantidade de dados em partes menores que podem ser operados em paralelo. Após o processamento, os dados são novamente combinados em um único conjunto.

A grande vantagem de uso das GPUs é que elas são placas dedicadas ao processamento de dados, não tendo outras atribuições, como acontece na CPU. Outra vantagem importante é o fato das GPUs serem muito mais acessíveis do que arquiteturas distribuídas, como *clusters*, que são caros e de difícil manutenção.

CUDA (*Compute Unified Device Architecture*) é uma plataforma de *software* para computação paralela que utiliza GPUs da NVIDIA. O modelo de programação CUDA é uma extensão da linguagem de programação C e permite ao programador disparar milhares de *threads* simultaneamente na GPU. Apesar de diversos autores classificarem a arquitetura CUDA como SIMD (*Single-Instruction, Multiple Data*) (JIMÉNEZ; Ruiz de Miras, 2012), a documentação da própria NVIDA chama atenção para o fato de CUDA ser SPMD (*Single-Program, Multiple Data*). Isso quer dizer que nas unidades de processamento paralelo (*threads*), o mesmo programa é executado em múltiplas partes do dado. Entretanto, o processamento dessas unidades não precisa executar a mesma instrução ao mesmo tempo, ao contrário da SIMD (NVIDIA; UIUC, 2015).

Não é objetivo desse trabalho descrever as arquiteturas de placas GPU e nem da plataforma CUDA. Para tanto, diversas literaturas podem ser consultadas (GULO, 2012; KLOCKNER et al., 2012; NVIDIA; UIUC, 2015; NVIDIA, 2016). Portanto, abaixo, o funcionamento básico da arquitetura CUDA foi sumarizado. A principal referência utilizada foi a documentação da própria NVIDIA (NVIDIA; UIUC, 2015).

Cada dispositivo CUDA (placas GPU) tem diversos muti-processadores (MPs), cada um consistindo de alguns processadores de *streaming* (SPs). Esses processadores são responsáveis por executar todas as *threads* de forma paralela.

Quando um código da CPU (também chamado código *host*) dispara a chamada função de *kernel*, o sistema CUDA gera uma grade de *threads*, que são organizadas numa hierarquia de dois níveis : bloco e *thread*. Todos os blocos têm o mesmo tamanho, isto é, a mesma quantidade de *threads*, especificada no código do *host*. Cada bloco recebe um conjunto de dados, que são tratados pelas *threads*.

A hierarquia de *threads* pode ser feita em 1, 2 ou 3 dimensões, dependendo da natureza do dado. A figura 4.7 ilustra uma grade de 2×3 blocos de *threads*. Cada bloco possui as dimensões 3×4 *threads*. A arquitetura ilustrada na figura 4.7 é a ideal para dados bi-dimensionais. Apesar dessa abstração, ao lidar com os dados na GPU, os dados estão armazenados de forma sequencial. Fica a cargo do programador realizar o mapeamento dos endereços.

Cada bloco tem o limite de 1024 *threads* e cada grade suporta cerca de 65 mil blocos. No entanto, essas configurações dependem do dispositivo e versão do CUDA utilizados.

Figura 4.7 - Divisão lógica de unidades de trabalho na GPU. FONTE : (NVIDIA; UIUC, 2015).

De forma genérica, pode-se dizer que um programa CUDA típico é formado pelos seguintes passos:

a) A CPU aloca memória na GPU
b) A CPU copia dados da CPU para a GPU
c) A CPU dispara *kernels* na GPU para processar o dado
d) A GPU processa os dados
e) A CPU copia os resultados da GPU

A GPU gerencia três tipos de memória:

- **Memória Global :** a mais lenta, com maior custo de acesso (*load*) e menor largura de banda. O conteúdo da memória global é vista por todos os blocos da grade.

- **Memória Compartilhada :** é colocada dentro do chip e, portanto, apresenta menor latência e maior largura de banda. O conteúdo é compartilhado apenas pelas *threads* do bloco;

- **Registradores :** também estão dentro do chipe e é a memória que apresenta menor latência e maior largura de banda. Acesso instantâneo ao valor da variável. O conteúdo é compartilhado apenas pelas *threads* do bloco.

A CPU transfere os dados para a memória global da GPU. Em geral, as *threads* cooperam para transferirem os dados da memória global para a memória compartilhada. Variáveis definidas dentro do código CUDA são armazenadas nos registradores. O tamanho das memórias compartilhada e dos registradores é, em geral, bem menor do que o da memória global.

4.3.1 Arquitetura Utilizada

Neste trabalho utilizou-se uma placa NVIDIA GeForce GTX 460, com *compute capability* 2.1 e CUDA *toolkit* 7. As principais características da GPU estão resumidas na tabela 4.10.

Tabela 4.10 - Características da placa GeForce GTX 460.

Número máximo de threads por SM	1536
Número máximo de blocos por SM	8
Número máximo de threads por bloco	1024
Quantidade de CUDA Cores	336
Número de multiprocessadores	7
Número máximo de threads simultâneas	10.752
Tamanho da memória global	963 MB
Tamanho da memória compartilhada por bloco	48 KB

Como o DFA-2D já havia sido implementado em Python, optou-se por utilizar o *wrapper* PyCUDA (KLÖCKNER, 2017). Basicamente, o PyCUDA conecta um programa Python a funções da API CUDA, que, como já mencionado, é escrita em C. Segundo Klockner et al. (2012), um código desenvolvido em Python terá o mesmo desempenho na GPU do que um código escrito em C, mas com muito menos esforço

por parte do programador.

Como já citado anteriormente, os testes empíricos foram realizados em uma máquina desktop, com processador Core i7 950, 3,07 GHZ, 8GB de memória RAM, rodando Linux Centos 7.2.15.11, totalmente dedicada.

4.3.2 Modelagem

Na linha 3 do algoritmo do DFA-2D (algoritmo 1) está relatado que, para cada escala s, a matriz $(m \times n)$ de entrada é dividida em sub-matrizes, gerando $\left\lfloor \frac{m}{s} \right\rfloor * \left\lfloor \frac{n}{s} \right\rfloor$ sub-matrizes, sendo m e n o número de linhas e colunas da matriz. Sendo assim, a ideia principal da abordagem paralela, é que os passos 6 a 10 do algoritmo 1 fossem realizados paralelamente entre as sub-matrizes. Para facilitar o entendimento, tais passos são repetidos na figura 4.8.

5: **for** cada sub-matriz $M_s(i)$ **do**
6: Obtenha o perfil $Y_s(i)$ a partir da integração da superfície $M_s(i)$:
$Y_s(i) = \sum_{k_1=1}^{i} \sum_{k_2=1}^{j} X_{v,w}(k_1, k_2)$
7: Obtenha a matriz destendenciada $z_s(i)$ como a diferença entre $Y_s(i)$ e a superfície $\tilde{Y}_s(i)$ ajustada por um plano (mínimos quadrados).
8: Compute a função de flutuação local: $F_s^2(i) = \frac{1}{s^2} \sum_{i=1}^{s} \sum_{j=1}^{s} (z_s(i))^2$
9: **end for**

Figura 4.8 - Passos do DFA-2D paralelizados com CUDA.

A primeira observação importante é que os passos descritos nas linhas de 6 a 8 não são, em si, paralelizáveis. Os motivos serão abordados a seguir. Então, a questão principal é que a paralelização do método se deu, na prática, pela paralelização do comando **for** da linha 5.

Abaixo serão discutidos os pontos principais dessa abordagem.

A linha 6 do algoritmo é a obtenção do perfil da superfície, via soma acumulada. A soma acumulada na GPU é um problema bastante estudado, por ser um problema estritamente sequencial. Algumas abordagens são citadas na literatura e até implementadas em algumas bibliotecas CUDA. No entanto, as funções implementadas por bibliotecas são funções *kernel* e não podem ser chamadas dentro do código da GPU. Nesse caso, o ganho com a utilização de uma função de biblioteca seria pequeno, uma vez que a transferência de dados seria grande. Então, optou-se por fazer a soma

acumulada dentro da GPU. A estratégia utilizada foi utilizar uma única *thread* para realizar essa tarefa.

Na linha 7, obtêm-se a matriz destendenciada. Para tanto, utilizou-se a fatoração LU. Na resolução de um sistema $Ax = b$, utilizando a fatoração LU, temos que $A * A^T \equiv A; A \equiv L * U$ e $A^T * B \equiv b$. Para solucionar o sistema, têm-se que : $L * y = b; U * x = y$.

Como para toda escala s, todas as sub-matrizes compartilham as mesmas matrizes L e U, as mesmas são computadas no código *host* e enviadas para a GPU. No código CUDA, três *threads* foram designadas para calcular $A^T * B \equiv b$, visto que a multiplicação também é uma tarefa sequencial.

Uma única thread executa as equações $L*y = b$ e $U*x = y$ e encontra os parâmetros a, b e c que resolvem o sistema $\tilde{Y}_{v,w}(i,j) = ai + bj + c$.

Por fim, alocando uma *thread* por linha, calcula-se a diferença entre $Y_{v,w}$ e $\tilde{Y}_{v,w}$ ao quadrado e a função de flutuação local é computada e retornada para o *host*.

Portanto, para cada bloco, são disparadas s *threads*, uma para cada linha da sub-matriz. Essa opção foi feita porque, como dito, não há muita paralelização no interior do processo e muitas *threads* ficariam bastante tempo ociosas durante quase todo processamento dos dados. Além disso, cada bloco tem limite de 1024 threads, o que faz com que não seja possível disparar uma *thread* por cada célula em matrizes maiores que 32×32. Além de calcular a função local, as demais *threads* são utilizadas para, de forma cooperativa, copiarem os dados da memória global para a memória compartilhada.

Por limitação no tamanho da memória compartilhada da placa utilizada, quando a escala s é maior que 110, é disparada uma segunda função de kernel, onde os dados permanecem na memória global durante todo o processamento.

A figura 4.9 mostra a chamada da função de kernel nomeada 'fit', onde:

- *d-L* são os parâmetros diferentes de zero da matriz L
- *d-U* são os parâmetros diferentes de zero da matriz U
- *d-mat* é a matriz de dados
- s é a escala atual
- *tam* é o tamanho da matriz de dados

Os parâmetros de *kernel* são defindidos com a variável $blockSize = (s, 1, 1)$ e $grid = (\lfloor \frac{m}{s} \rfloor, \lfloor \frac{m}{s} \rfloor, 1)$. Ou seja, a grade representa a sub-divisão da matriz nas sub-matrizes de tamanho s. E, como já mencionado, cada bloco tem s *threads*.

O código completo pode ser acessado em https://github.com/vanessavcos/DFA2D-PARALLEL.

```
#launch kernel
#get the kernel function from the compiled module
kernel = mod.get_function("fit")
#call the kernel on the card
kernel(
    #inputs
    d_L, d_U, d_mat,np.int32(s), np.int32(tam),
    #output
    d_vetF,
    # kernel parameters
    grid= grid, block=blockSize,
    #size of shared memory
    shared = sizeofSharedMemoryinBytes)
```

Figura 4.9 - Trecho do código PyCUDA que dispara a função de *kernel* nomeada 'fit', para o cálculo paralelo do DFA-2D.

A implementação do DFA-2D Multifractal Paralelo seguiu a mesma abordagem da versão monofractal. O que diferencia as versões é o cálculo da função global, que não foi paralelizada.

4.3.3 Ganho de performance com o DFA-2D Paralelo

Os testes de performance com a versão paralela implementada em PyCUDA foram feitas com os fractais fBm. Avaliou-se o tempo de processamento com diferentes tamanhos de matriz e diferentes expoentes de Hurst.

Verifica-se que a versão paralela melhorou consideravelmente o tempo de processamento, tanto na versão monofractal (tabela 4.11) quanto na multifractal (tabela 4.12), sem afetar a corretude dos resultados. Os tempos reportados estão em segundos e foram contabilizados desde o início de ambos procedimentos. No caso da versão paralela, o valor inclui também as transferências de dados. O *speed up* foi calculado como a divisão do tempo sequencial pelo tempo paralelo.

Tabela 4.11 - Comparação de tempo de processamento entre as implementações sequenciais em Python e a paralela em PyCuda para o DFA-2D Monofractal. Os tempos são dados em segundos. Média de 50 repetições.

Tamanho da matriz	**Tempo Sequencial (s)**	**Tempo Paralelo (s)**	***Speed Up***
64	0.09107 ± 0.0015	0.00909 ± 0.0005	10.1037
128	0.48455 ± 0.0025	0.02337 ± 0.0009	20.7370
256	2.35751 ± 0.0103	0.07386 ± 0.00152	31.9179
512	11.58794 ± 0.0715	0.52457 ± 0.0029	22.0905
1024	59.90978 ± 0.1484	3.03094 ± 0.0389	19.7660
2048	322.39272 ± 1.8915	27.5368 ± 0.03888	12.9642
4000	2443.151 ± 3.7664	320.7264 ± 0.0550	7.6176

Tabela 4.12 - Comparação de tempo de processamento entre as implementações sequenciais em Python e a paralela em PyCuda para o DFA-2D Multifractal. Os tempos são dados em segundos. Média de 50 repetições.

Tamanho da matriz	**Tempo Sequencial (s)**	**Tempo Paralelo (s)**	***Speed Up***
64	0.3443 ± 0.001	0.1691 ± 0.003	2.0356
128	1.1235 ± 0.007	0.2843 ± 0.003	3.9511
256	4.4307 ± 0.016	0.6015 ± 0.004	7.3662
512	19.2487 ± 0.043	1.8361 ± 0.006	10.4835
1024	89.9749 ± 0.1977	6.9600 ± 0.012	12.9273
2048	479.3076 ± 0.8609	40.5287 ± 0.03	11.8264
4000	2875.504 ± 29.97	344.4933 ± 0.188	8.3471

Verifica-se que o aumento no tempo está relacionado ao aumento do tamanho da matriz e que a implementação sequencial para matrizes de 2048×2048 já acarreta em um tempo bastante alto (cerca de 5 minutos). A abordagem paralela diminuiu esse tempo para cerca de 27 segundos. Destaca-se também que Yeh et al. (2012) reportam um tempo de processamento superior a três dias para processar uma imagem de tamanho 4080×3072 *pixels*. A implementação sequencial do DFA-2D feita nesse trabalho levou, em média, 6.78 horas para uma imagem de 4000×4000 *pixels*. Esse tempo foi reduzido para cerca de cinco minutos na versão paralela.

De fato, para aplicações de tempo real, como a relatada por Yeh et al. (2012), e aquelas classificadas como *Big Data*, a abordagem sequencial pode tornar-se impraticável.

O gráfico das figuras 4.10 e 4.11 comparam o tempo de processamento entre as abordagens sequencial e paralela do DFA-2D monofractal e multifractal, respectivamente. Apesar de haver ganho de tempo em todos os tamanhos de matriz, a diferença

torna-se significativa a partir da matriz de tamanho 1024.

	64	128	256	512	1024	2048	4000
Sequencial	0.0911	0.4845	2.3575	11.5879	59.9098	356.5330	2443.1513
Paralelo	0.0090	0.0234	0.0739	0.5246	3.0309	27.5014	320.7264

Figura 4.10 - Comparação do tempo de processamento entre as abordagens sequencial e paralela com o DFA-2D Monofractal. Os tempos são dados em segundos.

	64	128	256	512	1024	2048	4000
Sequencial	0.3443	1.1235	4.4307	19.2487	89.9749	479.3076	2875.5039
Paralelo	0.1691	0.2843	0.6015	1.8361	6.9601	40.5287	344.4933

Figura 4.11 - Comparação do tempo de processamento entre as abordagens sequencial e paralela com o DFA-2D Multifractal. Os tempos são dados em segundos.

5 RESPOSTA DO DFA-2D À SINAIS BI-DIMENSIONAIS TÍPICOS

Um questionamento levantado nesse trabalho é que a maioria dos estudos feitos com DFA, ou utilizam sinais fractais (fBm e fGn), ou utilizam o DFA apenas como um operador, sem se discutir de fato os resultados físicos de sua aplicação. Nesse sentido, esse capítulo visa analisar a resposta do DFA-2D em sinais conhecidos, como o ruído $1/f$ e dados caóticos como as redes de mapas acoplados (CML). As questões respondidas aqui serão:

- O DFA-2D é capaz de caracterizar diferentes tipos de ruídos da família $1/f$?
- A relação entre os expoentes α (DFA-2D) e β (*Power Spectrum*) é idêntica à definida com dados unidimensionais?
- Qual o comportamento do DFA-2D frente a dados caóticos como os dos mapas acoplados? O DFA-2D é capaz de caracterizar e diferenciar os diferentes modelos, tanto espacialmente, quanto no tempo?

Para cumprir os objetivos acima citados, os ruídos $1/f$ foram processados com o DFA-2D monofractal e os mapas acoplados com o DFA-2D mono e multifractal.

Os ruídos $1/f$ também foram processados com o *Power Spectrum* bi-dimensional (Anexo A.3) e os resultados foram comparados com os do DFA-2D. A relação entre α e β foi avaliada para sinais bi-dimensionais.

Uma dificuldade encontrada na análise de mapas acoplados foi que, tanto o DFA-2D, quanto o MFDFA-2D, quando utilizados no intuito de conhecer as dinâmicas que regem o dado, não devem se limitar a análise apenas dos valores numéricos entregues pelas técnicas. É necessário realizar uma inspeção visual nos gráficos. O desafio foi, portanto, fazer essa análise na presença de trezentas iterações no tempo. A metodologia definida foi realizar uma análise temporal e depois, com base nesse estudo, escolher algumas iterações para serem analisadas individualmente.

5.1 Ruídos $\frac{1}{f}$

A tabela 5.1 resume os dados utilizados nessa seção. Todas as matrizes tinham a dimensão de 500×500. Os ruídos foram comparados com os sinais fGn e fBm com expoente de Hurst 0.5, que também são conhecidos como ruído branco e *browniano*, respectivamente.

Tabela 5.1 - Resumo dos dados utilizados nos testes com ruídos da família $1/f$

Característica	**Quantidade**
Ruído Branco - $\beta = 0$	500
Ruído Rosa - $\beta = 1$	500
Ruído Vermelho - $\beta = 2$	500

A figura 5.1 e tabela 5.2 apresentam os gráficos *log-log* e os respectivos valores de α para os sinais analisados. Os valores são médias de quinhentas amostras de cada modelo.

Pelos valores da tabela 5.2, verifica-se que o DFA-2D foi capaz de descriminar os diferentes tipos de sinais. Verifica-se também que o α para os ruídos branco (fGn) e vermelho (fBm) apresentam valores limítrofes (aproximadamente 1 e 2, respectivamente) e revelam a relação entre eles, já que o ruído *browniano* é a integração do ruído branco. Um ponto importante é a avaliação do DFA-2D para ruídos rosa ($1/f$). O α recebe valor de aproximadamente 1.5, o que, para o DFA-1D revela sinais *brownianos*. Teoricamente o ruído rosa está exatamente entre o branco e o vermelho e essa teoria foi confirmada pelo DFA-2D.

É importante destacar que para ruídos, o DFA-2D retorna valores entre 1 e 2 e que, nesse caso, não há como derivar o expoente de Hurst.

O sinal fBm apresentou valor bastante próximo do esperado (≈ 2.5). O fGn dispersou um pouco do esperado (≈ 0.5), o que pode ser reflexo da metodologia utilizada para derivá-lo do fBm.

Tabela 5.2 - Valores de α (DFA-2D) e β (PSD-2D) para ruídos $1/f$, comparados com os valores para dados fGn e fBm com expoente de Hurst 0.5. A média e o desvio padrão são relativos a quinhentas amostras de cada modelo.

Sinal	**Parâmetro**	$\alpha_{médio}$	$\beta_{médio}$
fGn	Hurst = 0.5	0.4067 ± 0.0205	-0.9839 ± 0.0134
fBm	Hurst = 0.5	2.4794 ± 0.0676	2.7597 ± 0.094
Ruído Branco	$\beta = 0$	0.9992 ± 0.0207	-0.0009 ± 0.0083
Ruído Rosa	$\beta = 1$	1.4895 ± 0.0313	1.0022 ± 0.0080
Ruído Vermelho	$\beta = 2$	1.9820 ± 0.0434	2.0051 ± 0.0074

O gráfico *log-log* (figura 5.1) destaca a monofractalidade dos dados, já que a regressão linear é uma reta. Verifica-se também a diferença estrutural entre um ruído gaussiano

fracionário (fGn) e um ruído gaussiano ordinário (*white noise*). O fGn apresenta o mesmo valor da função de flutuação em todas as escalas, pois é um sinal estacionário.

Figura 5.1 - Gráfico *log-log* dos sinais típicos fBm e fGn com Hurst 0.5 e dos ruídos branco, rosa e vermelho. Os dados são médias de quinhentas amostras. As linhas e as equações destacadas referem-se ao fit dos dados. A inclinação dessa reta revela o valor do parâmetro α.

A tabela 5.2 apresenta ainda o valor do *power spectrum* 2D (PSD-2D - $\beta_{médio}$). Por ser um dado bi-dimensional, o expoente β é calculado por meio de uma média radial. E, neste caso, expressa a média do espectro de potência de Fourier 2D em todas as possíveis direções (PRUM et al., 1998). Para o cálculo dessa medida, o código Matlab distribuído por Ruzanski (2011) foi utilizado. Os gráficos do PSD-2D podem ser vistos na figura 5.2. Detalhes sobre o *power spectrum* podem ser vistos no Apêndice A.3.

Para os ruídos, o PSD-2D retorna o mesmo valor do parâmetro formador, ou seja, 0, 1 e 2 para os ruídos branco, rosa e vermelho, respectivamente (tabela 5.2). Para o fGn, o valor foi -1 e para o fBm, próximo de 2.7.

Comparando os valores de α e β, verifica-se que a relação $\beta \equiv 2\alpha - 1$ não se verifica. Tal relação foi validada experimentalmente, tanto para fBm, quanto para fGn unidimensionais (BULDYREV et al., 1995; HENEGHAN; MCDARBY, 2000).

Para sinais bi-dimensionais, pode-se dizer que, para os ruídos e sinais fGn, $\beta \equiv 2\alpha - 2$. Para sinais fBm, pode-se aproximar $\beta \equiv \alpha$. Para o sinal fBm com expoente de Hurst = 0.5, o espectro de potência radial 2D superestimou o valor de H e,

Figura 5.2 - Exemplos de espectro de potência radial para a)Ruído Branco ($\beta = 0$); b) Ruído Rosa ($\beta = 1$); c) Ruído Vermelho ($\beta = 2$); d) fGn (H=0.5); e)fBm (H=0.5).

consequentemente, o valor de α, já que o valor obtido para β foi 2.7, quando deveria ser igual a 2.5. Testes com outros valores de expoente de Hurst foram realizados e confirmaram a relação de $\beta \equiv \alpha$ para fBm bi-dimensional.

Considera-se que o DFA-2D foi capaz de classificar e caracterizar os diferentes regimes estudados, permitindo a identificação dos diferentes sinais. Outra contribuição importante é no discernimento teórico e prático do ruído branco. Foi possível perceber que sinais bi-dimensionais com $\beta = 0$ (ruído), $\alpha = 0.5$ (fGn) e $\alpha = 2.5$ (fBm), apesar de descorrelacionados, identificam, fisicamente, processos diferentes.

5.2 Redes de Mapas Acoplados - CML

Nesta sessão, mapas acoplados dos modelos Logístico, Dyadic Bernoulli e SOM serão avaliados, tanto com o DFA-2D, quanto com o MF-DFA2D. O objetivo é avaliar se os modelos avaliados apresentam ou não multifractalidade e, principalmente, estudar o significado do parâmetro α nesses casos.

A tabela 5.3 resume o conjunto de dados utilizado nessa seção. Para cada modelo foram geradas dez séries temporais, com trezentas iterações. Cada imagem tinha 500 linhas e 500 colunas.

Tabela 5.3 - Resumo dos dados utilizados nos testes com as redes de mapas acoplados.

Modelo	**Parâmetros**
Mapa Logístico	r = 4
Mapa Bernoulli	-
Mapa SOM	$\alpha = 0.6$, γ=0.2 $\beta = (\frac{0.8}{1+\alpha})$

Primeiro foi feita uma análise temporal. Os resultados apresentados são uma média das dez séries de cada modelo. O objetivo foi verificar se o DFA-2D era capaz de separar os diferentes modelos e, principalmente, detectar transições.

Na figura 5.3, observam-se as séries temporais de α para os três modelos. É nítido que o modelo Dyadic Bernoulli apresenta maior variação nas primeiras iterações, o que sugere que o modelo seja muito dependente da entrada. E, nesse sentido, o SOM é o menos afetado pela condição inicial na geração do modelo.

No mapa logístico, o valor do α sofre uma queda até por volta da iteração 50 e

Figura 5.3 - Análise temporal dos modelos de mapas acoplados com o DFA-2D. Na parte de baixo da figura, as séries com seus respectivos desvios padrão referente às dez amostras.

depois tende a subir linearmente, conforme o número de iterações aumenta. Até por volta da iteração 6, existe o regime fBm, com $\alpha > 2$. Posteriormente, as instâncias mantêm-se com α variando entre 1.3 e 1.7.

O mapa de Bernoulli apresenta regime fractal fBm até próximo da iteração 10. Também há um decaimento do valor de α até por volta da iteração 60. Depois disso, o α se estabiliza em torno de 1.2.

Portanto, tanto o mapa Logístico, quanto o de Bernoulli são caracterizados como ruídos pelo DFA-2D. O mapa Logístico aumenta o valor de α com o tempo e o de Bernoulli estabiliza-se depois da iteração 60.

