sid.inpe.br/mtc-m19/2012/02.24.17.32-TDI

LABORATÓRIO VIRTUAL PARA VISUALIZAÇÃO E CARACTERIZAÇÃO DO USO E COBERTURA DA TERRA UTILIZANDO IMAGENS DE SENSORIAMENTO REMOTO

Ramon Morais de Freitas

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Reinaldo Roberto Rosa, e Yosio Edemir Shimabukuro, aprovada em 16 de março de 2012.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3BDRG5P>

INPE
São José dos Campos
2012

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
Fax: (012) 3208-6919
E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (RE/DIR-204):
Presidente:
Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)
Membros:
Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)
Drª Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)
Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)
Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)
Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)
Drª Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação
Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)
BIBLIOTECA DIGITAL:
Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)
REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:
Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)
Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)
EDITORAÇÃO ELETRÔNICA:
Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)

sid.inpe.br/mtc-m19/2012/02.24.17.32-TDI

LABORATÓRIO VIRTUAL PARA VISUALIZAÇÃO E CARACTERIZAÇÃO DO USO E COBERTURA DA TERRA UTILIZANDO IMAGENS DE SENSORIAMENTO REMOTO

Ramon Morais de Freitas

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Reinaldo Roberto Rosa, e Yosio Edemir Shimabukuro, aprovada em 16 de março de 2012.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3BDRG5P>

INPE
São José dos Campos
2012

Dados Internacionais de Catalogação na Publicação (CIP)

F884l

Freitas, Ramon Morais de.
Laboratório virtual para visualização e caracterização do uso e cobertura da terra utilizando imagens de sensoriamento remoto / Ramon Morais de Freitas. – São José dos Campos : INPE, 2012.
xxxiv + 201 p. ; (sid.inpe.br/mtc-m19/2012/02.24.17.32-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.
Orientadores : Drs. Reinaldo Roberto Rosa, e Yosio Edemir Shimabukuro.

1. análise de séries temporais. 2. laboratório virtual de sensoriamento remoto. I.Título.

CDU 004.932:528.8

Copyright © 2012 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2012 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

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

Computação Aplicada

Dr. Fernando Manuel Ramos

Presidente / INPE / SJCampos - SP

Dr. Reinaldo Roberto Rosa

Orientador(a) / INPE / SJCampos - SP

Dr. Yosio Edemir Shimabukuro

Orientador(a) / INPE / São José dos Campos - SP

Dr. Jean Pierre Henry Balbaud Ometto

Membro da Banca / INPE / São José dos Campos - SP

Dr. Arcilan Trevenzoli Assireu

Convidado(a) / UNIFEI / Itajubá - MG

Dr. Tiago Garcia de Senna Carneiro

Convidado(a) / UFOP / Ouro Preto - MG

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

Aluno (a): Ramon Morais de Freitas

São José dos Campos, 16 de março de 2012

Les miroirs du temps - André Martins de Barros, 2002.

"Cada dia traz sua alegria e sua pena, e também sua lição proveitosa".

José Saramago
em *"História do cerco de Lisboa"*, 1987.

...aos milhares de brasileiros(as) que sequer tiveram a oportunidade de estudar ou aprender a ler, mas certamente possuem uma grande sabedoria.

AGRADECIMENTOS

Agradeço:

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES pela bolsa de doutorado.

Às agências FAPESP e o CNPq que indiretamente contribuíram para realização desse trabalho.

Ao INPE, PG-CAP, OBT e DSR pelo apoio logístico durante a realização da Pesquisa.

Ao meu orientador Dr. Reinaldo Roberto Rosa pela oportunidade de trabalho em conjunto, amizade e por compartilhar novos horizontes na pesquisa.

Ao meu orientador Dr. Yosio Shimabukuro por sempre me apoiar e acreditar nesta pesquisa, pela amizade, paciência, orientação e conselhos para vida.

A todos meus familiares que sempre me apoiaram nos estudos. Em especial a minha mãe e meu irmão, pela amizade, paciência e pelo apoio nos momentos difíceis.

Ao meu amor Patricia pelo companheirismo, incentivo e por me mostrar que as expressões artísticas são muitas vezes melhor e mais filosóficas que os números e as ciências exatas.

Aos amigos Egídio, Adami, André, Yuzo, Arley pela amizade, ajuda, conselhos e pelas conversas sobre as conjecturas INPEanas.

À amiga Liana Anderson pelas conversas sobre análise de dados ambientais, amazônia e sensoriamento remoto.

Aos amigos da MSA pelas conversas filosóficas, etílicas e demais conexões.

Aos meus amigos que passaram pelo laboratório e adjacências: Luciana Miura, Carlão (UENF), Enner, Thalita, Aragão, Fernandão, Doug, São, Sérgio, Valdete et alli, pelas conversas sobre sensoriamento e análise de dados.

A todos colegas e professores da turma CAP 2007.

Aos amigos do LAF em especial ao Dr. Bernardo Rudorff pelo suporte e fomento a pesquisa e desenvolvimento dentro do laboratório. Principalmente pelo incentivo ao

projeto derivado dessa tese: *Visualização de Séries-Temporais MODIS para análise de mudanças de uso e cobertura da terra.*

À Dra. Evlyn Novo pela amizade e incentivo.

Ao funcionário João Diehl pela disponibilização dos recursos nas estações de trabalho e rede.

Ao grande amigo Marcus Moreira pelo incentivo, amizade e as prescrições médicas.

Aos amigos volovelistas do Clube de Vôo a Vela do CTA (CVV-CTA) e aos amigos pilotos e entusiastas do Aeroclube de Rio Claro.

A todos os funcionários da DSR pelo acolhimento e amizade.

*À*s bibliotecárias do INPE, sempre atenciosas e solícitas.

A todos que não estão citados aqui, mas direta ou indiretamente contribuíram para o desenvolvimento desta tese, o meu muito obrigado.

RESUMO

Ao longo da última década, o conjunto de dados multitemporais de sensoriamento remoto tornou-se uma ferramenta eficaz e cada vez mais necessária para o monitoramento dos recursos naturais da Terra. Os conjuntos de dados de sensores remotos têm permitido o entendimento, caracterização e monitoramento espaço-temporal das mudanças no uso e cobertura da terra em diversas escalas temporais. O grande volume e a diversidade de fontes e formatos desses dados requerem estruturas computacionais específicas para visualização, análise e a transformação dos dados em informações úteis para pesquisadores, público em geral e os tomadores de decisão. Nesse contexto, a computação aplicada surge como um pilar fundamental para o suporte, a implementação, a visualização e a análise dos dados multitemporais de sensoriamento remoto. Assim, este trabalho propõe, discute e desenvolve o conceito de *laboratório virtual de sensoriamento remoto* que permite aos usuários terem acesso à visualização e análise de séries temporais de dados derivados de sensores remotos. Nesse trabalho, a arquitetura utiliza os globos virtuais como plataformas auxiliares de visualização dos dados incorporando métodos computacionais de análise de séries temporais provenientes de abordagens analíticas da física estatística. O estudo desses métodos envolveu a pesquisa, a seleção, a modificação, a implementação e a avaliação das seguintes técnicas: análise não tendencial de flutuações (DFA), análise de padrões gradiente (GPA) e gráfico de recorrência. Um estudo de caso foi realizado utilizando os operadores computacionais derivados das técnicas estudadas aplicadas nas séries temporais e espaço-temporais de índices de vegetação - *Enhanced Vegetation Index 2* (EVI2) derivados do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS) disponíveis no laboratório virtual. Os resultados apresentam uma nova abordagem para a caracterização do uso e cobertura da terra, indicando as áreas que ocorreram mudança em escalas locais, regionais ou continentais. Em escalas locais, a nova abordagem de visualização de séries temporais integradas em um ambiente virtual permite a inferência sobre a dinâmica do uso e cobertura da terra. Nesse contexto, a combinação de técnicas da computação aplicada e do sensoriamento remoto permitiu uma forma inédita de disponibilização de dados e ferramentas computacionais que são úteis nas mais diversas áreas do conhecimento como nas ciências agrárias, ciências florestais, engenharia ambiental entre outras áreas das geociências.

VIRTUAL LABORATORY FOR VISUALIZATION AND CHARACTERIZATION OF LAND USE AND LAND COVER USING REMOTE SENSING IMAGES

ABSTRACT

Over the last decade, the multitemporal remote sensing dataset has become an effective tool, and increasingly necessary to monitor the Earth's natural resources. The remote sensing datasets have allowed the understanding, characterization and monitoring the spatio-temporal land use and land cover changes in different time scales. The large data volume and the diversity of data sources and formats require specific computer data structures for visualization, analysis and transformation of the data into useful information for researchers, general public and decision makers. In this context, applied computer science emerged as a key pillar for the support, implementation, visualization and analysis of multitemporal remote sensing data. So , this work proposes, discusses and develops the concept of *Virtual Lab of remote sensing* that allows to access the visualization and analysis of time series data derived from remote sensors. In this work, the architecture uses the virtual globes as auxiliary platforms for data visualization incorporating computational methods for time series analysis from analytical approaches of statistical physics. The study of these methods involved the research, selection, modification, implementation and evaluation of the following techniques: Detrend Fluctuation Analysis (DFA) , Gradient Pattern Analysis (GPA) and Recurrence Plot (RP). A case study was performed using operators derived from computational techniques applied in the studied time series and spatio-temporal vegetation index (Enhanced Vegetation Index 2 (EVI2)) of Moderate Resolution Imaging Spectroradiometer (MODIS) available in the virtual lab. The results present a new approach to characterize the land use and land cover, indicating the areas where changes occurred at local, regional or continental scales. On local scales, the new approach to visualization of time series integrated in a virtual environment allows inferences about the dynamics of the land use and land cover. In this context, the combination of applied computing and remote sensing techniques has enabled an unprecedented form of accessing the data and computational tools that are useful in several areas of knowledge as in the agricultural sciences, forest sciences, environmental engineering and other areas of geosciences.

LISTA DE FIGURAS

Pág.

2.1 Exemplo de arquitetura multi-camadas de um laboratório virtual. 12
2.2 Arquitetura simplificada do Laboratório Virtual de Séries Temporais em Sensoriamento Remoto. 13
2.3 Mapa de cobertura da terra da América do Sul 15
2.4 Recortes de imagens MODIS que englobam a região da América do Sul. . 17
2.5 Resumo do algoritmo de decomposição proposto por Mallat a) Passos da tranformada discreta de *wavelet* b)Procedimento de limiarização dos coeficientes e c) Transformada inversa. 23
2.6 Funções *wavelet* a) Função de escala ; b) *Wavelet* mãe db8; c) coeficientes de decomposição filtro passa-baixa; d) coeficientes de decomposição filtro passa-alta;e) coeficientes de reconstrução filtro passa-baixa e f)coeficientes de reconstrução filtro passa-alta; 24
2.7 Exemplo de decomposição utilizando a transformada discreta *wavelet* de uma série EVI2 nos coeficientes de aproximação (a_m) e detalhes(d_m) e série filtrada eliminado os coeficientes de detalhes de maior frequência. . 25
2.8 Componentes do módulo de visualização no *website*: a) Google globo virtual usado para selecionar as coordenadas geográficas da área de interesse; b) diagrama polar de elevação em torno de dois círculos amostrados na imagem do Google, e c) série temporal EVI para o enredo selecionado (linha vermelha filtrada com *wavelet* e azul filtrada sem *wavelet*). 26
2.9 Visualização com imagens do GoogleEarth integradas às séries EVI para a coordenada selecionada. a) Imagem *GoogleEarth* b) Imagem *GoogleMaps* e C) Visualização conjunta das séries TRMM e EVI2 (linha verde filtrada com *wavelet* e vermelha filtrada sem *wavelet*). 27
2.10 Exemplos de extração de métricas e interpretação qualitativa das séries EVI2 apresentadas na Figura 2.8a. 28
2.11 Exemplo de série temporal de uma área de reflorestamento. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) Série EVI2 para coordenada selecionada. d)gráfico de área da precipitação acumulada mensal TRMM. 29
2.12 Exemplo de série temporal com detecção de desflorestamento. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) série EVI2 para coordenada selecionada. 30
2.13 Exemplo de série temporal de desflorestamento. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) série EVI2 para coordenada selecionada. 31

2.14 Exemplo de série temporal de cana-de-açúcar. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) Série EVI2 para coordenada selecionada. 32
2.15 Exemplo de desmatamento em cerrado: a) imagem GoogleMaps; b) gráfico polar de elevação ; c) série EVI2 para coordenada selecionada na figura a, ponto 1, cerrado; d)série EVI2 para coordenada selecionada na figura a, ponto 2, desmatamento em cerrado em 2007; e c)série EVI2 para coordenada selecionada na figura a, ponto 3, desmatamento em 2002. . . 34
2.16 Exemplo de série temporal EVI2 para verificação dos polígonos de desflorestamento PRODES. 35
2.17 a) Imagem híbrida do GoogleMaps integrada com polígono PRODES; b) Fotos aéreas do sobrevôo ao polígono no mês de dezembro de 2009 e imagem Landsat-TM do mês de agosto de 2009; c) Série temporal do EVI2 da coordenada selecionada. 36
2.18 Composição colorida de Imagem Landsat-TM integrada ao globo virtual a) imagem TM de 17/11/2002 ; b) imagem TM de 18/02/2005 ; c) imagem TM de 04/09/2007; d) imagem TM de 17/03/2009; e) Série temporal do EVI2 da coordenada selecionada dentro do polígono de expansão de cana-de-açúcar mapeado pelo projeto Canasat. Os pontos em vermelho nas séries EVI2 correspondem às datas das imagens Landsat-TM. 38
2.19 Identificação de pixels com grande amplitude por um curto período de tempo (Freitas et al., em preparação). 39

3.1 **1** e **2** fazem parte de uma ´´operação computacional", sobre o dado, uma vez que a técnica matemática está na forma de um algoritmo devidamente estruturado em um ambiente computacional (*software* e *hardware*). . . . 43
3.2 Fluxograma do método DFA 1-D. 47
3.3 Sequência de passos para aplicação do DFA em uma série estacionária tipo rúido branco. 50
3.4 Boxplots dos 1000 valores de α em função do comprimento da série. . . . 51
3.5 a) Concatenação de duas séries, sem memória (linha em vermelho) e com memória (linha azul). b) Valores de α para cada série de 1024 pontos construída a partir da posição x e o percentual utilizado das séries com memória (linha azul) e sem memória (linha em vermelho). 52
3.6 Valores de escala s *versus* a função de flutuação F(S) para série estacionária (a) e série não estacionária (b), utilizando polinômio linear para remoção da tendência. 53

3.7 Exemplo de cálculo do expoente α para matrizes: a) distribuição dos valores de α após a realização do método de *bootstrap*; b) *boxplot* dos valores de α com intervalo de 95% da mediana; c)Imagem a partir da qual o expoente α foi calculado, nesse caso um ruído branco representando uma superfície isotrópica e d)gráfico log-log das escalas (s) a partir da função de flutuação F(s). 57
3.8 Exemplo de cálculo do expoente α para matrizes: a) distribuição dos valores de α após a realização do método de *bootstrap*; b) *boxplot* dos valores de α com intervalo de 95% da mediana; c)Imagem a partir da qual o expoente α foi calculado, nesse caso um solução turbulenta da Equação de Ginzburg-Landau e d)gráfico log-log das escalas (s) a partir da função de flutuação F(s). 58
3.9 Exemplo de cálculo do expoente α para matrizes: a) distribuição dos valores de α após a realização do método de *bootstrap*; b) *boxplot* dos valores de α com intervalo de 95% da mediana; c)Imagem a partir da qual o expoente α foi calculado, nesse caso um sobreposição de superfícies periódicas e d)gráfico log-log das escalas (s) a partir da função de flutuação F(s). 59
3.10 Solução intermitente do mapa de Kaneko, cada imagem representa um passo da simulação, valores de amplitude representam a evolução de uma matriz escalar de índice de vegetação dado pela condição inicial. 59
3.11 Valores do expoente α calculados para cada imagem do modelo simulado. 60
3.12 Tempo médio para o cálculo dos valores de α em função da dimensão da matriz. 61
3.13 Síntese da análise padrões gradiente pesquisado nessa tese. 63
3.14 Exemplo de triângulação de Delaunay. 65
3.15 Exemplos da Triângulação de Delaunay do campo gradiente. a) Gradiente com pouca assimetria; b) Triângulação de Delaunay correspondente a; c)Gradiente de uma matriz aleatória ; D)Triângulação de Delaunay correspondente c. 65
3.16 Contorno e superfície de uma função SINC 2D sobre influência de ruídos localizados. A medida que as assimetrias bilaterais aumentam os valores de G_1A aumentam. 66
3.17 Aspectos fundamentais para o cálculo do primeiro momento do gradiente. 70
3.18 Distribuição de $G1_{A3x3}$ para a) superfície turbulenta gerada pela equação de Ginzburg-Landau; b)superfície periódica resultante do somatório de matrizes de seno e c) superfície isotrópica derivada de um ruído branco. . 72

3.19 (a) Série temporal com tendência; (b) Matriz correspondente; (c) Campo gradiente correspondente (9 vetores assimétricos) e (d) Campo de triangulação com 16 linhas conectoras. 73
3.20 a) Típico padrão de variabilidade de série temporal; b) Matriz correspondente; (c) Campo gradiente correspondente (9 vetores assimétricos) e d) Campo de triangulação com 20 linhas conectoras. 73
3.21 a) solução turbulenta da equação complexa de Ginzburg-Landau 2-D mostrando transição de fase, e b)evolução temporal do NFr_i para cada passo da solução. 76
3.22 a) Representação em imagem da solução do mapa de Kaneko ($a = 2.00$, $e = 0.3$) para duas condições iniciais: a) aleatória e b) envelope de senos. 77
3.23 Operadores computacionais para cada passo da simulação do mapa de Kaneko: a) NFr_i ; b) Diversidade de norma G_{2_D}; c) Diversidade de fase G_{3_D} e d) Norma de Frobenius $G_{2_{NF}}$ 78
3.24 Séries canônicas: a) série de senos; b) série de senos + ruído branco; c) Não estacionária + ruído branco; d) multi-periódica série de senos; e) E.C.G.L. - solução turbulenta; f) E.C.G.L. -intermitência; g) E.C.G.L. - turbulência em fase; h) E.C.G.L. - estruturas coerentes - pontos atípicos; i) ruído branco. 81
3.25 Matriz de distância : a) série de senos; b) série de senos + ruído branco; c) Não estacionária + ruído branco; d) multi-periódica série de senos; e) E.C.G.L. - solução turbulenta; f) E.C.G.L. -intermitência; g) E.C.G.L. - turbulência em fase; h) E.C.G.L. - estruturas coerentes - pontos atípicos; i) ruído branco. 82
3.26 Gráfico de recorrência global com limiar $\epsilon de 0,05$ desvio padrão : a) série de senos; b) série de senos + ruído branco; c) Não estacionária + ruído branco; d) multi-periódica série de senos; e) E.C.G.L. - solução turbulenta; f) E.C.G.L. -intermitência; g) E.C.G.L. - turbulência em fase; h) E.C.G.L. - estruturas coerentes - pontos atípicos; i) ruído branco. 83
3.27 Seqüência de passos para a triangulação do gráfico de recorrência. 84
3.28 Simulação dos valores do triângulo máximo encontrado em relação ao comprimento da série temporal. 86
3.29 Relação entre o número de triângulos e a área de cada triângulo para as nove séries canônicas utilizadas: a) séries ruído branco e b)passeio aleatório. 86

4.1 Séries temporais EVI2 a) um processo de desflorestamento na região amazônica; b) um desflorestamento no cerrado, e; c) uma área de floresta semi-decidual sem alterações. 91