O mapa SOM apresenta comportamento diferente dos demais. Há uma pequena queda no valor do α nas primeiras iterações, mas os valores vão subindo até que, por volta da iteração 20, estabiliza-se em torno de 2.3. Com a evolução temporal do sistema, o α tende a cair e se aproximar de 2. Dentro das trezentas iterações

analisadas, o mapa SOM produz uma rede que segue o regime fractal fBm. Na parte estável (depois da iteração 20), o expoente de Hurst aferido ($\hat{H}$) varia entre 0.05 e 0.3, o que caracteriza um sistema que possui memória de longo alcance, é anti-persistente e rugoso.

A tabela 5.4 apresenta as estatísticas gerais das séries para o valor de α. Nos três casos, os maiores valores ocorrem nas primeiras iterações.

Tabela 5.4 - Estatísticas básicas de α para os três modelos de mapas acoplados.

	Média	**Desvio Padrão**	**Mediana**	**Maior**	**Menor**	**Range**
Logístico	1.594	0.1854	1.595	2.953	1.336	1.617
Bernoulli	1.248	0.2098	1.185	2.974	1.156	1.818
SOM	2.246	0.0892	2.268	2.932	1.908	1.023

A figura 5.4 mostra as séries deslocadas em uma iteração no tempo. Verifica-se que os modelos são altamente diferenciáveis. Associado a essa informação visual, o teste estatístico de *Kruskal-Wallis* rejeita a hipótese de que os dados pertencem a distribuições iguais, com $p < 0.01$, assim como todos os testes de *Wilcoxon rank sum* entre os grupos.

Figura 5.4 - Série original versus série deslocada em uma unidade no tempo.

A seguir será apresentada uma análise de fractalidade para os mapas acoplados. Todas as imagens das séries foram processadas com o DFA-2D Multifractal (MFDFA-

2D). Um desafio encontrado nesse estudo foi em como avaliar a fractalidade e multifractalidade desse tipo de dado, visto que cada série possui trezentas imagens e analisa-las individualmente poderia tornar-se inviável. Sendo assim, foram analisadas algumas iterações, consideradas representantes do regime no entorno. As iterações 1, 5, 10, 15 e 25 foram escolhidas por estarem dentro do intervalo de tempo mais instável dos três mapas. A iteração 50 por se encontrar na região próxima a estabilização dos modelos e as iterações 100, 200 e 300 por estarem na região estável. A análise da iteração 1 é para servir como parâmetro de comparação, visto que os três modelos apresentam a mesma condição inicial (matriz gaussiana). As iterações analisadas são destacadas em amarelo na figura 5.5.

Figura 5.5 - Iterações escolhidas para análise de fractalidade dos modelos.

A figura 5.6 apresenta os resultados do MFDFA-2D para o mapa logístico. Na figura 5.6a, o gráfico de q versus $h(q)$ revela que há subconjuntos regidos por regimes diferentes nas iterações 1, 5, 10 e 15. Para essas iterações, o valor de $h(q)$ decresce com o aumento de q e, em todas elas, há uma estabilização do valor de $h(q)$ quando $q > 1$. A iteração 25 apresenta uma pequena variação, mas não chega a caracterizar multifractalidade. As iterações 50, 100, 200 e 300 apresentam comportamento totalmente linear para todos os momentos analisados. Todos os dados são médias das dez séries espaço-temporais de cada modelo.

Quando existe forte dependência não-linear entre $\tau(q)$ e q, há indícios de multifractalidade no dado (KANTELHARDT et al., 2002). Na figura 5.6b é possível verificar que as iterações 1, 5, 10, 15 e 25 apresentam uma relação não linear, o que caracteriza multifractalidade. A partir da iteração 50 não há regimes diferentes no sistema. A relação entre α e $f(\alpha)$ confirma os resultados (figura 5.6c).

Figura 5.6 - Resultado do MFDFA-2D para as iterações 1, 5, 10, 15, 25, 50, 100, 200 e 300 do mapa Logístico. Em a) Gráfico $q \times h(q)$. Em b) relação $q \times \tau(q)$. Em c) Espectro de singularidade. Os valores apresentados são médias das dez séries estudadas.

A tabela 5.5 apresenta os valores numéricos de $h(2)$, que equivale ao valor de α, a variação de $h(q)$ ($\Delta h(q)$) e a amplitude do espectro de singularidades ($\Delta\alpha$) para o mapa logístico. Os valores destacados na tabela garantem a inexistência de regimes diferentes nos dados. Para ser significativo, $\Delta h(q)$ e $\Delta\alpha$ precisam ser maiores que 1. Fan e Wu (2015) afirmam que esses dois parâmetros dão uma descrição completa das propriedades multifractais em séries temporais.

A figura 5.7 apresenta os resultados do MFDFA-2D para o mapa Dyadic Bernoulli. O comportamento do mapa de Bernoulli é muito parecido com o do mapa Logístico. E, da mesma, forma, é possível ver na tabela 5 que a partir da iteração 50 não há indícios de multifractalidade. As iterações 10 e 15 apresentam uma deformidade em sua curvatura no espectro de singularidades (Figura 5.7c), típica de sinais descorrelacionados. A iteração 1, por sua vez, apresenta a curvatura para cima, o que, segundo Fornari (2016), acontece quando o gráfico de q versus $\tau(q)$ possui uma

Tabela 5.5 - Valores numéricos para α, range da função de flutuação e amplitude do espectro de singularidades para as iterações analisadas no mapa Logístico.

Iteração	$\alpha(q=2)$	$\Delta h(q)$	$\Delta\alpha$
1	2.934	5.1227	5.5619
5	2.249	5.4592	5.9595
10	1.808	3.4116	3.8172
15	1.690	2.4620	2.8436
25	1.593	1.2847	1.6879
50	1.345	0.4508	0.7748
100	1.481	0.4976	0.8360
200	1.632	0.5170	0.8530
300	1.733	0.4758	0.7830

quebra na inclinação superior a 180 °.

Figura 5.7 - Resultado do MFDFA-2D para as iterações 1, 5, 10, 15, 25, 50, 100, 200 e 300 do mapa Dyadic Bernoulli. Em a) Gráfico $q \times h(q)$. Em b) relação $q \times \tau(q)$. Em c) Espectro de singularidade.

Tabela 5.6 - Valores numéricos para α, range da função de flutuação e amplitude do espectro de singularidades para as iterações analisadas no mapa Dyadic Bernoulli. Os valores apresentados são médias das dez séries estudadas.

Iteração	$\alpha(q=2)$	$\Delta h(q)$	$\Delta\alpha$
1	2.974	2.5773	4.1904
5	2.135	4.0686	4.5739
10	1.967	2.9413	3.4406
15	1.681	2.1681	205050
25	1.354	1.5578	1.8187
50	1.308	0.3768	0.5733
100	1.201	0.1738	0.3342
200	1.188	0.1678	0.3103
300	1.182	0.1822	0.3354

O mapa SOM tem comportamento bem diferente dos demais mapas analisados. Na tabela 5.7, os valores de $\Delta h(q)$ e $\Delta\alpha$ são grandes o suficiente para caracterizar as iterações do modelo SOM analisadas como multifractais. O valor de $\alpha(q=2)$ revela pertencer ao modelo fBm.

No gráfico de $q \times h(q)$ (figura 5.8a), as iterações 1, 5, 10 e 300 apresentam comportamento similar. Valores de *h(q)* praticamente constantes para $q < 0$. Próximo de zero há uma queda dos valores de *h(q)* e os valores estabilizam novamente para $q > 0$. Para as demais iterações (15, 25, 50, 100 e 200), *h(q)* tem uma elevação de valores para q próximo de zero. Para $q > 0$, todas as iterações mantêm-se estáveis na faixa de *h(q)* entre 2 e 3.

Sendo assim, *h(q)* mantêm-se quase constante para $q > 0$ e também para $q < 0$, mas com magnitudes diferentes. Esse comportamento faz com que o espectro de singularidades apresente faixas quase lineares nessas duas regiões. Um exemplo pode ser visto na figura 5.9.

Na falta da metodologia para realizar a análise de fractalidade nos mapas acoplados, considerando as diversas iterações nos modelos, os gráficos das figura 5.10, 5.11 e 5.12 sumarizam os resultados por meio de uma média entre todas as iterações pertencentes ao CML e também para as dez séries estudadas.

Na figura 5.10, a relação entre q e *h(q)*. Na figura 5.11, a relação entre q e $\tau(q)$ para cada modelo estudado e na figura 5.12, o espectro de singularidades.

Sendo assim, apesar de não ser possível caracterizar detalhadamente os regimes, os

Figura 5.8 - Resultado do MFDFA-2D para as iterações 1, 5, 10, 15, 25, 50, 100, 200 e 300 do mapa SOM. Em a) Gráfico $q \times h(q)$. Em b) relação $q \times \tau(q)$. Em c) Espectro de singularidade. Os valores apresentados são médias das dez séries estudadas.

Tabela 5.7 - Valores numéricos para α, range da função de flutuação e amplitude do espectro de singularidades para as iterações analisadas no mapa SOM.

Iteração	$\alpha(q=2)$	$\Delta h(q)$	$\Delta\alpha$
1	2.9316	7.5617	8.0154
5	1.9083	4.1158	4.6174
10	2.1413	3.8477	4.3565
15	2.2519	3.4781	3.9485
25	2.3159	3.2400	3.7272
50	2.3258	4.6731	2.8747
100	2.3067	5.5547	3.182
200	2.2538	7.6579	7.8571
300	2.0563	15.200	15.8670

Figura 5.9 - Espectro de singularidade característico para sinais que apresentam mono-fractalidade para $q < 0$ e $q > 0$. A interpretação física é que a série apresenta um comportamento de escala para $q < 0$ e outro para $q > 0$.

gráficos deixam claro que os mapas Logistico e de Bernoulli são muito parecidos e não exibem carácter multifractal. Já o mapa SOM tem um comportamento de escala bem diferente e apresenta carácter multifractal pela presença de flutuações em grandes e pequenas magnitudes.

Fan e Li (2015) analisaram alguns modelos de dados caóticos com o MFDFA-1D. Dentre eles, o mapa logístico com $r = 4$, cujo comportamento de escala foi idêntico ao do ruído branco ($H \approx 0.5$). Para a série 2D, o mapa logístico teve valor médio de $\alpha_{q=2} \approx 1.5$, o que caracteriza um ruído do tipo $1/f$ (ruído rosa).

Freitas (2012) utilizou o DFA-2D sobre um CML de mapa logístico com $r = 1.75$ e fator de acoplamento 0.3 que, segundo o autor, gera uma solução intermitente no mapa. O DFA-2D encontrou valor médio $\alpha_{q=2} \approx 0.8$ depois da iteração 10, o que configura um regime com memória de longo alcance persistente e estacionário (fGn). A análise da série temporal de α permitiu detectar a transição para o regime intermitente a partir da décima iteração. Assim como Freitas (2012), neste trabalho, a análise da série temporal de α permitiu detectar os momentos de mudanças no regime dos dados e também de separar um modelo do outro.

Figura 5.10 - Relação de h(q) e q média para os mapas estudados.

Figura 5.11 - Média dos valores de $\tau(q)$ para os mapas estudados.

Figura 5.12 - Espectro de singularidades médio para os mapas estudados.

O mapa de Bernoulli também pode ser caracterizado como ruído, com $\alpha_{q=2} \approx 1.2 \pm 0.2$. Ou seja, o regime pode ser caracterizado como um ruído rosa ou branco. Tanto o mapa Logístico, como o de Bernoulli apresentam as primeiras cinco iterações classificadas como fBm multifractal e, apesar dos modelos não serem fractais, o mapa Logístico apresenta maior amplitude no espectro do que o mapa de Bernoulli (figura 5.12).

O mapa SOM é classificado pelo DFA como um regime não estacionário (fBm) com correlação de longo alcance anti-persistente ($H \approx 0.3$) e multifractal. Dado os resultados encontrados por Fan e Li (2015) para CML 1D, o modelo SOM poderia ser comparado com o mapa de Lorentz, que também foi caracterizado como fBm ($\alpha_{q=2} \approx 1.2 \equiv H \approx 0.2$).

5.3 Considerações Finais

Com relação às questões levantadas no início do capítulo, pode-se dizer que todas foram respondidas. Foi provado que o DFA-2D é capaz de distinguir entre os diferentes sinais do tipo $1/f$ e que a consolidada relação entre α e β para dados unidimensionais não é válida para os bi-dimensionais.

Na análise de redes de mapas acoplados, como dito, a dificuldade é avaliar todas as iterações, ou pelo menos aquelas na parte instável do sistema. Neste trabalho definiu-

se como metodologia a escolha visual de algumas iterações para serem analisadas, a partir do estudo da série temporal de α. Os resultados encontrados na análise foram compatíveis com o que se espera dos modelos. Pode-se destacar, sobretudo, a caracterização multifractal do modelo SOM.

6 ANÁLISE E CLASSIFICAÇÃO DE GALÁXIAS SDSS UTILIZANDO DFA-2D MULTIFRACTAL

Nesta sessão, são apresentados os resultados do processamento de imagens de galáxias do *Sloan Digital Sky Survey* (SDSS) com o DFA-2D multifractal. O objetivo desse capítulo foi realizar um estudo de classificação de galáxias com o MF-DFA2D, utilizando para tanto, imagens de galáxias elípticas e espirais obtidas no projeto Galaxy Zoo. A hipótese testada foi que o MF-DFA2D é capaz de caracterizar e classificar imagens de galáxias elípticas e espirais, uma vez que o DFA revela a característica fractal dos sinais analisados e que já foi constatado na literatura que a morfologia das galáxias apresenta comportamento fractal.

Para fins comparativos, as imagens foram processadas com o índice de autocorrelação espacial Moran (MORAN, 1950) e com os descritores texturais de Haralick (HARALICK et al., 1973).

O poder de classificação do expoente α e de alguns descritores texturais foi avaliado utilizando o algoritmo de classificação de árvore de decisão J48 no WEKA (WAIKATO, 2016).

6.1 Métodos de Classificação de Galáxias

Em um trabalho pioneiro, Hubble (1936) propôs uma classificação de galáxias baseada em sua morfologia (Figura 3.5). No entanto, a classificação de galáxias no esquema proposto por Hubble é uma tarefa difícil, devido a dois motivos principais. O primeiro é a diversidade de classes possíveis no esquema Hubble. O segundo está relacionado a qualidade das imagens obtidas, que são muito ruidosas. Além disso, com o advento de modernos telescópios, uma quantidade enorme de dados de objetos celestiais tem sido capturada, tornando indispensável o estudo de metodologias que sejam capazes de classificar imagens de galáxias automaticamente com base em sua morfologia visual. E, de fato, diversas metodologias estão sendo desenvolvidas. Exemplos podem ser vistos em (BAZELL; AHA, 2001; ELFATTAH et al., 2014; GAUCI et al., 2010; SCHUTTER; SHAMIR, 2015; SHAMIR, 2009; PEDERSEN et al., 2013).

Em seu trabalho, Schutter e Shamir (2015) confirmam que são necessários diversos descritores de imagem para classificar as galáxias no esquema Hubble, mas que os descritores texturais têm mostrado carácter informartivo importante na separação de galáxias por sua morfologia. Para Thanki et al. (2009), é possível verificar visualmente diferença textural nas imagens de galáxias, uma vez que as elípticas parecem

ser menos rugosas que as espirais. Pedersen et al. (2013) afirmam que as diferentes texturas são causadas justamente pelas diferentes distribuições de estrelas e gás no interior da galáxia.

Pham (2010) afirma que a textura é um conceito fácil de reconhecer, mas muito difícil de definir e que, em termos gerais, a textura refere-se às características e aparência da superfície de um dado objeto, dado seu tamanho, forma, densidade, arranjo e proporção dos seus elementos fundamentais. Existem diferentes formas de analisar a textura em imagens digitais, tais como os métodos de Wavelet, Fourier e a teoria de Texton (GEBEJES; HUERTAS, 2013). No entanto, uma análise que se aproxime da forma como o sistema visual humano percebe a textura é mais interessante. Essa análise, mais simples, foi definida por Haralick e colaboradores em 1973 (HARALICK et al., 1973), e ainda é amplamente utilizada. Por meio de uma estatística de segundo grau, os descritores de Haralick capturam informações texturais dos objetos analisados.

Por outro lado, alguns autores sugerem que o uso de operadores fractais pode ser útil na distinção entre galáxias elípticas e espirais. Uma vez que a imagem de galáxia apresenta estrutura não estacionária no espaço e varia com a mudança de escala, o uso de uma abordagem fractal parece apropriado. A teoria dos fractais tem sido utilizada em estudos sobre a distribuição das galáxias pelo Universo (BALIAN; SCHAEFFER, 1988; GAITE; DOMÍNGUEZ; PÉREZ-MERCADER, 1999; PEEBLES, 1989; PIETRONERO, 1987) e também na classificação de galáxias (LEKSHMI; REVATHY; PRABHAKARAN NAYAR, 2003; SCHUTTER; SHAMIR, 2015; SHAMIR, 2009; THANKI; RHEE; LEPP, 2009).

Segundo Thanki et al. (2009), a motivação de utilizar técnicas fractais na classificação de galáxias é justamente o fato de ser possível verificar visualmente diferença textural nas imagens, uma vez que as elípticas parecem ser menos rugosas que as espirais. Neste sentido, os operadores texturais e fractais se complementam. De fato, operadores fractais têm obtido sucesso como descritores texturais (MYINT, 2003; SARKAR; CHAUDHURI, 1992). Os operadores texturais tendem a captar melhor as diferentes rugosidades na superfície das galáxias. Os operadores fractais possuem capacidade de classificá-las por seu regime formador. Na literatura, os trabalhos de classificação de galáxias em geral usam tanto os operadores texturais quanto os fractais.

Estudos demonstram o uso de diferentes operadores texturais na classificação de imagens de galáxias (AU et al., 2006; SCHUTTER; SHAMIR, 2015; GAUCI et al., 2010;

BAZELL; AHA, 2001; PEDERSEN et al., 2013) . Alguns desses trabalhos (SCHUTTER; SHAMIR, 2015; BAZELL; AHA, 2001) utilizaram os chamados descritores texturais de Haralick (HARALICK et al., 1973) em suas metodologias e obtiveram bons resultados. Por meio de uma estatística de segundo grau, os descritores de Haralick capturam informações texturais dos objetos analisados. Tais descritores são detalhados no Apêndice A.1.

Lekshmi et al. (2003) utilizaram assinatura fractal para classificar galáxias elípticas e espirais do *Nearby Galaxy Catalog*. No total, os autores utilizaram quatorze imagens de galáxias elípticas e noventa de espirais. Utilizando uma amostra de três galáxias elípticas e três espirais, foram geradas assinaturas fractais. Posteriormente, os algoritmos de vizinho mais próximo e redes neurais artificiais foram utilizados para classificar as demais, obtendo uma taxa de acerto de 92% e 95%, respectivamente. Os resultados mostraram que a assinatura fractal é uma técnica promissora na classificação de galáxias.

Thanki et al. (2009) processaram imagens de galáxias do SDSS com duas medidas fractais, a *capacity dimension* e a *correlation dimension* e concluíram que apenas a informação fractal não é suficiente para classificar as imagens, mas que em conjunto com outros parâmetros, pode ser altamente discriminatório. Tal afirmação é confirmada por Shamir (2009). Thanki et al. (2009) concluíram ainda que a dimensão fractal das galáxias elípticas é menor que a das espirais devido às formas menos complexas das elípticas. Característica corroborada por Schutter e Shamir (2015).

Shamir (2009) afirma ainda que descritores fractais podem detectar características na forma fractal de uma galáxia espiral que não se espera que existam em uma galáxia elíptica e que a fractalidade da galáxia espiral pode ser percebida visualmente.

Nos trabalhos acima citados, diferentes técnicas de obtenção da dimensão fractal foram utilizadas, mas nenhum utilizou o método DFA. Sendo assim, a motivação desse capítulo foi justamente verificar a resposta do DFA na classificação de galáxias elípticas e espirais. A escolha pela versão multifractal veio, na verdade, da observação dos gráficos *log-log* entre s e $F(s)$, que apresentavam alguns pontos de *crossover*. Os resultados são apresentados na próxima seção.

6.2 DFA-2D Multifractal na classificação de galáxias SDSS

Como relatado na seção 3.2, mil amostras de galáxias elípticas e mil de galáxias espirais foram obtidas no catálogo de imagens do projeto Galaxy Zoo. As imagens

passaram por um pré-processamento, onde foram recortadas de forma que a galáxia ficasse no centro da imagem. Após processar as duas mil imagens com o MFDFA-2D, verificou-se que algumas apresentaram valores muito discrepantes e foram removidas da amostra. Isso ocorreu porque o recorte é feito com base em informações sobre a localização da galáxia no céu e, por algum motivo, alguns falharam, conforme pode ser visto na figura 6.1.

Figura 6.1 - Amostras de imagens com erros que acarretaram em valores espúrios dos descritores de Haralick.

A tabela 6.1 apresenta os resultados numéricos do MF-DFA2D após a remoção dos dados espúrios. Verifica-se que tanto as galáxias elípticas, quanto as espirais apresentaram valor de $\alpha_{h(q=2)}$ entre 2 e 3, característica de modelo fBm. Na média, as galáxias elítpicas apresentam valor de $\alpha_{h(q=2)}$ menor que as espirais. O valor de $\alpha_{h(q=2)}$ para as galáxias espirais aproximou-se bastante de $\alpha_{h(q=2)} \approx 2.5$, o que caracteriza um sinal sem memória.

Tabela 6.1 - Valores de $\alpha_{h(q=2)}$ para as galáxias elípticas e espirais após a remoção dos dados espúrios.

	Amostras	**Média**	**Std**	**Mediana**	**Moda**	**Range**
Elíptica	902	2.3565	0.0989	2.3605	2.3181	0.4794
Espiral	872	2.4983	0.1217	2.5122	2.3907	0.5561

Na figura 6.2 são apresentados exemplos do processamento com o MF-DFA2D de uma imagem de galáxia elíptica (6.2a) e uma espiral (6.2b). Na primeira coluna da figura 6.2 apresenta-se o gráfico de q x h(q). O fato desse gráfico não ser linear,

indica que *h(q)* depende/varia com q, o que é sinal de multifractalidade. É possível verificar um pico de valores entre -5 e 5, aproximadamente. Fora desse intervalo, verifica-se que há uma certa estacionaridade dos valores de *h(q)*.

Figura 6.2 - Resultado do MF-DFA-2D para as galáxias elípticas (a) e espirais (b). Na primeira coluna, o gráfico $q \times h(q)$, na coluna do meio, o gráfico $q \times \tau(q)$ e na terceira coluna, o espectro de singularidades.

Na coluna do meio da figura 6.2 são apresentados os gráficos de $q \times \tau(q)$. É possível verificar que a relação também não é monotônica e que tal comportamento é mais presente nas galáxias elípticas do que nas espirais. Na terceira coluna são apresentados os espectros de singularidade. Neles, é possível ver um padrão chamado por alguns autores de *'zigzag'* e por outros de *'torção'*. Esse padrão leva o espectro a ter valores de $f(\alpha)$ maiores que o valor máximo esperado e acontece por dois motivos. O primeiro é devido a uma variação pequena de valores em torno de um *h(q)*, que quando passa pela transformada de Legendre, acarreta nessa deformação no topo da curva. Exemplo típico acontece sobre dados monofractais. O outro motivo é justamente pelo fato da relação entre q e *h(q)* não seguir um padrão monotônico decrescente (CZARNECKI; GRECH, 2010; MAKOWIEC et al., 2011).

Ludescher et al. (2011) verificaram empiricamente que dados unidimensionais multifractais acrescidos de ruídos brancos aditivos e tendências sazonais aditivas apresentaram variações espúrias no padrão entre q e *h(q)* semelhantes às encontradas para as galáxias.

Grech (2016) encontrou formas do perfil *h(q)* parecidos ao apresentado na primeira coluna da figura 6.2 em séries unidimensionais financeiras após a ocorrência de uma quebra (*crash*), no período nomeado pelo autor de zona dinâmica, onde há alta volatilidade nos dados. Czarnecki e Grech (2010) também encontraram esse padrão irregular no espectro de singularidades de dois índices do mercado de ações.

Makowiec et al. (2011) afirmam que no caso de formas irregulares no espectro multifractal, o mesmo precisa ser reinterpretado. Os autores sugerem uma análise dos dados sem a integração, pois, é sabido que o espectro de um sinal unidimensional integrado é idêntico ao do sinal sem integração, deslocado de 1 para a direita.

Testes empíricos realizados com o DFA-2D revelaram que o parâmetro α de dados fBm torna-se mais preciso sem utilizar o perfil oriundo da integração do sinal e tem seu valor decrescido de 2. Os mesmos testes revelaram que essa característica é válida apenas para dados fBm. Dados fGn e ruídos apresentam valores errôneos sem a integração, uma vez que o DFA opera sobre dados fBm (seção 4.2).

Neste sentido, considerando que os valores médios de α revelaram padrões fBm nas imagens de galáxias elípticas e espirais, as mesmas foram reprocessadas com o MF-DFA2D sem realizar o primeiro passo do algoritmo, que é a criação do perfil Y. Se a aproximação do expoente de Hurst para o sinal integrado e para o sinal sem integração forem semelhantes, significa que o dado é fBm e monofractal. Concordância entre os valores do expoente de Hurst e o valor de H_{max}[1] também é uma propriedade dos sinais monofractais (MAKOWIEC et al., 2011). Os resultados são apresentados na tabela 6.2.

Tabela 6.2 - Valores do expoente de Hurst ($H_{q=2}$) e do $Hmax(f(\alpha) = 2)$ obtidos pelo MF-DFA2D aplicado às imagens de galáxias elípticas e espirais. O processamento foi feito com dados integrados e sem integração (*Raw*).

	Hurst		H_{max}
	Raw	**Integrado**	**Integrado**
Elíptica	0.7688 ± 0.1062	0.3565 ± 0.0989	2.6875 ± 0.2159
Espiral	0.5212 ± 0.2010	0.4983 ± 0.1217	2.5417 ± 0.5028

Verifica-se na tabela 6.2 que para galáxias espirais houve concordância entre os

[1] H_{max} representa a singularidade mais provável do espectro de singularidades e é obtido como a média dos pontos no valor máximo do espectro que, para superfícies bi-dimensionais, é dado para $f(\alpha) = 2$.

resultados, o que sugere que as galáxias espirais seguem o modelo fBm, com expoente de Hurst próximo de 0.5, característico de um sinal não correlacionado.

As galáxias elípticas apresentaram expoentes de Hurst muito divergentes para os dados bruto e integrado. Esse comportamento pode indicar a presença de multifractalidade no sinal, apesar do formato irregular do espectro de singularidade.

A tabela 6.3 apresenta os principais indicadores de sinais multifractais obtidos para as galáxias elípticas e espirais. Fan e Li (2015) afirmam ΔH e ω[2] dão uma descrição completa das propriedades multifractais em séries temporais. $\Delta f(\alpha)$ complementa essa descrição, fornecendo o conteúdo de singularidades das séries. Os dados da tabela 6.3 reforçam a hipótese de que as galáxias espirais são monofractais e as elípticas são multifractais, uma vez que, para caracterizar significativamente comportamento multifractal, $\Delta h(q)$ e ω precisam ser maiores que 1.

Tabela 6.3 - Principais indicadores multifractais obtidos com o MF-DFA2D.

	ΔH	α_{min}	α_{max}	ω	$\mathbf{f}(\alpha_{min})$	$\mathbf{f}(\alpha_{max})$	$\Delta\mathbf{f}(\alpha)$
Elíptica	1.135	1.556	2.886	1.330	0.167	2.043	1.876
Espiral	0.736	1.759	2.665	0.907	0.081	2.035	1.955

Na presença de multifractalidade, Czarnecki e Grech (2010) afirmam que o DFA padrão ($H_{q=2}$) revela a regra de escala principal (aquela que possui as maiores ou mais frequente flutuações). No entanto esse valor pode ser distorcido devido às características do sinal. Sendo assim, considerando que é concenso na literatura que as galáxias elítpticas são menos rugosas que as espirais, avalia-se que o valor de H_{max} representa melhor a superfície das galáxias elípticas, uma vez que expoente de Hurst próximo a 0.6 retrata uma superfície com correlação de longo alcance persistente, menos rugosa do que uma superfície com expoente de Hurst próximo a 0.5.

Dada a relação $D \equiv 3 - H$, obtêm-se dimensão fractal média de 2.40 para as galáxias elípticas e 2.50 para as espirais. Tal resultado corrobora os de Thanki et al. (2009), no sentido de que a dimensão fractal das galáxias elípticas são menores que as das espirais. Tais autores reforçam, no entanto, que a escolha das imagens tem grande influência na efetividade da classificação das galáxias usando fractais. Ao traçar a assinatura fractal de galáxias elípticas e espirais, Lekshmi et al. (2003) concluíram que as espirais apresentavam comportamento mais caracteristicamente fractal do

[2] $\omega = \alpha_{max} - \alpha_{min}$

que as elípticas, o que pode ser corroborado com os resultados desse trabalho, uma vez que as elípticas foram classificadas como multifractais.

6.2.1 Comparação do MFDFA-2D com outros descritores texturais

Para fins de comparação, as imagens das galáxias foram processadas com o índice de correlação espacial de Moran e descritores texturais de Haralick. Detalhes das técnicas e das implementações utilizadas podem ser vistos no Apêndice A.2 e A.1, respectivamente.

O índice de Moran médio para as galáxias elípticas foi 0.9680 ± 0.0268 e para as espirais foi 0.9521 ± 0.0470. Esse resultado mostra que, na média, as galáxias elípticas são mais correlacionadas espacialmente do que as espirais; o que, em termos de textura, significa que as elípticas são menos rugosas que as espirais.

(HARALICK et al., 1973) definiram 14 descritores, a saber: Segundo Momento Angular (SMA), contraste, correlação, variância, homogeneidade, média da soma (*sum average*), variância da soma (*sum variance*), entropia da soma (*sum entropy*), entropia, variância da diferença (*difference variance*), entropia da diferença (*difference entropy*), medidas de correlação 1 (IMC-1) e 2 (IMC-2) e coeficiente máximo de correlação. O descritor coeficiente máximo de correlação é uma métrica pouco utilizada, tanto por seu custo computacional, quanto por sua instabilidade e não foi implementado.

Os descritores texturais refletem as mudanças regulares nos valores dos níveis de cinza em uma imagem. Tais mudanças são correlacionadas tanto estatisticamente, quanto espacialmente. Detalhes dessas medidas podem ser encontradas em (HARALICK et al., 1973; PHAM, 2010).

A tabela 6.4 apresenta os resultados de Haralick de treze descritores texturais de Haralick para as imagens de galáxias elípticas e espirais. Os valores obtidos pelos descritores foram compatíveis com o que se espera de imagens de galáxias elípticas (formas menos complexas e visualmente mais lisas) e galáxias espirais (formas mais complexas e visualmente mais rugosa).

Os descritores SMA e Homogeneidade são descritores de lisura, ou seja, quanto mais alto os valores, maior a homogeneidade da imagem. Verifica-se que ambos descritores acusaram valores mais altos para as galáxias elípticas. A entropia é uma medida de desordem e, quanto maior seu valor, mais rugosa a imagem deve ser. Nesse caso, o valor da entropia para a galáxia espiral foi maior que para a galáxia elíptica.

Tabela 6.4 - Resultado dos descritores texturais de Haralick para as galáxias elípticas e espirais

	Eliptica		Espiral	
	Média	**Desvio**	**Média**	**Desvio**
SMA	0.076	0.035	0.015	0.012
Contraste	9.840	4.542	43.001	43.369
Entropia	3.611	0.514	5.568	0.627
Variancia	187.102	77.419	519.370	280
Correlacao	0.972	0.009	0.958	0.038
Homogeneidade	0.638	0.089	0.565	0.102
Média da Soma	13.660	7.039	58.101	27.830
Entropia da Soma	2.972	0.318	4.310	0.430
Variância da Soma	783.566	306.646	2034.50	1102.8
Variância da Diferença	9.028	3.728	30.394	27.910
Entropia da Diferença	1.061	0.224	1.756	0.377
IMC_1	-0.510	0.096	-0.488	0.104
IMC_2	0.952	0.029	0.977	0.028

A entropia da soma e a entropia da diferença são medidas de aleatoriedade contida na imagem (PARK; CHEN, 2001). Tais descritores mostraram diferença significativa entre os tipos de galáxias, especialmente o primeiro. Para esses descritores, a galáxia espiral apresentou maiores valores, indicando, portanto, conter maior aleatoriedade espacial.

Os descritores que apresentaram desvio padrão muito alto foram descartados (contraste, variância, média da soma, variância da soma e variância da diferença), assim como aqueles que apresentaram valores muito próximos para as galáxias elípticas e espirais, sendo pouco significativos na separação entre ambas (correlação, IMC-1 e IMC-2).

Os descritores de correlação medem a dependência linear dos valores de nível de cinza na matriz de co-ocorrência. Para Baraldi e Parmiggiani (1995), o descritor de correlação só deve ser usado quando é esperada correlação linear entre os elementos da matriz.

Sendo assim, os descritores SMA, Entropia, Homogeneidade, Entropia da Soma e Entropia da Diferença foram selecionados para serem analisados com o algoritmo de árvore de decisão.

De forma geral, os descritores foram fiéis ao que se esperava como resultado, uma

vez que, como já citado, as diferentes formações internas dos constituintes das galáxias refletem em diferentes padrões visuais. Tais padrões foram capturados pelos descritores texturais de Haralick.

Para testar o poder discriminatório do α, utilizou-se o algoritmo de classificação por árvore de decisão J48, uma implementação própria do *software* Weka do algoritmo C4.5 (QUINLAN, 1992).

Os valores numéricos obtidos pelo DFA e pelos descritores de Haralick foram discretizados utilizando o filtro *discretize* do Weka, com 3 classes. Esse filtro cria classes nominais a partir dos valores numéricos dos atributos. O parâmetro *useEqualFrequency* também foi ativado. Esse parâmetro faz o balanceamento das amostras entre as classes. Dessa forma, as classes ficam com uma quantidade de amostras parecida.

Os resultados da classificação pelo algoritmo J4.8 podem ser vistos na tabela 6.5. Os valores apresentados são resultados de validação cruzada com 10 pastas. Foram realizados os seguintes testes:

- DFA com perfil Integrado
  - α
  - α juntamente com todos os descritores de Haralick
  - α juntamente com os descritores de Haralick selecionados (SMA, Entropia, Homogeneidade, Entropia da Soma e Entropia da Diferença)
- DFA sem integração do Perfil (Dados *Raw*)
  - α
  - α juntamente com todos os descritores de Haralick
  - α juntamente com os descritores de Haralick selecionados
- Haralick
  - Todos os descritores de Haralick
  - Apenas os descritores de Haralick selecionados

Em relação aos perfis integrado e *Raw*, verifica-se que o processamento sem a integração do perfil obteve melhores resultados de classificação. O uso conjunto do α com os descritores de Haralick melhoram substancialmente a classificação. A utilização de todos os descritores texturais de Haralick ou apenas os selecionados altera muito pouco a taxa de acerto das galáxias elípticas. No entanto, a utilização de todos os descritores melhora a classificação das galáxias espirais.

A melhor taxa de acerto obtida nos experimentos relatados na tabela 6.5 foi utili-

Tabela 6.5 - Matriz de confusão da classificação de galáxias elípticas e espirais com o algoritmo de árvore de decisão J48.

α - Perfil Integrado

	Elípticas	**Espirais**	**% Acerto**	**Média de Acerto**	**Kappa**
Elípticas	794	108	88.02%	72.04%	0.4376
Espirais	388	484	55.50%		

α - Perfil Integrado + Haralick

	Elípticas	**Espirais**	**% Acerto**		
Elípticas	858	44	95.13%	93.18%	0.8635
Espirais	77	795	91.17%		

α - Perfil Integrado + Haralick Selecionado

	Elípticas	**Espirais**	**% Acerto**		
Elípticas	794	108	88.03%	89.46%	0.7892
Espirais	79	793	90.94%		

α - Dados Raw

	Elípticas	**Espirais**	**% Acerto**		
Elípticas	846	56	93.79%	77.85%	0.5544
Espirais	337	535	61.35%		

α - Dados Raw + Haralick

	Elípticas	**Espirais**	**% Acerto**		
Elípticas	866	36	96.01%	92.90%	0.8578
Espirais	90	782	89.68%		

α - Dados Raw + Haralick Selecionado

	Elípticas	Espirais	% Acerto		
Elípticas	828	74	91.80%	90.98%	0.8195
Espirais	86	786	90.14%		

Haralick

	Elípticas	**Espirais**	**% Acerto**		
Elípticas	861	41	95.46%	93.35%	0.8668
Espirais	77	795	91.17%		

Haralick Selecionado

	Elípticas	**Espirais**	**% Acerto**		
Elípticas	879	23	97.45%	87.99%	0.7590
Espirais	190	682	78.21%		

H_{max} + Haralick

	Elípticas	Espirais	% Acerto	Média de Acerto	Kappa
Elípticas	851	51	94.35%	93.58%	0.8714
Espirais	63	809	92.77%		

SMA + H_{max} + α_{Int} + α_{Raw}

	Elípticas	**Espirais**	% Acerto		
Elípticas	812	90	95.13%	90.93%	0.8185
Espirais	71	801	91.17%		

zando os treze descritores de Haralick em conjunto com o H_{max} (93.58%). O resultado foi muito similar aos encontrados utilizando apenas o Haralick e o perfil integrado em conjunto com o Haralick. No entanto, verifica-se que a combinação do H_{max} com os descritores de Haralick forneceram a melhor taxa de acerto, inclusive para as galáxias espirais.

Apesar desse modelo ter melhor acurácia as regras de classificação ficaram muito mais complexas. O mesmo acontece para os modelos gerados sem o uso do filtro *discretize*. As taxas de acerto melhoram, mas os modelos ficam muito complexos, com o mesmo atributo repetindo-se várias vezes no mesmo ramo da árvore.

Sendo assim, o modelo do DFA sem integração com os descritores selecionados (α - Dados Raw + Haralick Selecionado), apesar de ter taxa de acerto um pouco menor (90.98%), apresenta a melhor relação entre acerto de classificação e complexidade do modelo. A figura 6.3 mostra a árvore gerada por esse modelo.

Figura 6.3 - Modelo de classificação de galáxias SDSS gerado pelo algoritmo J4.8 (α_{Raw} + descritores de Haralick selecionados).

Na figura 6.3 verifica-se que o algoritmo desprezou os descritores Entropia, Entropia da Diferença e Homogeneidade, utilizando apenas α_{Raw}, SMA e Entropia da Soma,

sendo SMA o atributo que oferece a maior separabilidade entre as classes e, portanto, aparece na raiz da árvore de decisão. O valor de separação do α_{Raw} é coerente com o que foi previamente apresentado. As espirais apresentam valores de α mais baixos do que as elípticas.

Segundo Baraldi e Parmiggiani (1995) a homogeneidade combina as características de SMA e contraste, sendo pouco efetivo e, portanto, deve ser evitado; e que o SMA deve ser preferido em relação à entropia. Essas características podem ter contribuído para o descarte dos atributos de homogeneidade e entropia pelo J4.8.

Utilizando descritores de Haralick para classificar imagens de galáxias, Schutter e Shamir (2015) fizeram uso dos descritores SMA e Entropia. Bazell e Aha (2001) utilizaram entropia e o descritor denominado por eles de P_{max}, que é o maior valor da matriz de co-ocorrência normalizada. Em ambos os trabalhos, as imagens de galáxias são do SDSS. No entanto, não há relato das razões de terem escolhido tais atributos.

Os modelos com o índice de Moran foram gerados, mas obtiveram baixas taxas de acerto e foram descartados.

Os descritores de Haralick são simples de serem calculados, mas podem ter um custo computacional elevado. O uso de algoritmos de classificação como o J4.8, que permite visualizar facilmente as regras geradas e os atributos desprezados, também contribui para reduzir a dimensionalidade dos dados e, consequentemente, reduzir o esforço de computação. Esse é um ponto importante desse trabalho, uma vez que os catálogos de imagens de galáxias são vistos como *big data*. Há uma grande quantidade de dados já no acervo e todos os dias novas galáxias são inseridas. Por isso ter uma metodologia que faça essa classificação automatica e rapidamente é tão relevante.

6.3 Considerações Finais

Apesar de os descritores texturais de Haralick separarem melhor galáxias elípticas de espirais, o operador fractal dado pelo DFA revela mais informações sobre os alvos do que classificadores texturais simples. Ele é capaz de caracterizar o modelo que rege o sistema e não apenas classificá-lo.

A forma não comportada dos gráficos $h(q)$, $\tau(q)$ e do espectro de singularidades é relatada por alguns autores como comum em dados reais e sempre configura-se como um desafio a mais na interpretação dos resultados. Acredita-se que o ruído presente nas imagens são causadores desse comportamento. Neste caso, computar e avaliar o

expoente α com e sem a integração do sinal pareceu ser uma boa metodologia, visto que a concordância entre os resultados caracteriza sinais fBm monofractais.

Corroborando com Shamir (2009) e Thanki et al. (2009), observou-se que a taxa de acerto da classificação utilizando apenas da informação fractal foi inferior àquela utilizando outros parâmetros texturais. No entanto, vale ressaltar mais uma vez, que técnicas fractais caracterizam o sistema, além de classificá-los e, nesse caso, conclui-se que o MF-DFA2D pode ser útil no estudo de morfologia de galáxias. E, como relatado na literatura, para classificar galáxias com base na morfologia visual do esquema Hubble são necessários diversos atributos, e não apenas os texturais e fractais. Nesse trabalho, a classificação foi feita em apenas duas classes : elípticas e espirais.

Algo interessente de ser notado é que o J4.8 selecionou os atributos SMA, que é um descritor de lisura; e entropia da soma que é um descritor de aleatoriedade. Como o MFDFA-2D caracterizou as galáxias espirais como sinais descorrelacionados, conclui-se que as informações são de fato complementares.

7 ANÁLISE DA RUGOSIDADE DE RELEVO UTILIZANDO DFA-2D

Este capítulo visa analisar o potencial do DFA-2D monofractal como quantificador de rugosidade de relevo utilizando, para tanto, os modelos digitais de elevação SRTM e GDEM descritos na secao 3.3. Implicaçõs práticas dessa caracterização do relevo estão relacionadas a aplicações de energia eólica, uma vez que os regimes de vento são fortemente influenciados pelas características locais de topografia e rugosidade do terreno.

Os MDEs também foram processados com os descritores texturais de Haralick e os resultados comparados aos do DFA-2D. Foi investigada a relação entre esses operadores e o coeficiente de rugosidade aerodinâmico, que é um parâmetro essencial nos modelos de extrapolação de vento.

Neste capítulo, o DFA-2D foi utilizado simplesmente como operador textural. Não houve a intenção de analisar e/ou diferenciar a origem do sinal, como na classificação das imagens de galáxias (Capítulo 6). O objetivo foi averiguar a efetividade do DFA-2D como um aferidor da rugosidade efetiva do relevo.

7.1 Vento e Superfície

A busca por fontes renováveis para geração de energia elétrica tem aumentado nos últimos anos, incentivada especialmente por preocupações com o meio ambiente, com os altos preços de combustíveis fósseis e com o aumento de consumo em regiões industriais e residenciais. Países do mundo inteiro têm incentivado e implementado soluções alternativas para geração de energia.

Neste cenário, a energia eólica tem se apresentado como alternativa promissora. Denomina-se energia eólica a energia cinética contida nas massas de ar em movimento (vento). Seu aproveitamento ocorre por meio de aerogeradores, nos quais a força do vento é captada por hélices ligadas a uma turbina que aciona um gerador elétrico. A quantidade de energia transferida é função da densidade do ar, da área coberta pela rotação das pás (hélices) e da velocidade do vento (MANWELL et al., 2002).

O vento é o deslocamento do ar causado pela diferença de pressão atmosférica proveniente da radiação solar que, por aquecer de forma não uniforme a superfície terrestre, cria os gradientes de pressão geradores dos ventos. O deslocamento é influenciado por efeitos locais como a rugosidade da superfície e orografia. Os ventos possuem ainda outras forças moduladoras que dão origem a uma componente hori-

Figura 7.1 - Perfil vertical da velocidade do vento desde a superfície até a altura do vento geostrófico. O comprimento de rugosidade (z_0) é a altura onde a velocidade é nula. FONTE : Adaptado de (MARTINS et al., 2008)

zontal e uma componente vertical (COSTA; LYRA, 2012; FIORIN et al., 2008).

A atmosfera da Terra é comumente dividida em camadas: troposfera, estratosfera, mesosfera e exosfera, sendo a troposfera a que se encontra mais próxima à superfície. A camada limite atmosférica (CLA) é a região da troposfera diretamente afetada pela presença da superfície da Terra.

Os ventos mais fortes, constantes e persistentes ocorrem em regiões situadas a cerca de 10 km acima da superfície da terra, onde o atrito com a superfície pode ser desprezado, mas onde não é possível a instalação de aerogeradores.

Sendo assim, o espaço de interesse para estudos de energia eólica encontra-se limitado a regiões da CLA onde o vento é diretamente afetado pela fricção na superfície, resultando em tensão de cisalhamento turbulenta à superfície. Essas forças de fricção fazem com que a velocidade do vento diminua sua intensidade à medida que se aproxima do solo, dando origem ao chamado perfil vertical de velocidade (CASTRO, 2005; LYRA, 2005).

A velocidade do vento dentro da CLA é, portanto, uma função da altura acima do solo, começando de zero na superfície até a velocidade do vento geostrófico, formando o perfil de velocidade do vento (figura 7.1).

Figura 7.2 - Perfil vertical do vento e a influência da rugosidade da superfície.

Cada tipo de rugosidade perturba o vento de um jeito diferente. Quanto mais rugosa a superfície, maior a turbulência, o que acarreta em maior variação da velocidade do vento com a altura. Por outro lado, rugosidades baixas fazem com que a variação de velocidade seja mínima, o que permite a instalação de torres anemométricas mais baixas (figura 7.2). Maiores detalhes sobre o impacto da instabilidade atmosférica, da rugosidade da superfície e da forma do terreno no perfil do vento podem ser obtidos em Manwell et al. (2009).

O Brasil possui um alto potencial eólico ainda pouco explorado, exceto por algumas iniciativas isoladas. Segundo o Ministério de Minas e Energia, apenas 4.7% (cerca de 21 GWh) da potência gerada atualmente no país advém da energia eólica (MME, 2016). No entanto, o Atlas do Potencial Eólico Brasileiro, elaborado pelo Centro de Pesquisas de Energia Elétrica (Cepel), mostra um potencial bruto de 143.5 GW, o que torna a energia eólica uma alternativa complementar importante à fonte hidrelétrica, predominante no sistema brasileiro.

A instalação de um parque eólico não é tarefa simples. As características operativas da usina eólica são fortemente dependentes do regime de vento local, que varia continuamente em função do tempo e da estação do ano, da altitude, da topografia do terreno, dos obstáculos no entorno da turbina, da condição meteorológica, entre outros (OLAOFE; FOLLY, 2013).

Além disso, a falta de medições de velocidade do vento em alturas relevantes para exploração de energia torna necessária a extrapolação dessas informações. Ou seja,

o potencial do vento estimado para turbinas que operem a 50-100m de altitude é normalmente avaliado a partir de velocidades do vento medidas entre 10 e 40 metros. Neste sentido, segundo Gualtieri e Secci (2011), a extrapolação da velocidade do vento é provavelmente um dos fatores mais críticos na avaliação da energia eólica em um determinado local.

Segundo Assireu et al. (2013), o relevo, em termos aerodinâmicos, é descrito pelos parâmetros de rugosidade, relacionados com a forma, altura e densidade de distribuição dos elementos na superfície. Um dos atributos aerodinâmicos utilizados para parametrizar a rugosidade do terreno é o coeficiente de rugosidade aerodinâmica (z_0), também chamado de comprimento de rugosidade superficial. Esse atributo corresponde à altura a partir do solo onde a velocidade do vento é igual a zero, considerando um perfil vertical cuja variação apresente comportamento logaritmo com a altitude (Figura 7.1).

Apesar da importância da rugosidade aerodinâmica nos modelos de extrapolação de vento, que objetivam estimar o potencial eólico de uma região, calcular a produção de energia, prever o comportamento médio dos ventos e a turbulência em todas as escalas, a mensuração desta variável é, na prática, considerada difícil, subjetiva e desprovida de consenso metodológico (ASSIREU et al., 2013; SILVA, 2013; TIAN et al., 2011).

O z_0 pode ser estimado via perfil logarítmico do vento, o que demanda a existência de uma torre anemométrica que meça a velocidade do vento simultaneamente em diferentes alturas. Apesar da simplicidade dessa metodologia, dentre outras desvantagens, pode-se destacar a dificuldade de obtenção dos dados de velocidade do vento e a dificuldade para extrapolar o parâmetro de rugosidade para áreas maiores. O comprimento de rugosidade calculado via perfil do vento é comumente denominado de comprimento de rugosidade aerodinâmica (z_{0-A}).

Em função das dificuldades acima mencionadas, abordagens metodológicas têm sido desenvolvidas de forma a relacionar a rugosidade física do terreno com a rugosidade aerodinâmica. Neste caso, quanto mais lisa a superfície, menor o valor de z_0, pois o fluxo de vento está exposto a uma menor turbulência. Esta segunda abordagem é mais desafiadora e apresenta um conjunto extenso de metodologias. A rugosidade obtida por essas metodologias é denominada comprimento de rugosidade efetiva (z_{0-E}).

As abordagens de obtenção da rugosidade efetiva foram divididas em duas categorias

por Hugenholtz et al. (2013). Uma que faz medições dos elementos de rugosidade *in-locu* e, portanto, em pequenas áreas. E a outra que utiliza técnicas de sensoriamento remoto para aferir as rugosidades, sendo capaz de abranger áreas maiores.

Autores como Manwell et al. (2002) e Wiernga (1993) compilaram informações de rugosidade efetiva em tabelas de referência, de acordo com diferentes usos da terra. Tais tabelas são amplamente utilizadas, especialmente pela facilidade de se obter o parâmetro de rugosidade. No entanto, apresentam alguns problemas, discutidos na próxima sessão.

Uma boa alternativa, então, é a obtenção do parâmetro de rugosidade z_{0-E}, por meio de medidas aferidas com modelos digitais de elevação oriundos de técnicas de sensoriamento remoto. A maioria dos trabalhos comunicados nessa área faz uso de modelos digitais de alta resolução espacial, tais como o LiDAR[1]. Hugenholtz et al. (2013) e Smith et al. (2016) confirmam que a estimação de z_{0-E} a partir da microtopografia da superfície apresenta boa conformidade com z_{0-A}.

No entanto, não se tem conhecimento de estudos que tenham avaliado o uso de modelos digitais de elevação globais para inferir a rugosidade aerodinâmica. Isso é compreensível, uma vez que a rugosidade aerodinâmica tem características locais, de pequena escala.

Sendo assim, esse capítulo de tese deseja investigar se os modelos digitais globais combinados com técnicas fractais, como o DFA-2D podem ser utilizados na caracterização do comprimento de rugosidade efetiva. Não é objetivo aferir z_{0-A}, mas apenas descobrir se a rugosidade obtida com o DFA-2D apresenta correlação com z_{0-A}.