4.2 Expoente α calculado para: a) um processo de desflorestamento na região amazônica; b) um desflorestamento no cerrado; e c) uma área de floresta semi-decidual sem alterações. 92
4.3 Boxplot dos valores α para as séries temporais amostradas. 92
4.4 Histograma dos valores de $G1_{A3x3}$ para as séries temporais amostradas por classe de uso. 94
4.5 Espectro de assimetria com valores de $G1_{A3x3}$ para as séries temporais: a) Série EVI2; b)espectro de assimetrias na escala *3x3* para série (a); c) histograma das assimetrias para série (a); d)Série EVI2 sem filtragem por *wavelets*; e) espectro de assimetrias na escala *3x3* para série (d); c) histograma das assimetrias para série (d). 95
4.6 Gráfico de recorrência para uma série temporal de floresta. a)Matriz de distância; b) gráfico de recorrência e c)série temporal. 96
4.7 Gráfico de recorrência para uma série temporal de desflorestamento em floresta. a)Matriz de distância; b) gráfico de recorrência e c)série temporal. 97
4.8 Gráfico de recorrência para uma série temporal de desflorestamento em cerrado. a)Matriz de distância; b) gráfico de recorrência e c)série temporal. 98
4.9 Distribuição por amostra dos valores de a)número de triângulos; b) área do máximo triângulo são normalizados e c)gráfico de dispersão entre o número de triângulos e área do máximo triângulo são normalizados . . . 99
4.10 Gráfico de recorrência para uma série temporal de agricultura com mudança do tipo de cultura. a)Matriz de distância; b) gráfico de recorrência e c)série temporal 100
4.11 Síntese da metodologia para espacialização de pontos de mudança. 105
4.12 Distribuição dos valores de $\beta 1$ para os valores de anomalia. 106
4.13 Mapa da distribuição de *hotspots* de pontos de mudança. 107
4.14 Sequencia de imagens para uma área agrícola. 109
4.15 Distribuição das 9 variáveis para área de agricultura. 110
4.16 Sequencia de imagens para uma área de instalação de reservatório hidrelétrico. 111
4.17 Distribuição das 9 variáveis da área de instalação de reservatório hidrelétrico. 112
4.18 Sequencia de imagens para uma área de reflorestamento. 113
4.19 Distribuição das 9 variáveis da área de reflorestamento. 114
4.20 Sequencia de imagens para uma área de desflorestamento. 115
4.21 Distribuição das 9 variáveis da área de desflorestamento. 116
4.22 Distribuição dos valores médios de 2000 a 2010 na escala de 32x32km para os valores de: a) α do DFA-2-D e b) média de EVI2. 117

A.1 Função de autocorrelação (c e d) para a série temporal sem memória (a) e uma série temporal com memória de longo alcance (c). 146
A.2 Série estacionária com respectivo espectro de potência (a,b) e série não estacionária (c, d). 148
A.3 Série periódica com respectivo espectro de potência série série estacionária (a,b) e não estacionária (c, d). 151
A.4 a) Valores de α para umidade relativa. b) Valores do índice χ para umidade relativa separados por região geográfica. Fonte: Chen et al. (2007). . 155
A.5 a) Valores de α para 100 anos de dados observados de temperatura. b) Valores de α para os dados simulados 1000 anos.Fonte: Fraedrich e Blender (2003). 156
A.6 a) Série temporal de NDVI após uma queimada. Região de Ventimiglia, norte da Itália. Fonte: Telesca e Lasaponara (2006). b) Valores do expoente α para vegetação queimada e não queimada. Fonte: Telesca e Lasaponara (2005) 158
A.7 Valores de α ao longo de um dia. Fonte: Pin et al. (2008) 160
A.8 a) Imagem da granulação solar devido à atividade magnetohidrodinâmica c) emulsão de água e óleo observada por um microscópio eletrônico. b e d) comportamento da função ${}_a vr_\theta(n)$ e α_θ ao longo das escalas em pixels (n). Fonte: Alvarez-Ramirez et al. (2006) 162

B.1 a)Série simulada com 8 períodos, N=300, $\sigma^2 = 0.1$; b)Série simulada com 10 períodos, N=100, μ =0, $\sigma^2 = 0.3$ 164
B.2 a)Envelope de senos simulados 2-D com 3,5 períodos, N=100, $\sigma^2 = 0$; ;b)Envelope de senos simulados 2-D com 3,5 períodos, N=100, $\sigma^2 = 0.3$;c) Projeção do eixo Z no plano xy do envelope de senos 2-D, com 0.5 períodos, N=100, $\sigma^2 = 0$; d) Projeção do eixo Z no plano xy do envelope de senos 2-D, com 0,5 períodos, N=100 e $\sigma^2 = 0.3$. 165
B.3 Solução da grade de mapas acoplados mostrando o surgimento de padrões intermitentes na seqüência da solução 166
B.4 Solução da ECGL 1-D pelo código proposto por Winterbottom(2005), $\alpha_{gl} = 2,5$, $\mu_{gl} = 1$ e $\beta_{gl} = -1,97$, condições de contorno periódica. 169
B.5 Solução da ECGL 2-D pelo código proposto por Winterbottom(2005), $\alpha_{gl} = 0$, $\mu_{gl} = 1$ e $\beta_{gl} = -1,5$, condições de contorno periódica. 170

C.1 Passo 1 171
C.2 Passo 2 172
C.3 Passo 3 172
C.4 Passo 4 173

D.1 Fluxograma simplificado da estrutura de implementação computacional. 175

E.1 Série temporal com desmatamento em cerrado, a partir de 2008 transformação em área agrícola. 178
E.2 Série temporal com desmatamento em cerrado, a partir de 2007 transformação em área agrícola. 179
E.3 Série temporal para uma área de floresta. 180
E.4 Série temporal para uma área de agricultura intensa, pivô central. 181
E.5 Série temporal para uma área de agricultura intensa, pivô central. 182
E.6 Série temporal para uma área de cana-de-açúcar. 183
E.7 Série temporal com desflorestamento em floresta em 2007 e uso agrícola a partir de 2009. 184
E.8 Série temporal com desflorestamento em floresta em 2004 e conversão para agrícola em 2007. 185
E.9 Série temporal com desflorestamento em floresta em 2007 com posterior regeneração. 186
E.10 Série temporal com desflorestamento em floresta em 2007 com uso agrícola a partir de 2009. 187
E.11 Série temporal para uma área convertida de caatinga para culturas permanentes. 188

A.1 a) Fotografia de uma plataforma de petróleo (Fonte: Petrobrás); b)Localização desta plataforma na imagem RADARSAT-1 c); recorte de imagem centrado na plataforma; e d) amostragem radial a partir do centro da imagem (plataforma) 196
A.2 a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando que a superfície é isotrópica num raio entre 6,5 km 196
A.3 a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando que a superfície se torna anisotrópica num raio entre 6,5 a 13 km 197
A.4 a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando a anisotropia num região de baixa intensidade de vento 197

A.5 a) Recorte de imagem centralizado na plataforma b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando uma superfície totalmente isotrópica 198
A.6 a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando a anisotropia devido ocorrência de resíduo oleoso 198

LISTA DE TABELAS

Pág.

2.1 Principais algoritmos e ferramentas para análise de séries temporais propostos para futuras implementações do Laboratório Virtual de Sensoriamento Remoto 41

3.1 Tabela com os possfveis valores de G_{1A} para o tamanho de uma matriz 3x3. 70

3.2 Características dos Operadores Computacionais 88

A.1 Classificação do expoente β. FONTE: Adaptado de Malamud e Turcotte (1999). 148

LISTA DE ABREVIATURAS E SIGLAS

2D	–	Bidimensional
3D	–	Tridimensional
6S	–	Second Simulation of the Satellite Signal in the Solar Spectrum Radiative Code
ANSI C	–	American National Standards Institute C
API	–	Application Programming Interface
AVHRR	–	Advanced Very High Resolution Radiometer
CDF	–	Common Data Format
DAAC	–	Distributed Active Archive System
DFA	–	Detrended Fluctuation Analysis
DFA-2D	–	Detrended Fluctuation Analysis for 2-D
DWT	–	Discrete Wavelets Transform
ECGL	–	Equação Complexa de Ginzburg-Landau
EOS	–	Earth Observing System
ESA	–	European Space Agency
EVI	–	Enhanced Vegetation Index
EVI2	–	Enhanced Vegetation Index 2
FA	–	Fluctuation Analysis
FAC	–	Função Coeficiente de Autocorrelação
FFT	–	Fast Fourier Transform
FFTW	–	Fast Fourier Transform in the west
FITS	–	Flexible Image Transport System
GeoTIFF	–	Georeferencing Tagged Image File Format
GPA	–	Gradient Pattern Analysis
GR	–	Gráfico de Recorrência
HDF	–	Hierarchical Data Format
HTML	–	HyperText Markup Language
IF	–	Inequalidade de Frobenius
JAXA	–	Japan Aerospace Exploration Agency
mfB	–	movimento fracionário Browniano
LAPACK	–	Linear Algebra PACKage
MODIS	–	Moderate Resolution Imaging Spectroradiometer
MRT	–	MODIS Reprojection Tool
NASA	–	National Aeronautics and Space Administration
NCEP	–	National Centers for Environmental Prediction
NCAR	–	National Center for Atmospheric Research
NDVI	–	Normalized Difference Vegetation Index
NDWI	–	Normalized Difference Water Index
NOAA	–	National Oceanic and Atmospheric Administration
NPP	–	NPOESS Preparatory Project
NPOESS	–	National Polar-orbiting Operational Environmental Satellite System

PHP	–	PHP: Hypertext Preprocessor
PRODES	–	Projeto de Monitoramento do Desflorestamento na Amazônia Legal
PSD	–	Power Spectrum Density
R/S	–	Rescaled Range Statistic
RADAR	–	RAdio Detection And Ranging
SET	–	Série Espaço-Temporal
SIG	–	Sistema de Informação Geográfica
SOC	–	Self-Organized Criticality
SPOT	–	Satellite Pour l'Observation de la Terre
SRTM	–	Shuttle Radar Topography Mission
ST	–	Série Temporal
TRMM	–	Tropical Rainfall Measuring Mission
USGS	–	United States Geological Survey
VIIRS	–	Visible Infrared Imaging Radiometer Suite
WIST	–	Warehouse Inventory Search Tool

LISTA DE SÍMBOLOS

$\lfloor \rfloor$	–	função de arredondamento chão
$\rceil \lceil$	–	função de arredondamento teto
$\hat{\beta}$	–	estimador dos valores de coeficientes lineares
$\| \|_F$	–	Norma de Frobenius
$\| . \|$	–	Norma de distância
$\overline{A}$	–	matriz transposta de A
$*$	–	complexo conjugado
$(\downarrow 2)$	–	representa a operação de sub-amostragem *downsampling*
$(\overline{\psi})$	–	representa o complexo conjugado dessa função
$(\uparrow 2)$	–	operação de super-amostragem *upsampling*
α	–	expoente de escala do método DFA
α_θ	–	expoente de escala do método DFA anisotrópico em séries espaciais
α_{kaneko}	–	parâmetro de controle caótico do mapa de Kaneko
α_{gl}	–	parâmetro fenomenológicos ECGL
β	–	expoente de escala do espectro de potência
β_{gl}	–	parâmetro fenomenológicos ECGL
β_0	–	coeficiente linear da regressão
β_1	–	coeficiente de ajuste angular da regressão
χ	–	índice de umidade relativa
ϵ	–	erro ou limiar de distância plote de recorrência
$\epsilon_{\nu,w}$	–	matriz de resíduos
ϵ_m	–	é uma matriz de ruídos gaussiano de média 0 e variância
ϵ_v	–	é o vetor de ruído gaussiano de média 0 e variância σ^2 e dimensão *1xn*
ϵ_k	–	fator de acoplamento da grade de mapas acoplados (mapa de kaneko)
γ	–	expoente de escala função de densidade espectral
μ_{gl}	–	parâmetro de ordem equação ECGL
μ_1	–	Momento de primeira ordem média
μ_2	–	Momento de segunda ordem variância
μ_3	–	Momento de terceira ordem assimetria
μ_4	–	Momento de quarta ordem curtose
μ_k	–	parâmetro de não linearidade do mapa de kaneko
∇	–	operador gradiente
ν	–	segmento da sub-matriz
$\Phi_(m,k)$	–	é a função base de escala e e translação
ψ	–	função *wavelet* mãe
σ	–	desvio padrão
σ^2	–	variância
$\Theta(.)$	–	é a função de Heaviside
θ_n	–	ângulo
a	–	coeficiente *wavelets*

A_{xx}	–	parâmetro de segunda ordem
$area_{max}$	–	área máxima de triângulação
a	–	parâmetro de escala - wavelets
b	–	parâmetro de translação - wavelets
c	–	coeficiente de aproximação wavelets
$C(s)$	–	função de auto-correlação
d	–	coeficiente de detalhes
$D_{i,j}$	–	distância euclidiana
d_m	–	distância na matriz m
dx	–	componente em x do gradiente bi-dimensional
dy	–	componente em y do gradiente bi-dimensional
$EVI2_{an}$	–	anomalia das séries de EVI2
f	–	frequência
$F(\nu, w, s)$	–	Função de flutuação do DFA
$f(x)$	–	é dado pelo mapa logístico
G_1	–	primeiro momento gradiente
G_2	–	segundo momento gradiente
G_3	–	terceiro momento gradiente
G_4	–	quarto momento gradiente
G_{1A}	–	coeficiente de assimetria do primeiro momento do GPA
G_{1A3x3}	–	coeficiente de assimetria na escala 3x3
G_{2_D}	–	Diversidade de norma
G_{3_D}	–	Diversidade de fase
G_{2_NF}	–	Norma de Frobenius de G_2
H	–	expoente de Hurst
h	–	coeficiente de aproximação, transformada *wavelet*
Hz	–	Hertz
i	–	índice de posição linha
IQR	–	distância inter quartil
j	–	índice de posição coluna
k	–	índice de posição
k_e	–	número de pontos do envoltório convexo
L	–	filtros passa-baixa (L)
$L\infty$	–	norma máxima
L_1	–	norma de Manhattan
L_2	–	norma euclidiana
lag	–	escala
lxl	–	dimensão de matriz quadrada
m	–	nível de escala - transformada discreta *wavelets*
M	–	matriz bi-dimentional de amplitudes
Ms	–	é a matriz de envelopes senoidais
mB	–	série temporal representando movimento Browniano

n	–	número de elementos
n_e	–	número de arestas
n_t	–	número de triângulos
n_f	–	número de faces
N	–	número total de elementos
N_s	–	número de escala DFA
Nc	–	número de arestas
NF_i	–	índice de Frobenius
NIR	–	valor de reflectância na banda do infravermelho próximo
Nv	–	número de vértices
$O()$	–	ordem de complexidade
P	–	é o número de períodos
P_ϕ	–	proporção de fases
P_r	–	proporção de normas
QR	–	decomposição ortogonal triangular (QR)
R	–	raio
Red	–	banda do infravermelho próximo
Rs	–	raio amostral
$Sous$	–	escala
$S(f)$	–	Função de densidade espectral
s_{max}	–	escala máxima
s_{min}	–	escala mínima
S_r	–	escala por raio amostral
t	–	tempo
T_ϕ	–	tolerância em fase
T_r	–	tolerância em raio
TT	–	Total de Triângulos
U	–	superfície de ajuste
u_s^t	–	matriz de amplitude do mapa de Kaneko
$V_{(m,k)}$	–	coeficientes de escala *wavelets*
$\vec{x}$	–	trajetória no espaço de fases
W	–	watts
$W_{(m,k)}$	–	coeficientes da transformada discreta *wavelets*
X	–	série temporal ou espacial unidimensional
Y	–	vetor unidimensional
Z	–	vetor unidimensional

SUMÁRIO

Pág.

1 INTRODUÇÃO 1

2 VISUALIZAÇÃO DE SÉRIES TEMPORAIS EM SENSORIAMENTO REMOTO 9

2.1 Laboratório Virtual de Séries Temporais de Sensoriamento Remoto . . . 10

2.2 Conjuntos de dados 14

2.2.1 Abrangência Geográfica 14

2.2.2 Sensor MODIS 15

2.2.3 Produto MOD09 - Refletância de Superfície 16

2.2.4 Dados TRMM 18

2.3 Procedimentos de Filtragem 18

2.4 Ferramentas de Visualização 23

2.5 Interpretação das séries temporais 26

2.6 Aplicações do Laboratório Virtual no Contexto de Aplicações do INPE . 33

2.7 Propostas de novos algoritmos 40

3 ANÁLISE COMPUTACIONAL DE SÉRIES TEMPORAIS . . . 43

3.1 O Método *Detrended Fluctuation Analysis* (DFA) 44

3.1.1 Persistência e Correlações de Longo Alcance 45

3.1.2 Algoritmo do DFA 1-D 46

3.1.3 Experimentos Numéricos 49

3.1.4 Método DFA-2D 52

3.1.5 Algoritmo do DFA 2-D 54

3.1.6 Resultados da Implementação do DFA-2D 56

3.2 Análise de Padrões de Gradiente 62

3.2.1 Novos Aspectos computacionais do cálculo de G_{1A} 66

3.2.2 Exemplo do cálculo de G_{1A3x3} 71

3.2.3 Definição de outros operadores computacionais 71

3.3 Gráfico de Recorrência Global 77

3.4 Síntese dos Operadores Computacionais Propostos 87

4 ESTUDO DE CASO PARA DETECÇÃO DE MUDANÇA DE USO E COBERTURA DA TERRA 89

4.1 Detecção de Mudanças de Uso e Cobertura da Terra com Séries Temporais 90
4.2 Detecção de Mudanças de Uso e Cobertura da Terra com Séries Espaço-Temporais 99

5 CONCLUSÕES 119

Referências Bibliográficas 123

APÊNDICE A - ANÁLISE DE SÉRIES TEMPORAIS E MÉTODO DFA

A.1 Revisão Bibliográfica do Método DFA 143
A.1.1 Análise de Séries Temporais em Sensoriamento Remoto 143
A.1.2 Espectro de Potência e Correlações de Longo Alcance 145
A.1.3 Avaliação e Variantes do DFA 150
A.1.4 Exemplos de Aplicações 154
A.1.5 Extensões para outras dimensões 160

APÊNDICE B - MODELOS NUMÉRICOS UTILIZADOS PARA GERAÇÃO DE DADOS SINTÉTICOS 163

B.1 Envelope de senos 163
B.2 Grade de Mapas Acoplados - Mapa de Kaneko 164
B.3 Equação de Ginzburg-Landau 167

APÊNDICE C - ELABORAÇÃO DAS IMAGENS MODIS. 171

APÊNDICE D - FLUXOGRAMA DAS IMPLEMENTAÇÕES COMPUTACIONAIS 175

APÊNDICE E - EXEMPLOS DE VISUALIZAÇÃO E INTERPRETAÇÃO DAS SÉRIES EVI2
. 177

ANEXO A - Exemplo de Aplicação de Método DFA anisotrópico para monitoramento ambiental. 189

A.1 Introdução 190
A.2 Método DFA 191
A.3 Procedimento Proposto 193
A.4 Experimentos 194
A.4.1 Dados 194
A.4.2 Metodologia 195

A.4.3 Resultados e Discussão 195
A.5 Conclusões 198

ANEXO B - Publicações Relacionadas ao Trabalho de Tese. 201

1 INTRODUÇÃO

Our ideas must be as broad as Nature if they are to interpret Nature.
(*Sir Arthur Ignatius Conan Doyle*)

Ao longo da última década a utilização de dados multitemporais[1] de sensoriamento remoto tornou-se uma ferramenta eficaz e cada vez mais necessária para o monitoramento dos recursos naturais da Terra. Os conjuntos de séries temporais derivadas de sensoriamento remoto têm permitido o monitoramento espaço-temporal das mudanças no uso e cobertura da terra [2] em escalas temporais anteriormente não observadas. Recentemente, tais estudos vêm ganhando maior atenção da comunidade científica, uma vez que os ecossistemas terrestres exercem um controle nos ciclos biogeoquímicos e hidrológicos, os quais influenciam significativamente os sistemas climáticos através de propriedades radiativas dos gases de efeito estufa e outros gases reativos (DEFRIES et al., 2002; FRIEDL et al., 2002).

O monitoramento de sistemas ambientais por satélite tem atraído grande atenção por parte do governo, centros de pesquisa, bancos de desenvolvimento e organizações não governamentais. Um exemplo é a região Amazônica que desperta o interesse mundial, não apenas em função de sua biodiversidade e riqueza de recursos naturais, mas também no que se refere aos desequilíbrios ecossistêmicos, assunto que vem sendo extensivamente abordado na temática das mudanças globais. (MORAN, 1981; SKOLE e TUCKER, 1993; DEFRIES, 2002; ROBERTS et al., 2003; DEFRIES 2008).