Sendo assim, propõe-se aplicar o DFA-2D sobre modelos digitais de elevação globais (SRTM e GDEM) para avaliar sua relação com o parâmetro de rugosidade aerodinâmica. A hipótese testada é que o DFA-2D é capaz de gerar parâmetros de rugosidade efetiva baseados no modelo digital de elevação da região, e não apenas no uso da terra.

Estudos de caso foram realizados em quatro diferentes áreas.

[1]Do inglês *Light Detection and Ranging*, é uma tecnologia óptica de detecção remota (sensor remoto ativo) que, por meio de varredura a laser, gera modelos de terreno precisos definidos por um conjunto tridimensional de pontos (nuvem de pontos).

7.2 Comprimento de Rugosidade

Nos estudos de energia eólica, além das variações de estabilidade atmosférica, a extrapolação da velocidade do vento em função da altura depende da rugosidade aerodinâmica da superfície (z_0) (MANWELL et al., 2009). Por sua vez, este parâmetro depende das características físicas do terreno, tais como altura, geometria, densidade e disposição dos elementos de rugosidade na superfície; e também de características do vento, tais como intensidade e turbulência. Sendo assim, o mesmo pode ser aferido pelo perfil vertical do vento ou pela quantificação da rugosidade da superfície.

Algo importante de se destacar é que o comprimento de rugosidade é um parâmetro difícil de ser aferido, justamente por depender do vento e da superfície em questão. Bañuelos-Ruedas et al. (2011) afirmam que o comprimento de rugosidade pode variar durante o dia e a noite e também durante o ano. Como exemplo, os autores dissertam sobre torres de vento que ficam em terras agrícolas, cuja altura dos dosséis das culturas ali plantadas podem se alterar durante o ano.

Outro exemplo pode ser ilustrado da seguinte maneira : Imaginando um campo aberto, coberto com grama alta. Em condições normais, a grama oferece certa resistência ao escoamento do vento na região. Mas, se a velocidade do vento aumenta muito, ocasionada por mudanças bruscas no tempo, por exemplo, então a força do vento pode fazer 'deitar' essa grama, o que pode favorecer ainda mais o escoamento do vento, ao invés de oferecer resistência.

Os exemplos ilustram que o comprimento de rugosidade depende de diversas variáveis e que, por isso, a maioria dos estudos pressupõem condições atmosféricas neutras e regiões planas. Na prática, a rugosidade da superfície altera as condições de turbulência no escoamento do vento. A estabilidade influencia na forma do perfil, divergindo da lei logarítmica. Silva (2013) avaliou o impacto da rugosidade no aumento da intensidade de turbulência em um parque eólico e concluiu, por meio de simulações, que o aumento do comprimento de rugosidade é suficiente para elevar a intensidade de turbulência geral média do parque. Sendo assim, é imprescindível que as áreas mais turbulentas devam ser conhecidas e evitadas em parques eólicos.

As próximas sessões descrevem a obtenção do comprimento de rugosidade pelo perfil do vento e pela quantificação da rugosidade do terreno.

7.2.1 Comprimento de Rugosidade Aerodinâmica

A rugosidade da superfície é descrita pelo comprimento de rugosidade (z_0) e deslocamento do plano (d) (LYRA; PEREIRA, 2007). z_0 é a altura na qual a velocidade horizontal do vento tende a zero, ou seja, é o limite inferior de validade do perfil médio. Manwell et al. (2009) explicam que esse limite inferior é necessário porque superfícies naturais nunca são completamente uniformes, ou lisas.

Sendo assim, d e z_0 são constantes de integração do perfil vertical da velocidade horizontal do vento na equação 7.1, conhecida como perfil logarítmico de vento:

$$U_z = \left(\frac{u_*}{k}\right) ln \left(\frac{z-d}{z_0}\right) \tag{7.1}$$

Onde:

- U_z=velocidade horizontal média do vento à altura z
- k é a constante de von Kármán (k=0.4)
- u_* é a velocidade de fricção
- d é o deslocamento do plano

Nesse caso, a rugosidade descreve a resistência da superfície para o fluxo de vento e, portanto, é definida para estratificação térmica neutra, onde o fluxo vertical não é aumentado ou diminuído pelos efeitos de flutuabilidade (WIERNGA, 1993; BAÑUELOS-RUEDAS et al., 2010).

Além da lei logarítmica, existem outros métodos de extrapolação da velocidade do vento, tais como a lei de potência e a lei log-linear (BAÑUELOS-RUEDAS et al., 2010; GUALTIERI; SECCI, 2011).

A partir da lei logarítmica, Bauer et al. (1992) sugerem um método gráfico para obtenção do z_{0-A}. Isto é, com os dados de intensidade do vento (U_z) medidos simultaneamente a diferentes alturas (z), constrói-se o perfil do vento, considerando o deslocamento do plano (d) igual a zero. Por regressão linear, o perfil criado é ajustado a lei logarítmica do vento (equação 7.1). Caso o perfil obtido se ajuste ao perfil logarítmico com determinado grau de certeza (r^2), o z_{0-A} é estimado como o coeficiente linear (intercepto) no gráfico velocidade $\times$ log(altura). Caso contrário, o perfil é descartado. Nessa metodologia, a velocidade de fricção corresponde ao coeficiente angular (inclinação da reta) no gráfico velocidade $\times$ log(altura) (BAUER et al., 1992; LANCASTER, 2004; LYRA, 2005).

Camelo et al. (2010) ressaltam que a equação 7.1 se aplica para terrenos planos e homogêneos e possui limitação em relação à dependência do tipo de terreno, da velocidade do vento, modificações na rugosidade da superfície e também da influência de variáveis meteorológicas, como a temperatura do ar e pressão atmosférica. Bauer et al. (1992) destacam três fontes de incerteza na estimação de z_0 pelo perfil de velocidade dos ventos. São elas : problemas nos equipamentos, uso das equações de perfil de velocidade em condições fora daquelas assumidas pela equação e mau uso dos procedimentos estatísticos para derivar z_0. Paul-Limoges et al. (2013) destacam ainda que a obtenção de z_0 por perfil de vento é cara, específica para cada local e limitada a área onde as medidas foram coletadas, o que não permite a determinação desse parâmetro em grandes escalas.

Em função das desvantagens acima citadas, a rugosidade física do terreno tem sido estudada, de forma a relacioná-la com a rugosidade aerodinâmica. O chamado comprimento de rugosidade efetiva z_{0-E} é detalhado na próxima sessão.

7.2.2 Comprimento de Rugosidade Efetiva

Conforme destacado anteriormente, duas abordagens são encontradas na literatura para aferição do comprimento de rugosidade efetiva. Uma é baseada em estudos feitos *in-locu*, onde a rugosidade da região é determinada visualmente. Para tanto, são traçados raios no entorno do equipamento anemométrico, com área bastante reduzida. Exemplo dessa metodologia pode ser acompanhada em Lancaster (2004), onde a rugosidade da superfície foi estimada medindo comprimento, largura e altura do alvo desejado (rochas), num raio de 30 metros da torre anemométrica. A velocidade e direção do vento foram coletadas na alturas 0.25, 0.5, 1, 2, e 4m. É nítido que essa metodologia proverá informações mais precisas, porém muito locais e de difícil expansão.

A segunda abordagem calcula a rugosidade física do terreno a partir de modelos digitais de elevação obtidos por sensoriamento remoto. Neste caso, os produtos mais utilizados são baseados no escaneamento a laser da topografia da região, tanto os aerotransportados (LiDAR) (BROWN; HUGENHOLTZ, 2012; PAUL-LIMOGES et al., 2013; TIAN et al., 2011), quanto os terrestres (TLS) (HUGENHOLTZ et al., 2013; NIELD et al., 2013).

Uma breve, mas interessante comparação entre as abordagens é feita por Hugenholtz et al. (2013).

Com objetivo de auxiliar a estimação do z_{0-E} por meio da rugosidade da superfície sem derivar as informações do MDE, Wiernga (1993) compilou o trabalho de diversos autores em uma tabela de referência. Da mesma forma, Manwell et al. (2002), também propuseram uma relação entre o tipo do terreno e o parâmetro z_0, chamada, nesses casos, de rugosidade típica. No entanto, o tipo do terreno é definido, na verdade, pelo uso da terra no local, conforme exemplificado nas tabelas 7.1 e 7.2. Neste caso, uma prática muito comum entre os autores é a avaliação do uso da terra da região via imagens de sensoriamento remoto e a posterior classificação do grau de rugosidade utilizando as tabelas.

Tabela 7.1 - Valores aproximados de comprimento de rugosidade da superfície para vários tipos de terreno. FONTE : Adaptado de (MANWELL et al., 2002)

Descrição do Terreno	z_0 **(mm)**
Terreno muito liso, gelo, lama	0.01
Mar aberto em condições de calmaria	0.20
Mar agitado	0.50
Neve	3
Gramado	8
Pasto Rugoso	10
Campo de pousio	30
Plantações	50
Poucas árvores	100
Muitas árvores, poucas construções	250
Florestas	500
Subúrbios	1.500
Centros de cidades com construções altas	3.000

Quatro pontos são merecedores de destaque na adoção dessa metodologia: O primeiro ponto é que a classificação da região em um terreno descrito nas tabelas de referência é, por vezes, subjetiva ao especialista (BARTHELMIE et al., 1993; SILVA, 2013). Além disso, muitas vezes o terreno analisado não possui apenas um uso da terra e, nesse caso, os autores sugerem que se faça uma média entre as rugosidades típicas encontradas no local. Barthelmie et al. (1993) ressaltam ainda que existe uma questão a respeito do peso que deve ser dado a rugosidade que está próxima à torre anemométrica, e a rugosidade que está distante.

O segundo ponto é que nem sempre o uso da terra do local estudado se enquadra em algum uso descrito nas tabelas. Silva (2013) cita, por exemplo, a necessidade do estudo e inclusão dos biomas brasileiros em tais tabelas de referência.

Tabela 7.2 - Comprimento de rugosidade de tipos de superfícies homogêneas (z_0). FONTE : Adaptado de (WIERNGA, 1993)

Descrição do Terreno	z_0 **(m)**
Mar, areia solta e neve	≈ 0.0002
Concreto, deserto, mar calmo	0.0002 - 0.0005
Neve	0.0001 - 0.0007
Gelo áspero	0.001 - 0.012
Campo de pousio	0.001 - 0.004
Grama baixa e musgo	0.008 - 0.03
Grama alta e arbustos	0.02 - 0.06
Plantações baixas	0.04 - 0.09
Plantações altas (grãos)	0.12 - 0.18
Cerrado	0.35 - 0.45
Floresta de pinheiros	0.8 - 1.6
Floresta tropical	1.7 - 2.3
Subúrbios	0.4 - 0.7
Centro de cidades	0.7 - 1.5

O terceiro ponto é que mesmo quando a área analisada se enquadra em algum uso descrito nas tabelas, a generalização do tipo de uso pode ser um problema, conforme relatam Hasager et al. (2017). Os autores afirmam que em áreas de florestas a estrutura da floresta depende, dentre outras variáveis, da altura e densidade das árvores, da existência ou não de clarões e dos tipos de folhas, e que o método de designar uma rugosidade típica para o uso da terra 'floresta' nem sempre é suficiente. Outro exemplo é dado por Andreas (2011), que pressupunha que superfícies de oceano congeladas pudessem ser modeladas com os mesmos parâmetros, mas descobriu que valores diferentes de z_0 são necessários para os oceanos congelados no Ártico e na Antártica.

O quarto ponto é o que mais tem relação com o estudo dessa tese e está relacionado ao fato de as rugosidades típicas estarem associadas a um uso da terra, desconsiderando o relevo da região, isto é, elas valem para regiões planas onde, de fato, o uso da terra influencia bastante o regime dos ventos. No entanto, conforme ressaltam Mann et al. (2017), modelar o fluxo do vento em uma superfície caracterizada por topografia complexa e vegetação alta continua sendo um grande desafio para os modelos existentes dentro da comunidade de energia eólica.

No Brasil, especialmente no Sul de Minas Gerais, onde novos parques eólicos tem sido sugeridos, o relevo em si é mais importante que o uso da terra, e as tabelas

de rugosidade aerodinâmica típicas sugeridas por Manwell et al. (2002) e Wiernga (1993) não se adequam satisfatoriamente aos modelos de vento.

Neste caso, a alternativa são os estudos de rugosidade via MDE. As principais referências da literatura são (HUGENHOLTZ et al., 2013; NIELD et al., 2013; PAUL-LIMOGES et al., 2013; SMITH et al., 2016; TIAN et al., 2011). Paul-Limoges et al. (2013) também fazem uma boa revisão de trabalhos relacionados.

Nield et al. (2013) fazem uma revisão das metodologias disponíveis para calcular o parâmetro de rugosidade efetivo, incluindo modelos 1D, 2D e 3D, baseados em estatísticas simples, análise de Fourier ou Wavelets. Os autores calcularam z_{0-A} a partir da lei logarítmica e z_{0-E} a partir de diferentes métricas geométricas (1D, 2D e 3D), em 20 estações diferentes, tendo como fonte de dados um MDE TLS com resolução espacial de 0.01m. Eles concluíram que com um modelo de alta resolução qualquer métrica é capaz de aferir z_{0-E}, e que, nesses casos, a escolha de um modelo depende do padrão de variabilidade da superfície.

Paul-Limoges et al. (2013) estudaram o comprimento de rugosidade efetiva em uma área de floresta, antes e depois do corte das árvores. O MDE LiDAR utilizado por eles tinha 0.15m de resolução espacial e para calcular z_{0-E} foi utilizada a equação de altura da rugosidade. O z_{0-A} foi obtido pela equação logarítmica. Os autores contavam apenas com uma medida de velocidade do vento, obtida por um anemômetro ultrassônico colocado a 4.5 metros acima da superfície. Os autores obtiveram alta correlação entre z_{0-A} e z_{0-E}.

Em trabalho parecido aos anteriores, mas em menor escala Hugenholtz et al. (2013) estimaram a rugosidade efetiva de sete diferentes áreas não vegetada, mas com diferentes rugosidades, utilizando escaneamento a laser (TLS) e os métodos de altura média e altura quadrática média da raiz (*root-mean square height* - RMSH). A rugosidade aerodinâmica foi obtida pelo método gráfico de Bauer et al. (1992), a partir de medições feitas por anemômetros colocados a 0.35, 0.50, 1, 2 e 3 metros de altitude em relação ao solo. Os autores encontram alta correlação entre z_{0-A} e z_{0-E}.

Outros trabalhos utilizando MDE TLS e LiDAR para estimar z_{0-E} podem ser vistos em (SMITH et al., 2016; TIAN et al., 2011).

Pela descrição dos trabalhos acima, é possível perceber que as regiões estudadas são relativamente pequenas e que a resolução dos MDEs utilizados é da ordem de centímetros. Os métodos geométricos de obtenção de z_{0-E} são baseados na diferença

de altitudes presentes nos MDEs. A próxima sessão apresenta o uso do DFA-2D em modelos MDEs globais, cuja resolução espacial é de 30 metros.

7.3 Comprimento de Rugosidade Efetiva por meio do DFA-2D em MDEs globais

Nessa seção os resultados do processamento dos MDEs globais (SRTM e ASTER-GDEM) pelo DFA-2D são apresentados e comparados com a rugosidade aerodinâmica obtida pelo perfil dos ventos.

A rugosidade aeorodinâmica foi obtida pelo método gráfico de Bauer et al. (1992), combinada com algumas restrições que já descartam o perfil em alguns casos e evita a computação desnecessária do ajuste. Sendo assim, o perfil de vento foi considerado apto nas seguintes condições:

- A maior intensidade do vento não pode ser aquela obtida no anemômetro de menor altura da estação. Essa restrição foi adaptada do código de Jager e Andreas (1996). Fica claro que se a maior velocidade for a do anemômetro mais baixo, o perfil não é logaritmico.
- A intensidade do vento na menor altitude da estação não deve ser menor que 3m/s. Essa restrição foi adaptada do trabalho de Lancaster (2004). O autor explica que na ausência de medidas de temperatura, essa restrição diminui potencial erros introduzidos pela instabilidade termal.
- O erro máximo no *fit* obtido por mínimos quadrados (r^2) entre velocidade do vento $\times$ log(Altura) deve ser de 5% (BAUER et al., 1992).

A computação do comprimento de rugosidade aerodinâmico foi realizada em Python. Para tal, todos os dados que ainda não eram, foram convertidos para médias de 10 minutos. O método de Bauer foi aplicado sobre essas médias.

Posteriormente, os dados foram agrupados por quadrantes com base na direção predominante do vento medido no anemômetro de menor altura. Os quadrantes considerados foram: Quadrante 1 : entre 0 e 90 °; Quadrante 2 : entre 90 e 180 °; Quadrante 3 : entre 180 e 270 °; Quadrante 4 : entre 270 e 360 °. A rugosidade aerodinâmica do quadrante foi considerada como a média de todos os perfis pertencentes àquele quadrante.

A rugosidade aerodinâmica foi utilizada nas análises comparativas e qualitativas

com a rugosidade efetiva computada pelo DFA-2D sobre os modelos SRTM e ASTER/GDEM.

Para cada área de estudo, as cenas SRTM e ASTER/GDEM correspondentes foram obtidas no catálogo Earth Explorer.

Os MDEs foram recortados em matrizes quadradas de dimensões 256 × 256 pixels, de forma que a torre anemométrica da região de interesse ficasse no centro da imagem. Para o tratamento dos dados geográficos no Python, utilizou-se a biblioteca GDAL[2].

Diferentes análises foram realizadas, incluindo a subdivisão da imagem em 16 quadrantes com dimensões 64 × 64 pixels e a análise concêntrica a partir da torre. No entanto, os melhores resultados foram obtidos com os quadrantes de 128 × 128 pixels. Os demais resultados foram descartados.

O DFA-2D foi aplicado a cada quadrante. Como o parâmetro α tem significado oposto ao da rugosidade aerodinâmica, ou seja, quanto menor o valor de α, maior a rugosidade, o parâmetro z_{0-E} de rugosidade efetiva foi calculado conforme equacao 7.2.

$$z_{0-E} = \frac{1}{\alpha} \tag{7.2}$$

Posteriormente, os quadrantes foram classificados em uma escala de maior para menor rugosidade, seguindo os valores de z_{0-A} e z_{0-E}. Os resultados foram comparados e análises foram feitas considerando outras informações, tais como a rosa dos ventos e imagem histórica da região.

Os resultados são apresentados a seguir, separadamente por área de estudo.

7.3.1 Área de Estudo NREL

A área de estudo do *National Renewable Energy Laboratory* - NREL compreende a região da torre do Centro Nacional de Tecnologia do Vento (*National Wind Technology Center* - NWTC), situado no estado do Colorado nos Estados Unidos da América, no sopé das Montanhas Rochosas. A torre que coleta os dados de vento está localizada nas coordenadas 39° 54' 38.34"N e 105° 14' 5.28"O, numa altitude de 1855m (figura 7.3). Os dados de intensidade e direção do vento são da torre M2, que possui 135 metros de altura e coleta dados de vento a 2, 5, 10, 20, 50 e 80 metros.

[2]https://pypi.python.org/pypi/GDAL

Foram obtidos dados de 2010 a 2016. A área foi escolhida porque os dados de vento disponibilizados são extremamente confiáveis e com uma série bastante longa. Além disso, a região é plana, com pouca cobertura vegetal e com pouca presença de corpos d'água.

Figura 7.3 - Localização espacial da Torre M2 do NREL.

A figura 7.4 mostra os MDEs ASTER/GDEM (7.4a) e SRTM (7.4b) da região do NREL e as indicações dos quadrantes. Cada quadrante tem 128 × 128 pixels, que na região corresponde a cerca de 3km por 4km (12km^2). A área total é de 48.14km^2.

Figura 7.4 - Representação dos quadrantes processados com os MDEs a)GDEM; b)SRTM para região do NREL.

A tabela 7.3 apresenta o comparativo do valor de α para os quatro quadrantes da reigão. Todos os quadrantes, em ambos MDEs, foram classificados como fBm, pois apresentam α entre 2 e 3.

Tabela 7.3 - Valores numéricos de rugosidade efetiva para os quatro quadrantes do NREL, utilizando os MDEs ASTER/GDEM e SRTM.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-E}_GDEM	0.3968	0.4203	0.3883	0.3748
z_{0-E}_SRTM	0.3783	0.4062	0.3803	0.3726

Numa escala de rugosidade da maior para a menor, a classificação do GDEM é dada pela seguinte ordem de quadrantes : $Q2 > Q1 > Q3 > Q4$. Já o SRTM classificou da seguinte forma : $Q2 > Q3 > Q1 > Q4$.

De forma geral, o SRTM detectou que os quadrantes 1, 3 e 4 possuem rugosidades parecidas (próximas de 2.6), enquanto o quadrante 2 possui rugosidade mais acentuada que os demais ($\alpha \approx 2.46$). O GDEM classificou os quadrantes 1 e 3 com características mais similares, o quadrante 4 o mais suave e o 2 o mais rugoso.

Verifica-se que ambos MDEs acusaram o quadrante 4 como o de rugosidade mais suave e o quadrante 2 como o de maior rugosidade. Houve discordância a respeito do quadrante 1. O GDEM o classificou com rugosidade superior à classificação do SRTM. Na figura 7.4b, é possível verificar que o quadrante 1 do SRTM apresenta uma mancha no canto superior direito que refere-se a um corpo d'água que realmente existe no local. Tal corpo d'água não é visível no GDEM e pode ter influenciado o DFA-2D com SRTM em classificar o quadrante 1 com rugosidade inferior ao do DFA-2D com ASTER/GDEM.

Observando a imagem histórica da região do NREL (figura 7.5a), no ano de 2002, que é próxima do ano de obtenção dos dados do SRTM e no intervalo de coleta do GDEM, verifica-se que o quadrante 2 apresenta elementos de rugosidades explícitos. Há áreas construídas e também uma grande área de mineração, com montes de areia e corpos d´água oriundos dessa atividade. O quadrante 1 se destaca pela presença de corpos d´água. Os quadrantes 3 e 4 são mais similares, a não ser pela presença de uma vegetação esparsa no quadrante 4. Esses quadrantes são os que apresentam maior variabilidade de altitude.

Os dados de vento coletados na região eram a cada minuto. Como forma de padronizar os dados em todas as estações estudadas, o conjunto do NREL foi modificado para uma média de 10 minutos, compondo uma série temporal de 341.717 amostras. A figura 7.5b mostra a rosa dos ventos da região, obtida entre os anos de 2010 e 2016, para a altura de 2m (média de 10 minutos). A direção preferencial do vento na região vem dos quadrantes 3 (SO) e 4 (NO).

A tabela 7.4 mostra os valores quantitativos da amostragem de perfis de vento em cada quadrante e suas respectivas porcentagens.

Figura 7.5 - a) Imagem de satélite da região de estudo NREL, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os anos 2010 e 2016, medidos a 2 metros.

Tabela 7.4 - Quantidade de amostras e respectivos percentuais por quadrantes na região do NREL, medidas a 2 metros.

	Quadrantes				**Total**
	1	**2**	**3**	**4**	
Quantidade de Amostras	36.704 (10.74%)	79.250 (23.19%)	119.064 (34.84%)	106.699 (31.22%)	341.717 (100%)

Ao aplicar o algoritmo de cálculo da rugosidade aerodinâmica (z_{0-A}), aproximadamente 23% das amostras de vento foram consideradas aptas (78.567). A figura 7.6 mostra o perfil característico do vento na região do NREL. Os pontos no gráfico são as médias de velocidade considerando apenas os perfis aptos nas distintas alturas. Em 7.6a apresenta-se o gráfico de velocidade $\times$ altura. Em 7.6b velocidade $\times$

log(altura), cuja forma linear confirma o ajuste logarítmico dos perfis.

Figura 7.6 - Representação do perfil vertical da velocidade média do vento na região do NREL. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura)

A tabela 7.5 apresenta os valores de z_{0-A} estimados a partir do perfil do vento na região. Os valores são as médias obtidas para cada quadrante. Para complementar a análise, são apresentados também a quantidade de amostras em cada quadrante, o desvio padrão de z_{0-A} e a porcentagem de amostras em relação ao total apresentado na tabela 7.4. Os quadrantes 3 e 4 apresentaram a maior parte dos perfis aptos. Os quadrantes 1 e 2 foram os mais turbulentos, com menor quantidade de perfis aptos.

Comparando as tabelas 7.4 e 7.5, verifica-se que o quadrante quatro foi aquele com maior quantidade de perfis que seguem a lei logaritmica ($\approx 44\%$). O quadrante 2 foi o que teve menor quantidade de perfis aptos ($\approx 6\%$) e os demais não passaram de 20% a quantidade de perfis aptos. Esses valores revelam que mesmo em regiões que atendem a todo o arcabouço teórico do perfil logarítmico, como no NREL, com relevo plano e cobertura da terra uniforme, é pequena a percentagem de tempo no qual o regime do vento é considerado logaritmico. Com isso, a obtenção do comprimento de rugosidade aerodinâmico pelo perfil de vento é limitada, o que reforça a necessidade de investigações de técnicas que visem parametrizar o comprimento de rugosidade, a partir de atributos do terreno.

Neste trabalho z_{0-A} será dado em metros. Quanto mais alto o valor da rugosidade aerodinâmica, maior a rugosidade da superfície deveria ser. Nesse caso, ordenando da maior para a menor rugosidade, obtêm-se a seguinte ordem dos quadrantes : $Q2 > Q1 > Q3 > Q4$. Ou seja, os quadrantes 3 e 4, que são de direção predominante do vento, são de fato os menos rugosos. Enquanto os quadrantes 1 e 2 foram os mais rugosos, sendo o 2 o mais rugoso. Essa é a mesma classificação de rugosidade efetiva

Tabela 7.5 - Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas, para a região do NREL.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-A}(m)	0.0320	0.0455	0.0286	0.0193
Quantidade de Amostras Logaritmicas	4.760	4.964	21.964	46.879
Desvio Padrão	0.0516	0.0683	0.0467	0.0370
% em relação ao total de amostras do quadrante	12.97%	6.26%	18.45%	43.94%

inferida pelo DFA-2D, com o MDE ASTER/GDEM. Esse resultado mostra que a suavização do quadrante 1 pelo SRTM pode impactar no valor da rugosidade efetiva.

A figura 7.7 mostra a distribuição de z_{0-A} nos quadrantes. A figura corrobora os resultados da tabela 7.5. Os quadrantes 3 e 4 apresentam maior quantidade de perfis, com menor z_{0-A}. Já os quadrantes 1 e 2 apresentam poucas amostras e com maiores intensidades de z_{0-A}.

Figura 7.7 - Comprimento de Rugosidade Aerodinâmica na região do NREL, calculada pelo perfil vertical dos ventos.

A tabela 7.6 mostra a matriz de correlação entre os valores de α, aferidos dos MDEs GDEM e SRTM como z_{0-E} e a rugosidade aerodinâmica (z_{0-A}), calculada a partir

do perfil de vento. A correlação entre z_{0-A} e z_{0-E} é maior com o MDE GDEM do que com o SRTM.

Sendo assim, para a região NREL, a configuração que melhor ajustou o comprimento de rugosidade efetiva z_{0-E} às informações de vento foi o α aferido com o MDE ASTER/GDEM.

Tabela 7.6 - Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região do NREL.

	z_{0-E}__**GDEM**	z_{0-E}__**SRTM**
z_{0-E}__**SRTM**	0.935769	
z_{0-A}	0.998208	0.947317

As próximas três áreas de estudo são no Brasil, sendo uma em Minas Gerais (Guapé) e duas no Ceará (Camocim e Paracuru) (figura 7.8). Nessas regiões os quadrantes de 128 $\times$ 128 pixels cobrem uma área aproximada de 14km^2. A área total é de 59.20km^2.

Figura 7.8 - Localização das áreas de Guapé, Camocim e Paracuru.

7.3.2 Área de Estudo de Guapé

A área de estudo de Guapé (figura 7.9) está localizada no reservatório da Usina Hidrelétrica de Furnas, próximo à cidade de Guapé, Minas Gerais. Os dados de intensidade e direção do vento são provenientes de uma torre de 42m instalada na coordenada 20°44'26.07"S; 45° 57' 27.03"O; as medidas foram realizadas a 3, 20 e 40 metros de altura entre os meses de Maio e Agosto de 2014.

Figura 7.9 - Localização espacial da torre de Guapé/MG.

Os MDEs utilizados são apresentados na figura 7.10. É perceptível que todos os quadrantes apresentam elevado percentual de cobertura de água.

Os resultados do DFA-2D no processamento dos MDEs para região de Guapé são apresentados na tabela 5. Verifica-se que todos os quadrantes tiveram valores entre 2 e 3, sendo classificados fBm. A escala de rugosidade obtida com o DFA-2D para ambos MDEs foi idêntica, apesar de os mesmos apresentarem valores numericamente diferentes : $Q1 > Q3 > Q2 > Q4$, sendo Q1 o mais rugoso e Q4 o menos rugoso.

É interessante notar que mesmo tendo suas áreas bastante ocupadas com água, os valores de α obtidos nos quadrantes 1 e 3 são baixos, especialmente com os dados do GDEM, indicando alta rugosidade. Esses quadrantes apresentam alta quantidade de recortes, ou seja, bordas da represa. Essa interseção entre terra e água leva, consequentemente, a altas diferenças nas cotas de altitude, dando a esses quadrantes um comportamento anti-persistente.

Figura 7.10 - Representação dos quadrantes processados com os MDEs a)ASTER/GDEM; b)SRTM para região de Guapé/MG

Tabela 7.7 - Valores numéricos de rugosidade efetiva para os quatro quadrantes de Guapé, utilizando os MDEs GDEM e SRTM.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-E}_GDEM	0.4234	0.4000	0.4213	0.3839
z_{0-E}_SRTM	0.4099	0.3829	0.3984	0.3788

De forma geral, percebe-se que os valores de z_{0-E} obtidos com o GDEM são maiores do que os do SRTM, o que indica maior rugosidade. As áreas de terra nos quadrantes 2 e 3 são quase em sua completude cobertos por plantações de café. O GDEM parece, portanto, ser mais afetado pelo dossel da cultura do que o SRTM. De fato, o radar tem maior capacidade de penetrar a cultura além do dossel, quando comparado com o sensor passivo. No entanto, considerando uma aplicação de energia eólica e, especialmente, o operador de rugosidade aerodinâmica, evidenciar a presença de dosséis configura uma informação importante no processo de extrapolação do perfil de vento.

Os dados de vento coletados na região eram a cada minuto e foram transformados em uma média de 10 minutos, compondo uma série temporal de 6.193 amostras. A tabela 7.8 mostra os valores quantitativos da amostragem de perfis de vento em cada quadrante e suas respectivas porcentagens.

A figura 7.11 apresenta a imagem da região no ano de 2002 (7.11a) e a rosa dos ventos (7.11b) medidos a 3 metros. As direções predominantes do vento na região são NE (Q1) e NO (Q4). Fernandes (2015) destaca que nesses quadrantes existe uma grande

Tabela 7.8 - Quantidade de amostras e respectivos percentuais por quadrantes na região de Guapé, medidos a 3 metros.

	Quadrantes				**Total**
	1	**2**	**3**	**4**	
Quantidade de Amostras	3.046 (49.18%)	885 (14.29%)	1.329 (21.46%)	933 (15.07%)	6.193 (100%)

pista para atuação do vento (*fetch*) que propicia o escoamento do mesmo e que a ilha presente no quadrante 4 diminui esse *fetch* em comparação com o quadrante 1 e altera o regime de ventos na região.

Figura 7.11 - a) Imagem de satélite da região de estudo de Guapé, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os meses de Maio e Agosto de 2014, medidos a 3m.

Ao aplicar o algoritmo de cálculo da rugosidade aerodinâmica (z_{0-A}), aproximadamente 9.56% das amostras de vento foram consideradas aptas (592). A figura 7.12 mostra o perfil característico do vento na região de Guapé.

Em estudo anterior feito na região com o mesmo conjunto de dados de vento utilizado neste trabalho, Fernandes (2015) e Reis (2016) verificaram que os ventos na região são bastante instáveis e, nesse caso, os perfis são considerados inaptos pelo algoritmo utilizado para o cálculo de z_{0_A}, o que pode explicar a baixa taxa de perfis aptos.

A tabela 7.9 apresenta os valores de z_{0-A} estimados a partir do perfil de vento região, assim como a quantidade de amostras em cada quadrante, o desvio padrão e a porcentagem de amostras em relação ao total apresentado na tabela 7.8.

Ordenando da maior para a menor rugosidade, têm-se a seguinte escala de z_{0-A} : $Q1 > Q2 > Q3 > Q4$. Comparando com os valores obtidos pelo DFA-2D, verifica-se

Figura 7.12 - Representação do perfil vertical da velocidade média do vento na região de Guapé. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura).

Tabela 7.9 - Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas, para a região de Guapé.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-A}(m)	0.0323	0.0101	0.0019	0.0010
Quantidade de Amostras Logaritmicas	552	5	9	26
Desvio Padrão	0.0479	0.0081	0.0048	0.0014
% em relação ao total de amostras do quadrante	18.12%	0.56%	0.68%	2.79%

que ambos MDEs classificaram Q1 como o mais rugoso e Q4 como o menos rugoso, assim como z_{0-A}. No entanto, houve discordância na classificação dos quadrantes dois e três. O DFA-2D classificou o quadrante três mais rugoso que o dois, enquanto z_{0-A} classificou o dois como sendo mais rugoso que o três. Essa inversão pelo DFA-2D pode ser explicada pelo motivo já anteriormente citado, de que quanto maior a interseção terra/água, mais anti-persistente será o sinal para o DFA-2D, ou seja, com maior rugosidade.

A figura 7.13 mostra a distribuição de z_{0-A} nos quadrantes. A baixa amostragem nos quadrantes 2 e 3, apresentada na tabela 7.9 é percebida na figura. Verifica-se também que a rugosidade aerodinâmica no local é bem baixa, variando quase em sua totalidade de amostras entre 0 e 0.1.

A tabela 7.10 mostra a matriz de correlação entre os valores de z_{0-E}, aferidos dos MDEs GDEM e SRTM e a rugosidade aerodinâmica (z_{0-A}), calculada a partir do

Figura 7.13 - Comprimento de Rugosidade Aerodinâmica na região de Guapé, calculada pelo perfil vertical dos ventos.

perfil de vento. A força da correlação desses operadores com o MDE GDEM é maior que a do SRTM, com valor próximo de 73%.

Tabela 7.10 - Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região de Guapé.

	z_{0-E}__**GDEM**	z_{0-E}__**SRTM**
z_{0-E}__**SRTM**	0.9343	
z_{0-A}	0.5546	0.7283

As outras duas áreas estão localizadas no Nordeste do Brasil, no Estado do Ceará, nos municípios de Camocim e Paracuru (figura 7.8). A instalação dessas duas torres anemométricas teve como objetivo estudos de viabilidade técnica e econômica, por parte da Secretaria de Infraestrutura do Estado do Ceará, visando à instalação de turbinas eólicas nas regiões. As torres coletam dados de intensidade do vento em três diferentes alturas, 20.6m, 40.2m e 60.4m e da direção do vento nas alturas de 40.2m e 60m. Os dados foram coletados entre 2004 e 2006.

Os pontos foram considerados adequados e representativos para os estudos eólicos nos locais pela secretaria de infraestrutura do Ceará. As informações anemométricas e as descrições dos equipamentos e instalação das torres são disponibilizadas pela

Secretaria de Infraestrutura do Estado do Ceará[3] . Uma análise dos dados de vento dessa região foi feita por Camelo et al. (2010).

7.3.3 Área de Estudo Camocim

A torre de Camocim está localizada nas coordenadas geográficas 02°51'56.7"S e 40°53'09.20"O. O ponto possui cota de nível aproximada de 10 metros. A figura 7.14 apresenta os MDEs da área de Camocim, e a tabela 7.11 os resultados do DFA-2D aplicado aos respectivos MDEs. Nessa área de estudo, os resultados obtidos com o MDE SRTM foram diferentes dos obtidos com o ASTER GDEM.

No SRTM, os quadrantes 1 e 4 não retornaram valores numéricos para α. Isso acontece porque tais quadrantes são cobertos majoritariamente por água e apresenta todas as cotas de altitude idênticas. Nesses casos, a função global é zero e o logaritmo de zero não existe. No entanto, a interpretação física desse resultado é que a região é estacionária, constante e muito lisa. Os quadrantes 2 e 3 foram classificados como fBm, anti-persistentes, sendo Q3 mais rugoso que Q2.

Figura 7.14 - Representação dos quadrantes processados com os MDEs a)ASTER/GDEM; b)SRTM para a região de Camocim/CE.

Tabela 7.11 - Valores numéricos de rugosidade efetiva para os quatro quadrantes de Camocim, utilizando o MDE GDEM e SRTM.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-E}_GDEM	0.4657	-	-	-
z_{0-E}_SRTM	*nan*	2.4055	2.2396	*nan*

[3] http://www.seinfra.ce.gov.br/index.php/downloads/category/6-energia

No MDE GDEM, com exceção do quadrante 1, todos os demais foram classificados como ruído pelo DFA-2D, isto é, o α apresentou valores entre 1 e 2, o que indica que a imagem não é auto-similar. Algo importante de ser citado aqui é que para os quadrantes que possuem grande parte de sua área ocupada por água, os resultados de ruído podem ser oriundos do fato de o DFA não ser uma técnica que deva ser aplicada em séries estacionárias. Nesses casos, a parte estacionária da imagem pode estar enviesando o valor de α.

O fato do DFA-2D ter retornado valores de α fora da faixa contemplada por modelos fBm e fGn com o MDE ASTER/GDEM não permite a derivação do expoente de Hurst e, consequentemente, a noção de persistência que infere a rugosidade. Sendo assim, para a área de Camocim, apenas os valores obtidos com o SRTM serão comparados com a rugosidade aerodinâmica. A classificação da rugosidade efetiva pelo DFA-2D com o SRTM é a seguinte : $Q3 > Q2 > [Q1, Q4]$, sendo Q3 o quadrante mais rugoso. Como Q1 e Q4 retornaram *not a number* (*nan*), pode-se considerar o máximo de auto-correlação positiva, correspondente a $\alpha \approx 3$.

Em uma inspeção visual e comparando a imagem da região no ano de 2002 (figura 7.15a), o MDE do SRTM parece ser mais fiel à realidade da região. O rio que corta o quadrante 2 é imperceptível no MDE do GDEM. Por outro lado, o GDEM destaca pequenas ilhas e uma área de interseção maior com o mar e mais recortada que o SRTM.

Em Camocim, a direção preferencial do vento na região é a leste (E) e Sudeste (SE), incluídas nos quadrantes Q1 e Q2 (figura 7.15b). A tabela 7.12 apresenta os valores quantitativos e respectivos percentuais de amostras por quadrante. A maioria das amostras encontra-se nos quadrantes 1 e 2, que é a região de predominância do vento. A análise desses dados permite afirmar que nos quadrantes 3 e 4 há pouca incidência de vento. Especialmente no quadrante 4, praticamente não há registros de vento para 20.6 metros.

Tabela 7.12 - Quantidade de amostras e respectivos percentuais por quadrantes na região de Camocim, medidas a 20.6 metros.

	Quadrantes				**Total**
	1	**2**	**3**	**4**	
Quantidade de Amostras	33.404 (36.90%)	55.876 (61.72%)	1.071 (1.18%)	184 (0.20%)	90.535 (100%)

Figura 7.15 - : a) Imagem de satélite da região de Camocim, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os anos 2004 e 2006, medidos a 20.6m.

Do total de amostras de perfil de vento (90.535), 63.57% foram consideradas aptas para o cálculo da rugosidade aerodinâmica. O perfil característico do vento considerado apto na região de Camocim pode ser visto na figura 7.16.

Figura 7.16 - Representação do perfil vertical da velocidade média do vento na região de Camocim. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura).

A tabela 7.13 apresenta os valores de rugosidade aerodinâmica aferidos via perfil de vento da região de Camocim. A classificação de rugosidade dos quadrantes segue a seguinte ordem, do mais rugoso para o menos rugoso : $Q3 > Q4 > Q2 > Q1$.

Ao contrário das regiões de estudo anteriores (NREL e Guapé), a região de Camocim apresenta alta percentagem de perfis de vento logaritmico, com exceção do quadrante 4. O fato do anemômetro mais baixo estar localizado a 20.6 metros de altura pode

Tabela 7.13 - Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas na região de Camocim.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-A} (m)	0.0253	1.1192	3.4372	1.1479
Quantidade de Amostras Logaritmicas	18.526	38.600	416	9
Desvio Padrão	0.3127	1.4873	3.0968	2.6827
% em relação ao total de amostras do quadrante	55.46	69.08	38.84	4.89

ter influenciado esse alta taxa de perfis aptos. A baixa amostragem de perfis aptos no quadrante quatro pode ser explicada pela turbulência causada pelo efeito rampa (**??**), presente na região. Outro motivo aparente é mudança abrupta na condição da superfície (descontinuidade de rugosidade) (**??**), ou seja, o vento fluía sobre o mar e encontra a faixa de areia e topografia do litoral. Essas duas características citadas (efeito rampa e descontinuidade de rugosidade) podem ser exploradas pela análise do perfil do relevo, apresentado na figura 7.17. O anemômetro mais baixo na região de Camocim está localizado a 20.6 m de altura.

Na figura 7.17a, o 'x' amarelo corresponde ao local da torre anemométrica. A linha vermelha corresponde ao segmento utilizado para análise do perfil apresentado na figura 7.17b. A figura 7.17c mostra o subsegmento que tem início no mar e chega até a posição da torre. Verifica-se portanto, que o vento que sopra predominantemente de sudeste na região é favorecido pelo efeito rampa, uma vez que há um desnível de aproximadamente 15 metros. Por outro lado, o vento que sopra do mar em direção à torre no quadrante 4, sofre com a mudança brusca de superfície evidenciada na figura 7.17c, e é também desfavorecido pelo efeito rampa, uma vez há um aclive de aproximadamente 4 metros.

O gráfico da figura 7.18 apresenta a distribuição polar da rugosidade aerodinâmica em Camocim, considerando a direção do anemômetro mais próximo da superfície (20.6 m). Verifica-se que a rugosidade aerodinâmica no quadrante 2 é superior à do quadrante 1 e que devido a baixa amostragem nos quadrantes 3 e 4 7.13, os mesmos não aparecem no gráfico.

A tabela 7.13 e a figura 7.18 sugerem que o comprimento de rugosidade aerodinâmcia no local é bastante alto, especialmente quando comparado ao das áreas anteriores

Figura 7.17 - Perfil do relevo para os quadrantes 2 e 4 na região de Camocim.

(NREL e Guapé), que variavam entre 0 e 0.2. Em Camocim, z_{0-A} apresenta valores maiores que 1 para os quadrantes 2, 3 e 4. Esse valor, seguindo a tabela de Wiernga (1993), refere-se a áreas florestadas e cidades. Já na tabela de Manwell et al. (2002), refere-se a subúrbios e centros urbanos. De fato, o quadrante 2 poderia ser caracterizado como subúrbio, uma vez que apresenta áreas urbanas e regiões de floresta. Já o quadrante 3, tem cerca de 50% de sua área coberto por vegetação e o restante por areia e água. O quadrante 4 volta a ser o ponto de discrepância, uma vez que é coberto majoritariamente por água e, portanto, deveria apresentar z_{0-A} próximo de zero, de forma similiar ao quadrante 1.

Em comparação com os valores de z_{0-E} aferidos pelo SRTM, houve concordância na classificação de rugosidade dos quadrantes Q3 como mais rugoso, e Q1 como um dos menos rugosos. Houve discordância entre os quadrantes 2 e 4. A rugosidade aerodinâmica nesses quadrantes teve valores muito parecidos ($z_{0-A} \approx 1.1$), o que

Figura 7.18 - Comprimento de Rugosidade Aerodinâmica na região de Camocim, calculada pelo perfil vertical dos ventos.

não corrobora os resultados obtidos com o DFA-2D, nem com a inspeção visual da imagem da região. O quadrante 2 é ocupado com mata, rio e área urbana. Enquanto o quadrante 4 é quase que completamente ocupado pelo mar (figura 7.15a). A explicação dessa incompatibilidade de resultados é que, para essa região, z_{0-A} no quadrante 4 sofre as interferências já acima discutidas : efeito rampa e mudança abrupta de rugosidade, o que irá gerar uma turbulência no vento. Soma-se ainda o fato de o anemômetro mais baixo estar a uma altura de 20.6 metros.

A matriz de correlação é apresentada na tabela 7.14. Optou-se por utilizar os valores obtidos com o GDEM para avaliar se, mesmo não sendo fBm ou fGn haveria correlação com os valores de rugosidade aerodinâmica. No entanto, verifica-se pela tabela que a correlação foi inferior ao valor obtido com o SRTM. Para o cálculo da correlação, os valores de *not a number* foram substituídos por 3. A correlação entre z_{0-A} e z_{0-E} mostrou-se próxima de 0.8.

Tabela 7.14 - Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região de Camocim.

	z_{0-E}__**GDEM**	z_{0-E}__**SRTM**
z_{0-E}__**SRTM**	0.5946	
z_{0-A}	0.5405	0.7831

7.3.4 Área de Estudo Paracuru

A torre de Paracuru está localizada nas coordenadas geográficas 03 °24'42.4"S e 38 °59'02.8"O. O ponto possui cota de nível aproximada de 20 metros. Os MDEs utilizados na área de Paracuru são apresentados na figura 7.19. O SRTM da região apresenta falhas, nos quadrantes 3 e 4.

Figura 7.19 - Representação dos quadrantes processados com os MDEs a)GDEM; b)SRTM para a região de Paracuru/CE.

Os valores aferidos pelo DFA-2D são apresentados na tabela 7.15. Apesar da grande parte coberta com água, em nenhum dos MDEs, a técnica apontou valores não numéricos. O quadrante 4 apresentou valor muito próximo de 2 para o MDE SRTM e será assim considerado. Em todos os casos, o valor de α corresponde ao modelo fBm. Na escala de rugosidade, não houve total concordância entre os valores obtidos com os diferentes MDEs. Com o GDEM, a escala é dada por : $Q3 > Q2 > Q1 > Q4$. Q1 e Q4 apresentaram valores muito parecidos ($z_{0-E} \approx 0.41$). Com o SRTM, o DFA-2D retornou a seguinte escala de rugosidade : $Q4 > Q3 > Q2 > Q1$, tendo Q3 e Q4 valores próximos a 0.5, indicando uma rugosidade muito alta. No entanto, no SRTM, Q4 e Q1 apresentam rugosidades antagônicas. Ou seja, Q1 é o mais suave e Q4 o mais rugoso. Vale ressaltar que os quadrantes 3 e 4 do SRTM apresentam falta de dados no modelo digital.

A figura 7.20 mostra a imagem histórica do local e a rosa dos ventos da região, com ventos medidos o mais próximo da superfície. Em Paracuru, os ventos foram medidos a 20.6, 40.2 e 60.4 metros de altura. O vento na região é predominantemente de Leste (E), englobando os quadrantes 1 e 2 (figura 7.20b). A região apresenta algumas dunas, principalmente no quadrante 3 e não é possível perceber características de área urbana na região. Os valores quantitativos de amostras por quadrante são

Tabela 7.15 - Valores numéricos de rugosidade efetiva para os quatro quadrantes de Paracuru, utilizando o MDE GDEM e SRTM.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-E}_GDEM	0.4092	0.4793	0.4921	0.4087
z_{0-E}_SRTM	0.4194	0.4569	0.4767	0.5000

apresentados na tabela 7.16.

Figura 7.20 - a) Imagem de satélite da região de Paracuru, do ano de 2002. FONTE : Google Earth; b) Rosa dos ventos da região entre os anos 2004 e 2006, medidas a 20.6 metros.

Tabela 7.16 - Quantidade de amostras e respectivos percentuais por quadrantes na região de Paracuru, medidas a 20.6 metros.

	Quadrantes				**Total**
	1	**2**	**3**	**4**	
Quantidade de Amostras	57.633 (64.30%)	30.396 (33.91%)	1.135 (1.27%)	465 (0.52%)	89.629 (100%)

Do total de amostras (89.629), 40.77% foram consideradas aptas para o cálculo da rugosidade aerodinâmica. Desses, a maior parte encontra-se nos quadrantes 1 e 2 (aproximadamente 99% das amostras). A figura 7.21 apresenta o perfil característico do vento considerado apto na região de Paracuru.

Os valores de z_{0-A} obtidos na área são apresentados na tabela 7.17, assim como a

Figura 7.21 - Representação do perfil vertical da velocidade média do vento na região de Paracuru. Em a) Velocidade × Altura. Em b) Velocidade × Log(Altura).

quantidade de amostras e o desvio padrão da média. Na escala de rugosidade, temos : $Q3 > Q2 > Q4 > Q1$. Os quadrantes 1 e 4 foram os menos rugosos e com valores de z_{0-A} típicos para regiões cobertas com água. O valor de z_{0-A} para o quadrante 2 é, segundo Manwell et al. (2002) e Wiernga (1993) como cerrado, ou campo de pousio, compatível com o local, que também apresenta parte de sua área coberta com água. Já o quadrante 3 foi considerado o mais rugoso, com valor de z_{0-A} maior que 1, que segundo as tabelas de referência é compatível com área de floresta de pinos ou subúrbio. A região, como já mencionado, é formada por uma vegetação, tipo cerrado e algumas dunas.

Comparando os valores da tabela 7.17 com os da tabela 7.15, verifica-se que há concordância com os resultados obtidos pelo DFA-2D com o GDEM, considerando que a rugosidade dos quadrantes 1 e 4 foram muito parecidas.

Tabela 7.17 - Média dos comprimentos de rugosidade aerodinâmica (z_{0-A}) por quadrante, calculados a partir da intensidade do vento em diferentes alturas na região de Paracuru.

	Quadrantes			
	1	**2**	**3**	**4**
z_{0-A} (m)	0.0043	0.3618	1.1194	0.0793
Quantidade de Amostras Logaritmicas	20.672	15.663	161	43
Desvio Padrão	0.0758	0.7778	1.1999	0.2036
% em relação ao total de amostras do quadrante	35.87	51.53	34.62	9.24

Na comparação com os valores obtidos pelo SRTM, observa-se que houve concordância sobre a rugosidade de Q3 ser maior que a de Q2, que é maior que a de Q1. A discordância veio na classificação do quadrante 4, considerado pelo z_{0-E} do SRTM como o mais rugoso. A razão pode ser àquela já discutida anteriormente, que como a região possui muita interseção entre água e terra, essas mudanças de cota de altitude afetam o DFA-2D. Além disso, o grande espaço ocupado com água, cuja cota é única, pode enviesar o resultado. A figura 7.22 mostra a distribuição polar de z_{0-A}.

Figura 7.22 - Comprimento de Rugosidade Aerodinâmica na região de Paracuru, calculada pelo perfil vertical dos ventos.

A tabela 7.18 apresenta a matriz de correlação entre as variáveis estudadas. A rugosidade aerodinâmica apresenta correlação de 83% com a rugosidade efetiva aferida pelo DFA-2D com o ASTER/GDEM.