Os sistemas ambientais representam entidades organizadas na superfície terrestre, sendo geralmente abordados sob duas perspectivas: ecológica e geográfica. A ecológica focaliza as comunidades biológicas e o seu habitat, enquanto a segunda focaliza a organização dos elementos físicos e geográficos no contexto espacial. Ambas as abordagens procuram entender os aspectos estruturais, funcionais e dinâmicos dos ecossistemas e geosistemas. Os sistemas ambientais em virtude da variedade de elementos de seus componentes e dos fluxos de interação entre si constitui um exemplo canônico de Sistemas Dinâmicos Complexos. [3] (CHRISTOFOLETTI, 2000).

[1]Em sensoriamento remoto a nomenclatura *dados multitemporais* representa um conjunto de imagens e/ou outros tipos de dados adquiridos em mais de uma data, não necessariamente espaçados equidistantes no tempo.

[2]As mudanças relacionadas ao uso estão associadas às atividades humanas e as de cobertura aos parâmetros biofísicos da vegetação. Por exemplo, uma área tem uso agrícola e a cobertura da terra em um determinado período está sem vegetação.

[3]Sistemas Dinâmicos Complexos apresentam processos e regimes cuja variabilidade nos domínios usuais é não-linear e com amplo espectro de escalas. A não-linearidade pode ser interpretada como

Entender processos como: desflorestamento e mudança do uso e cobertura da terra; extinção de espécies de pássaros; interação de poluentes no ar ou na água; emissões de gases de efeito estufa, e padrões de queimadas em florestas envolvem escalas espaciais e temporais específicas. Nesse contexto, a transdiciplinaridade entre a matemática, física, química, biologia, geografia e principalmente a computação é fundamental no entendimento e modelagem dos processos temporais e espaciais que governam a dinâmica de sistemas ambientais. Sendo que a computação aplicada estabelece um papel agregador não só pelo suporte na modelagem, implementação e simulação mas também na combinação de conhecimento para a manipulação, processamento e disponibilização, e extração de informações de dados ambientais.

Do ponto de vista da computação científica existem duas abordagens mais comuns no estudo dos sistemas ambientais. A primeira abordagem consiste na simulação e modelagem de um problema ambiental com o objetivo de gerar cenários para possíveis previsões. Em geral, isso pode ser realizado através do estudo de processos regidos por equações diferenciais e/ou modelos puramente estocásticos que descrevem o comportamento de um conjunto de variáveis no tempo e/ou no espaço. Um exemplo são os modelos computacionais utilizados para a previsão do tempo e do clima. A segunda abordagem, utilizada neste trabalho, consiste na análise de séries temporais ambientais para caracterização de um processo que ocorra no tempo e/ou no espaço. Esse tipo de abordagem é importante para entender o padrão de variabilidade de uma série temporal através de correlações, tendências, ciclos e variações sazonais que podem auxiliar na elaboração de modelos e na própria previsão de um evento ou fenômeno estudado.

Ainda no contexto matemático e computacional, os dados ambientais podem ser tratados por diferentes técnicas estatísticas, técnicas mais específicas para processamento de sinais, reconhecimento de padrões e recentemente incorporando conceitos de computação inteligente. A escolha da técnica mais apropriada é função do tipo de problema, do custo computacional e da disponibilidade e arranjo dos dados. Dentro da computação aplicada aos sistemas ambientais e no contexto espacial destacam-se principalmente as pesquisas nas áreas de geofísica, meteorologia e sensoriamento remoto.

Nos últimos anos, na área de sensoriamento remoto, os estudos de mudanças de uso e cobertura da terra em escalas regionais ou globais, utilizando imagens multi-

---

as irregularidades da variabilidade (tanto em amplitude, como em freqüência e recorrência), uma propriedade comum encontrada em vários tipos de sinais (ROSA et al., 1998).

temporais derivadas de sensores a bordo dos satélites de observação da Terra têm despertado grande interesse na comunidade científica. O conjunto dessas imagens tem permitido a comunidade científica de geociências novas interpretações e conclusões sobre os processos que ocorrem na superfície terrestre. Sensores orbitais como *Advanced Very High Resolution Radiometer* (AVHRR), *Moderate Resolution Imaging Spectroradiometer* (MODIS), *Satellite Pour l'Observation de la Terre* (SPOT VEGETATION), dentre outros têm permitido a essa comunidade não só a visualização mas a análise das causas e efeitos decorrentes das mudanças de uso e cobertura da terra (DEFRIES, 2008). Os índices de vegetação derivados de sensores orbitais são comumente utilizados para as análises de mudanças no uso e cobertura da terra, pois apresentam informações de fácil interpretação, e são sensíveis a alterações na cobertura vegetal. Os índices de vegetação variam entre -1 e 1, valores menores que 0 referem-se a áreas não vegetadas como solo exposto, água ou sem cobertura vegetal, valores próximos a 1 em áreas de vegetação densa e homogênea (SKOLE e TUCKER, 1993; DEFRIES, 2002; DEFRIES, 2008). O pressuposto básico para utilização desses índices assume que se há alteração no regime dinâmico de variabilidade da vegetação existe alguma mudança de origem antrópica ou não no uso e cobertura da terra. Um exemplo clássico é a supressão da vegetação como num processo de desflorestamento, em que os valores de índice de vegetação diminuem abruptamente quando não existe nenhum tipo de vegetação, o chamado corte raso. Mais de uma dezena de índices de vegetação são utilizados para ressaltar os contrastes entre tipos e estádios fenológicos da vegetação, dependendo do tipo de aplicação. Em geral, a utilização desses índices requer conhecimento adicional em sensoriamento remoto para a seleção e uso de bandas espectrais adequadas para o seu cálculo (SKOLE e TUCKER, 1993; DEFRIES, 2002; DEFRIES, 2008).

Grandes centros de pesquisa espacial e de distribuição de dados de observação da Terra como a *National Aeronautics and Space Administration* (NASA), *United States Geological Survey* (USGS), Instituto Nacional de Pesquisas Espaciais (INPE), *Japan Aerospace Exploration Agency* (JAXA), *European Space Agency* (ESA) e *Oak Ridge National Laboratory* (ORNL) possuem portais de distribuição e disseminação de dados provenientes de sensores remotos. No entanto, apesar do fácil acesso a estes dados, a utilização das séries temporais derivadas de imagens de sensoriamento remoto requer conhecimento prévio dos usuários referente a todas as etapas de pré-processamento, bem como infra-estrutura computacional com capacidade de armazenamento e processamento dos dados que não estão disponíveis em muitos laboratórios de pesquisa. Esses fatos restringem a utilização dos dados às comunidades de usuários e pesquisadores especializados na área de sensoriamento remoto,

dotados do suporte computacional adequado e principalmente do conhecimento técnico/científico necessário ao processamento e manipulação dos dados. Uma nova abordagem nos últimos anos tenta popularizar a espacialização dos dados através da utilização dos globos virtuais [4] como o *GoogleEarth* (BUTLER, 2006). Entretanto, a integração das séries temporais derivadas de sensoriamento remoto com os globos virtuais está limitada ao desenvolvimento de arquiteturas computacionais específicas e à capacidade de armazenamento e visualização rápida de dados. Nesse sentido, tal abordagem pode permitir um incremento nas aplicações e estudos da dinâmica de mudança de uso e cobertura da terra, monitoramento ambiental, estudos de ciclos fenológicos da vegetação, agricultura e principalmente no compartilhamento e utilização de dados pela comunidade de não especialistas em sensoriamento remoto.

No campo da análise das séries temporais o reconhecimento de padrões tem sido largamente utilizado em sensoriamento remoto nos algoritmos de classificação de imagens (MATHER, 1999) os quais, em maioria, utilizam abordagem estatística baseada em momentos e funções estatísticas na classificação de dados. No âmbito da análise estatística de séries temporais, não abordada neste trabalho, destaca-se uma vasta gama de modelos estatísticos para análise de séries temporais como por exemplo,*AutoRegressive model* (AR) , *AutoRegressive Moving Average* (ARMA) , *Autoregressive Conditional Heteroskedasticity* (ARCH) , *AutoRegressive Integrated Moving Average* (ARIMA), *Fractional AutoRegressive Integrated Moving Average* (FARIMA), *Seasonal AutoregRessive Integrated Moving Average* (SARIMA), dentre outros (BOX et al., 1994; MORETTIN e TOLOI, 2004; AGUIRRE, 2007).

Outra linha de pesquisa, foco desta tese, utiliza abordagens da física estatística para análise e caracterização dinâmica de séries temporais e espaço-temporais. Em geral, o desenvolvimento e/ou aplicação das técnicas baseiam-se em conceitos da teoria de sistemas complexos e sua abordagem a partir do cálculo de dimensões generalizadas [5](KANTZ e SCHREIBER, 1997). Um dos objetivos básicos no desenvolvimento de novos métodos de caracterização de séries temporais e espaço-temporais da física estatística é a análise de fenômenos críticos, como por exemplo, transições de fase, auto-organização e eventos extremos. Os estudos desses fenômenos passam por conceitos avançados da física moderna como escala, universalidade e grupos de renormalização (STANLEY, 1999). A abordagem da física estatística busca interpretar

[4]Nomenclatura utilizada para ambientes virtuais de representação da superfície da terra que possuem ampla capacidade de visualização de dados geográficos em multiresoluções e diferentes formatos.

[5]Conjunto de dimensões utilizadas na teoria de sistemas dinâmicos para descrever a dinâmica de um sistema complexo (HENTSCHEL e PROCACCIA, 1983)

o resultado obtido em uma análise de flutuações considerando uma fenomenologia física universal subjacente, como por exemplo, regimes dinâmicos de periodicidade, caos, transições de fase em geral relacionados *à* processos intermitentes e turbulentos (KANTZ e SCHREIBER, 1997).

Uma variedade de métodos e técnicas derivadas de abordagens da física estatística tem sido utilizada na análise de não linearidades e riquezas de escalas em séries temporais geradas a partir de sistemas dinâmicos complexos (KANTZ e SCHREIBER, 1997). Dentre as técnicas e métodos que permitem analisar não linearidades e escalas podemos citar os algoritmos para cálculos das dimensões generalizadas (dimensão de correlação e dimensão Hausdorff) e seus expoentes dinâmicos (Lyapunov, Hurst, etc). Técnicas e métodos mais recentes permitem análises complementares em escalas mais finas como o método *Detrended Fluctuation Analysis* (DFA) (Peng et al., 1994), gráfico de recorrência - *Recurrence Plot* (RP) (ECKMANN et al., 1987; MARWAN et al., 2008) e análise de padrões gradientes - *Gradient Pattern Analysis* (GPA) (ROSA et al. 1999), a qual tem sido empregada com grande sucesso para caracterizar regimes complexos como intermitência e turbulência durante a evolução de padrões espaços-temporais em sistemas não-lineares (ROSA et al. 1999; RAMOS et al. 2000; ASSIREU et al., 2002; ROSA et al., 2008). As aplicações ultrapassam as fronteiras da física teórica, alcançando as mais diversas áreas da ciência, como economia, ecologia, fisiologia, genética, meteorologia, oceanografia, ciências dos materiais e engenharias. Sendo que, na área de sensoriamento as aplicações não são comuns, uma vez que essa área tradicionalmente utiliza métodos da estatística convencional e reconhecimento de padrões estatísticos por semelhanças. Nesse sentido, esse trabalho traz uma novidade para área, ao combinar métodos de física estatística computacional para os estudos de processos dinâmicos da vegetação em imagens de sensoriamento remoto. Os processos dinâmicos da vegetação podem ser estudados em sensoriamento remoto devido ao incremento e a disponibilidade de imagens de sensores orbitais que possuem uma alta resolução temporal[6], ou seja, uma grande capacidade de aquisição e geração continua de imagens de toda superfície terrestre em curtos períodos de tempo.

Neste contexto, considerando a crescente disponibilidade de dados multitemporais, bem como, a possibilidade de integração entre diferentes métodos e técnicas de computação aplicada, a tese teve como motivação a pesquisa de uma nova abordagem

[6]Em sensoriamento remoto o termo resolução temporal está relacionado ao tempo de revisita de um sensor ao imagear uma mesma área. Em geral, esse tempo é função do tipo de órbita do satélite onde está o sensor e qual o ângulo de visada do imageamento.

para a visualização, análise e caracterização do uso e da cobertura da terra. Esse estudo[7] inclui discutir como disponibilizar dados e ferramentas em ambientes virtuais para aplicações de sensoriamento remoto.

Essa tese é desenvolvida considerando as seguintes motivações:

- De que forma a integração de séries temporais derivadas de sensoriamento remoto com o globo virtual pode ser útil para visualização e caracterização do uso e cobertura da terra?

- A pesquisa de novos operadores computacionais a partir da técnica de análise de padrões de gradiente, método DFA e gráfico de recorrência pode ser útil para o estudo de séries temporais derivadas de dados de sensoriamento remoto?

A partir dessas indagações o objetivo geral deste trabalho é estudar a combinação dos métodos computacionais da física estatística para suporte *à* visualização analítica e caracterização temporal e espaço-temporal do uso e cobertura da terra utilizando dados de sensoriamento remoto.

Para atingir o objetivo geral foram definidas as seguintes atividades:

- Definição e implementação de operadores computacionais a partir do método DFA, análise de padrões de gradiente e gráfico de recorrência;
- Definição e implementação computacional de modelos simplificados para geração de dados canônicos utilizados para o estudo dos operadores computacionais propostos;
- Definição e implementação de uma arquitetura computacional para integração de grandes volumes de dados derivados de sensoriamento remoto em globos virtuais;
- Aplicação dos operadores computacionais para caracterização do uso e cobertura da terra;
- Disponibilização e divulgação dos métodos computacionais estudados e implementados.

[7] Optou-se por empregar o termo estudos, uma vez que o mesmo pode explorar a pesquisa, seleção, modificação, implementação e avaliação dos métodos e ferramentas computacionais necessárias para o desenvolvimento desse trabalho.

Essa tese está dividida em mais 4 capítulos, abordando os seguintes conteúdos: O Capítulo 2 apresenta os dados de sensoriamento remoto utilizados e os métodos para integração de dados derivados de sensoriamento remoto ao globo virtual. Esse capítulo enfatiza os aspectos da visualização de dados em sensoriamento remoto, propondo um novo conceito baseado em laboratórios virtuais. O Capítulo 3 descreve os aspectos computacionais dos métodos propostos baseados em exemplos de aplicações em dados canônicos. O Capítulo 4 apresenta um estudo de caso combinando os métodos abordados no Capitulo 3 com os dados apresentados no Capítulo 2, enfatizando a caracterização da dinâmica de uso e cobertura da terra baseadas em séries temporais e espaço-temporais de índices de vegetação. O último capítulo apresenta as conclusões e perspectivas futuras para o trabalho proposto. Além dos capítulos formais, informações suplementares foram colocadas em Apêndices. O Apêndice A compreende uma revisão bibliográfica da análise de séries temporais e método DFA aplicado aos estudos ambientais. O Apêndice B descreve os modelos para geração de padrões dinâmicos, cada um no seu formalismo simplificado, utilizados para teste das técnicas estudadas e selecionadas para uso em séries temporais e imagens de sensoriamento remoto. O Apêndice C ilustra a formação das séries temporais a partir de imagens, o Apêndice D apresenta a estrutura das implementações computacionais realizadas e o Apêndice E contém exemplos das séries temporais com diferentes tipos de mudança do uso e cobertura da terra. Os Anexos A e B contém os artigos publicados e relacionados com o desenvolvimento da tese.

2 VISUALIZAÇÃO DE SÉRIES TEMPORAIS EM SENSORIAMENTO REMOTO

Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.
(*Albert Einstein*)

Esse Capítulo enfatiza um aspecto fundamental na análise de séries temporais que é a visualização de dados multitemporais de sensoriamento remoto. Nesse capítulo são apresentadas as etapas de pré-processamento e disponibilização das séries temporais bem como uma proposta de arquitetura computacional que permita a aplicação dos métodos que será apresentada no Capítulo 3.

Ao longo das últimas décadas, imagens multitemporais de satélites de observação da Terra transformaram-se em uma fonte primordial de informação para o monitoramento do planeta Terra, especialmente para estudar o uso da terra e mudanças de cobertura vegetal (LAMBIN e LINDERMAN, 2006). Sensores remotos orbitais, tais como: o AVHRR a bordo dos satélites *National Oceanic and Atmospheric Administration* (NOAA), o *Vegetation* a bordo do SPOT e o MODIS a bordo das plataformas Terra (EOS-AM1) e Aqua (EOS-PM1), têm sido responsáveis pela construção de um grande conjunto de dados ao longo dos últimos anos. Todos estes sensores adquirem imagens em com uma frequência quase diária, que é uma característica importante para sensores ópticos projetados para observar as mudanças de uso e cobertura da terra. Para o sensor MODIS um avanço significativo foi alcançado através da melhoria das resoluções espacial e espectral. Além disso, um consórcio internacional de cientistas tem se concentrado em fornecer dados validados MODIS com boa qualidade radiométrica e geométrica, desde o lançamento do satélite Terra (JUSTICE et al., 2002). Nos últimos 11 anos, milhões de gigabytes de dados MODIS foram gerados, forçando a comunidade de usuários de sensoriamento remoto procurar novas abordagens para o processamento, visualização e análise desse conjunto de imagens.

Por outro lado, o desenvolvimento de ferramentas *web* com base em globos virtuais tem proporcionado livre acesso a imagens de alta resolução espacial e mapas geográficos derivados de satélites de sensoriamento remoto. O desenvolvimento destes globos virtuais permite aos pesquisadores e público em geral visualizar dados geoespaciais, entender a multi-escala da geografia, processar os dados e publicar informações (BUTLER, 2006; BALLAGH et al., 2011; CHIANG et al., 2011). A visualização de dados de sensoriamento remoto, para fins científicos conjuntos, tem um grande

potencial para uma melhor compreensão da complexa variabilidade da dinâmica espaço-temporal dos ecossistemas terrestres. Esta ferramenta é útil para os cientistas compreenderem de forma mais eficiente os diferentes fenômenos incorporados em um grande volume de dados (NIELSON, 1991). No entanto, o pré-processamento e a extração de informações a partir desses conjuntos de dados necessitam de software específico e conhecimento técnico para colocá-los disponíveis para os usuários finais de uma forma amigável e acessível. A integração de séries temporais para os estudos sobre as mudanças de uso e cobertura da terra usando globos virtuais como o *Google Maps* (http://maps.google.com/), o *GoogleEarth* (http://earth.google.com/) e o Microsoft Virtual Earth (http://www.microsoft.com/maps/) ainda não estão facilmente acessíveis aos usuários devido as restrições no armazenamento de dados e a falta de uma arquitetura específica computacional para a integração e visualização das séries temporais desses dados. Aliado a esses fatos o amplo espectro de métodos computacionais para a análise dos dados muitas vezes estão disponíveis em diferentes softwares e requerem suporte computacional de hardware adequado para o processamento.

Nesse contexto, esse capítulo tem o objetivo de apresentar os resultados de uma proposta para facilitar a visualização e análise de grandes volumes de dados derivados de sensoriamento remoto. A primeira seção descreve uma arquitetura computacional para integração de séries temporais a um globo virtual. A segunda apresenta como a visualização de múltiplas fontes de dados pode contribuir para a caracterização do uso e cobertura da terra. Por último, na terceira seção são apresentados alguns exemplos práticos e potenciais aplicações da metodologia proposta neste trabalho de doutorado.

2.1 Laboratório Virtual de Séries Temporais de Sensoriamento Remoto

O conceito de Laboratório Virtual (LV) é associado a vários significados dependendo das aplicações. Ainda não existe na literatura um conceito fechado sobre o seu real significado. O conceito adotado nesse trabalho considera o Laboratório Virtual como um ambiente virtual para análise e visualização de dados. Sendo necessariamente suportado por métodos computacionais com o objetivo de diminuir a dependência de usuários não especializados a instalações físicas, de hardware e software.