Tabela 7.18 - Matriz de Correlação entre as aferições da rugosidade efetiva e rugosidade aerodinâmica para a região de Paracuru.

	z_{0-E}__**GDEM**	z_{0-E}__**SRTM**
z_{0-E}__**SRTM**	0.1734	
z_{0-A}	0.8362	0.2654

7.4 Análise comparativa utilizando os descritores texturais de Haralick

Para fins de comparação, os descritores texturais de Haralick (Apêndice A.1) foram computados para os quadrantes em todas as áreas. A tabela 7.19 apresenta os resultados da matriz de correlação entre os descritores texturais de Haralick e os valores obtidos para z_{0-A} e z_{0-E} com os MDEs que tiveram melhor performance junto ao DFA-2D. É possível perceber que não existe consenso a respeito de um descritor, ou um conjunto deles que descreva adequadamente a rugosidade aerodinâmica, nem a rugosidade efetiva obtida com o DFA-2D.

Isto porque, nota-se na tabela 7.19 que para cada área de estudo, um conjunto diferente de descritores de Haralick apresentaram melhor correlação com os comprimentos de rugosidade efetiva e aerodinâmica.

Além da computação de um único valor para todo o quadrante, diversas variações de tamanho de janela e vizinhança foram experimentadas. No entanto, a conclusão foi idêntica em todos os casos.

Tabela 7.19 - Matriz de correlação entre os valores obtidos com os descritores texturais de Haralick para as áreas estudadas, e os parâmetros de rugosidade aerodinâmica e efetiva, considerando o MDE de melhor perfomance com a análise do DFA-2D. As células destacadas indicam correlação maior que 90%.

NREL - GDEM

	SMA	**Contraste**	**Entropia**	**Variancia**	**Correlacao**	**Homogeneidade**	**Sum_Avg**	**Sum_Entropy**	**Sum_Variance**	**Diff_Variance**	**Diff_Entropy**	**IMC_1**	**IMC_2**	**z0_A**
z_{0-A}	0.99967	0.503194	-0.982407	-0.796653	-0.899193	-0.516575	0.848966	-0.968355	-0.799855	0.44461	-0.079803	0.525046	-0.855327	
z_{0-E}	-0.997032	-0.548519	0.981861	0.76003	0.922049	0.565582	-0.832371	0.965329	0.763491	-0.490335	0.034085	-0.573106	0.883066	-0.998277

Guape - GDEM

	SMA	**Contraste**	**Entropia**	**Variancia**	**Correlacao**	**Homogeneidade**	**Sum_Avg**	**Sum_Entropy**	**Sum_Variance**	**Diff_Variance**	**Diff_Entropy**	**IMC_1**	**IMC_2**
z_{0-E}	0.738605	-0.881343	-0.512461	0.612901	0.856412	0.705729	-0.699584	-0.588693	0.619375	-0.886164	-0.867575	-0.998363	0.586442
IT													

Guape - SRTM

	SMA	**Contraste**	**Entropia**	**Variancia**	**Correlacao**	**Homogeneidade**	**Sum_Avg**	**Sum_Entropy**	**Sum_Variance**	**Diff_Variance**	**Diff_Entropy**	**IMC_1**	**IMC_2**
z_{0-E}	0.148963	-0.850657	0.201162	0.224779	0.974272	0.217533	-0.011984	0.102335	0.235117	-0.815756	-0.307931	-0.684979	0.724804
IT													

Camocim - SRTM

	SMA	**Contraste**	**Entropia**	**Variancia**	**Correlacao**	**Homogeneidade**	**Sum_Avg**	**Sum_Entropy**	**Sum_Variance**	**Diff_Variance**	**Diff_Entropy**	**IMC_1**	**IMC_2**	**z0_A**
z_{0-A}	-0.654049	0.757809	0.674439	0.000505	-0.869324	-0.752389	0.916666	0.63877	-0.084466	0.311467	0.677058	0.740064	0.326118	
z_{0-E}	0.982506	-0.485085	-0.98704	0.370389	0.978192	0.998569	-0.863888	-0.977856	0.448637	0.046928	-0.987888	-0.997046	-0.842576	-0.783024

Paracuru - GDEM

	SMA	**Contraste**	**Entropia**	**Variancia**	**Correlacao**	**Homogeneidade**	**Sum_Avg**	**Sum_Entropy**	**Sum_Variance**	**Diff_Variance**	**Diff_Entropy**	**IMC_1**	**IMC_2**	**z0_A**
z_{0-A}	-0.741573	0.601977	0.87205	-0.567043	-0.915026	-0.901463	0.821638	0.835017	-0.573269	-0.716191	0.849951	0.954787	0.417735	
z_{0-E}	0.84796	-0.523784	-0.899418	0.326747	0.606075	0.922938	-0.909105	-0.891011	0.331785	0.803271	-0.87146	-0.871722	-0.554294	-0.847782

7.5 Considerações Finais

Esse capítulo analisou o uso do DFA-2D na inferência de rugosidade efetiva, obtida sobre modelos digitais de elevação, e sua correlação com a rugosidade aerodinâmica, obtida por perfis de vento medidos a diferentes alturas. De forma geral, o α inferiu satisfatoriamente a rugosidade da superfície, revelando boa correlação com a rugosidade aerodinâmica. Na região do NREL, onde o terreno é mais plano e a série de vento era maior e com mais alturas para definição do perfil, a correlação foi maior que 90% em ambos MDEs, sendo mais alta com o GDEM.

As demais regiões apresentam relevo mais recortado, com mudanças bruscas de rugosidade da superfície causadas pela interseção entre terra e água. O DFA-2D é impactado por áreas totalmente homogêneas, pois é uma técnica *Random Walk* e precisa de variabilidade para funcionar. A homogeneidade das áreas ocupadas por água nas regiões estudadas de fato impactou alguns resultados, associando ao α valores que não caracterizam superfícies auto-afim, tornando inviável a derivação do expoente de Hurst e, consequentemente, inferir a rugosidade da região.

A região com coeficiente de correlação mais baixo entre z_{0-A} e z_{0-E} foi na região de Guapé, que apresenta um conjunto de características que, tanto afetam o valor de z_{0-A}, tais como descontinuidade de rugosidade e efeito rampa; quanto impactam o funcionamento do DFA-2D, com a grande área de borda da represa e homogeneidade na cota de altitude dentro da represa.

Apesar disso, em todas as áreas, o DFA-2D foi capaz de apontar o quadrante mais rugoso e, em quase todas, o menos rugoso.

Aparentemente, o MDE ASTER/GDEM é mais ruidoso que o SRTM para as regiões estudadas e essa característica favoreceu o ASTER/GDEM nas regiões do NREL e Paracuru. Já para as regiões de Guapé e Camocim houve melhor correlação com o SRTM. Essas duas regiões são as que apresentam maior quantidade de recortes na interseção entre terra e água. Então, para esses casos, a suavização do modelo SRTM parece melhorar a eficiência do DFA-2D na caracterização da rugosidade efetiva.

O perfil logaritmico considerado pela comunidade científica da área de engenharia do vento mostrou-se ocorrer num percentual do tempo nunca maior do que 50%. A única exceção foi para a região de Camocim, onde 63% dos perfis foram considerados aptos. Guapé é a região que apresenta relevo mais complexo dentre as regiões estudadas e também foi a região de menor percentagem de perfis aptos (cerca de 10%). Como

já mencionado, a obtenção do comprimento de rugosidade aerodinâmica pelo perfil vertical do vento é cara e de abrangência local. Ressalta-se novamente, portanto, a importância de novas metodologias que qualifiquem os quadrantes a partir do relevo. E, neste ponto, as metodologias que utilizam produtos de alta resolução espacial apresentam as mesmas limitações acima citadas, isto é, são caras e abrangem uma área pequena. São necessários mais estudos em áreas com uma maior diversidade de terrenos para validar a metodologia proposta e testada nesse capítulo. No entanto, o DFA-2D mostrou-se como uma técnica promissora nesse tipo de estudo.

8 CONCLUSÕES

Neste trabalho a técnica de Análise de Flutuação Destendenciada bi-dimensional foi investigada em suas três componentes mais críticas : operacional, interpretação do parâmetro α e performance.

Na componente operacional o objetivo foi tratar questões abertas e discrepantes na literatura acerca do algoritmo em si. Foram realizados testes empíricos para comprovar os resultados apresentados. Acredita-se que as discussões levantadas foram todas respondidas e que os resultados contribuem para um bom entendimento da técnica e de sua implementação.

A interpretação do parâmetro α é uma questão levantada por alguns autores e confirmada nesse trabalho, especialmente nos estudos de caso com dados reais. Neste sentido, o processamento de conjuntos de dados canônicos, mas não sabidamente fractais, teve o propósito de auxiliar na interpretação do DFA-2D mono e multifractal. Tais experimentos apontaram também desafios no uso da técnica, especialmente em sua versão multifractal, quando os conjuntos de dados não apresentam relação monotônica entre q e $h(q)$, e também quando as imagens fazem parte de uma série temporal longa. Não foi encontrada na literatura metodologia que tratasse definitivamente as situações acima citadas. Esse trabalho de tese pretendeu, portanto, trazer contribuições no tratamento de tais situações.

O outro ponto crítico do DFA-2D tratado nesse trabalho foi a performance. A paralelização do método utilizando GPGPU-CUDA diminuiu significativamente o tempo de processamento. Neste ponto é importante ressaltar que a paralelização aconteceu mais no processamento simultâneo de diversas sub-matrizes, do que na paralelização do procedimento interno em cada sub-matriz. Outras modelagens podem melhorar ainda mais o ganho o que, em ambientes de grande fluxo de dados ou em cenários onde o processamento da imagem precisa ser rápido, é essencial. Outro ponto é que a placa GPU utilizada nesse trabalho possui limitações que culminaram em algumas adaptações do código. Mesmo assim, o ganho foi substancial.

Os estudos de caso com dados reais mostraram que o DFA-2D é um bom aferidor de textura e dimensão fractal em imagens digitais, constituindo-se, portanto como uma boa métrica para classificação automática de padrões.

No estudo de caso para classificação de galáxias em elípticas e espirais, os ruídos nas imagens podem ser os causadores da torção presente no espectro de singularidades.

Para esse conjunto de dados, a classificação foi melhor utilizando os descritores texturais de Haralick. No entanto, tais descritores não são capazes de caracterizar o processo físico subjacente. O DFA-2D caracterizou galáxias espirais como fBm não correlacionado ($\alpha \approx 2.5$) e as elípticas como fBm que apresenta persistência ($\alpha \approx 2.6$), corroborando a literatura.

No uso do DFA-2D como indicador de comprimento de rugosidade efetiva do terreno, por meio do processamento de MDEs globais, apesar da baixa amostragem, os resultados foram considerados satisfatórios. Houve boa concordância com a rugosidade aerodinâmica, especialmente na indicação dos quadrantes mais e menos rugosos. Dada a dificuldade de aferir o comprimento de rugosidade por técnicas convencionais (anemométricas e com produtos de sensoriamento remoto de alta resolução), essa metodologia pode auxiliar cientistas e técnicos a terem uma ideia inicial desse parâmetro com base no relevo, utilizando dados distribuídos gratuitamente.

Todos os códigos implementados nessa tese estão disponíveis no repositório do GitHub[1]. Apesar de popular, não se tem conhecimento de nenhum código do DFA-2D disponível, como existe para a versão unidimensional.

Neste trabalho, as imagens tratadas (sintéticas e reais) não receberam nenhum tratamento, isto é, não passaram por nenhum tipo de filtro. Como linhas de pesquisa para trabalhos futuros sugere-se aplicação de filtros para avaliar se há melhorias nas classificações, ou se tais filtros interferem na caracterização dos sinais. Os filtros podem ser úteis especialmente na classificação das galáxias, uma vez que as imagens são muito ruidosas.

Avalia-se que os objetivos propostos foram alcançados e que essa tese traz contribuições inéditas em estudos com o DFA-2D.

[1] https://github.com/vanessavcos

REFERÊNCIAS BIBLIOGRÁFICAS

ALPATOV, A. V.; VIKHROV, S. P.; GRISHANKINA, N. V. Revealing the surface interface correlations in a-Si:H films by 2D detrended fluctuation analysis. **Semiconductors**, v. 47, n. 3, p. 365–371, mar 2013. ISSN 1063-7826. Disponível em: <http://link.springer.com/10.1134/S1063782613030020>. 14, 29

ALVAREZ-RAMIREZ, J.; RODRIGUEZ, E.; CERVANTES, I.; CARLOS, J. Scaling properties of image textures : A detrending fluctuation analysis approach. **Physica A**, v. 361, p. 677–698, 2006. 12, 18, 28, 29

ANDREAS, E. L. A relationship between the aerodynamic and physical roughness of winter sea ice. **Q. J. R. Meteorol. Soc.**, v. 137, n. 659, p. 1581–1588, 2011. ISSN 00359009. 108

ASSIREU, A. T.; PEREIRA, R. d. R.; SOUZA, V. C. O. d.; SOUZA, A. F. d.; FRANCHITO, S. H. Aplicação do operador de fragmentação assimétrica (fa) para estimativas da rugosidade superficial a partir de mapas digitais de elevação. In: EPIPHANIO, J. C. N.; GALVÃO, L. S. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2013. p. 9055–9062. ISBN 978-85-17-00066-9 (Internet) and 978-85-17-00065-2 (DVD). Disponível em: <http://urlib.net/dpi.inpe.br/marte2/2013/05.29.00.32.55>. Acesso em: 01 ago. 2017. 102

AU, K.; GENOVESE, C.; CONNOLLY, A. **Inferring Galaxy Morphology Through Texture Analysis**. [s.n.], 2006. 7 p. Disponível em: <http://repository.cmu.edu/statistics/119/>. 3, 40, 86, 87

BAÑUELOS-RUEDAS, F.; ANGELES-CAMACHO, C.; RIOS-MARCUELLO, S. Analysis and validation of the methodology used in the extrapolation of wind speed data at different heights. **Renew. Sustain. Energy Rev.**, v. 14, n. 8, p. 2383–2391, 2010. ISSN 13640321. 105

BAÑUELOS-RUEDAS, F.; CAMACHO, C. Á.; RIOS-MARCUELLO, S. Methodologies used in the extrapolation of wind speed data at different heights and its impact in the wind energy resource assessment in a region. In: SUVIRE, G. O. (Ed.). **Wind Farm**: technical regulations, potential estimation and siting assessment. [S.l.]: InTech, 2011. cap. 4. ISBN 978-953-307-483-2. 104

BARALDI, A.; PARMIGGIANI, F. An Investigation of the Textural Characteristics Associated with Gray Level Co-occurrence Matrix Statistical

Parameters. **IEEE Trans. Geosci. Remote Sens.**, v. 33, n. 2, p. 293–304, 1995. 93, 97, 163

BARRERA, E.; GONZALEZ, F.; RODRIGUEZ, E.; ALVAREZ-RAMIREZ, J. Correlation of optical properties with the fractal microstructure of black molybdenum coatings. **Appl. Surf. Sci.**, v. 256, n. 6, p. 1756–1763, 2010. ISSN 01694332. 30

BARTHELMIE, R. J.; PALUTIKOF, J. P.; DAVIES, T. D. Estimation of sector roughness lengths and the effect on prediction of the vertical wind speed profile. **Boundary-Layer Meteorol.**, v. 66, n. 1, p. 19–47, 1993. ISSN 00068314. 107

BASHAN, A.; BARTSCH, R.; KANTELHARDT, J. W.; HAVLIN, S. Comparison of detrending methods for fluctuation analysis. **Physica A**, v. 387, n. 21, p. 5080–5090, 2008. 18

BATISTA, C. A. **Métodos emergentes de física-estatística aplicados a séries temporais**. Dissertação (Mestrado) — Universidade Federal Rural de Pernambuco, 2006. 6

BAUER, B. O.; SHERMAN, D. J.; WOLCOTT, J. F. Sources of uncertainty ain shear stress and roughness length derived from velocity profiles. **Prof. Geogr.**, v. 44, n. June, p. 453–464, 1992. 105, 106, 109, 110

BAZELL, D.; AHA, D. Ensembles of Classifiers for Morphological Galaxy Classification. **Astrophys. J.**, v. 548, n. 1, p. 219–223, 2001. ISSN 0004-637X. 85, 86, 87, 97

BOROWSKA, M.; SZARMACH, J.; OCZERETKO, E. Fractal texture analysis of the healing process after bone loss. **Comput. Med. Imaging Graph.**, v. 46, p. 191–196, dec 2015. ISSN 08956111. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0895611115001172>. 9

BROWN, O. W.; HUGENHOLTZ, C. H. Estimating aerodynamic roughness (z o) in mixed-grassland prairie with airborne LiDAR. **Can. J. Remote Sens.**, v. 37, n. 4, p. 1–21, 2012. 106

BRYCE, R. M.; SPRAGUE, K. B. Revisiting detrended fluctuation analysis. **Sci. Rep.**, v. 2, n. 235, p. 315, jan 2012. ISSN 2045-2322. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3303145{&}tool=pmcentrez{&}rendertype=abstract>. 1, 19

BULDYREV, S. V.; GOLDBERGER, A. L.; HAVLIN, S.; PENG, C.-K.; SIMONS, M.; STANLEY, H. E. Generalized Lévy-walk model for DNA nucleotide sequences. **Phys. Rev. E**, v. 47, n. 6, p. 4514–4523, jun 1993. ISSN 1063-651X. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.47.4514>. 14

BULDYREV, S. V.; GOLDBERGER, A. L.; HAVLIN, S.; MANTEGNA, R. N.; MATSA, M. E.; PENG, C.-K.; SIMONS, M.; STANLEY, H. E. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. **Phys. Rev. E**, v. 51, n. 5, p. 5084–5091, may 1995. ISSN 1063-651X. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.51.5084>. 71

CAI, X.; LANGTANGEN, H. P.; MOE, H. On the performance of the python programming language for serial and parallel scientific computations. **Sci. Program.**, v. 13, n. 1, p. 31–56, 2005. ISSN 1058-9244. Disponível em: <http://www.hindawi.com/journals/sp/2005/619804/abs/>. 48, 49

CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. Análise espacial de áreas. In: Suzana Druck, Marília Sá Carvalho, Gilberto Câmara, A. M. V. M. (Ed.). **Análise espacial de dados geográficos**. Brasília/DF: Embrapa, 2004. cap. 5. ISBN 85-7383-260-6. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf>. 57, 164

CAMELO, N.; HENRIQUE, P.; MARIA, S. D.; MARQUES, P. C. Métodos de extrapolação de velocidade do vento para regiões litorâneas do nordeste brasileiro. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 16., Belém. **Anais...** Rio de Janeiro: SBMet, 2010. p. 5. 106, 123

CARBONE, A. Algorithm to estimate the Hurst exponent of high-dimensional fractals. **Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.**, v. 76, n. 5, p. 1–7, 2007. ISSN 15393755. 11

CASTRO, R. M. **Introdução a energia eólica**. Lisboa - Portugal: [s.n.], 2005. 82 p. 100

CAVALCANTE, H. L. D. d. S. **Bifurcações com intermitência e sincronismo de sistemas caóticos: circuitos eletrônicos e lasers**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2003. Disponível em: <http://repositorio.ufpe.br/handle/123456789/6520>. 38

CHEN, Z.; HU, K.; STANLEY, H. E. Effect of nonstationarities on detrended fluctuation analysis. **Phys. Rev. E**, v. 65, n. 4, p. 15, 2002. 1, 19

CLARKE, K. C. Computation of thye Fractal Dimension of Topographic Surfaces Using the Triangular Prism Surface Area Method. **Comput. Geosci.**, v. 12, p. 713–722, 1986. 10

COSTA, G. B.; LYRA, R. F. d. F. Análise dos padrões de vento no estado de Alagoas. **Rev. Bras. Meteorol.**, v. 27, n. 1, p. 31–38, 2012. 100

CRUZ, T. Leis de Escala e Dimensão Fractal em Filmes Finos: Microscopia de Força Atômica e Voltametria Cíclica. **Physicae**, v. 1, n. 1, p. 29–36, 2000. ISSN 16799569. 6

CZARNECKI; GRECH, D. Multifractal dynamics of stock markets. **Acta Phys. Pol. A**, v. 117, n. 4, p. 623–629, 2010. ISSN 05874246. 22, 23, 27, 89, 90, 91

DELIGNIERES, D.; RAMDANI, S.; LEMOINE, L.; TORRE, K.; FORTES, M.; NINOT, G. Fractal analyses for 'short' time series: A re-assessment of classical methods. **J. Math. Psychol.**, v. 50, n. 6, p. 525–544, 2006. ISSN 00222496. 15, 33, 34

EINSTEIN, A. J.; WU, H.-S.; GIL, J. Detrended Fluctuation Analysis of chromatin texture for diagnosis in breast cytology. **Fractals**, World Scientific Publishing Company, v. 10, n. 1, p. 19–25, apr 2012. Disponível em: <http://www.worldscientific.com/doi/abs/10.1142/S0218348X02000999>. 29

EKE, A.; HERMÁN, P.; BASSINGTHWAIGHTE, J. B.; RAYMOND, G. M.; PERCIVAL, D. B.; CANNON, M.; BALLA, I.; IKRÉNYI, C. Physiological time series: Distinguishing fractal noises from motions. **Pflugers Arch. Eur. J. Physiol.**, v. 439, n. 4, p. 403–415, 2000. ISSN 00316768. 15, 34

ELFATTAH, M. A.; EL-BENDARY, N.; Abu Elsoud, M. A.; HASSANIEN, A. E.; TOLBA, M. F. An intelligent approach for galaxies images classification. In: INTERNATIONAL CONFERENCE ON HYBRID INTELLIGENT SYSTEMS, 13. (HIS 2013), 2013, Gammarth, Tunisia. **Proceedings...** USA: IEEE, 2014. p. 167–172. ISBN 9781479924394. 3, 85

FAN, C.; LI, L. Characteristics analysis of nonstationary signals based on multifractal detrended fluctuation analysis method. In: CHINESE CONTROL AND DECISION CONFERENCE, 27. (CCDC), Quingdao, China. **Procedings...** IEEE, 2015. p. 1614–1618. ISBN 978-1-4799-7017-9. Disponível em: <http://ieeexplore.ieee.org/document/7162177/>. 81, 83, 91

FAN, Q.; WU, Y. Quantifying two-dimensional nonstationary signal with power-law correlations by detrended fluctuation analysis. **Eur. Phys. J. B**, v. 88, n. 8, p. 199, 2015. ISSN 1434-6028. Disponível em: <http://link.springer.com/10.1140/epjb/e2015-60332-x>. 77

FERNANDES, L. C. **Técnicas de reconhecimento de padrão aplicado a modelo digital de terreno e influências para o escoamento do vento**. 2015. 32p p. Monografia. 119, 120

FILHO, P. L. d. P. **Reconhecimento de Espécies Florestais através de imagens macroscópicas**. 116 p. Tese (Doutorado) — Universidade Federal do Paraná, 2012. 8

FIORIN, D. V.; SCHUCH, N.; MARTINS, F.; BRACKMANN, R.; CECONI, M.; PEREIRA, E.; GUARNIERI, R. Comportamento estatístico sazonal dos ventos na região do observatório espacial do sul. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 15., São Paulo/SP. **Anais...** Rio de Janeiro: SBMet, 2008. p. 1–5. 100

FORNARI, G. **Estudo de técnicas fractal e multifractal para análise de irregularidades na ionosfera equatorial**. 133 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, 2016. Disponível em: <http://urlib.net/8JMKD3MGP3W34P/3LHFT8S>. 13, 28, 54, 77

FORTIN, C.; OHLEY, W.; GEWIRTZT, H. Two-dimensional Fractional Brownian Motion in the analysis of medical images. In: INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, Paris. **Procedings...** New York: IEEE, 1992. p. 975–976. ISBN 0-7803-0785-2. Disponível em: <http://ieeexplore.ieee.org/document/5761333/>. 11, 34

FREITAS, R. M. d. **Laboratório virtual para visualização e caracterização do uso e cobertura da terra utilizando imagens de sensoriamento remoto**. 235 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012-03-16 2012. Disponível em: <http://urlib.net/sid.inpe.br/mtc-m19/2012/02.24.17.32>. Acesso em: 01 ago. 2017. 30, 54, 81

FREITAS, R. M. d.; GENOVEZ, P.; BENTZ, C. M.; ROSA, R. R.; SHIMABUKURO, Y. E. Análise de anisotropia de imagens utilizando o método dfa: Um estudo de caso na área de exploração de petróleo. In: EPIPHANIO, J. C. N.; GALVÃO, L. S. (Ed.). **Anais...** São José dos Campos: Instituto Nacional

de Pesquisas Espaciais (INPE), 2009. p. 6463–6470. ISBN 978-85-17-00044-7. Disponível em: <http://urlib.net/dpi.inpe.br/sbsr@80/2008/11.14.18.20>. Acesso em: 01 ago. 2017. 13, 29

GAUCI, A.; ADAMI, K. Z.; ABELA, J. Machine Learning for Galaxy Morphology Classification. **Computer (Long. Beach. Calif).**, v. 8, n. May, p. 1–8, 2010. Disponível em: <http://arxiv.org/abs/1005.0390>. 85, 86, 87

GEBEJES, A.; HUERTAS, R. Texture characterization based on grey-level co-occurrence matrix. In: PROCEEDINGS IN CONFERENCE OF INFORMATICS AND MANAGEMENT SCIENCES (ICTIC), Catania. **Proceedings...** [S.l.]: EDIS - Publishing Institution of the University of Zilina, 2013. p. 375–378. ISBN 978-80-554-0648-0. ISSN 1339-9144. 86, 159, 163

GILDEN; GILDEN, D.; THORNTON, T.; MALLON, M. **1/F Noise in Human Cognition**. 1995. 1837–1839 p. Disponível em: <http://www.sciencemag.org/content/267/5205/1837>. 36

GLERIA, I.; MATSUSHITA, R.; SILVA, S. Sistemas complexos, criticalidade e leis de potência. **Rev. Bras. Ensino da Física**, v. 26, n. 2, p. 99–108, 2004. ISSN 0102-4744. 1, 5

GNEITING, T.; SCHLATHER, M. Stochastic models which separate fractal dimension and Hurst effect. **SIAM Rev.**, v. 46, n. 2, p. 269–282, 2001. ISSN 0036-1445. Disponível em: <http://arxiv.org/abs/physics/0109031>. 10, 11

GOLDBERGER, A. L.; AMARAL, L. A. N.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals. **Circulation**, v. 101, n. 23, p. e215–e220, jun 2000. ISSN 0009-7322. Disponível em: <http://circ.ahajournals.org/cgi/doi/10.1161/01.CIR.101.23.e215>. 54

GÓMEZ-GUTIÉRREZ, Á.; SCHNABEL, S.; CONTADOR, F. L.; MARÍN, R. G. Testing the quality of open-access DEMs and their derived attributes in Spain : SRTM , GDEM and PNOA DEM. **Geomorphometry**, p. 53–56, 2011. 44

GONG, P.; MARCEAU, D. J.; HOWARTH, P. J. A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. **Remote Sens. Environ.**, v. 40, n. 2, p. 137–151, 1992. ISSN 00344257. 159, 162

GONZALES-BARRON, U.; BUTLER, F. Fractal texture analysis of bread crumb digital images. **Eur. Food Res. Technol.**, v. 226, n. 4, p. 721–729, 2008. ISSN 14382377. 9

GRECH, D. Alternative measure of multifractal content and its application in finance. **Chaos, Solitons and Fractals**, v. 88, p. 183–195, 2016. ISSN 09600779. 90

GRIGORIEV, R. O. Optimal Control of Spatiotemporal Chaos in Coupled Map Lattices. In: **Solid Mech. its Appl.** [S.l.: s.n.], 1999. p. 543–550. 37

GROHMANN, C. H. Aplicações dos Modelos de Elevação SRTM em Geomorfologia. **Rev. Geogr. Acadêmica**, v. 2, n. 2, p. 73–83, 2008. 42

GU, G.-F.; ZHOU, W.-x. Detrended fluctuation analysis for fractals and multifractals in higher dimensions. **Phys. Rev.**, v. 74, n. 6, p. 1–7, 2006. xvii, 2, 18, 20, 21, 25, 29, 30, 35, 47, 50, 52, 53

GUALTIERI, G.; SECCI, S. Comparing methods to calculate atmospheric stability-dependent wind speed profiles: A case study on coastal location. **Renew. Energy**, Elsevier Ltd, v. 36, n. 8, p. 2189–2204, aug 2011. ISSN 09601481. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0960148111000450>. 102, 105

GULO, C. A. S. J. **Técnicas de paralelização em GPGPU aplicadas em algoritmo para remoção de ruído multiplicativo**. 80 p. Dissertação (Mestrado) — Unversidade Estadual Paulistana, 2012. 60

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. The WEKA data mining software: an update. **SIGKDD Explor.**, v. 11, p. 10–18, 2009. 166

HALLEY, J. M.; KUNIN, W. E. Extinction Risk and the 1/f Family of Noise Models. **Theor. Popul. Biol.**, v. 56, n. 3, p. 215–30, 1999. ISSN 0040-5809. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/10607517>. 36

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural Features for Image Classification. **IEEE Trans. Syst. Man Cybern.**, v. 3, p. 610–620, 1973. 85, 86, 87, 92, 159, 160, 162

HASAGER, C. B.; BADGER, M.; DELLWIK, E.; FLOORS, R. R.; HAHMANN, A. N.; MANN, J. On the spatial and temporal resolution of land cover products for applied use in wind resource mapping. In: **WorldCover**. Rome, Italy: [s.n.], 2017. 108

HASEGAWA, H. H.; SAPHIR, W. C. Unitarity and irreversibility in chaotic systems. **Phys. Rev. A**, v. 46, n. 12, p. 7401–7423, dec 1992. ISSN 1050-2947. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevA.46.7401>. 38

HAYAKAWA, Y. S.; OGUCHI, T.; LIN, Z. Comparison of new and existing global digital elevation models: ASTER G-DEM and SRTM-3. **Geophys. Res. Lett.**, v. 35, n. 17, p. 1–5, 2008. ISSN 00948276. 44

HENEGHAN, C.; LOWEN, S.; TEICH, M. Two-dimensional fractional brownian motion: wavelet analysis and synthesis. In: IEEE SOUTHWEST SYMPOSIUM ON IMAGE ANALYSIS AND INTERPRETATION, San Antonio,TX. **Proceedings...** [S.l.], 1996. p. 213–217. ISBN 0-7803-3200-8. ISSN 0780332008. 34, 35

HENEGHAN, C.; MCDARBY, G. Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. **Phys. Rev. E**, v. 62, n. 5, p. 6103–6110, 2000. ISSN 1063-651X. 1, 19, 71

HÖLL, M.; KANTZ, H. The fluctuation function of the detrended fluctuation analysis - Investigation on the AR (1) process. **Eur. Phys. J. B**, v. 88, n. 126, p. 1–15, 2015. ISSN 1434-6028. 19

_____. The relationship between the detrendend fluctuation analysis and the autocorrelation function of a signal. **Eur. Phys. J. B**, v. 88, n. 12, p. 1–7, 2015. ISSN 14346036. 1, 19

HUBBLE, E. P. **The realm of the nebulae**. [S.l.]: Yale University Press, 1936. 39, 85

HUGENHOLTZ, C. H.; BROWN, O. W.; BARCHYN, T. E. Estimating aerodynamic roughness (z0) from terrestrial laser scanning point cloud data over un-vegetated surfaces. **Aeolian Res.**, Elsevier B.V., v. 10, p. 161–169, 2013. ISSN 18759637. Disponível em: <http://dx.doi.org/10.1016/j.aeolia.2013.03.004>. 103, 106, 109

HURST, H. Long term storage capacity of reservoirs. **Trans. thr Am. Soc. Civ. Eng.**, v. 116, p. 770–799, 1951. 1, 7, 10

IBANEZ, D. M. **Análise do relevo da Amazônia Central com o emprego de modelo digital de elevação e geometria fractal**. Tese (Doutorado) — Universidade de São Paulo, 2012. 10

IEEE. **The 2016 Top Programming Languages**. 2016. Disponível em: <http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages>. xiii, 48, 49

INRIA. **FRACLAB, A Fractal Analysis Toolbox for Signal and Image Processing**. 2016. Disponível em: <http://fraclab.saclay.inria.fr/>. 35

JAGER, D.; ANDREAS, A. **NREL National Wind Technology Center (NWTC)**: M2 tower; boulder, colorado (data). nrel report no. da-5500-56489. NREL, 1996. Disponível em: <https://www.nrel.gov/midc/nwtc{_}m2/>. 110

JIMÉNEZ, J.; Ruiz de Miras, J. Fast box-counting algorithm on GPU. **Comput. Methods Programs Biomed.**, Elsevier Ireland Ltd, v. 108, n. 3, p. 1229–1242, 2012. ISSN 01692607. Disponível em: <http://dx.doi.org/10.1016/j.cmpb.2012.07.005>. 59, 60

KANEKO, K. Overview of coupled map lattices. **Chaos An Interdiscip. J. Nonlinear Sci.**, v. 2, n. 3, p. 279–282, jul 1992. ISSN 1054-1500. Disponível em: <http://aip.scitation.org/doi/10.1063/1.165869>. 37

KANTELHARDT, J. W. Fractal and Multifractal Time Series. In: Robert A. Meyers (Ed.). **Encycl. Complex. Syst. Sci.** 1. ed. [S.l.]: Springer-Verlag New York, 2009. cap. Fractals a, p. MCXX, 10398. ISBN 978-0-387-75888-6. 1, 5, 14

KANTELHARDT, J. W.; KOSCIELNY-BUNDE, E.; REGO, H. H. A. Detecting long-range correlations with detrended fluctuation analysis. **Physica A**, v. 295, p. 441–454, 2001. 18

KANTELHARDT, J. W.; ZSCHIEGNER, S. A.; KOSCIELNY-BUNDE, E.; HAVLIN, S.; BUNDE, A.; STANLEY, H. E. Multifractal detrended fluctuation analysis of nonstationary time series. **Physica A**, v. 316, n. 1-4, p. 87–114, 2002. ISSN 03784371. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378437102013833>. 1, 15, 23, 25, 53, 76

KAPLAN, L.; KUO, C.-C. Texture roughness analysis and synthesis via extended self-similar (ESS) model. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 17,

n. 11, p. 1043–1056, 1995. ISSN 01628828. Disponível em: <http://ieeexplore.ieee.org/document/473230/>. 8

KAPLAN, L. M.; KUO, C.-C. An improved method for 2-D self-similar image synthesis. **IEEE Trans. Image Process.**, v. 5, n. 5, p. 754–761, 1996. ISSN 1057-7149. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/18285164>. 34, 35

KAVASSERI, R. G.; NAGARAJAN, R. Evidence of crossover phenomena in wind speed data. **Circuits Syst. I Regul. Pap. IEEE Trans.**, v. 51, n. 11, p. 2255–2262, 2004. 1, 12, 20

KLÖCKNER, A. **PyCuda**. 2017. Disponível em: <https://mathema.tician.de/software/pycuda/>. 62

KLOCKNER, A.; PINTO, N.; LEE, Y.; CATANZARO, B.; IVANOV, P.; FASIH, A. PyCUDA and PyOpenCL: A scripting-based approach to GPU run-time code generation. **Parallel Comput.**, v. 38, n. 3, p. 157–174, 2012. ISSN 01678191. 60, 62

LACASA, L.; LUQUE, B.; LUQUE, J.; NUÑO, J. C. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. **EPL (Europhysics Lett.**, v. 86, n. 3, p. 30001, 2009. ISSN 0295-5075. Disponível em: <http://iopscience.iop.org/0295-5075/86/3/30001/fulltext/>. 11

LANCASTER, N. Relations between aerodynamic and surface roughness in a hyper-arid cold desert: McMurdo Dry Valleys, Antarctica. **Earth Surf. Process. Landforms**, v. 29, n. 7, p. 853–867, 2004. ISSN 01979337. 105, 106, 110

LANDAU, E. C.; GUIMARÃES, D. P. Análise comparativa entre os modelos digitais de elevação aster, srtm e topodata. In: EPIPHANIO, J. C. N.; GALVÃO, L. S. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2011. p. 4003–4010. ISBN 978-85-17-00056-0 (Internet) and 978-85-17-00057-7 (DVD). Disponível em: <http://urlib.net/dpi.inpe.br/marte/2011/07.18.13.40>. Acesso em: 01 ago. 2017. 43

LANGTANGEN, H. P. **Python scripting for computational science**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. (Texts in Computational Science and Engineering, v. 3). ISSN 1611-0994. ISBN 978-3-540-73915-9. Disponível em: <http://link.springer.com/10.1007/978-3-540-73916-6>. 48

LEKSHMI, S.; REVATHY, K.; Prabhakaran Nayar, S. R. Galaxy classification using fractal signature. **Astron. Astrophys.**, v. 405, n. 3, p. 1163–1167, 2003. ISSN 0004-6361. 3, 87, 91

LENNON, J. J. Red-shifts and red herrings in geographical ecology. **Ecography (Cop.).**, v. 23, n. 1, p. 101–113, 2000. ISSN 09067590. 37

LI, J.; DU, Q.; SUN, C. An improved box-counting method for image fractal dimension estimation. **Pattern Recognit.**, v. 42, n. 11, p. 2460–2469, nov 2009. ISSN 00313203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320309000843>. 9

LUDESCHER, J.; BOGACHEV, M. I.; KANTELHARDT, J. W.; SCHUMANN, A. Y.; BUNDE, A. On spurious and corrupted multifractality: The effects of additive noise, short-term memory and periodic trends. **Phys. A Stat. Mech. its Appl.**, Elsevier B.V., v. 390, n. 13, p. 2480–2490, 2011. ISSN 03784371. Disponível em: <http://dx.doi.org/10.1016/j.physa.2011.03.008>. 89

LUNDAHL, T.; OHLEY, W. J.; KAY, S. M.; SIFFERT, R. Fractional brownian motion: a maximum likelihood estimator and its application to image texture. **IEEE Trans. Med. Imaging**, v. 5, n. 3, p. 152–61, 1986. ISSN 0278-0062. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/18244001>. 34

LYRA, G. B. **Parâmetros de rugosidade aerodinâmica sobre vegetação esparsa**. 75 p. Tese (Doutorado) — Universidade de São Paulo, 2005. 100, 105

LYRA, G. B.; PEREIRA, A. R. Aerodynamic Roughness Parameters Over Sparse Vegetation in a Semi-Arid Region. **Rev. Bras. Meteorol.**, v. 22, n. 2, p. 262–272, 2007. 105

MAKOWIEC, D.; RYNKIEWICZ, a.; GAŁASKA, R.; WDOWCZYK-SZULC, J.; ŻARCZYŃSKA-BUCHOWIECKA, M. Reading multifractal spectra: Aging by multifractal analysis of heart rate. **EPL (Europhysics Lett.**, v. 94, n. 6, p. 68005, 2011. ISSN 0295-5075. Disponível em: <http://stacks.iop.org/0295-5075/94/i=6/a=68005?key=crossref.40da0a1f1d23425dc9b5d4d73459c98f>. xiii, 23, 26, 27, 89, 90

MALATHI, G.; SHANTHI, V. Statistical measurement of ultrasound placenta images using segmentation approach. **J. Inf. Hiding Multimed. Signal Process.**, v. 2, n. 4, p. 332–343, 2011. ISSN 20734212. 163

MANDELBROT, B. B.; Van Ness, J. W. Fractional Brownian Motions, Fractional Noises and Applications. **SIAM Rev.**, v. 10, n. 4, p. 422–437, oct 1968. ISSN 0036-1445. Disponível em: <http://epubs.siam.org/doi/abs/10.1137/1010093>. 33, 34

MANN, J.; ANGELOU, N.; ARNQVIST, J.; CALLIES, D.; CANTERO, E.; ARROYO, R. C.; COURTNEY, M.; CUXART, J.; DELLWIK, E.; GOTTSCHALL, J.; IVANELL, S.; KÜHN, P.; LEA, G.; MATOS, J. C.; PALMA, J. M. L. M.; PAUSCHER, L.; PEÑA, A.; RODRIGO, J. S.; SÖDERBERG, S.; VASILJEVIC, N.; RODRIGUES, C. V. Complex terrain experiments in the New European Wind Atlas. **Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.**, v. 375, n. 2091, p. 20160101, apr 2017. ISSN 1364-503X. Disponível em: <http://rsta.royalsocietypublishing.org/lookup/doi/10.1098/rsta.2016.0101>. 108

MANWELL, J.; MCGOWAN, J.; ROGERS, A. **Wind energy**: theory, design and application. Chichester, UK: John Wiley & Sons, 2002. xviii, 99, 103, 107, 109, 127, 131

MANWELL, J. F.; MCGOWAN, J. G.; ROGERS, A. L. **Wind energy explained**. Chichester, UK: John Wiley & Sons, 2009. 677 p. ISBN 9781119994367. 101, 104, 105

MARAUN, D.; RUST, H. W.; TIMMER, J. Tempting long-memory - on the interpretation of DFA results. **Nonlinear Process. Geophys.**, v. 11, p. 495–503, 2004. 18

MARINHO, E. B. S.; BASSREI, A.; ANDRADE, R. F. S. Comparação de estatísticas em análise de séries temporais : aplicação a dados climatológicos de Salvador - BA. In: INTERNATIONAL CONGRESS OF THE BRAZILIAN GEOPHYSICAL SOCIETY, 14., Rio de Janeiro. **Proceedings...** Rio de Janeiro: Sociedade Brasileira de Geofísica, 2015. Disponível em: <https://doi.org/10.1190/sbgf2015-292>. 6

MARTINS, F.; GUARNIERI, R.; PEREIRA, E. O aproveitamento da energia eólica. **Rev. Bras. Ensino Física**, v. 30, n. 1, p. 1304.1–1304.13, 2008. ISSN 1806-1117. xv, 100

MARTINS, M. H. A.; CLARKE, R. T.; HAERTEL, V. Emprego da dimensão fractal para separar classes de textura presentes numa aerofoto da cidade de Porto Alegre. **Bol. Parana. Geociências**, v. 52, p. 7–18, 2003. 10

MCGAUGHEY, D. R.; AITKEN, G. J. M. Generating two-dimensional fractional Brownian motion using the fractional Gaussian process (FGp) algorithm. **Phys. A Stat. Mech. its Appl.**, v. 311, n. 3-4, p. 369–380, 2002. ISSN 03784371. 34, 35

MENDONÇA, M. B. d. M. de; GARCIA, C. A. d. A.; NOGUEIRA, R. d. A.; GOMES, M. A. F.; VALENÇA, M. M.; ORÉFICE, F. Análise fractal da vasculatura retínica: métodos de segmentação e de cálculo dimensional. **Arq. Bras. Oftalmol.**, Conselho Brasileiro de Oftalmologia, v. 70, n. 3, p. 413–422, jun 2007. ISSN 0004-2749. 5

MIELNICZUK, J.; WOJDYŁŁO, P. Estimation of Hurst exponent revisited. **Comput. Stat. Data Anal.**, v. 51, n. 9, p. 4510–4525, 2007. ISSN 01679473. 12

MIETUS, J.; PENG, C.-K.; MOODY, G. **Detrended Fluctuation Analysis (DFA)**. Physionet, 2005. Disponível em: <http://physionet.incor.usp.br/physiotools/dfa/>. 54

MIYAZAKI, S.; HIGUCHI, Y. Chaos as Irregular Hopping between Unstable Periodic Orbits and. **Forma**, v. 28, p. 1–5, 2013. 38

MME, M. d. M. e. E. **Resenha energética brasileira**: Exercício de 2015. Brasília: Ministério das Minas e Energia, 2016. Disponível em: <http://www.mme.gov.br/documents/10584/3580498/02+-+Resenha+Energ{é}tica+Brasileira+2016+-+Ano+Base+2015+(PDF)/66e011ce-f34b-419e-adf1-8a3853c95fd4;version=1.0>. 101

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, v. 37, p. 17–23, 1950. 48, 57, 85

MUZY, J. F.; BACRY, E.; ARNEODO, A. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. **Phys. Rev. E**, v. 47, n. 2, p. 875–884, feb 1993. ISSN 1063-651X. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.47.875>. 22

MYINT, S. W. Fractal approaches in texture analysis and classification of remotely sensed data: Comparisons with spatial autocorrelation techniques and simple descriptive statistics. **Int. J. Remote Sens.**, v. 24, n. 9, p. 1925–1947, 2003. ISSN 0143-1161. 9, 10, 86

NIE, Q.; XU, J.; MAN, W.; SUN, F. Detrended fluctuation analysis of spatial patterns on urban impervious surface. **Environ. Earth Sci.**, Springer Berlin

Heidelberg, v. 74, n. 3, p. 2531–2538, 2015. ISSN 18666299. Disponível em: <http://dx.doi.org/10.1007/s12665-015-4271-4>. 31

NIELD, J. M.; KING, J.; WIGGS, G. F. S.; LEYLAND, J.; BRYANT, R. G.; CHIVERRELL, R. C.; DARBY, S. E.; ECKARDT, F. D.; THOMAS, D. S. G.; VIRCAVS, L. H.; WASHINGTON, R. Estimating aerodynamic roughness over complex surface terrain. **J. Geophys. Res. Atmos.**, v. 118, n. 23, p. 12948–12961, 2013. ISSN 21698996. 106, 109

NVIDIA. **NVIDIA CUDA programming documentation**. 2016. Disponível em: <http://docs.nvidia.com/cuda/>. 60

NVIDIA; UIUC, U. o. I. **GPU Teaching Kit : Accelerated Computing**. 3 edition. ed. NVIDIA, 2015. Disponível em: <https://developer.nvidia.com/teaching-kits>. xiv, 60, 61

OHANIAN, P. P.; DUBES, R. C. Performance evaluation for four classes of textural features. **Pattern Recognit.**, v. 25, n. 8, p. 819–833, 1992. ISSN 00313203. 8, 9

OLAOFE, Z. O.; FOLLY, K. a. Wind energy analysis based on turbine and developed site power curves: A case-study of Darling City. **Renew. Energy**, Elsevier Ltd, v. 53, p. 306–318, may 2013. ISSN 09601481. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0960148112006945>. 101

OLSEN, L. A Multifractal Formalism. **Adv. Math. (N. Y).**, v. 116, n. 1, p. 82–196, nov 1995. ISSN 00018708. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0001870885710663>. 22

PANTIC, I.; NESIC, Z.; Paunovic Pantic, J.; RADOJEVIĆ-ŠKODRIĆ, S.; CETKOVIC, M.; Basta Jovanovic, G. Fractal analysis and Gray level co-occurrence matrix method for evaluation of reperfusion injury in kidney medulla. **J. Theor. Biol.**, v. 397, p. 61–67, 2016. ISSN 00225193. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0022519316001405>. 9

PARISI, G. **Field theory, disorder and simulations**. Singapore: World Scientific, 1992. 1

PARK, B.; CHEN, Y. R. Co-occurrence Matrix Texture Features of Multi-spectral Images on Poultry Carcasses. **J. Agric. Eng. Res.**, v. 78, n. 2, p. 127–139, 2001. ISSN 00218634. Disponível em: <http:

//www.sciencedirect.com/science/article/pii/S0021863400906292>. 93, 162, 163

PAUL-LIMOGES, E.; CHRISTEN, A.; COOPS, N. C.; BLACK, T. A.; TROFYMOW, J. A. Estimation of aerodynamic roughness of a harvested Douglas-fir forest using airborne LiDAR. **Remote Sens. Environ.**, Elsevier Inc., v. 136, p. 225–233, 2013. ISSN 00344257. Disponível em: <http://dx.doi.org/10.1016/j.rse.2013.05.007>. 106, 109

PEDERSEN, K. S.; STENSBO-SMIDT, K.; ZIRM, A.; IGEL, C. Shape index descriptors applied to texture-based galaxy analysis. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, Salt Lake City. **Proceedings...** Salt Lake City, UT: Scientific Computing and Imaging Institute University of Utah, 2013. p. 2440–2447. ISBN 9781479928392. ISSN 1550-5499. 85, 86, 87

PENG, C.; BULDYREV, S.; GOLDBERGER, A.; HAVLIN, S.; SCIORTINO, F.; SIMONS, M.; STANLEY, H. Long-range correlations in nucleotide sequences. **Phys. Rev. E**, v. 49, p. 1685–1689, 1994. 1, 2, 12, 13, 14, 18

PENG, C. K.; HAVLIN, S.; STANLEY, H. E.; GOLDBERGER, A. L. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. **Chaos**, v. 5, p. 82–87, 1995. 12, 28

PENTTINEN, A.; VIRTAMO, J. Simulation of Two-Dimensional Fractional Gaussian Noise. **Methodol. Comput. Appl. Probab.**, v. 6, n. 1, p. 99–107, mar 2004. ISSN 1387-5841. Disponível em: <http://link.springer.com/10.1023/B:MCAP.0000012417.10460.03>. 35

PHAM, T. A. **Optimization of texture feature extraction algorithm**. Dissertação (Mestrado) — Delft University of Technology, 2010. 86, 92, 162

PRUM, R. O.; TORRES, R. H.; WILLIAMSON, S.; DYCK, J. Coherent light scattering by blue feather barbs. **Nature**, v. 396, n. 6706, p. 28–29, nov 1998. ISSN 00280836. Disponível em: <http://www.nature.com/doifinder/10.1038/23838>. 71, 165

QIAN, H.; RAYMOND, G. M.; BASSINGTHWAIGHTE, J. B. On two-dimensional fractional Brownian motion and fractional Brownian random field. **J. Phys. A. Math. Gen.**, v. 31, n. 28, p. L527–L535, jul 1998. ISSN 0305-4470. Disponível em: <http://stacks.iop.org/0305-4470/31/i=28/a=002?key=crossref.b3ae893804f24d0eb7bcab273b326c99>. 35

QUILICI-GONZALEZ, J. A.; ZAMPIROLLI, F. d. A. **Sistemas inteligentes e mineração de dados**. Santo André/ SP: Triunfal Gráfica e Editora, 2014. 165

QUINLAN, J. R. **C4.5 Programs for Machine Learning**. San Mateo, CA: Morgan Kaufmann, 1992. 94, 166

RAMOTOROKO, C. D.; RANGANAI, R. T.; NYABEZE, P. Extension of the Archaean Madibe-Kraaipan granite-greenstone terrane in southeast Botswana: Constraints from gravity and magnetic data. **J. African Earth Sci.**, v. 123, p. 39–56, nov 2016. ISSN 1464343X. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1464343X16302151>. 165

REA, W.; OXLEY, L.; REALE, M.; BROWN, J. Estimators for Long Range Dependence: An Empirical Study. **Fractals An Interdiscip. J. Complex Geom. Nat.**, v. 3, n. 1980, p. 0–16, 2009. Disponível em: <http://arxiv.org/abs/0901.0762>. 12

REIS, A. L. dos. **Influência da rugosidade do terreno para o perfil vertical do vento**. 2016. 77 p. Monografia. 120