Os laboratórios virtuais são hoje uma realidade para a comunidade científica, engenharias e também para os educadores. No campo da educação os LV são utilizados na maioria das disciplinas demonstrando conceitos da física, reações químicas e biologia em um ambiente puramente computacional (Rzepa e Tonge 1998,Afsarmanesh et al.

2000, Afsarmanesh et al. 2001, YARON, 2010). Aplicações mais complexas podem ser encontradas na área de engenharia nuclear em que simulações são necessárias para análise de riscos, contingenciamento e treinamento de recursos humanos que operam sistemas de alto risco (Tlaczala et al.,2008).

Grande parte do desenvolvimento dos LV deve-se especialmente a redução de custos que um ambiente simulado propicia e principalmente a um incremento na capacidade de transferência de dados na internet que permite ter usuários e componentes descentralizados. Na área de ciência e tecnologia, uma arquitetura computacional robusta para o desenvolvimento de laboratórios virtuais científicos foi proposta por Afsarmanesh et al. (2001). O trabalho pioneiro é uma iniciativa do projeto *Virtual Laboratory* da Universidade de Amsterdam, onde esses autores propõem uma arquitetura multi-camadas para integração dos vários ambientes e componentes de um laboratório virtual conforme apresentado na Figura 2.1. De uma maneira geral, o desenvolvimento de uma arquitetura dessa maneira permite a integração completa da aquisição, processamento, visualização e análise dos dados. Entretanto, apesar de ser o ambiente ideal de um laboratório virtual exige um esforço de recursos humanos e computacional para sua implementação, desenvolvimento e manutenção.

Na área de sensoriamento remoto a utilização dos laboratórios virtuais ainda são restritas e específicas. O que existe são soluções integradas em softwares proprietários que exploram pouco o conceito de laboratório virtual e se concentram mais na parte de processamento distribuído. Esse avanço é muito mais notável nas ferramentas de geoprocessamento e análise espacial do que propriamente nas de sensoriamento remoto. Ferramentas como ArcGis Explorer (explorer.arcgis.com) são exemplos de softwares de geoprocessamento que foram desenvolvidos para visualização de dados georeferenciados utilizando a internet. No entanto, com a popularização dos globos virtuais esses ambientes muitas vezes estão condicionados a usuários especialistas. A exploração de dados temporais e espaço-temporais, seja em softwares de geoprocessamento ou em globo virtuais, ainda é rara porque depende do desenvolvimento de arquiteturas de software para essa finalidade. Nesse contexto, esse trabalho busca contribuir com uma arquitetura computacional simplificada para visualização e análise de dados multitemporais em sensoriamento remoto.

Uma visão macro do Laboratório Virtual de Séries Temporais em Sensoriamento Remoto, proposto por Freitas et al. (2011), é dividida em cinco componentes modulares que são apresentados na Figura 2.2. Esta arquitetura embora bem mais simplificada que a arquitetura proposta por Afsarmanesh et al. (2000) é mais rea-

Figura 2.1 - Exemplo de arquitetura multi-camadas de um laboratório virtual.
Fonte: Afsarmanesh et al. (2000)

lista e customizada para o tratamento de grande volume de dados como é típico em sensoriamento remoto e enfatiza a visualização de dados integrada ao globo virtual.

O componente de *conjunto de dados* inclui as estruturas de *hardware* e *software* para o armazenamento de dados de sensoriamento remoto. Nesse contexto, a utilização de bancos de dados não é adequada para o armazenamento, considerando que sistemas de arquivos específicos para dados científicos como os formatos *Hierarchical Data Format* (HDF), *Common Data Format* (CDF) e *Flexible Image Transport System* (FITS) não suportam um grande volume de dados. Nesse sentido, o módulo de conjunto de dados é construído para transformar os dados espaciais em séries temporais indexadas geograficamente elaborando assim uma estrutura de arquivos específica

Figura 2.2 - Arquitetura simplificada do Laboratório Virtual de Séries Temporais em Sensoriamento Remoto.

para esse fim. O módulo de gerenciamento de dados realiza a conexão entre todos os componentes do laboratório sendo a parte crucial para garantir a rápida recuperação e visualização das séries temporais. O módulo de algoritmo e o módulo de análise devem desenvolver a capacidade de inserção de novos algoritmos bem como todo suporte necessário para elaboração de análise das séries temporais. O módulo de visualização estabelece a interface entre o laboratório e o usuário final utilizando todas as facilidades do globo virtual.

Em uma etapa inicial todo o laboratório é integrado no ambiente MatLab, utilizando as rotinas e bibliotecas matemáticas para suporte a visualização e análise dos

dados integrada ao globo virtual. Após a etapa de teste, com intuito de aumentar a interatividade e a popularização das séries MODIS, foi elaborado um portal de visualização das séries temporais utilizando recursos da linguagem *PHP: Hypertext Preprocessor* (PHP) e *HyperText Markup Language* (HTML). Esse portal consiste na interface principal de interação entre os usuários e os dados. Sua principal característica é a utilização de um globo virtual como elemento base, através do qual, tudo está integrado e georreferenciado, para tal utilizou-se o Google Maps (http://maps.google.com/) como globo virtual. Ferramentas de suporte estatístico das séries temporais, a utilização e o intercâmbio de outras bases de dados matriciais e vetoriais foram os requisitos para elaboração desse portal. O software aberto *MapServer* (http://mapserver.org/) é utilizado como sistema gerenciador e integrador de outras bases de dados que não as séries temporais.

O sistema gerenciador de dados implementado em linguagem *ANSI C* está dentro de um servidor de dados e rodando no sistema operacional Linux e também integrado *à* internet, garantindo que outros ambientes de visualização, como o de prototipagem em ambiente MatLab, também possa compartilhar a mesma base de dados. Isso garante a flexibilidade para usuários que queiram analisar suas séries temporais em outros ambientes fora do laboratório virtual. Como este trabalho possui o enfoque na análise de uso e cobertura da terra, os conjuntos de dados do sensor MODIS e do sensor TRMM foram integrados ao laboratório virtual. A etapa de pré-processamento desses dados bem como a área de estudo são descritas na próxima seção.

2.2 Conjuntos de dados

2.2.1 Abrangência Geográfica

O conjunto de dados do laboratório virtual integra todo o continente da América do Sul, que compreende cerca de 18 milhões de km^2, representando 12% da superfície terrestre da Terra. A região é caracterizada por diferentes biomas como a floresta tropical e sazonal, caatinga, pastagens, pantanal, cerrado e outros (EVA et al., 2004). Nesse continente ocorre em alguns biomas uma intensa pressão antrópica que gera mudanças no uso e cobertura da terra como, por exemplo, a conversão de florestas naturais e áreas de cerrado em pastos ou áreas agrícolas. Na sub-região tropical, há intensificação da agricultura relacionada com a produção de alimentos e de biocombustíveis (LAPOLA et al., 2010). A Figura 2.3 apresenta um mapa com os principais tipos de cobertura da terra da América do Sul para o ano de 2000.

Figura 2.3 - Mapa de cobertura da terra da América do Sul
Fonte: Adpatado: Eva et al.(2004).

2.2.2 Sensor MODIS

Lançado em dezembro de 1999, com início de operação em fevereiro de 2000, o sensor MODIS tem como objetivo geral atender três segmentos do projeto Earth Observing System (EOS): atmosfera, oceano e continente. Este sensor permitiu avanços nos trabalhos que necessitam de maior resolução temporal (quase diária), aliada a moderada resolução espacial (250 m). A qualidade do georeferenciamento das imagens MODIS (WOLFE et al., 2002) permite a elaboração de séries temporais com

precisão. Este projeto combina as características das imagens (qualidade geométrica, radiométrica, espectral e temporal) com algoritmos de geração de produtos validados (JUSTICE et al., 2002) que estimam, por exemplo, parâmetros biofísicos da vegetação. Nesse sentido, o objetivo dos produtos voltados para aplicações terrestres gerados pelo MODIS está, basicamente, no monitoramento da cobertura vegetal do planeta, identificando as mudanças e a variabilidade no tempo (JUSTICE et al., 2002).

2.2.3 Produto MOD09 - Refletância de Superfície

Este produto refere-se às imagens de refletância espectral de superfície para as bandas 1 a 7, centradas nos seguintes comprimentos de onda: 648 ηm, 858 ηm, 470 ηm, 555 ηm, 1240 ηm, 1640 ηm, e 2130 ηm, sendo computado a partir dos níveis 1 A de processamento (JUSTICE et al., 2002). Estes produtos são disponibilizados com resolução espacial de 250 m para as bandas do vermelho e infravermelho próximo e 500 metros para as demais bandas. As imagens também estão disponíveis como produtos diário e em composições de 8, 16 e 30 dias. A correção atmosférica deste produto é realizada com o modelo desenvolvido por Tanré et al. (1990) e aperfeiçoado por Vermote et al. (1997), denominado *Second Simulation of the Satellite Signal in the Solar Spectrum Radiative Code* (6S). A correção dos efeitos atmosféricos nas bandas do sensor MODIS utiliza informações de produtos e bandas específicas para correção atmosférica, como os produtos MOD35 (Máscara de Nuvens), MOD04 (Aerossol), MOD07 (Perfis Atmosféricos) tornando este produto operacional uma vez que os dados de entrada para o 6S são previamente estimados com os outros produtos (KING et al., 2002; JUSTICE et al., 2002). Os produtos de refletância de superfície são utilizados para elaboração dos índices de vegetação. Esses índices apresentam algumas relações indiretas com os parâmetros biofísicos da vegetação, permitindo algumas inferências sobre a fenologia e cobertura vegetal.

As imagens MODIS foram adquiridas no portal *Warehouse Inventory Search Tool* (WIST) (https://wist.echo.nasa.gov). Três produtos foram selecionados na sua versão atual de processamento v005, conhecida também como coleção 5. O produto MOD13Q1 é a composição de 16 dias e resolução espacial de 250 m e os produtos MOD09Q1 e MOD09A1 são as bandas de reflectância de superfície com 250m e 500m de resolução espacial e 8 dias de resolução temporal, respectivamente. Os dados foram adquiridos no período de julho de 2000 a junho de 2011. A área de estudo está dividida em 32 *tiles* MODIS com 1.200 x 1.200 km cada, conforme ilustrado na Figura 2.4. Os *tiles* são divisões espaciais adotadas para distribui-

ção dos arquivos do produto MODIS para aplicações na superfície terrestre. Um *tile* compreende a informação de 10x10 graus na projeção sinusoidal. Um total de 8.736 tiles foi adquirido, o que correspondente a 4,5 TB de dados brutos. As imagens estão disponíveis na projeção sinusoidal (datum WGS84). Todas as bandas foram re-projetados para o sistema de coordenadas geográficas com o mesmo datum e convertidos a partir do HDF para o formato *Georeferencing Tagged Image File Format* (GeoTIFF) para garantir a portabilidade de dados entre softwares. Para esse procedimento foi utilizado o software *MODIS Reprojection Tool* (MRT) (http://lpdaac.usgs.gov/lpdaac/tools/modis$_r eprojection_t ool$).

Figura 2.4 - Recortes de imagens MODIS que englobam a região da América do Sul.

Como o objetivo é estudar a dinâmica da vegetação, as bandas espectrais dos produtos reflectância de superfície foram utilizadas para o cálculo do índice de vegetação. O índice de vegetação *Enhanced Vegetation Index 2* (EVI2) (JIANG et al., 2008) foi escolhido para o presente estudo, pois ressalta as variações da vegetação corrigindo os problemas de saturação e ruídos como outros índices tradicionais como o *Normalized Difference Vegetation Index* (NDVI) e *Enhanced Vegetation Index* (EVI). Esse índice não está disponível como um produto MODIS, o que torna sua disponibilização e distribuição inédita dentro da gama de produtos MODIS para estudos da superfície terrestre. O EVI2 é calculado usando a reflectância de superfície das

bandas do vermelho (*Red*) e infravermelho próximo (*NIR*) disponíveis no produto MOD13Q1(16 dias) ou nos produtos MOD09A1 (8 dias) e é definido por:

$$EVI2 = 2,5\frac{(NIR - Red)}{(NIR + 2,4 * Red + 1)}. \tag{2.1}$$

Além do cálculo do EVI2 (Eq. 2.1), as bandas auxiliares com informação de ângulo de visada zenital do sensor e a banda azul (reflectância de superfície) foram utilizadas para uma etapa de pré-filtragem dos dados brutos, como será explicado na seção de filtragem dos dados MODIS. Além do EVI2 serão inclusos outros índices espectrais como o *Normalized Difference Water Index* (NDWI), bem como a utilização de imagens MOD09 com 8 dias de resolução temporal. Nesse trabalho foram utilizadas apenas as imagens MOD13 com 16 dias de resolução temporal. Entretanto, ressalta-se que a mesma metodologia é aplicável às imagens MOD09 com 8 dias de resolução temporal.

2.2.4 Dados TRMM

A série de estimativa de precipitação acumulada mensal do produto *3B43 V6* do *Tropical Rainfall Measuring Mission* (TRMM) foi utilizada como um dado auxiliar. Esses são disponibilizados pelo *Distributed Active Archive System*, daac.gsfc.nasa.gov (DAAC) da NASA (HUFFMAN et al., 2007). O pixel TRMM tem uma área mínima de aproximadamente de 0,25 graus, o que requer baixa capacidade de armazenamento. Os dados de precipitação foram incorporados ao laboratório virtual como um dado marginal que representa os períodos secos e chuvosos. Ressalta-se que, embora amplamente utilizado na literatura, em escalas locais como proposto nesse trabalho os valores absolutos de precipitação devem ser analisados e avaliados com cuidado. De qualquer forma, é a melhor medida de precipitação que se pode disponibilizar em grande escala, além de ser uma importante informação para os estudos de fenologia da vegetação.

2.3 Procedimentos de Filtragem

Os dados ópticos em sensoriamento remoto são freqüentemente afetados por cobertura de nuvens e ruído do sensor os quais interferem na capacidade de caracterizar a dinâmica espacial-temporal de cobertura do solo. Dessa forma para se construir uma série temporal contínua e consistente é necessário filtrar os dados. Os procedimentos de filtragem foram realizados em duas etapas, gerando duas séries temporais de

valores de EVI2 denominadas: 1) sem a filtragem por transformada *wavelets*[1]; e 2) com transformada *wavelet*. O primeiro procedimento de filtragem utilizou limiares rígidos seguindo as metodologias propostas por Sakamoto et al. (2005) e Thayn e Price (2008). Esses autores propõem que os dados sejam excluídos da série temporal original se a reflectância na faixa azul for maior do que 10% ou se o ângulo zenital for maior que $32,5$. Esses limiares fundamentam-se na capacidade de detecção de nuvens na banda do azul e também na eliminação de pixels que estão em um ângulo muito distante da visada vertical do sensor, os ditos pixels *off-nadir*.

A partir de eliminação dos *pixels* inválidos, a interpolação linear é realizada utilizando a data diária de observação de cada pixel válido, garantindo séries igualmente espaçadas no tempo. Um parâmetro de qualidade é gerado para cada série temporal contando o percentual de *pixels* interpolados. O Apêndice C descreve o procedimento de elaboração de uma série temporal a partir das imagens composição.

Um procedimento de filtragem adicional é realizado com o objetivo de suavizar as regiões de altas freqüências das séries temporais. Essa metodologia seguiu os procedimentos propostos por Freitas e Shimabukuro (2008). A decomposição do sinal pela transformada *wavelet* permite visualizar todas as escalas de um sinal e assim eliminar as altas freqüências que são, no caso das imagens EVI2, tipicamente associadas com a presença de ruído. A escolha do método *wavelet* foi devido a sua robustez para o tratamento de séries não-estacionárias como as apresentadas nesse trabalho. Uma abordagem detalhada sobre esse método pode ser encontrada em Meyer (1992), Morettin (1999) e Domingues et al. (2005) que descrevem os fundamentos e aplicações da transformada *wavelet*.

A transformada *wavelet* é dada por (DAUBECHIES, 1992; MEYER,1992):

$$w(a,b) \equiv \int_{\infty}^{-\infty} \psi^*_{a,b}(t) f(t) dt = \frac{1}{\sqrt{a}} \int_{\infty}^{-\infty} \psi^*_{a,b}(\frac{t-b}{a}) f(t) dt, \tag{2.2}$$

$$\psi_{a,b}(t) \equiv \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a}), a > 0, -\infty < b < \infty, \tag{2.3}$$

Onde a é o parâmetro de escala, b é o parâmetro de translação, *f(t)* é a função a ser transformada, ψ é a chamada *wavelet* mãe e ψ^* representa o complexo conjugado

[1]Nesse trabalho adotou-se a nomenclatura em inglês por ser universalmente conhecida e porque não existe uma definição formal nos dicionários oficiais de língua portuguesa. Embora existam traduções livres comumente utilizadas como ondaletas ou ondeletas.

dessa função.

As Equação 2.2 e 2.3 são contínuas, sendo indicadas para análise de séries ou sinais contínuos. Para a filtragem de sinais discretos é necessário usar a transformada discreta de *wavelet* (*Discrete Wavelets Transform* (DWT)). Nessa função não é viável calcular todos os possíveis valores dos parâmetros a e b nos conjuntos dos números reais. O procedimento comum é aplicar uma restrição e calcular os parâmetros em escala diádica, essa é semelhante *à* notação musical em que potências de dois estão relacionadas com os intervalos (oitavas) e a durações das notas (GALVÃO et al.,2001). Com essa transformação o resultado da aplicação da transformada é um conjunto de coeficientes *wavelet* indexados por m (nível de escala) e n(índice de translação) e dado por:

$$a = 2^m,\ b = n2^m,\ em\ que,\ m \in Z\ e\ n \in Z, \tag{2.4}$$

A transformada discreta de *wavelet* é definida por:

$$DWT(m,k) = 2^{m/2}\psi(2^m t - k), \tag{2.5}$$

Onde ψ é a *wavelet* mãe e $\psi_{m,k}$ são as chamadas *wavelets* filhas. O nível m determina quantas *wavelets* são necessárias para cobrir uma *wavelet* mãe e o número k determina a posição da *wavelet* no tempo t. Com isso é possível decompor qualquer sinal arbitrário *f(t)* dentro de suas componentes *wavelets*. Essa aproximação é similar a análise harmônica exceto que, em vez do sinal ser decomposto em diferentes freqüências e amplitudes, é decomposto em diferentes escalas (níveis), diferentes posições correspondentes a amplitude das *wavelets* dada por :

$$f_m(t) = A_m(t) + \sum_{m}^{i=1} D_m(t), \tag{2.6}$$

$$D_m(t) = \sum_{k=-\infty}^{\infty} W_{m,k}\psi_{m,k}(t), \tag{2.7}$$

$$A_m(t) = \sum_{k=-\infty}^{\infty} V_{m,k\phi_{m,k}(t)}, \tag{2.8}$$

onde $W_{(m,k)}$ são os coeficientes das *wavelets*, $\Phi_{(m,k)}$ é a função base de escala e e translação e $V_{(m,k)}$ são os coeficientes de escala.

A implementação da função discreta (Equação 2.5) é realizada através de algoritmos recursivos de filtros passa-baixa (L) e de passa-alta(H), esse é o algoritmo conhecido como pirâmide e foi proposto por Mallat (1989). Para cada sinal decomposto na escala m existe um coeficiente de detalhes (d) definido por filtros passa-alta, e aproximações (c) definido por filtros passa-baixa que estão associados à *wavelet* mãe. Os passos da decomposição do algoritmo de Mallat está descrito na Figura 2.5a. O símbolo ($\downarrow$ 2) representa a operação de sub-amostragem e consiste em eliminar todos os coeficientes de índice para (0,2,...) de uma seqüência. Já o símbolo ($\uparrow$ 2) é o contrário, representando a operação de inserção de zeros entre os pontos de uma seqüência. Para cada nível em escala m gera-se recursivamente dois conjuntos de coeficientes dos detalhes d_m e os de aproximação c_m, esses coeficientes são definidos por:

$$c_{m+1}(n) = \sum_{k=0}^{K-1} l(k) C_m(2n - k + 1), \tag{2.9}$$

$$d_{m+1}(n) = \sum_{k=0}^{K-1} h(k) C_m(2n - k + 1), \tag{2.10}$$

em que as seqüências *l(k)* e *h(k)*, de K pontos representam, respectivamente, as filtragens passa-baixas e passa-altas. O número de coeficientes depende do número de níveis de escala, algumas vezes chamado de níveis de resolução. A representação da *wavelet* mãe $\psi(x)$ não possui uma formulação matemática explícita e é apresentada de forma numérica a partir das seqüências de filtros *l(k)* e *h(k)*. Os coeficientes de detalhes estão associados às características locais do sinal, representando a variabilidade em alta-frequência. A aproximação é o contrário, representando a alta-frequência, ou seja, uma suavização do sinal. Em analogia grosseira pode-se relacionar com a visão humana, quando olhamos um objeto de longe, vemos a aproximação desse e quanto mais nos aproximamos do objeto percebemos os detalhes (GALVÃO, et al. 2001).

Uma propriedade da transformada discreta de *wavelet* é sua capacidade de inversão. Nesse sentido, pode-se reconstruir a série observada eliminando alguns detalhes e obtendo assim uma série suavizada (Figura 2.5b). A reconstrução dos coeficientes

c_{m-1} a partir de c_m e d_m é dada por:

$$c_{m-1}(n) = \sum_{k=0}^{N_m-1} \left[\overline{l}(n-2k+K-2)c_m(k) + \overline{h}(n-2k+K-2)d_m(k))\right], \quad (2.11)$$

em que as sequências d_m e c_m possuem N_m pontos cada uma e $\overline{l}(k) = \overline{h}(k) = 0$ para $k < 0$ e $k \leq K$

Para alguns tipos de *wavelets* as sequências dos filtros de inversão são iguais dadas por:

$$\overline{l}(k) = l(K-1-k) \; e \; \overline{h}(k) = h(K-1-k), \quad (2.12)$$

essa inversão permite limiarizar alguns coeficientes de detalhes e reconstruir um sinal suavizado, procedimento ilustrado na Figura 2.5c. Uma gama de heurísticas estão disponíveis para definir a limiarização dos coeficientes nas diferentes escalas. Nesse trabalho foi adotado o limiar rígido eliminando, os detalhes das duas últimas decomposições, esse procedimento mostra-se satisfatório suavizando as séries de EVI2. O procedimento de filtragem consiste em zerar os coeficientes de detalhe que, em módulo, são menores que o limiar estabelecido em cada nível m de decomposição (Figura 2.5b).

A escolha da *wavelet* mãe ainda é um problema em aberto dentro da análise *wavelets* e depende do tipo de sinal estudado, e em geral, é função da facilidade de implementação. Para esse trabalho, foi selecionada a *wavelet* Daubechies (db8) (DAUBECHIES, 1992) como função mãe (Figura 2.6). Esta escolha foi fundamentada em dois aspectos: primeiro em testes empíricos avaliando o erro médio quadrático e a forma da suavização entre a decomposição e o sinal original; e segundo que essa mesma família de *wavelets* é amplamente encontrada na literatura para análise de sinais ambientais e de séries temporais de índices de vegetação (SAKAMOTO et al., 2005).

Para cada conjunto de pixels é elaborado um perfil de séries temporais nas imagens ordenadas no tempo. Cada série temporal é decomposta em oito diferentes escalas e reconstruída utilizando as seis maiores escalas correspondentes às freqüências mais baixas. As freqüências mais altas foram eliminadas porque são geralmente associadas

Figura 2.5 - Resumo do algoritmo de decomposição proposto por Mallat a) Passos da tranformada discreta de *wavelet* b)Procedimento de limiarização dos coeficientes e c) Transformada inversa.
Fonte: Adaptado de Galvão et al. (2001)

com a presença de ruídos de alta frequência do sensor e respostas espectrais de nuvens e suas sombras. A Figura 2.7 ilustra esse procedimento de decomposição.

2.4 Ferramentas de Visualização

O conjunto de dados gerados atualmente conta com mais de 500 milhões de séries temporais EVI2 filtradas com e sem transformada *wavelet* para todo o continente Sul-Americano. A elaboração desse conjunto de dados teve um esforço computacional significativo envolvendo mais de 60 dias de processamento por meio de três computadores pessoais (PC) com a configuração processador Intel Core i7, HD de 2 TB e 8 GB de memória RAM e sistema operacional Linux *Mandriva*. Todos os procedimentos computacionais foram realizados nos ambientes MatLab, utilizando a plataforma *ANSI C* para integração das séries temporais ao laboratório virtual.

As séries temporais EVI2 foram integradas ao globo virtual (*GoogleMaps*) usando o gerenciador de conjunto de dados. Para visualizar a série temporal EVI2 no globo virtual um *website* foi elaborado e disponibilizado em http://www.dsr.inpe.br/laf/series.html (Figura 2.8) usando o *Google Maps* e a *Go-*

Figura 2.6 - Funções *wavelet* a) Função de escala ; b) *Wavelet* mãe db8; c) coeficientes de decomposição filtro passa-baixa; d) coeficientes de decomposição filtro passa-alta;e) coeficientes de reconstrução filtro passa-baixa e f)coeficientes de reconstrução filtro passa-alta;

ogle Visualization Application Programming Interface para suporte *à* visualização do globo virtual e das séries temporais. Para cada chamada de uma coordenada geográfica do globo virtual as duas séries temporais EVI2 são instantaneamente recuperadas. A série temporal recuperada em cada chamada, *click*, refere-se a um pixel MODIS observado ao longo do tempo. A integração com o globo virtual mostra espaço geográfico estático com imagens de alta resolução espacial fornecidas pelo *Google Maps* (Figura 2.8a). No entanto, a análise dessas séries temporais deve ser cautelosa devido à resolução espacial diferente das imagens, a de alta resolução disponível no *googlemaps* e pixels MODIS. Cada pixel MODIS representa aproximadamente uma área de 6,25ha (250 x 250 m), enquanto a imagem de alta resolução espacial fornecida pelo Google Maps é usada apenas para localizar o pixel MODIS. Além da visualização das imagens, uma ferramenta foi construída para avaliar a anisotropia de elevação em torno do ponto selecionado (balão vermelho). Esta ferramenta utiliza as informações do modelo de elevação disponíveis na *Google Maps API*. A visualização de anisotropia ao entorno do terreno é disponibilizada através de um simples gráfico polar de elevação em torno de dois círculos amostrais, permitindo uma visão rápida da topografia em torno do ponto selecionado (Figura 2.8b). Esta ferramenta permite interatividade e oferece a opção de variar as distâncias en-

Figura 2.7 - Exemplo de decomposição utilizando a transformada discreta *wavelet* de uma série EVI2 nos coeficientes de aproximação (a_m) e detalhes(d_m) e série filtrada eliminado os coeficientes de detalhes de maior frequência.

tre o centro da coordenada selecionada e os círculos amostrados. Na Figura 2.8c é possível visualizar os 11 anos de séries temporais EVI2 utilizando todos os recursos interativos fornecidos pelas funcionalidades do *Google Charts API*. As linhas vermelhas e azuis representam as séries temporais filtradas com *wavelet* e sem *wavelet*, respectivamente.

Uma forma alternativa para a visualização e análise dos dados foi desenvolvida numa etapa anterior *à* elaboração do *website*. Essa integração e visualização dos dados utiliza o gerenciador de dados do laboratório virtual e recursos das ferramentas *Google Maps* e do *Google Earth API* como suporte *à* visualização. Toda essa implementação computacional utiliza os recursos do ambiente MatLab. Usando essa função é possível visualizar a série temporal de um pixel do MODIS e TRMM para qualquer coordenada geográfica na América do Sul, um exemplo dessa visualização é apresentado na Figura 2.9.

Figura 2.8 - Componentes do módulo de visualização no *website*: a) Google globo virtual usado para selecionar as coordenadas geográficas da área de interesse; b) diagrama polar de elevação em torno de dois círculos amostrados na imagem do Google, e c) série temporal EVI para o enredo selecionado (linha vermelha filtrada com *wavelet* e azul filtrada sem *wavelet*).

2.5 Interpretação das séries temporais

A Figura 2.10 apresenta em detalhes algumas métricas extraídas e a interpretação qualitativa das séries mostradas utilizando as setas em preto. A série em azul no gráfico EVI2 (sem filtro, Figura 2.10a) representa os valores observados em 1 pixel nas imagens MODIS ao longo de 11 anos, de fevereiro de 2000 a dezembro de 2010. A série em vermelho (com filtro) é resultado de uma suavização da série em azul utilizando a transformada *wavelet*. Analisando a fase e amplitude das séries de EVI2 é possível inferir sobre o tipo de uso e a cobertura da terra num determinado período de tempo (Figura 2.10a). Grandes amplitudes (EVI2 > 0,6) durante um curto período de tempo estão associadas principalmente a culturas anuais como soja e milho, por exemplo. Ciclos com duração de 1 ano ou mais estão associados a pastagem, algumas formações naturais ou a culturas semi-perenes ou perenes

Figura 2.9 - Visualização com imagens do GoogleEarth integradas *às* séries EVI para a coordenada selecionada. a) Imagem *GoogleEarth* b) Imagem *GoogleMaps* e C) Visualização conjunta das séries TRMM e EVI2 (linha verde filtrada com *wavelet* e vermelha filtrada sem *wavelet*).

como cana-de-açúcar, café, dentre outras. Este exemplo está relacionado a uma área convertida para cana-de-açúcar a partir de 2007. A interpretação das séries de EVI2 deve considerar também a contextualização regional da área observada, ou seja, área agrícola, pecuária intensiva, agro florestal, etc. Em geral, as séries apresentam as variações de fenologia e/ou as variações no uso da terra. Essas variações podem estar relacionadas ao tipo de manejo como no exemplo da Figura 2.10 que mostra uma área de agricultura irrigada (pivô central) com intenso cultivo de culturas anuais de 2000 a 2007 e a conversão para cana-de-açúcar de 2007 a 2010. A série temporal de precipitação acumulada mensal (Figura 2.10b) é formada por 1 pixel TRMM cuja posição é a mais próxima da coordenada geográfica do polígono em azul mostrada na Figura 2.8. Embora o gráfico da série de precipitação acumulada represente uma estimativa com uma resolução espacial baixa ($0,25°$ ou $\approx 25km$) esta informação auxilia na interpretação dos dados EVI2, principalmente nos estudos relacionados à fenologia da vegetação e verificação dos regimes pluviométricos secos e chuvosos. Observando essa série é possível definir bem o período das estações secas e chuvosas, e os picos de máxima precipitação.

Comparando as séries de EVI2 e precipitação TRMM pôde-se observar que, durante os períodos secos observados entre os anos de 2001 a 2006, ocorreu o desenvol-

vimento de culturas agrícolas, caracterizando assim a necessidade de irrigação da cultura nesse período. O regime pluviométrico da região também pôde ser observado, mostrando que há uma relação entre os períodos mais secos e a colheita da cana-de-açúcar nos anos entre 2007 a 2010 (Figura 2.10b).

Figura 2.10 - Exemplos de extração de métricas e interpretação qualitativa das séries EVI2 apresentadas na Figura 2.8a.

O exemplo da Figura 2.11 representa uma área de reflorestamento na qual pôde-se verificar que o regime de variabilidade da série temporal é diferente do apresentado na Figura anterior (2.10). Nessa série identificou-se um ciclo de 7 anos entre 2000 e o inicio de 2007. Esse ciclo representa o período que o eucalipto foi plantado ou rebrotado até a época do seu corte. A interpretação do corte do eucalipto em 2007 foi devido a queda abrupta do índice EVI2, indicando a supressão da vegetação. A afirmação de que era uma plantação de eucalipto deve-se principalmente ao conhecimento da área da região geográfica do município de Caçapava em São Paulo. De qualquer forma, nas imagens de alta resolução disponíveis no *GoogleMaps* foi possível identificar que é uma área de reflorestamento. Porém somente com a imagem disponível no *GoogleMaps* não consegue-se inferir, por exemplo, a idade de desenvolvimento da cultura. A série de precipitação do TRMM foi auxiliar e indicou os

períodos secos e chuvosos entre os anos analisados.

Figura 2.11 - Exemplo de série temporal de uma área de reflorestamento. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) Série EVI2 para coordenada selecionada. d)gráfico de área da precipitação acumulada mensal TRMM.

A Figura 2.12a mostra uma região de floresta com várias áreas desmatadas no município de Feliz Natal no estado do Mato Grosso. A interpretação do gráfico polar permite inferir que a elevação em torno do ponto selecionado $(11°55'S, 54°10'W)$ varia entre 354 e 360 metros, indicando um terreno razoavelmente plano (Figura

2.12b). As séries temporais EVI2 são apresentadas na Figura 2.12c e indicam uma diminuição nos valores de EVI2 no ano de 2004, relacionada com a perda significativa de biomassa devido ao processo de desmatamento. De 2005 a 2007 não havia quase nenhuma vegetação em processo de regeneração, isso pode ser afirmado devido aos baixos valores na série de EVI2 durante esse período. Para os anos de 2007 e 2008, uma resposta típica espectral de áreas agrícolas pôde ser observada, caracterizada por um incremento rápido seguido por uma diminuição dos valores de índice de vegetação, indicando de forma bem definida os ciclos curtos de crescimento de culturas agrícolas anuais.

Figura 2.12 - Exemplo de série temporal com detecção de desflorestamento. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) série EVI2 para coordenada selecionada.

A sazonalidade de um ponto selecionado em uma área que foi desmatada em 2004, no Parque Nacional do Xingu, Mato Grosso, Brasil é mostrada no gráfico da Figura 2.13. A análise visual das séries temporais indica uma conversão da floresta para pastagem. Também pode ser observado que o processo de desmatamento começou no primeiro trimestre e terminou no último trimestre de 2004. É interessante notar

que as duas áreas observadas nas Figuras 2.12 e 2.13 apresentam diferentes tipos de mudanças de uso que podem ser facilmente observados na análise das séries temporais do EVI2 de 11 anos do MODIS.

Figura 2.13 - Exemplo de série temporal de desflorestamento. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) série EVI2 para coordenada selecionada.

Ressalta-se que as duas séries temporais geram curvas semelhantes para os gráficos selecionados das Figuras 2.12c e 2.13c. Na série suavizada (vermelha) há pouca diferença significativa em termos de análise de mudança de uso e cobertura da terra. No entanto, com suavização filtrada a série temporal EVI2 pode ser usada em procedimentos de classificação e mineração dos dados bem como em outros procedimentos em que o sinal de alta freqüência não é interessante.

A Figura 2.14a apresenta uma região com intenso cultivo da cana-de-açúcar no município de Brotas, no estado de São Paulo. A Figura 2.14c mostra a dinâmica de nove anos de colheita de cana. Com algum conhecimento sobre o cultivo da cana, diversas informações podem ser extraídas a partir das séries temporais. Uma breve descrição das práticas de cultivo de cana nessa área pode ser dada pelo analista

da seguinte forma. No início da série temporal os baixos valores de EVI2 indicam solo exposto sobre o qual a cana foi plantada no início de 2001. Esta planta de cana cresceu por cerca de 18 meses, sendo colhida pela primeira vez por volta de julho de 2002. Após o primeiro corte, a soqueira da cana foi colhida uma vez por ano, por volta de julho de 2003 a 2007. A área de cana foi renovada no final de 2007, quando ocorreu o rodízio de cultura agrícola com uma safra de verão anual seguida pela cana plantada nova no final de 2008 e colhida em meados de 2009. Maiores informações sobre esta área podem ser obtidas no site do projeto Canasat ($http://www.dsr.inpe.br/laf/canasat/$) (RUDORFF et al., 2010).

A descrição acima indica que com um mínimo de conhecimento técnico sobre práticas agrícolas na colheita da cana-de-açúcar, é possível ao usuário recuperar os 11 anos de história de talhões agrícolas. Isto pode ser de grande interesse para as empresas certificadoras de qualidade que precisam saber o histórico de alterações de uso e cobertura da terra.

Figura 2.14 - Exemplo de série temporal de cana-de-açúcar. a) imagem GoogleMaps; b) gráfico polar de elevação ; c) Série EVI2 para coordenada selecionada.

A Figura 2.15 mostra uma região de cerrado na fronteira agrícola localizada no oeste do estado da Bahia. A região era originalmente coberta por cerrado e tem sido gradualmente convertida para uso agrícola. Esta região é caracterizada por grandes plantações de soja, milho, algodão e café. Três séries ao longo do tempo ilustram o comportamento típico da mudança de uso e cobertura da terra nessa região (Figuras 2.15c, 2.15d e 2.15e). Figura 2.15c mostra o comportamento típico de uma área de cerrado com baixos valores de EVI2 e pequenas amplitudes que não se alteraram ao longo dos 11 anos. A Figura 2.15d mostra a série de EVI2 de uma área de cerrado que foi convertida para a agricultura durante a safra de 2007/2008. A Figura 2.15 e mostra a conversão de cerrado para a agricultura após 2002. Nesses exemplos é possível observar que o regime de variabilidade se altera significativamente para alguns tipos de uso e cobertura da terra. Além disso, é possível analisar qualitativamente os períodos aproximados de plantio, máximo desenvolvimento e colheita para determinadas áreas.

2.6 Aplicações do Laboratório Virtual no Contexto de Aplicações do INPE

O Laboratório Virtual permite a integração com outras fontes de dados, como por exemplo, formatos vetoriais que podem ser incorporados na análise. O exemplo da Figura 2.16 mostra como essa integração pode auxiliar na auditoria de dados e no entendimento do uso da terra após um desflorestamento mapeado pelo Projeto de Monitoramento do Desflorestamento na Amazônia Legal (PRODES) no município de União do Sul-MT. Ressalta-se que devido a diferença de resolução espacial dos dados utilizados pelo PRODES e a resolução espacial dos dados MODIS, somente os desmatamentos maiores do que 6,25 ha podem ser verificados. No exemplo da Figura 2.16 é apresentada uma área que foi mapeada como desflorestamento pelo PRODES no ano de 2007. Esta ferramenta permite listar os dados vetoriais disponíveis no portal (Figura 2.16b). Nesse exemplo são listados os polígonos PRODES dos anos de 2007 a 2010. Analisando a série temporal (Figura 2.16c) observamos que o desflorestamento tem início no ano de 2007, como mapeado pelo PRODES. A série temporal permite estimar que o desflorestamento dessa área ocorreu no intervalo entre março e dezembro de 2007, realçado em tons laranjas. Além disso, é possível estimar o tipo de uso da terra após o desflorestamento. Neste caso observamos uma resposta típica de regeneração da área com o incremento gradual do índice EVI2 ao longo do período de 2008 a 2010. As pequenas variações da amplitude entre 2009 e 2010 permitem concluir que há uma recuperação da biomassa, e não ocorrendo nenhuma atividade de limpeza, grande queimada ou preparação do solo para o uso

Figura 2.15 - Exemplo de desmatamento em cerrado: a) imagem GoogleMaps; b) gráfico polar de elevação ; c) série EVI2 para coordenada selecionada na figura a, ponto 1, cerrado; d)série EVI2 para coordenada selecionada na figura a, ponto 2, desmatamento em cerrado em 2007; e c)série EVI2 para coordenada selecionada na figura a, ponto 3, desmatamento em 2002.

agrícola.

Figura 2.16 - Exemplo de série temporal EVI2 para verificação dos polígonos de desflorestamento PRODES.

Na Figura 2.17 são apresentados como a integração com outras fontes de dados, tais como fotografias aéreas e imagem TM, pode auxiliar na interpretação das séries temporais e permitir uma tomada de decisão para otimização de custos em um projeto. Este exemplo refere-se ao projeto da Moratória da Soja (RUDORFF et al., 2011). Neste projeto, inicialmente a série temporal de EVI2 é utilizada para verificar se o polígono tem probabilidade de ser cultura agrícola no ano em questão. Com isto pode-se selecionar os polígonos para verificação através de sobrevoos com conseqüente redução do custo de operação. Além disso, permite verificar se o polígono selecionado com a cultura de soja está dentro do escopo da moratória, ou seja, se este polígono representa uma área que foi desflorestada após o dia 24 de julho de 2006, que é o caso desse exemplo. Entretanto, devido ao comportamento similar entre diversos cultivos anuais, como por exemplo, as culturas de arroz de sequeiro e soja, esta interface apenas indica este polígono como área agrícola. Dessa maneira,

o sobrevoo foi realizado confirmando a presença de soja em parte da área na safra de 2009/10 (Figura 2.17b). Portanto, a série temporal possibilitou verificar que essa área localizada no município de Querência-MT está no contexto da moratória da soja, uma vez que esta área foi desflorestada em data posterior a estabelecida e ocorreu o plantio de soja. Pode-se também afirmar que neste caso, a conversão para área agrícola demorou 3 anos (Figura 2.17c).