RODRIGUES, T.; DEBIASI, P.; SOUZA, R. Avaliação da adequação dos produtos ASTER GDEM no auxílio ao mapeamento sistemático brasileiro. In: SIMPÓSIO BRAS. CIÊNCIAS GEODÉSICAS E TECNOL. DA GEOINFORMAÇÃO, 3., Recife. **Anais...** Recife: UFPE, 2010. 43

ROSA, R.; BARONI, M.; ZANIBONI, G.; Ferreira da Silva, A.; ROMAN, L.; PONTES, J.; BOLZAN, M. Structural complexity of disordered surfaces: Analyzing the porous silicon SFM patterns. **Phys. A Stat. Mech. its Appl.**, v. 386, n. 2, p. 666–673, dec 2007. ISSN 03784371. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0378437107009557>. 1

RUZANSKI, E. **Radially averaged power spectral density - Matlab code**. 2011. Disponível em: <https://www.mathworks.com/matlabcentral/fileexchange/23636-radially-averaged-power-spectrum-of-2d-real-valued-matrix>. 71, 165

SANTOS, A. G. **Estrutura fractal em séries temporais: uma investigação quanto à hipótese de passeio aleatório no mercado à vista de commodities agrícolas brasileiro**. 151 p. Tese (Dissertação (Mestrado)) — Fundação Getúlio Vargas, 2013. 5, 6

SARKAR, N.; CHAUDHURI, B. An efficient approach to estimate fractal dimension of textural images. **Pattern Recognit.**, v. 25, n. 9, p. 1035–1041, sep 1992. ISSN 00313203. Disponível em: <http://www.sciencedirect.com/science/article/pii/003132039290066R>. 9, 86

SAUTTER, R. **CML**. 2017. Disponível em: <https://github.com/rsautter/CML-CoupledMapLattice>. 39

SCHMITTBUHL, J.; VILOTTE, J.-p.; ROUX, S. Reliability of self-affine measurements. **Phys. Rev. E**, v. 51, n. 1, p. 131–147, jan 1995. ISSN 1063-651X. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.51.131>. 8, 9

SCHUTTER, A.; SHAMIR, L. Galaxy morphology - An unsupervised machine learning approach. **Astron. Comput.**, Elsevier B.V., v. 12, p. 60–66, 2015. ISSN 22131337. Disponível em: <http://dx.doi.org/10.1016/j.ascom.2015.05.002>. 85, 86, 87, 97

SCIPY.ORG. **Numpy and SciPy Documentation**. 2016. Disponível em: <https://docs.scipy.org/doc/>. 48, 50, 163

SHAMIR, L. Automatic morphological classification of galaxy images. **Mon. Not. R. Astron. Soc.**, v. 399, n. 3, p. 1367–1372, 2009. ISSN 00358711. 85, 87, 98

SHAO, Y.-h.; GU, G.-f.; JIANG, Z.-q.; ZHOU, W.-x.; SORNETTE, D. Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. **Sci. Rep.**, v. 2, p. 1–5, 2012. 19

SHARMA, A. K.; SAHNI, S. A Comparative Study of Classification Algorithms for Spam Email Data Analysis. **IJCSE**, v. 3, n. 5, p. 1890–1895, 2011. 166

SILVA, R. V. **Estudo da sensibilidade da estimativa de vento no modelo WASP® relativa à rugosidade superficial**. 112 p. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2013. Disponível em: <http://repositorio.ufpe.br/handle/123456789/13183>. 102, 104, 107

SMITH, M. W.; QUINCEY, D. J.; DIXON, T.; BINGHAM, R. G.; CARRIVICK, J. L.; IRVINE-FYNN, T. D. L.; RIPPIN, D. M. Aerodynamic roughness of glacial ice surfaces derived from high-resolution topographic data. **J. Geophys. Res. Earth Surf.**, v. 121, n. 4, p. 748–766, 2016. ISSN 21699011. 103, 109

SOUZA, J.; SANTOS, A.; GUARIEIRO, L.; MORET, M. Fractal aspects in O2 enriched combustion. **Phys. A Stat. Mech. its Appl.**, v. 434, p. 268–272, sep 2015. ISSN 03784371. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378437115003891>. 14

SRINIVASAN, G.; SHOBHA, G. Statistical texture analysis. In: PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY. **Proceedings...** 2008. v. 36, n. December, p. 1264–1269. ISBN 9180671780. Disponível em: <http://staff.fh-hagenberg.at/wbackfri/Teaching/FBA/Uebungen/UE07charRecog/StatTextAnalysisSrinivasan08.pdf>. 9

THANKI, S.; RHEE, G.; LEPP, S. Fractal Dimension of Galaxy Isophotes. **Astron. J.**, v. 138, n. 3, p. 941–950, 2009. ISSN 0004-6256. Disponível em: <http://stacks.iop.org/1538-3881/138/i=3/a=941?key=crossref.c962757a989370fb19307e0eef7ce8fc>. 3, 40, 85, 86, 87, 91, 98

TIAN, X.; LI, Z. Y.; TOL, C. van der; SU, Z.; LI, X.; HE, Q. S.; BAO, Y. F.; CHEN, E. X.; LI, L. H. Estimating zero-plane displacement height and aerodynamic roughness length using synthesis of LiDAR and SPOT-5 data. **Remote Sens. Environ.**, v. 115, n. 9, p. 2330–2341, 2011. ISSN 00344257. 102, 106, 109

USGS. **Earth explorer**. 2016. Disponível em: <http://earthexplorer.usgs.gov/>. 43, 44

USGS, U. S. G. S. **Routine ASTER Global Digital Elevation Model**. 2017. Disponível em: <https://lpdaac.usgs.gov/dataset{_}discovery/aster>. 43

VARGAS-OLMOS, C.; MURGUÍA, J. S.; RAMÍREZ-TORRES, M. T.; Mejía Carlos, M.; ROSU, H. C.; GONZÁLEZ-AGUILAR, H. Two-dimensional DFA scaling analysis applied to encrypted images. **Int. J. Mod. Phys. C**, v. 26, n. 08, p. 1550093, 2015. ISSN 0129-1831. Disponível em: <http://www.worldscientific.com/doi/abs/10.1142/S012918311550093X>. 30, 53

VELAZQUEZ-CAMILO, O.; BOLAÑOS-REYNOSO, E.; RODRIGUEZ, E.; ALVAREZ-RAMIREZ, J. Fractal analysis of crystallization slurry images. **J. Cryst. Growth**, v. 312, n. 6, p. 842–850, 2010. ISSN 00220248. 30

VOSS, R. F. Evolution of long-range fractal correlations and 1/f noise in DNA base sequences. **Phys. Rev. Lett.**, v. 68, n. 25, p. 3805–3808, 1992. ISSN 00319007. 36

WAIKATO, U. of. **Weka - Waikato Enviroment for Knowledge Analysis**. 2016. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>. 85

WANG, F.; LI, Z.-s.; LI, J.-w. Applied Surface Science Local multifractal detrended fluctuation analysis for non-stationary image ' s texture segmentation. **Appl. Surf. Sci.**, v. 322, p. 116–125, 2014. 25

WANG, F.; LIAO, D.-w.; LI, J.-w.; LIAO, G.-p. Two-dimensional multifractal detrended fluctuation analysis for plant identification. **Plant Methods**, v. 11, p. 12, 2015. ISSN 1746-4811. Disponível em: <http://www.plantmethods.com/content/11/1/12>. 25

_____. _____. **Plant Methods**, v. 11, n. 12, p. 1–11, 2015. 31

WIERNGA, J. Representative roughness parameters for homogeneous terrain. **Boundary-Layer Meteorol.**, v. 63, n. 4, p. 323–363, 1993. ISSN 00068314. xviii, 103, 105, 107, 108, 109, 127, 131

WITT, A.; MALAMUD, B. D. Quantification of Long-Range Persistence in Geophysical Time Series: Conventional and Benchmark-Based Improvement Techniques. **Surv. Geophys.**, v. 34, n. 5, p. 541–651, 2013. ISSN 0169-3298. Disponível em: <http://link.springer.com/10.1007/s10712-012-9217-8>. 7, 20

XIE, X. A review of recent advances in surface defect detection using texture analysis techniques. **. . . Lett. Comput. Vis. Image Anal.**, v. 7, n. 3, p. 1–22, 2008. ISSN 1577-5097. 8, 10

XIONG, H.; SHANG, P. Detrended fluctuation analysis of multivariate time series. **Commun. Nonlinear Sci. Numer. Simul.**, v. 42, p. 12–21, jan 2017. ISSN 10075704. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1007570416301502>. 2

YADAV, R. P.; DWIVEDI, S.; MITTAL, A. K.; KUMAR, M.; PANDEY, A. C. Fractal and multifractal analysis of LiF thin film surface. **Appl. Surf. Sci.**, Elsevier B.V., v. 261, p. 547–553, 2012. ISSN 01694332. Disponível em: <http://dx.doi.org/10.1016/j.apsusc.2012.08.053>. 30

YEARSLEY, J. **spatialPattern**. 2016. Disponível em: <https://www.mathworks.com/matlabcentral/fileexchange/5091-generate-spatial-data>. 37

YEH, R.-g.; LIN, C.-w.; ABBOD, M. F.; SHIEH, J.-s. Two-Dimensional Matrix Algorithm Using Detrended Fluctuation Analysis to Distinguish Burkitt and Diffuse Large B-Cell Lymphoma. **Comput. Math. Methods Med.**, v. 2012, p. 1–8, 2012. ISSN 1748-670X. Disponível em: <http://www.hindawi.com/journals/cmmm/2012/947191/>. 29, 54, 66

ZHOU, Y.; LEUNG, Y.; YU, Z. G. Relationships of exponents in two-dimensional multifractal detrended fluctuation analysis. **Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.**, v. 87, n. 1, p. 4–7, 2013. ISSN 15393755. 21

ZUÑIGA, A. G.; FLORINDO, J. B.; BRUNO, O. M. Gabor wavelets combined with volumetric fractal dimension applied to texture analysis. **Pattern Recognit. Lett.**, v. 36, n. 1, p. 135–143, 2014. ISSN 01678655. 8, 9

APÊNDICE A - TÉCNICAS COMPLEMENTARES

Este apêndice detalha os conceitos e a implementação das técnicas que foram usadas para complementar a compreensão e ter parâmetros de comparação com o DFA. São elas : Descritores de Haralick, Índice de Moran e *Power Spectrum*.

O apêndice apresenta também o algoritmo de árvore de decisão J4.8 utilizado nas tarefas de classificação realizadas na tese.

A.1 Matriz de co-ocorrência dos níveis de cinza e descritores texturais

Há diversas formas de analisar a textura em imagens digitais, tais como os métodos de Wavelet, Fourier e a teoria de Texton (GEBEJES; HUERTAS, 2013). No entanto, uma análise que se aproxime da forma como o sistema visual humano percebe a textura é mais interessante. Essa análise, mais simples, foi definida por Haralick e colaboradores em 1973, e ainda é amplamente utilizada.

Para medir as características texturais de um objeto, (HARALICK et al., 1973) propuseram o método da matriz de co-ocorrência dos níveis de cinza (*Gray Level Cooccurrence Matrix Method* - GLCM), composto por duas fases : na primeira, constrói-se a chamada Matriz de Co-ocorrência, a qual é uma tabulação de diferentes combinações de valores de intensidade dos pixels que ocorrem em uma imagem; na segunda fase, são calculadas as métricas a partir da matriz de co-ocorrência, que descrevem a informação espacial do alvo, a partir da posição relativa dos elementos na matriz (GONG et al., 1992). A GLCM é atualmente uma técnica conhecida e consolidada para extração de textura em imagens digitais.

A matriz de co-ocorrência é uma matriz quadrada que tem como tamanho a quantidade de níveis de cinza da imagem a ser analisada. As combinações de ocorrência entre os níveis de cinza são calculadas nas direções 0, 45, 90 e 135 °. Cada par i, j da matriz de co-ocorrência apresenta a quantidade de vezes em que o tom de cinza i apareceu na direção θ, a uma distância d, do tom de cinza j. A figura A.1 ilustra a construção da matriz de co-ocorrência. A figura A.1a apresenta uma matriz quantizada em três níveis de cinza (0,1,2). A figura A.1b mostra a matriz de co-ocorrência gerada a partir da figura A.1a, nas direções 0 °. O valor $P(i, j)$ armazenado em uma linha i e coluna j da matriz da figura A.1b consiste no número de vezes em que pixels de valor i possuem vizinho de valor j à esquerda ou à direita e a uma distância de 1 pixel. Depois de somados os valores, a matriz é normalizada (figura A.1c). Cada direção analisada gera uma matriz de co-ocorrência diferente.

a)

0	0	0	1	2
1	1	0	1	1
2	2	2	0	0
1	1	0	2	0
0	0	1	0	1

b)

	0	1	2
0	8	7	3
1	7	6	1
2	3	1	4

c)

	0	1	2
0	0,20	0,175	0,075
1	0,175	0,15	0,025
2	0,075	0,025	0,10

Figura A.1 - Exemplo de geração de uma matriz de co-ocorrência. Em a) Imagem quantizada em 3 níveis de cinza - 0, 1 e 2; Em b) Matriz de co-ocorrência da imagem, na direção 0° e distância de 1 pixel; Em c)Matriz de Co-Ocorrência normalizada pela soma dos valores da matriz.

Após gerada a matriz de co-ocorrência, ela é usada no cálculo dos descritores texturais. (HARALICK et al., 1973) definiram 14 descritores, a saber: Segundo Momento Angular (SMA), contraste, correlação, variância, homogeneidade, média da soma (*sum average*), variância da soma (*sum variance*), entropia da soma (*sum entropy*), entropia, variância da diferença (*difference variance*), entropia da diferença (*difference entropy*), medidas de correlação 1 (IMC-1) e 2 (IMC-2) e coeficiente máximo de correlação. Os descritores texturais refletem as mudanças regulares nos valores dos níveis de cinza em uma imagem. Tais mudanças são correlacionadas tanto estatisticamente, quanto espacialmente. As equações A.1 a A.14 apresenta os descritores acima citados.

$$sma = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (p(i,j))^2 \tag{A.1}$$

$$contraste = \sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} p(i,j)) \right\}, n = |i-j| \tag{A.2}$$

$$correlação = \frac{\sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{A.3}$$

$$variância = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (i-\mu)^2 p(i,j)) \tag{A.4}$$

$$homogeneidade = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} \frac{p(i,j))}{1+(i-j)^2} \tag{A.5}$$

$$mediadasoma = \sum_{i=2}^{2Ng} i * p_{x+y}(i) \tag{A.6}$$

$$variânciadasoma = \sum_{i=2}^{2Ng} (i - A.8)^2 * p_{x+y}(i) \tag{A.7}$$

$$entropiadasoma = -\sum_{i=2}^{2Ng} p_{x+y}(i) * log_2 \{p_{x+y}(i)\} \tag{A.8}$$

$$entropia = -\sum_{i=0}^{Ng} \sum_{j=0}^{Ng} p(i,j) * log_2(p(i,j)) \tag{A.9}$$

$$variânciadadiferença = -\sum_{i=0}^{Ng} (i - f'_{10})^2 * p_{x-y}(i) f'_{10} = \sum_{i=0}^{Ng} p_{x-y}(i) \tag{A.10}$$

$$entropiadadiferença = -\sum_{i=2}^{2Ng} p_{x-y}(i) * log_2 \{p_{x-y}(i)\} \tag{A.11}$$

$$MedidadeCorrelação1 = \frac{HXY - HXY1}{max\{HX, HY\}} \tag{A.12}$$

$$MedidadeCorrelação2 = (1 - exp[-2(HXY2 - HXY)])^{\frac{1}{2}} \tag{A.13}$$

$$coeficientemáximodecorrelação = (Segundomaiorauto{-}vetordeQ)^{\frac{1}{2}} Q(i,j) = \sum_k \frac{p(i,k) * p(j,k)}{p_x(i) * p_y(k)} \tag{A.14}$$

O Segundo Momento Angular (SMA) varia de 0 a 1, sendo 1 para imagem constante. Ou seja, quanto mais lisa a textura, mais alto o valor do SMA (aproxima-se mais de 1). Essa métrica é também chamada de uniformidade ou energia.

Também conhecida como Momento Inverso da Diferença, a homogeneidade atinge seu valor máximo quando a concentração dos valores na diagonal da matriz de co-ocorrência for máxima. Ou seja, quanto mais constante a imagem (lisa), maior

o valor da homogeneidade. Por outro lado, a entropia mede a desordem em uma imagem. Valores baixos (próximos de zero) são atribuídos a imagens constantes (lisas). Quanto mais rugosa a imagem, maior o valor de entropia.

O contraste também mede a presença de transição abrupta de níveis de cinza. Se este valor for alto, indica uma maior variabilidade local na imagem e, portanto, uma maior rugosidade. A variância é uma medida de dispersão, similar a entropia. Segundo Pham (2010), a variância responde a seguinte pergunta : "'Qual a dispersão da diferença entre a referência (média) e os pixels vizinhos nessa janela?"' Em uma imagem homogênea, a variância deve ser próximo de zero.

A medida de correlação mostra a dependência linear dos valores de nível de cinza na matriz de co-ocorrência, sendo 0 para pixels descorrelacionados e 1 para pixels perfeitamente correlacionados. Para uma imagem com grandes áreas de intensidades de nível de cinza similares, a correlação é muito maior do que em uma imagem ruidosa.

Os descritores média da soma (*sum-average*) e variância da soma (*sum-variance*) são a média e variância da matriz de co-ocorrência no domínio espacial. A entropia da soma (*sum-entropy*) e entropia da diferença (*difference-entropy*) são medidas de aleatoriedade contida na imagem. O descritor variância da diferença (*difference-variance*) fornece uma indicação de disparidade na imagem (PARK; CHEN, 2001).

O coeficiente máximo de correlação é uma métrica pouco utilizada, tanto por seu custo computacional, quanto por sua instabilidade.

Haralick et al. (1973) afirmam que após extrair os 14 descritores texturais para cada direção de vizinhança na matriz de co-ocorrência, a média e o range (diferença entre o maior e menor valor do descritor) podem ser calculados e, como existe uma grande sobreposição entre os descritores, alguma técnica de seleção de atributos pode ser utilizada para reduzir a dimensionalidade antes de uma próxima etapa de processamento dos dados.

Gong et al. (1992) chamam atenção para o fato de diversas variáveis estarem envolvidas no GLCM, tais como : tamanho da janela na imagem, nível de quantização, distância entre os pixels, direção da avaliação da vizinhança, resolução espacial e espectral da imagem e os descritores utilizados. Para Park e Chen (2001), a determinação desses parâmetros de forma a ter vantagens tanto estatísticas, quanto das propriedades estruturais da textura é difícil sem experimentos empíricos.

Baraldi e Parmiggiani (1995)avaliaram os descritores segundo momento angular (SMA), contraste, correlação, entropia, variância e homogeneidade e concluíram que a homogeneidade deve ser evitada, porque esse descritor combina as características de SMA e contraste, sendo pouco efetivo. O descritor de correlação só deve ser usado quando é esperada correlação linear entre os elementos da matriz. SMA e Entropia são descritores opostos a Contraste e Variância. Desses, a energia (SMA) é preferida em relação à entropia, e o contraste em relação à variância.

Park e Chen (2001) avaliaram os treze primeiros descritores de Haralick em imagens na faixa do visível e infra-vermelho próximo para classificar carcaças de aves saudáveis e doentes. Dos descritores avaliados, os autores concluíram que a variância, média da soma, variância da soma e entropia da soma foram os mais significativos para o problema abordado.

Malathi e Shanthi (2011) afirmam que a homogeneidade, contraste, entropia e energia são sensíveis à mudança da direção de avaliação da vizinhança.

Gebejes e Huertas (2013) avaliaram se os valores dos descritores contraste, homogeneidade, dissimilaridade, energia e entropia se alteram quando a escala da imagem muda. Os autores concluíram que os descritores dissimilaridade, homogeneidade e energia são invariantes à escala. A entropia e o contraste apresentam pequenas alterações em seus valores.

A.1.1 Implementação

O código para o cálculo dos descritores de Haralick foi desenvolvido utilizando a linguagem Python[1] e, especialmente, as bibliotecas Numpy e Scipy (SCIPY.ORG, 2016). O código gerado está disponível em `https://github.com/vanessavcos/Haralick`.

Um aspecto importante da implementação da matriz de co-ocorrência para as imagens de galáxias e dos Modelos Digitais de Elevação foi que, como as imagens são obtidas em um formato próprio (fit e geotiff) e, após o pré-processamento, foram salvas como arquivos texto (e não em formato de imagem), os valores dos *pixels* eram números reais. Sendo assim, considerou-se apenas a parte inteira desses valores. Além disso, para reduzir a quantidade de 'níveis de cinza' na matriz de co-ocorrência, optou-se por listar à priori os diferentes valores presentes na imagem e criar a matriz apenas com esses valores.

[1] https://www.python.org/

Os descritores foram computados nas quatro direçoes (0, 45, 90, 135 °) e posteriormente foram calculadas a média e o *range*. O descritor coeficiente máximo de correlação não foi implementado.

Os descritores de Haralick são simples de serem calculados, mas podem ter um custo computacional elevado. A estratégia de varrer os valores distintos antes de montar a matriz de co-ocorrência e a implementação em Python tornou a computação bem rápida.

A.2 Índice de Moran

O índice de Moran permite testar a hipótese de dependência espacial nos dados. A hipótese nula é de independência espacial, com valor igual a zero. Valores positivos (entre 0 e +1) indicam correlação direta e negativos, (entre 0 e −1) correlação inversa. A hipótese implícita do cálculo do índice de Moran é a estacionariedade, pois quando há não-estacionariedade de primeira ordem (tendência) nos dados, os vizinhos tenderão a ter valores mais parecidos que áreas distantes, pois cada valor é comparado à média global, inflacionando o índice (CÂMARA et al., 2004).

Numa matriz de vizinhança normalizada o índice de Moran é dado por:

$$I = \frac{\sum_{i \neq j}^{n} w_{i,j}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{(z)})^2} \tag{A.15}$$

Onde:

- w_{ij} é o elemento na matriz de vizinhança para o par i e j
- $\bar{(z)}$ é o desvio em relação à média
- n é o número de observações

Maiores detalhes sobre o índice de Moran podem ser obtidos em Câmara et al. (2004).

A implementação do índice de Moran utilizada nessa tese foi a do módulo Pysal [2]. Pysal é uma biblioteca Python *open source* para análise exploratória de dados espaciais e geocomputação.

[2] https://pysal.readthedocs.io/en/latest/users/index.html

A.3 Função de Densidade Espectral de Potências

Para saber se a série possui influências sazonais torna-se importante conhecer a correlação entre os instantes de tempo distantes entre si por s ou múltiplos de s, onde s representa o período da sazonalidade.

A função de autocorrelação faz isso, medindo o grau de correlação de uma variável, em um dado instante, consigo mesma, em um instante de tempo posterior. Ela permite que se analise o grau de irregularidade de um sinal.

A função de densidade espectral de potências (*Power Spectrum*) também é útil para detectar periodicidades numa série temporal, ou seja, a presença de ciclos ou oscilações e suas amplitudes. Em outras palavras, ela mostra o 'poder' do sinal em cada frequência. Uma das maneiras de se calcular a densidade espectral de potências consiste em usar o Teorema de Wiener Khinchin, do qual resulta que a função densidade espectral é a transformada de Fourier da função de autocovariância.

Se um sinal $a(t)$ possui uma transformada de Fourier $A(t)$, sua função de densidade espectral de potências $S(k) = |A(t)|^2$. Computacionalmente, utiliza-se a transformada de fourier discreta (FFT), que é uma transformada de um série finita de sinais discretos.

Em dados bi-dimensionais, o expoente do *Power Spectrum* (β) é calculado por meio de uma média radial. E, neste caso, expressa a média do espectro de potência de Fourier 2D em todas as possíveis direções (PRUM et al., 1998; RAMOTOROKO et al., 2016).

O cálculo do *Power Spectrum* radial utilizado nesse trabalho foi realizado com o código Matlab distribuído por Ruzanski (2011).

A.4 Árvore de Decisão - Algoritmo J4.8

Algoritmos de árvores de decisão são classificadores que utilizam a estrutura de árvore para modelar a relação entre os atributos e os resultados potenciais. A interpretação da árvore deve começar do seu nó raiz e passar pelos nós de decisão em direção às folhas. Os nós de decisão dividem os dados em galhos que levam a sucessivas decisões ou a uma decisão final que são as folhas da árvore. Pode-se ter mais de duas escolhas por nó de decisão (QUILICI-GONZALEZ; ZAMPIROLLI, 2014).

O algoritmo J4.8 é uma implementação própria do *software* Weka do algoritmo C4.5

(QUINLAN, 1992). O J4.8 implementa uma versão posterior e mais eficiente (C4.5 versão 8).

A abordagem de decisão do J4.8 é a mais utilizada em problemas de classificação. Assim como todos os algoritmos indutores de árvores, o J4.8 utiliza o critério de ganho de informação obtida na escolha do atributo para subdivisão (HALL et al., 2009; QUINLAN, 1992), ou seja, "'dividir para conquistar"'. A cada iteração do processo de indução da árvore, conjunto de dados é dividido em subconjuntos visando obter nesses conjuntos a homogeneidade. A ideia básica consiste em dividir os dados em intervalos com base nos valores de atributo para esse item encontrado na amostra de treinamento (formação da árvore) (SHARMA; SAHNI, 2011).

Cada nó interno da estrutura da árvore contém um teste condicional, onde o resultado determina para qual galho da árvore deve-se ir partindo daquele nó. Já os nós folhas contêm os nomes das classes de classificação. Durante a classificação, quando o resultado de um teste condicional alcança um nó folha, sabe-se que aquela será sua classe pertencente.

Uma característica relevante do J4.8 é a capacidade de trabalhar com atributos contendo dados numéricos e alfanuméricos.

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.