Figura 2.17 - a) Imagem híbrida do GoogleMaps integrada com polígono PRODES; b) Fotos aéreas do sobrevôo ao polígono no mês de dezembro de 2009 e imagem Landsat-TM do mês de agosto de 2009; c) Série temporal do EVI2 da coordenada selecionada.

Outro exemplo é a estimativa do uso anterior ao cultivo da cana-de-açúcar. Para isto foram integrados ao laboratório virtual polígonos de cana-de-açúcar previamente mapeados pelo projeto Canasat (em preto na Figura 2.18) e uma série de imagens do sensor Landsat-TM georreferenciadas. Essas imagens ajudam na interpretação da série temporal de EVI2 e permite inferir quais classes de uso da terra estavam sendo convertidas para a cultura da cana-de-açúcar. Neste projeto, com auxílio das imagens TM e das séries temporais, os fotointérpretes classificaram qual o tipo de

uso em cada ano-safra. Esta verificação foi realizada utilizando uma ferramenta web integrada ao portal das séries temporais. Neste exemplo, nos anos-safra de 2000 a 2003 a área era ocupada por pastagem, que pode ser comprovada com a imagem TM de 17/11/2003 (Fig. 2.18a). A baixa amplitude na série temporal de EVI2 neste período e a interpretação da composição colorida da imagem TM permitem chegar a essa interpretação. Durante os anos de 2003 a 2008 a área foi convertida para o uso agrícola, conforme pode ser observado nas imagens TM (Fig. 2.18b e 2.18c). A série EVI2 durante esse período apresenta as grandes amplitudes e ciclos curtos, mostrando mais de duas safras ao longo de um ano. No ano de 2008 ocorre a conversão para a cana-de-açúcar com um ciclo de 1 ano que pode ser confirmado na série temporal de EVI2. Analisando somente a imagem TM de 17/03/2009 não é possível concluir que se trata do plantio de cana-de-açúcar, nesse contexto a série temporal de EVI2 auxilia o fotointérprete a confirmação desse uso, pois a série temporal permite verificar a duração do ciclo. Esse caso é um exemplo de como as informações associadas às séries temporais podem auxiliar na interpretação de imagens de melhor resolução, mas que estão disponíveis para poucas datas ou em datas específicas.

Além de imagens TM, qualquer outra fonte de dados matriciais pode ser integrada ao laboratório virtual desde que esteja devidamente georreferenciada, considerando as questões de compatibilidade de escalas espaciais. Porém a integração de diferentes fontes de dados em diferentes escalas temporais e espaciais possibilita melhor contextualização da área e conseqüentemente o entendimento das séries temporais.

Trabalhos futuros envolvem a disponibilização de novos índices de vegetação, temperatura de superfície e a atualização periódica da base de imagens MODIS e TRMM, além do desenvolvimento de ferramentas de análise e classificação das séries MODIS.

O mapeamento de pixels com grande amplitude em um curto período de tempo é proposto por Freitas et al. (em preparação). Esse conceito poderá ser muito útil para identificação de áreas agrícolas bem como todos os pixels que possuam grande amplitude por um curto período de tempo. A Figura 2.19 mostra um exemplo desses pixels mapeados integrados ao globo virtual. Esse algoritmo tem a característica de ser interativo, permitindo ao usuário a alteração dos parâmetros de forma das séries, como amplitude, fase, posição do máximo valor entre outros parâmetros e visualizar no globo virtual, permitindo uma classificação preliminar dessas áreas. Na Figura 2.19c é possível observar os pontos em vermelho como pixels mapeados de grande amplitude por um curto período de tempo. As aplicações não se restringem a

Figura 2.18 - Composição colorida de Imagem Landsat-TM integrada ao globo virtual a) imagem TM de 17/11/2002 ; b) imagem TM de 18/02/2005 ; c) imagem TM de 04/09/2007; d) imagem TM de 17/03/2009; e) Série temporal do EVI2 da coordenada selecionada dentro do polígono de expansão de cana-de-açúcar mapeado pelo projeto Canasat. Os pontos em vermelho nas séries EVI2 correspondem às datas das imagens Landsat-TM.

agricultura. A identificação de macrofitas e outras vegetações naturais que possuem um ciclo fenológico curto são outro exemplo de aplicação. Nesta Figura 2.19, por exemplo, podemos observar áreas ao longo da planície rio amazonas com a vegetação que apresentam esse tipo de variabilidade temporal.

Além dos exemplos apresentados nessa seção, o Apêndice E contém exemplos ilustrativos de mudanças do uso e cobertura da Terra utilizando a função de visualização implementada em MatLab e os recursos da GoogleMaps e GoogleEarth API.

Os exemplos apresentados nessa seção demonstram o potencial de aplicações das séries temporais derivadas do laboratório virtual em ambientes virtuais de educação e computação colaborativa. Principalmente as baseadas em sistemas de informa-

a) SPAHAP (red colour) integrated at Virtual Laboratory of Remote Sensing Time Series ;
b) MODIS EVI2 time series illustrating the land use and land cover dynamics at one SPAHAP classified pixel.

Figura 2.19 - Identificação de pixels com grande amplitude por um curto período de tempo (Freitas et al., em preparação).

ção geográfica como os projetos do tipo: *Volunteered Geographic Information* (VGI) (Goodchild 2007), *Participatory Geographic Information Systems*(PGIS) (Rambaldi et al. 2006), e também abordagens *Citizen Science* [2] em que o foco seja monitoramento ambiental ou a resposta a questões científicas relacionadas, por exemplo, ao uso e cobertura da terra. Em contrapartida, as informações coletadas em campo por colaboradores como fotografias e descrições físicas da paisagem, auxiliariam em muito a interpretações das séries temporais e a elaboração de uma base de dados consistente e amplamente validada.

[2]O termo *Citizen Science* é traduzido livremente aqui como ciência cidadã. Uma revisão sobre os conceitos e aplicações na área de ecologia e geociências podem ser encontradas em Bonney et al. (2009) e Newman et al. (2011).

2.7 Propostas de novos algoritmos

Trabalhos futuros envolvem a elaboração de uma interface amigável para o intercâmbio e aplicação de novos algoritmos, hoje realizada exclusivamente em ambiente MatLab. A Tabela 2.1 descreve sucintamente os principais algoritmos com a principal aplicação a serem utilizados dentro do laboratório virtual de sensoriamento remoto.

O próximo capítulo descreve os aspectos computacionais das técnicas de Análise de Padrões Gradiente, DFA e Gráficos de Recorrência como integrantes do módulo de algoritmos do Laboratório Virtual de Sensoriamento Remoto.

Tabela 2.1 - Principais algoritmos e ferramentas para análise de séries temporais propostos para futuras implementações do Laboratório Virtual de Sensoriamento Remoto

Ferramentas e algoritmos de análise	**Aplicações**
Ferramentas de análise no domínio temporal e da frequência (Análise de Fourier, Transformadas Wavelets e Chirplets) (MALLAT, S., 1999; BOASHASH, B., 2003; LE PENNEC e MALLAT,2005)	Compreender a variabilidade de séries temporais com base em ferramentas de análise espectral e temporal
Análise de Padrões Gradientes (GPA) (ROSA et al., 2000; RAMOS et al., 2000; ASSIREU et al.,2002; FREITAS et al.,2010)	Análise não-linear de séries temporais com base em conceitos de assimetria e do campo gradiente
Detrended Fluctuation Analysis (DFA) (PENG, C.K. et al. 1994, KANTELHARDT J.W. et al. 2001)	Análise não-linear de séries temporais com base em conceitos de invariância de escala
Gráficos de Recorrência (ECKMAN et al. 1987; MARWAN et al. 2007ab.)	Análise não-linear de séries temporais com base em conceitos de sistemas dinâmicos
Ferramentas de Análise de Agrupamentos (DUDA e HART, 1973 e HARTIGAN 1985)	Análise Exploratória e classificação do conjunto de dados
Árvores de Decisão e Máquinas de Vetores de Suporte (SVM) (YUAN e SHAW , 1995;YANG, T., 2006, THEODORIDIS e KOUTROUMBAS , 2009)	Análise temporal e classificação do uso e cobertura da terra
Ferramentas de integração com outros formatos de arquivos típicos de Sistemas de Informação Geográfica (LONGLEY et al. 2005,LONGLEY,LONGLEY, P. 2008)	Ferramentas de suporte para análise suplementar de dados
Ferramentas de análise estatística por região (LONGLEY, P. 2008)	Análise espacial e estatística local e regional
Ferramentas para recorte geográfico e download dos dados e uso de metafile de metadados (BRETHERTON e SINGLEY, 1994)	Dados para auxiliar a interpretação da imagem e análise do uso e cobertura da terra
Ferramentas de extração da sazonalidade (JÖNSSON e EKLUNDH, 2004)	Calculo de métricas de sazonalidade com ciclos anuais, mensais, anomalias e ciclos fenológicos

3 ANÁLISE COMPUTACIONAL DE SÉRIES TEMPORAIS

It is hardly an exaggeration to say that we will introduce almost as many analysis algorithms as there are signals. . . signals are so rich and complex that a single analysis method. . . cannot serve them all.

(*Yves Meyer*)

A crescente disponibilidade de dados em diferentes formatos e especificações técnicas tornou necessária outras abordagens vinculadas aos tradicionais métodos de processamento e armazenamento de dados. Nesse contexto a computação aplicada surge como um pilar fundamental para a popularização, suporte, implementação, visualização e análise de um grande volumes de dados. A computação aplicada como elo integrador proporcionou o aprimoramento de várias metodologias. Como exemplo, é comum encontrar aplicação de métodos computacionais que surgiram em grupos de pesquisa de biologia genética, física, astronomia, meteorologia, geofísica e vice-versa.

A pesquisa desenvolvida nessa tese possui uma abordagem similar utilizando um conjunto de métodos que surgiram em grupos de pesquisa em sistemas complexos e física estatística computacional com o principal objetivo de analisar séries temporais e espaço-temporais de sensoriamento remoto. O conceito de operador computacional discutido nesse trabalho é o resultado de um método intrinsecamente computacional que sintetiza em um único valor a análise de um vetor unidimensional ou matrizes multidimensionais, seja nos domínios temporal, espacial, espectral ou qualquer outro domínio abstrato (ver Figura 3.1). Esse conceito está intimamente relacionado à redução da dimensionalidade de dados em que operações computacionais são necessárias para transformação de um conjunto de dados em uma informação. Em síntese, um operador computacional quando aplicado em um conjunto de dados gera um valor único que pode ser passível de interpretação.

Figura 3.1 - **1** e **2** fazem parte de uma ´´operação computacional", sobre o dado, uma vez que a técnica matemática está na forma de um algoritmo devidamente estruturado em um ambiente computacional (*software* e *hardware*).

Nessa pesquisa a definição de séries temporais é utilizada unicamente para representação unidimensional enquanto séries espaço-temporais representam a evolução de um padrão bi-dimensional no tempo conforme será descrito nas seções posteriores.

Esse capítulo contém três seções que descrevem os aspectos computacionais necessários para implementação do método de análise de flutuações *destendenciadas* [1] (DFA), do gráfico de recorrência e da técnica de análise de padrões de gradiente (GPA). A primeira seção contextualiza a análise de séries temporais e descreve o método DFA, juntamente com os aspectos computacionais da implementação para avaliações de séries temporais, espaço-temporais ou uma única imagem ou campo bidimensional. A segunda seção fornece detalhamento técnico sobre os operadores computacionais propostos para análise de padrões de gradiente, juntamente com uma discussão crítica sobre essa técnica. A terceira seção desse capítulo apresenta os operadores derivados dos *plotes* de recorrência como ferramenta de suporte para visualização e análise de tendências em séries temporais curtas.

Os exemplos de aplicações utilizados nesse capítulo estão restritos aos dados canônicos descritos no apêndice B. As aplicações em sensoriamento remoto das técnicas de análise são descritas no Capítulo 4 que apresenta exemplos de aplicações na caracterização da mudança do uso e cobertura da Terra.

3.1 O Método *Detrended Fluctuation Analysis* (DFA)

O método DFA foi proposto por Peng et al. (1994) e tem se estabelecido como uma importante ferramenta para detecção de correlações de longo alcance em séries temporais com não estacionaridade [2]. Originalmente o método DFA é um aperfeiçoamento do método *Fluctuation Analysis* (FA) (PENG et al., 1992), também baseado na teoria de caminhos aleatórios (*random walk theory*) (MANDELBROT, 1983; ADDISON, 1997). Inicialmente proposto para aplicações em análise seqüêncial de DNA, o método DFA tem sido aplicado, como um caracterizador de correlações de longo alcance em uma série temporal, em diversas áreas do conhecimento (ver Apêndice

[1] O termo *destendenciada* não existe no dicionário oficial de língua portuguesa, nesse sentido Veronese (2011) propõe a tradução de *não tendencial* para o termo *detrend*. Nesse trabalho adotamos a sigla em inglês DFA para referência ao método pois essa é a forma conhecida do método na literatura mesmo em língua portuguesa.

[2] O conceito de estacionaridade é definido de maneira diferente dependendo da área de aplicação. Na análise estatística de séries temporais, o conceito não é aplicado as séries temporais e sim aos modelos que descrevem as séries, existindo várias classes de estacionaridades. Nesse trabalho adotou-se o conceito de série temporal estacionária para séries em que os momentos estatísticos não apresentem tendência ao longo do tempo. De modo contrário é a definição de uma série não estacionária.

A).

3.1.1 Persistência e Correlações de Longo Alcance

A definição formal do termo série temporal é função da área do conhecimento em estudo. Diferentes definições são propostas, por exemplo, na estatística, matemática e engenharia eletrônica. Nesse trabalho defini-se série temporal-ST como um sinal discreto $X_{i,N}$ unidimensional e finito, de comprimento N que armazena os valores de uma variável X_i variando de $i = 1, ..., N$ equidistantes espaçados no tempo ou em outro domínio equivalente, sendo que $(i \in \mathbb{N}\ ,\ \{X_{i,N}\} \in \Re)$ com $\sum_N^{i=1}(|X_i|^2 < \infty)$, o que garante a "energia finita" e portanto a existência de uma faixa dinâmica da série temporal. Do mesmo modo uma série espaçotemporal-SET é definida neste trabalho como um conjunto de matrizes de amplitudes $\{A_{N,M,T}\}$ em que $(N, M, T \in \mathbb{N}\ ,\ \{A_{N,M,T}\} \in \Re,\ T > 1)$ com o $\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{t=1}^{T}|A_{i,j,t}|^2 < \infty$ para também garantir a existência de uma faixa dinâmica de flutuação na série espaço-temporal. Para série espaço-temporal N e M indicam respectivamente, o total de linha e coluna da matriz e T o total de matrizes da série espaço-temporal representando o domínio temporal.

Uma série temporal pode ser caracterizada pela combinação de qualquer um dos seguintes componentes: (i) tendência, (ii) um ou mais componentes periódicos, e; (iii)componentes estocásticos. A tendência é um componente a longo-prazo que incrementa ou decrementa os valores das séries temporais. Muitas séries também possuem componentes periódicas que podem ou não ser regulares, enquanto que a componente estocástica não inclui qualquer tendência ou periodicidade. Em geral, para quantificar a componente estocástica de uma série temporal é necessário especificar a distribuição estatística dos valores e a persistência ou anti-persistência (MALAMUD e TURCOTTE, 1999). A componente estocástica de uma série temporal é persistente se os valores adjacentes forem correlacionados positivamente e anti-persistente se os valores forem anti-correlacionados. A persistência também é conhecida como memória ou correlação interna de um processo, sendo às vezes reclassificada de acordo com sua intensidade, fraca ou forte, e também com sua duração de curto ou longo-alcance (MALAMUD e TURCOTTE, 1999; CHATFIELD, 2004). Em muitos casos a persistência (ou antipersistência) de uma série temporal pode ser quantificada utilizando a função de autocorrelação (F.A.C.)(Eq. 3.1).

O cálculo das correlações entre os valores de X em diferentes escalas s é definido pela função de auto-correlação, dada por:

$$C(s) = \frac{< \overline{X}_i \overline{X}_{i+s} >}{< \overline{X}_i^2 >} = \frac{1}{(N-s) < \overline{X}_i^2 >} \sum_{i=1}^{N-s} \overline{X}_i \overline{X}_{i+s}, \tag{3.1}$$

Onde, $< X >= \frac{1}{N} \sum_{i=1}^{N} X_i$ é a média e $\overline{X}_i \equiv X_i - < X >$ os desvios em relação a esta. Se o valor de X_i não é correlacionado, C(s) é zero para $s > 0$. Correlações de curto alcance de X_i são descritas por um C(s) decaindo exponencialmente até a escala que não existe mais correlação. Para a correlação de longo alcance, o valor de C(s) obedece a uma lei de potência do tipo:

$$C(s) \propto s^{-\gamma} \tag{3.2}$$

Onde vale o seguinte intervalo $0 < \gamma \leq 1$. O cálculo direto do valor de γ não é trivial devido a ruídos, tendências ou até mesmo sazonalidades presentes no sinal que se deseja estudar. Por exemplo, a definição de C(s) falha quando a série é não estacionária e possui correlações de longo alcance. A solução para esse tipo de problema é determinar o valor de γ indiretamente (KANTELHARDT et al., 2001). Nesse sentido o método DFA tornou-se uma abordagem alternativa para estimativa de persistência e correlações de longo alcance em séries com a presença de ruídos, tendências e sazonalidades. A descrição apresentada nessa tese está centrada nos aspectos computacionais do algoritmo DFA e não no formalismo matemático dos métodos de estimação das funções de autocorrelação, espectros de potência e métodos correlatos. O apêndice A apresenta uma visão geral sobre essas abordagens.

3.1.2 Algoritmo do DFA 1-D

A abordagem do DFA é fundamentada na análise de flutuações de uma ST em diferentes escalas. O método DFA pode ser aplicado diretamente em qualquer tipo de série temporal, independente da origem do sinal, seja ele contaminado ou não por ruídos e/ou possuir tendências (PENG et al., 1994; STANLEY et al., 1999; KANTELHARDT et al., 2001). O conceito básico do DFA é eliminar a tendência de um sinal em diferentes escalas, assumindo que são analisadas as flutuações intrínsecas do dado. Entenda-se como flutuações, a medida de variabilidade do sinal associada à variância de cada segmento da série, em diferentes escalas.

O algoritmo do DFA é composto por 6 passos apresentados no fluxograma da Figura 3.2 e descritos a seguir.

Figura 3.2 - Fluxograma do método DFA 1-D.

a) A partir de uma série temporal ou sinal, representado por X, onde (X_i) varie entre $i = 1, ..., N$ eqüidistantes espaçados no tempo ou no espaço, calcular a *função perfil*. Esta função é obtida pela soma cumulativa da série temporal. Nesse passo também ocorre a subtração da média da série temporal $<X>$ dada por:

$$Y(i) = \sum_{i=1}^{N} X_i - < X > \tag{3.3}$$

b) Subdividir o perfil Y(i) em N_s segmentos não sobrepostos de mesma escala s, em que

$$N_s \equiv [N/s] \tag{3.4}$$

c) calcular a tendência local de cada segmento (ν) através do ajuste polinomial de mínimos quadrados. Em cada segmento ν calcular a série sem

tendência através da diferença entre o perfil Y(i) e o polinômio ajustado $p_\nu(i)$. Em geral, usa-se o polinômio de grau linear (p=1), obtendo-se o DFA1 de primeira ordem, que retira tendências lineares da função perfil Y(i). Polinômios de maior grau, como por exemplo, quadráticos (p=2), podem ser utilizados nesse passo para retirar tendências de ordens maiores obtendo-se, para cada grau de polinômio p, um valor da função $Y_s(i)$.

$$Y_s(i) = Y(i) - p_\nu(i) \tag{3.5}$$

d) calcular a variância para cada segmento(ν) definida por:

$$F_s^2(\nu) = \left\langle Y_s^2(i) Y(i) \right\rangle = \frac{1}{s} \sum_{i=1}^{s} Y_s^2 \left[((\nu - 1)s + i) \right] \tag{3.6}$$

e) calcular a função de flutuação DFA utilizando o total de segmentos de cada escala s, conforme descrito a seguir:

$$F(s) = \left[\frac{1}{N_s} \sum_{\nu=1}^{N_s} F_s^2(\nu) \right]^{\frac{1}{2}} \tag{3.7}$$

f) Se a série original apresenta correlações de longo alcance, o valores de $F^{(n)}(s)$ seguem uma lei de potência, em que :

$$F^n(s) \propto s^\alpha \tag{3.8}$$

Calcular na escala *loglog* o valor de α em que

$$log(F^n(s)) = \alpha log(s) \tag{3.9}$$

Na prática isso permite o cálculo do expoente α através do ajuste linear em escala logarítmica de s em função de $F^{(n)}(s)$. Onde n é a ordem do polinômio utilizado para retirar a tendência da série. Para grandes valores de s vale a relação com expoente da autocorrelação γ, onde α=1-$\gamma/2$ para $0 < \gamma < 1$ (TAQQU et al., 1995). Essa relação foi estabelecida empiricamente associando os valores de γ de séries simuladas com o histograma de valores do coeficiente de α calculados pelo método DFA.

O expoente α também é conhecido como expoente de *Hurst H* (TAQQU et al., 1995, MALAMUD e TURCOTTE, 1999; ver apêndice A) que por sua vez está relacionado

com expoente β do espectro de potência [3] $\beta = 2\alpha - 1$. O expoente de flutuação pode ser classificado de acordo com uma faixa dinâmica de valores (KANTELHARDT et al. 2001; BASHAN et al. 2008; ZHENG et al., 2008):

$\alpha < \frac{1}{2}$: anti-correlacionado, sinal anti-persistente.
$\alpha \cong \frac{1}{2}$: sem correlação, ruído branco, não apresenta memória.
$\alpha > \frac{1}{2}$: apresenta correlações de longo alcance.
$\alpha \cong 1$: ruído tipo 1/f. [4]
$\alpha > 1$: não estacionário, como passeio aleatório.
$\alpha \cong \frac{3}{2}$: ruído tipo browniano[5]

A Figura 3.3 apresenta a seqüência de passos para aplicação do DFA na série sem correlação, ruído branco [6]. Neste exemplo são apresentadas as decomposições em 4 escalas diferentes utilizando o polinômio linear.

3.1.3 Experimentos Numéricos

Para uma simples avaliação do desempenho do DFA em séries curtas é realizado um experimento numérico que mostra a variabilidade de α em função do comprimento da série. Nesse experimento gerou-se 1000 séries pseudo-aleatórias para cada comprimento variando entre 2^3 e 2^{12}. As séries em teoria apresentam o valor de $\alpha =$ 0,5, ou seja, representam um ruído branco. A Figura 3.4 apresenta o *boxplot* dos 1000 valores de α em função do comprimento da série. Os resultados apresentados concordam com Bashan et al. (2008), indicando que a variabilidade de α em séries muito curtas (N $<$ 100) é maior. Isso mostra que o tratamento estatístico de séries muito curtas ainda é um problema em aberto para o DFA, embora o método apresente uma estimativa bem mais robusta em relação a técnicas como o espectro de potência como mostrado por Veronese et al. (2011). Uma avaliação mais detalhada

[3] O teorema de Wiener-Khinchin define a relação direta entre o espectro de potência e F.A.C., dado por $S(f) = \int_{-\infty}^{\infty} C(s)\, e^{-2\,\pi\, i\, f}\; d\tau =$, S(f) é normalmente representada com unidade de potência por unidade de frequência, por exemplo Watts por Hertz (W/Hz) o que explica o nome densidade espectral de potência dado em inglês é conhecido com *Power Spectral Density* (PSD)(MALAMUD e TURCOTTE, 1999), onde S(f) é a chamada função de densidade espectral. Mais detalhes sobre esse teorema e informações de outros métodos de estimativa da densidade espectral podem ser obtidos em Koopmans (1995).

[4] Também conhecido como ruído rosa, é amplamente encontrado na natureza, nesse tipo de ruído a distribuição da magnitude do sinal é inversamente proporcional à freqüência (Malamud e Turcotte, 1999).

[5] Exemplo clássico de um processo difusivo e amplamente encontrado na natureza é a de um integração do ruído branco, também conhecido em processos estocásticos como processo de Wiener (MALAMUD e TURCOTTE, 1999).

[6] Ou ruído gaussiano, tem a distribuição uniforme de frequência no espectro de potência.

Figura 3.3 - Sequência de passos para aplicação do DFA em uma série estacionária tipo rúido branco.

da influência do comprimento da série no valor de α é descrita por Coronado e Carpena (2005). Esses autores concluiram que o DFA é o método mais robusto quando comparado com a análise de correlação e a análise R/S de Hurst.

Com o objetivo de avaliar a sensibilidade do DFA à presença de correlações de longo alcance foi realizado outro experimento numérico. Utilizou-se uma série sem memória, ruído branco (média 0,5 e desvio padrão 1) com 1024 pontos concatenados com uma série normalizada de um perfil topográfico derivado de um imageamento realizado com o RADAR da missão SRTM[7] , representando a variabilidade da altimetria ao longo de um perfil de latitude localizado numa região de transição entre cerrado e floresta no estado de Mato Grosso. Basicamente é um exemplo típico de passeio aleatório (*random walk*), comumente observado nos perfis de relevo existentes na natureza (ADDISON, 1997; TURCOTTE, 1997).

A sensibilidade à mudança da dinâmica da variabilidade é realizada variando a posição inicial da série, ou seja, para cada ponto percorrido em x montou-se uma

[7] *Shuttle Radar Topography Mission* - SRTM, Missão liderada pela NASA e o DLR para mapeamento do relevo do planeta usando RADAR interferométrico a bordo da espaçonave *space shuttle*.

Figura 3.4 - Boxplots dos 1000 valores de α em função do comprimento da série.

série de 1024 pontos a partir desse ponto inicial. Desta forma, pode-se verificar a variabilidade de α entre pontos vizinhos além do comportamento dos valores de α com a inclusão de pontos da série com memória de longo alcance. O eixo x da Figura 3.5b representa o ponto inicial de cada série montada a partir da posição x da série Y(i) (Figura 3.5a). O eixo y da esquerda apresenta o percentual utilizado de cada série construída para o cálculo de α e o eixo y da direita mostra os valores de α. É interessante ressaltar que o comportamento dos valores de α modifica abruptamente com incremento de 10 a 20% da série não estacionária (azul) indicando a presença de correlações de longo alcance. A partir do incremento de 50% da série de memória de longo alcance os valores de α tornan-se assintóticos e oscilam entre $1,22$ e $1,28$, caracterizando o passeio aleatório típico de relevos. Quanto às condições iniciais, nesse experimento não se verifica a presença de saltos na série dos valores de α, isso indica a sensibilidade do método a presença de correlações de longo alcance.

A Figura 3.6 mostra os valores de escala s *versus* a função de flutuação F(S) em diferentes escalas. Para ambos os exemplos, ruído branco (Fig. 3.5, série vermelha) e perfil de amplitudes(Fig. 3.5, série azul), o procedimento de excluir a tendência da série é calculado utilizando um polinômio linear. É interessante ressaltar que o

Figura 3.5 - a) Concatenação de duas séries, sem memória (linha em vermelho) e com memória (linha azul). b) Valores de α para cada série de 1024 pontos construída a partir da posição x e o percentual utilizado das séries com memória (linha azul) e sem memória (linha em vermelho).

método DFA pode ser aplicado em outras escalas e com polinômios de diferentes graus retirando assim tendências de ordens maiores. Esse procedimento é realizado no algoritmo alterando a ordem do polinômio p que é ajustado para o procedimento da remoção da tendência.

O apêndice A contém uma revisão das principais aplicações do DFA 1-D nas comunidades de meteorologia, oceanografia, ecologia e geofísica. Um exemplo de aplicação do DFA-1D para estudo de campo de anisotropia de imagens é descrito por Freitas et al. (2010) e está disponível no anexo A. Essa técnica é apenas uma extensão do DFA-1D aplicado *à* imagens. Os aspectos computacionais para aplicação do DFA em matrizes 2-D são descritos na próxima seção.

3.1.4 Método DFA-2D

Do ponto de vista da coleta e tratamento dos dados a abordagem unidimensional é mais utilizada. Investigar as correlações de longo alcance no domínio maior que

Figura 3.6 - Valores de escala s *versus* a função de flutuação F(S) para série estacionária (a) e série não estacionária (b), utilizando polinômio linear para remoção da tendência.

o unidimensional não é uma tarefa trivial, uma vez que geralmente as correlações em um campo maior que o 1-D podem ser heterogêneas ao longo de determinadas direções. A capacidade de analisar imagens e campos escalares ou vetoriais tridimensionais pode auxiliar, por exemplo, na identificação dos tipos de modelos para descrever a variabilidade dos dados. Porém, técnicas robustas, computacionalmente de fácil implementação e que descrevam as relações de dependência entre escalas em diferentes dimensões ainda estão em desenvolvimento. Segundo Gu e Zhou (2006), a transformada *wavelets* é um exemplo típico de uma técnica robusta, porém, com dificuldades na implementação computacional e na escolha da melhor parametrização.

No contexto de imagens, a análise de correlações está diretamente ligada à textura. Esta é relacionada à forma da estrutura do campo bi-dimensional e a sua caracterização pode ser realizada por meio da análise de rugosidade (KAPLAN, 1999). Na área de reconhecimento de padrões e processamento de imagens a análise de textura é fundamental para entender, caracterizar e classificar uma imagem, tendo várias aplicações na área de sensoriamento remoto e medicina (KAPLAN, 1999). Segundo esse autor as texturas naturais como nuvens, modelos de terrenos, fumaça entre outros podem ser caracterizadas pelo expoente de Hurst. Este expoente controla a rugosidade visual de um processo em todas as escalas. Um exemplo típico desse processo é o movimento fracionário Browniano conhecido como *fractional Brownian motion*(fBm) popularizado por Mandelbrot (1983).

Ressalta-se que existe na literatura outras abordagens tradicionais de processamento de imagens para descritores de textura (HARALICK et al., 1973; DUTRA, 1990;

SCHOWENGERD, 1997; MATHER, 1999) e as abordagens de correlações espaciais clássicas (Moran, 1950; Geary, 1954).

3.1.5 Algoritmo do DFA 2-D

A metodologia básica para generalização do DFA multidimensional foi inicialmente proposta por Gu e Zhou (2006). Nesse trabalho apresenta-se uma versão de implementação para a estimativa de α em matrizes bidimensionais, bem como a inclusão de um intervalo de confiança para os valores de α.

Dada uma imagem ou campo bi-dimensional A de dimensão *(MxN)* em que $i = 1, 2, 3, ..., M$ e $j = 1, 2, 3, ..., N$, tem-se os seguintes passos para a estimativa de α:

a) Particionar a matriz A em $N_s x M_s$ segmentos quadráticos de mesmo tamanho s x s, onde $M_s = [M/s]$ e $N_s = [N/s]$. Cada segmento fica definido por $A_{\nu,w}(i,j) = A(l_1 + i, l_2 + j)$ para $1 <= i, j <= s$, onde $l_1 = (\nu - 1)s$ e $l_2 = (w - 1)s$

b) Para cada segmento $A_{\nu,w}$, identificado por ν e w, a soma cumulativa é calculada por

$$U_{\nu,w}(i,j) = \sum_{k_1=1}^{i} \sum_{k_2=1}^{j} A_{\nu,w}(k_1, k_2); \tag{3.10}$$

c) Uma superfície de tendência é calculada utilizando uma função polinomial bivariada dada no caso linear, representado por um plano,

$$\widetilde{U}_{\nu,w}(i,j) = ai + bj + c; \tag{3.11}$$

onde $1 >= i, j <= s$ e *a,b,c* sendo esses parâmetros livres a serem estimados pelo método de mínimos quadrados, utilizando por exemplo, operadores aritméticos para decomposição ortogonal triangular(QR) que são numericamente mais estáveis (ANDERSON et al., 1999). Esses algoritmos estão disponíveis na biblioteca Linear Algebra PACKage (LAPACK) (http://www.netlib.org/lapack/). Além dessa função (3.11) outras superfícies podem ser utilizadas para a estimativa da superfície de tendência, como superfícies de ordem quadrática. No algoritmo implementado está também disponível uma função quadrática tipo $U_{\nu,w}(i,j) = ai^2 + bj^2 + c$. Todos os testes realizados nessa tese utilizaram a função linear (plano)

para remoção da tendência como recomendado por Gu e Zhou (2006).

d) O próximo passo é obter a matriz de resíduos do ajuste da superfície.

$$\epsilon_{\nu,w}(i,j) = U_{\nu,w} - \widetilde{U}_{\nu,w}(i,j); \tag{3.12}$$

e) A função de flutuação da superfície sem tendência é $F(\nu, w, s)$ para cada segmento $A_{\nu,w}$ é definida como a variância amostral da matriz de resíduos $\epsilon\nu, w(i,j)$ dada por:

$$F^2(\nu, w, s) = \frac{1}{s^2} \sum_{i=1}^{s} \sum_{j=1}^{s} \epsilon_{\nu,w}(i,j)^2 \tag{3.13}$$

Nesse caso a média dos resíduos é zero devido ao procedimento de remoção da tendência.

f) A função de flutuação para cada escala s é dada pela média da flutuação de todos segmentos dada por:

$$F^2(s) = \frac{1}{M_s N_s} \sum_{\nu=1}^{M_s} \sum_{w=1}^{N_s} F^2(\nu, w, s) \tag{3.14}$$

g) Variando o valor de s numa faixa de $s_{min} = 6$ até $s_{max} = min(M, N)/4$ pode-se determinar uma lei de escala entre os valores de s e a função de flutuação *F(s)* dada por:

$$F^2(s) \approx s^{\alpha} \tag{3.15}$$

A escolha de s_{min} e s_{max} é determinada empiricamente. Desde que M ou N não sejam múltiplos da escala s, dois perfis da superfície podem restar em cada direção. O procedimento computacional adotado para esse caso é iniciar a partição dos segmentos opostos tendo como origem a posição da matriz *i=M*, *j=M*, aumentando o número de segmentos s analisados e utilizando todos os sub-segmentos da matriz A.

Na escala *log-log* é possível estimar o valor de α usando a relação linear entre os valores de escala **s** e a função *F(s)*. O procedimento proposto para estimativa dos valores de α utilizou o método de regressão linear robusta, que consiste na minimização da soma quadrada dos resíduos tirando os pesos de *outliers* e usando pesos bi-quadrados (STREET et al., 1988; DUMOUCHEL e O'BRIEN, 1989). Os algoritmos de regressão linear robusta tendem a minimizar os efeitos de *outliers* na

estimativa do coeficiente angular. Essa regressão utiliza o algoritmo interativo de mínimos quadrados ponderado (*iteratively reweighted least squares*), em que o peso de cada iteração é ajustado com uma função biquadrada da iteração anterior. O valor de *p-value* é utilizado como um parâmetro de qualidade para cada estimativa de α, ou seja, para cada estimativa de α tem-se associado a estatística t para cada reta. O algoritmo para o cálculo da regressão robusta utiliza a biblioteca de ajuste de curvas do software MatLab que possui o algoritmo proposto por DuMouchel e O'Brien (1989) já implementado.

O método de *bootstrap* é utilizado para estimativa do intervalo de confiança de α e valor de significância para cada estimativa de α. Esse método consiste na estimação para vários valores de α a partir da simulação de subconjuntos de s e *F(s)*. Esse procedimento é comumente utilizado na estatística de poucas amostras e na estimativa de intervalos de confiança (EFRON 1982; EFRON e TIBSHIRANI 1993). A técnica *bootstrap* não assume a normalidade dos dados e pode ser aplicada em pequenas amostras. As amostras para o cálculo da regressão são escolhidas de forma aleatória e as estimativas recalculadas. A simulação com 1000 iterações permite a elaboração de histogramas dos valores de média, variância, mediana, assimetria e curtose, permitindo estabelecer o intervalo de confiança para as estimativas da regressão.

A implementação computacional do DFA 2-D foi realizada em ambiente MatLab e será disponibilizada dentro de uma *Toolbox* juntamente com as outras funções descritas nesse trabalho.

3.1.6 Resultados da Implementação do DFA-2D

A verificação da robustez do método DFA-2D para caracterização de padrões espaciais e espaço-temporais distintos é realizada utilizando dados canônicos[8] que estão descritos no apêndice B. Esses modelos são implementados com objetivo de gerar padrões espaço-temporais tipicamente utilizados nos estudos em física-estatística computacional. Ressalta-se que o objetivo desse trabalho não se concentrou nos modelos, mas nos distintos padrões espaço-temporais gerados e como os operadores computacionais caracterizam esses padrões. Com os 3 modelos implementados e descritos no Apêndice B podem-se gerar séries temporais e espaço-temporais de distintos padrões de variabilidade como ruídos, periodicidade, processos caos e turbulência.

[8]Séries temporais e espaço-temporais geradas a partir de modelos simplificados de sistemas físicos, são consideradas dados canônicos, representando os mais diversos padrões de variabilidade em sistemas complexos.

A Figura 3.7 apresenta a estimativa dos valores de α utilizando o algoritmo implementado. A Figura 3.7a apresenta a distribuição dos valores de α após a realização do método de *bootstrap*. Na Figura 3.7b a estimativa do intervalo de confiança a 95% α bem como a representação do boxplot da distribuição de α. Na Figura 3.7c é apresentada a representação da imagem a partir da qual o α é calculado, nesse caso uma superfície isotrópica de um ruído branco. A Figura 3.7d disponibiliza a estimativa de α utilizando o ajuste linear na escala log-log das escalas (s) a partir da função de flutuação F(s). É interessante ressaltar a coerência entre a estimativa de α e os valores observados na imagem os quais indicam através dos valores de α que a imagem não possui correlação.

Figura 3.7 - Exemplo de cálculo do expoente α para matrizes: a) distribuição dos valores de α após a realização do método de *bootstrap*; b) *boxplot* dos valores de α com intervalo de 95% da mediana; c)Imagem a partir da qual o expoente α foi calculado, nesse caso um ruído branco representando uma superfície isotrópica e d)gráfico log-log das escalas (s) a partir da função de flutuação F(s).

No exemplo da Figura 3.8 tem-se as mesmas representações, voltadas para a solução da equação complexa de Ginzburg-Landau [9] para o caso 2-D, sendo considerado um exemplo de solução de um processo turbulento. Nessa imagem deve existir correlações de longo alcance típico de um processo turbulento, indicando a presença de distintas e desordenadas estruturas espaciais cujas flutuações em diferentes escalas estão em interação. A Figura 3.9 é um exemplo de uma estrutura periódica formada pela sobreposição de matrizes de senos horizontais e verticais gerado pelo modelo de

[9] A Equação de Ginzburg-Landau é utilizada em física-estatística para descrever transições de fase entre diferentes padrões de variabilidade

superfícies periódicas (apêndice B).

Figura 3.8 - Exemplo de cálculo do expoente α para matrizes: a) distribuição dos valores de α após a realização do método de *bootstrap*; b) *boxplot* dos valores de α com intervalo de 95% da mediana; c)Imagem a partir da qual o expoente α foi calculado, nesse caso um solução turbulenta da Equação de Ginzburg-Landau e d)gráfico log-log das escalas (s) a partir da função de flutuação F(s).

Os valores de média e variância são normalizados para 0 e 1. Nesses exemplos é possível observar a separação entre as imagens simuladas. Para o ruído branco o valor de α é de aproximadamente $0,5$, no caso do exemplo de solução de um processo turbulento gerado pela equação de Ginzburg-Landau o valor é $1,38$ e para os modelos de senos concatenados esse valor é muito mais alto $\alpha = 3,49$. Esse valor bem maior que $1,5$ indica que a série é determinística, não tendo componentes estocásticos.

No resultado a seguir é apresentado um teste do DFA para detectar mudança entre regimes de variabilidade espaço-temporal. Na simulação da grade acoplada a partir de uma condição inicial o sistema evolui fragmentando-se e chegando a um regime intermitente. Essa transição pode estar relacionada não somente aos valores de amplitude da matriz, mas também aos padrões e geometrias presentes na condição inicial. A Figura 3.10 apresenta a seqüência de imagens geradas a partir do modelo do mapa de Kaneko [10]. Cada imagem simulada possui a dimensão de 128x128 *pi-*

[10]Modelo tradicionalmente utilizado em sistemas dinâmicos para representação de sistemas não-lineares em sistemas espacialmente estendidos. Foi inicialmente proposto pelo físico Kunihiko Ka-

Figura 3.9 - Exemplo de cálculo do expoente α para matrizes: a) distribuição dos valores de α após a realização do método de *bootstrap*; b) *boxplot* dos valores de α com intervalo de 95% da mediana; c)Imagem a partir da qual o expoente α foi calculado, nesse caso um sobreposição de superfícies periódicas e d)gráfico log-log das escalas (s) a partir da função de flutuação F(s).

xels com valores arbitrários de amplitude baseados na condição inicial estabelecida através de uma imagem índice de vegetação adquirida na região amazônica. Essa imagem é escolhida por representar um padrão não isotrópico na distribuição de amplitudes, sendo e o único objetivo ilustrar a variabilidade dos valores de α em cada passo da simulação.

Figura 3.10 - Solução intermitente do mapa de Kaneko, cada imagem representa um passo da simulação, valores de amplitude representam a evolução de uma matriz escalar de índice de vegetação dado pela condição inicial.

neko no início dos anos 80 nos estudos de caos espaço-temporal. Em português também é conhecido como Grades de Mapas Acoplados - (*Coupled Map Lattices*) (CML).

A Figura 3.11 apresenta a evolução da grade de mapas acoplados (mapa de Kaneko) com fator de acoplamento $0,3$ e $\alpha_{kaneko} = 1.852$, produzindo uma solução intermitente no mapa. A partir da condição inicial estabelecida pela imagem de índice de vegetação há uma fragmentação na imagem com o passar dos passos da simulação. Isso representa uma transição da variabilidade espacial entre a estrutura inicial e a final da grade de mapas acoplados. Os valores de α estimados pelo DFA conseguem detectar melhor a transição para um regime intermitente a partir da décima imagem. Os momentos estatísticos como média e variância (linhas em vermelho e verde) não ressaltam bem essa transição uma vez que são medidas baseadas somente nos valores de amplitude da imagem. A descrição dos modelos simplificados para geração dos dados canônicos encontra-se no Apêndice B.

Figura 3.11 - Valores do expoente α calculados para cada imagem do modelo simulado.

Com o objetivo de verificar a relação entre o tamanho da matriz e o tempo de execução do algoritmo implementado foi realizado um simples experimento dimensionado com 1000 cálculos de valores de α para cada dimensão da matriz. Os valores utilizados consideraram a simulação de superfície isotrópica (ruído branco) com média 0 e variância 1. A Figura 3.12 apresenta o tempo médio para o cálculo dos valores de α em função da dimensão da matriz. O teste descrito foi executado em um computador pessoal com a seguinte configuração: processador Intel Core i7, HD de 2 TB e 8 GB de memória RAM. Este teste simplificado é uma estimativa do tempo computacional exigido para o processamento de matrizes com maior dimensão em termos de linhas e colunas de pixels [512x512 pixels], sendo este elevado cerca de 7 segundos, 14 vezes mais lento que para as matrizes de 128x128 elementos. Esse tempo deve ser considerado principalmente quando o objetivo consiste em analisar

uma seqüência de várias imagens no tempo ou várias partes de uma imagem de elevada dimensão, considerando a possibilidade de disponibilizar a análise em um ambiente virtual como proposto por (Dantas et al., 2011 e Freitas et al., 2011).

Figura 3.12 - Tempo médio para o cálculo dos valores de α em função da dimensão da matriz.

Um aspecto muito importante do método DFA é sua parametrização. Ressalta-se que o método deve ser utilizado em um conjunto de dados com os mesmos parâmetros uma vez que qualquer alteração dos parâmetros de escala bem como função de remoção da tendência naturalmente influenciará os valores estimados de α. Maraun et al.(2004) discute o uso do DFA em termos de especificidade e sensibilidade, um algoritmo é especifico se ele tem a capacidade de negar uma certa propriedade, no caso do DFA a presença de correlação de longo alcance e é sensível se ele consegue detectar essa propriedade. O algoritmo ideal deve ter as duas propriedades a sensibilidade e a especificidade. Embora esses autores concluam que o DFA não é um método sensível, isso não invalida o uso do DFA para alguns tipos de aplicação. Um exemplo típico é o caso dessa tese em que os valores de α são utilizados de forma comparativa e como um operador computacional, e não necessariamente para afirmar unicamente a presença de correlações de longo alcance em uma única série. A variabilidade das estimativas do expoente de escala (α) é discutida por Franzke et al. (2012). Esses autores comparam 5 métodos de estimativa de expoente de escala com diferentes modelos sintéticos geradores de correlações de longo alcance. Os resultados mostram uma discordância entre os métodos e que existe uma sensibilidade de acordo com o tipo de correlação de longo alcance gerada. Isso demonstra que a estimativa de expoentes de escala com um método robusto e universal ainda é um problema em aberto.

A técnica descrita na próxima seção é uma alternativa *à* análise estatística de padrões espaciais e espaço-temporais. Essa técnica baseia-se na análise geométrica dos mesmos caracterizando a complexidade do padrão espacial a partir do grau de assimetrias bilateral, calculado a partir do campo gradiente de uma imagem (matriz escalar).

3.2 Análise de Padrões de Gradiente

A Análise de Padrões de Gradiente - *Gradient Pattern Analysis* (GPA) permite caracterizar padrões de variabilidade em séries curtas, com ou sem estacionaridade, observadas ou simuladas nos domínios temporal e espaço-temporal. Esta técnica tem sido utilizada com sucesso na caracterização de regimes complexos, como: (i) turbulência em plasmas (ROSA et al., 1999); (ii) análise de séries temporais aplicadas à oceanografia (ASSIREU et al., 2002); (iii) caracterização de padrões estruturais em nano ciência (BRITO, 2004; BARONI et al., 2006); (iv) classificação de flutuações de densidade na simulação de estruturas cosmológicas (ANDRADE et al., 2006); (v) detecção e caracterização de desflorestamento na Amazônia, utilizando séries espaço temporais de imagens de sensoriamento remoto (FREITAS, 2007), e; (vi) caracterização dinâmica em física solar (ROSA et al., 2008) e (vii) análise de séries limnológicas (VALÉRIO et al. 2012).

O GPA é um método baseado na geometria convexa e utiliza o campo gradiente de uma matriz para caracterizar padrões complexos possivelmente relacionados *à* processos dinâmicos subjacentes, como por exemplo, reação-difusão, caos espaço-temporal e turbulência. A técnica baseia-se na aplicação de operadores computacionais que quantificam quebras de simetria e desordem de fase em escalas locais e globais de um padrão espaço-temporal qualquer (ROSA et al., 1999; ROSA et al., 2000; ASSIREU et al., 2002; COSTA JÚNIOR, 2004).

Matematicamente o GPA pode ser representado por quatro medidas (chamadas de momentos gradiente) extraídas a partir de um campo escalar ϵ_{ij}. Este campo escalar representa a matriz de amplitudes de uma dada variável a partir da qual é produzido o campo vetorial gradiente $G = \nabla\epsilon_{i,j}$, composto por V vetores $\mathbf{r}$, onde cada vetor $r_{i,j}$, localizado na posição (i,j) do campo gradiente é caracterizado por sua norma e fase $(r_{i,j}, \phi_{i,j})$. Este campo escalar pode então ser descrito pela composição de quatro momentos gradientes: (i) momento gradiente de primeira ordem (G_1), medida global da distribuição de todos os vetores, ou seja, de todas as normas e suas respectivas fases localizadas no campo gradiente; (ii) momento gradiente de segunda ordem (G_2), estimador da desordem existente no módulo dos vetores; (iii)

momento gradiente de terceira ordem (G_3), estimador da desordem existente na fase dos vetores, e; (iv) quarto momento (G_4), medida global da norma dos vetores e da fase dos vetores (ROSA et al., 1999). O quarto momento gradiente G_4 é uma representação complexa do módulo, $|G_4|$, e da fase, $\phi G4$, de um campo gradiente (ROSA et al., 2003).

A lógica da definição de momentos gradiente está ilustrada na Figura 3.13. Em síntese, valores de G_1 conseguem capturar pequenas flutuações não lineares de um sinal, já o $|G_4|$ é uma medida global de todo o campo gradiente e permite quantificar variabilidade global de todo o sinal (RAMOS et al., 2000; ROSA et al., 2003). Esse trabalho está restrito a definição de operadores computacionais derivados de primeiro, segundo e terceiro momento do gradiente. O operadores computacionais propostos por Ramos et al. (2000) baseados no conceito de entropia de Shannon como o *Complex Entropic Form* (*CEF*) e *Generalized Complex Entropic Form* (GEF) serão explorados em trabalhos futuros.

Figura 3.13 - Síntese da análise padrões gradiente pesquisado nessa tese.
Fonte: Adaptada de Rosa et al. (2003).

A medida de quebras de simetrias do campo gradiente relacionada às flutuações locais das normas e fases de cada vetor pode ser obtida por meio do operador computa-

cional de Fragmentação Assimétrica A (ROSA et al., 1999), atualmente denominado coeficiente de assimetria gradiente G_{1A} (ROSA et al., 2003). Padrões bidimensionais sem fragmentação ou padrões fragmentados com simetria total possuem G_{1A} nulo, ou seja, sem assimetria na distribuição das flutuações. Esta operação sobre uma matriz gera uma medida da assimetria global do padrão espacial, quantificando no tempo as possíveis quebras de simetria. O coeficiente G_{1A} (equação (3.16)) é espacialmente caracterizado pela não-linearidade das flutuações e suas respectivas conexões.

O coeficiente de assimetria do gradiente é definido a partir do primeiro momento do gradiente e definido como:

$$G_{1A} = (Nc - Nv)/Nv \qquad (Nc \geq Nv > 0) \tag{3.16}$$

Em que Nv é o número de vetores assimétricos e Nc é o número de conectividades gerado pela triangulação de *Delaunay* do campo gradiente assimétrico, tomando um ponto local de cada vetor assimétrico como vértice. Neste contexto, a triangulação de Delaunay função de Nc e Nv é um campo fracionário com dimensão menor que a dimensão da grade (igual a 2). Quando não há nenhuma correlação assimétrica no padrão, o número total de vetores assimétricos é zero, e então, por definição $G_{1A} = 0$. Para uma determinada grade de tamanho lxl o padrão totalmente desordenado gera o mais alto valor para G_{1A}, padrões convectivos compostos por estruturas em forma de vórtices e labirintos apresentam valores específicos não-nulos para G_{1A} (ROSA et al., 1999).

O que a triangulação de Delaunay faz é maximizar o menor ângulo de todos os triângulos, esse procedimento busca evitar triângulos com ângulos internos muito pequenos (Ver Figura 3.14). Essa triangulação foi estabelecida pelo matemático russo Boris Delaunay em 1934 e é amplamente utilizada nas mais diversas aplicações, como processos de visualização 3-D, teoria dos grafos e análise espacial de dados. Maiores detalhes sobre os métodos computacionais de triangulação de Delaunay podem ser encontrados em Okabe et al. (1992) e em Berg (2008).

A Figura 3.15 apresenta dois padrões gradientes com o mesmo número de flutuações, porém com diferente energia de correlação geométrica, o que gera valores diferentes para G_{1A}. Na Figura 3.16 é apresentado o contorno e a superfície de uma modulação da função SINC 2D [11] sobre a influência de ruídos localizados. Os valores de G_{1A}

[11] Dado por $SINC2D(i,j) = (sen(\sqrt{X_{i,j}^2 + Y_{i,j}^2}))/(\sqrt{X_{i,j}^2 + Y_{i,j}^2}))$ em que X e Y são matrizes de posição ao longo da direção x e y e centradas em 0.

Figura 3.14 - Exemplo de triângulação de Delaunay.

aumentam a medida que o ruído aumenta devido *às* quebras de simetria em relação a condição inicial.

Figura 3.15 - Exemplos da Triângulação de Delaunay do campo gradiente. a) Gradiente com pouca assimetria; b) Triângulação de Delaunay correspondente a; c)Gradiente de uma matriz aleatória ; D)Triângulação de Delaunay correspondente c.

Figura 3.16 - Contorno e superfície de uma função SINC 2D sobre influência de ruídos localizados. A medida que as assimetrias bilaterais aumentam os valores de G_1A aumentam.

3.2.1 Novos Aspectos computacionais do cálculo de G_{1A}

A primeira consideração computacional a ser mencionada é a não obrigatoriedade de realizar a triangulação. A existência de uma relação entre o envelope convexo e o número de conexões da triangulação de Delaunay dispensa a necessidade de realizar a triangulação. Com isso o algoritmo para o cálculo de G_{1A} torna-se mais simples e computacionalmente mais rápido. Isso de certa forma invalida o uso do G_{1A} para matrizes com a mesma borda e núcleo ou centro com valores assimétricos, ou seja, conhecendo apenas os pares da borda da matriz é possível estimar o número de conexões em todo o interior do campo de triangulação. Essa relação é dada calculando o envoltório convexo (*Convex Hull*) que no caso do campo gradiente representa as arestas de conexão entre todos os vértices externos do campo de triangulação.

Esse teorema estabelece que sendo P um conjunto de n pontos no plano, não colineares, e k_e o número de pontos do envoltório convexo de P, qualquer triangulação de P tem $2n - 2 - k_e$ triângulos e $3n - 3 - k_e$ arestas. A prova desse teorema vem da característica de Euler[12] que é dada por:

[12]Invariante topológica definida por Leonhard Euler em seu trabalho *"Demonstratio nonnullarum insignium proprieatatum, quibus solida hedris planis inclusa sunt praedita"* apresentada em 1752 e publicada em 1758 traduzido por "propriedade geral dos sólidos limitados por faces planas"

$$n - n_e + n_f = 2, \tag{3.17}$$

em que, n é o numero de pontos da triangulação, n_e o número de arestas e n_f o número de faces. Para o caso da triangulação n_f é sempre igual o número de triângulos n_t mais *1*. Ou seja, $n_f = n_t + 1$, pois considera-se a face de cada triangulo mais a face do plano de triangulação. Como cada aresta é incidente exatamente em duas faces, o número de arestas é $n_e = (3n_t + k_e)/2$. Usando a fórmula da característica de Euler (Eq. 3.17) e substituindo os valores de n_e e n_f tem-se,

$$n_t = 2n - k_e - 2, \tag{3.18}$$

o que implica no número de arestas dependente do número total de pontos n e dos pontos do envoltório convexo k_e sendo dado por:

$$n_e = 3n - k_e - 3, \tag{3.19}$$

Como o número de arestas é utilizado no cálculo de G_{1A}, temos que $N_c = n_e$, tornando N_c apenas função do número de vértices do envoltório convexo k_e, uma vez que $n = N_v$ (Eq. 3.16).

Pela característica de Euler, computacionalmente o cálculo do valor de G_{1A} torna-se menos custoso uma vez que não há a necessidade de triangulação completa. A visualização da triangulação torna-se computacionalmente cara e às vezes inviável principalmente no processo de visualização em matrizes de elevada dimensão. O cálculo de G_{1A} é sensível ao envoltório convexo e torna-se um problema clássico de busca, ou seja, é necessário remover pares simétricos para completar a triangulação. Qualquer algoritmo proposto tem que estabelecer alguns critérios para definir o que são pares simétricos e como eliminá-los. Nesse contexto, o algoritmo proposto nesse trabalho segue a seqüência de passos indicada a seguir:

a) Definição de tolerância de norma (T_r) e fase (T_ϕ) para definição do que são pares simétricos;

(MARTINES, 2009). Inicialmente essa propriedade foi publicada em uma carta direcionada ao seu amigo Goldbach em 1750. Outros detalhes históricos e a demonstração completa dessa característica é descrita por Martines (2009).

b) Escolha da lista de pares de vetores de norma e fase a serem comparados;

c) Calculo da distância euclidiana entre os pares de vetores, considerando a posição de cada par de vetores na matriz ;

d) Escolha do critério de distância (Dp) para eliminação de pares simétricos;

e) Eliminação dos pares de vetores considerados simétricos;

f) Cálculo do envoltório convexo da matriz com os pares eliminados;

g) Cálculo do valor de $G1_A$ considerando o número de conexões e o número de vetores assimétricos.

Por definição é assumido que o ponto de triangulação é a ponta do vetor e valores nulos são considerados pontos da triangulação. O critério de distância Dp é necessário caso ocorra a simetria entre um número ímpar de vetores simétricos. Esse critério define a ordem de eliminação considerando os mais próximos ou mais distantes. O algoritmo acima apresenta vantagem sobre qualquer outro já implementado, uma vez que considera todos os pares de combinação. Por sua vez isso limita a aplicação do operador em matrizes de grandes dimensões caso a complexidade computacional seja exponencial em buscas exaustivas $O(c^e)$. Como se pode armazenar previamente os pares de busca, não é necessário calcular os pares de distância cada vez que o algoritmo é executado. Outros padrões de busca de pares assimétricos podem ser facilmente implementados e, nesse sentido, é necessária somente a atualização dos pares de busca. Ou seja, o cálculo do valor de G_{1A} torna-se modular e depende apenas dos pares de busca e do critério de tolerância estabelecido. Na prática a utilização da busca exaustiva torna-se inviável em função da combinação 2 a 2 e do tamanho da matriz, sendo seu uso limitado a matrizes $n < 64$.

A definição de um critério de distância ou de outro tipo de busca, permite que o algoritmo possa ser utilizado em matrizes de maior dimensão a partir de raios de busca menores, limitando o espaço de busca. Essa parametrização do raio de distância depende do tipo de problema em questão e está relacionado a qual distância será considerada para iniciar a busca dos pares de vetores simétricos. Uma infinidade de padrões de busca pode ser considerada como critério de comparação e eliminação de vetores simétricos, sendo necessário que o usuário forneça os pares ou seqüência de pares a serem comparados. Outro aspecto importante é o armazenamento dos pares de matrizes a serem comparados, eliminando a necessidade de percorrer a matriz a cada cálculo do G1. Nesse sentido, é necessário assumir que a matriz 2-D possui

uma indexação, ou seja, ela pode ser transformada em um vetor unidimensional e ser indexada.

Para o cálculo do envoltório convexo é utilizado às mesmas bibliotecas de algoritmos típicos de geometria computacional. Podemos citar os algoritmos do embrulho de presentes (gift wrapping) e Graham Scan (BERG et al., 2008). A versão implementada utiliza a biblioteca Qhull ($http://www.qhull.org/$), resultado da implementação oficial do algoritmo *Quickhull* (BARBER et al., 1996). Esse último algoritmo tem complexidade computacional $O(nlogn)$ e é tradicionalmente o mais utilizado em problemas de geometria computacional.

Como discutido anteriormente um aspecto muito importante a ser observado é a dependência dos valores de G_{1A} em relação à borda da matriz. Ou seja, duas matrizes de núcleo completamente diferentes, mas com mesma borda e igual número de vetores assimétricos terão o mesmo valor de G_{1A}. Nesse contexto, ressalta-se a grande importância da eliminação de pares simétricos ou da utilização de G_{1A} em sub-matrizes elementares de 3x3 elementos. Dessa maneira, propõe-se a forma de aplicação dos valores de G_{1A} considerando uma subdivisão da matriz em elementos de 3x3. Considerando esse aspecto, para uma dada matriz, será obtido um subconjunto de valores de G_{1A}, ou seja, uma distribuição de G_{1A} será analisada em uma escala mais fina 3x3.

O espectro de valores de G_{1A3x3} pode ser calculado através da equação (3.19), variando n de 3 a 9, fornecendo possíveis valores de G_{1A} na escala *3x3* (Tabela 3.1). Quando a dimensão da matriz é aumentada esse espectro de valores aumenta exponencialmente, ficando restrito ao valor limite de 2 como provado por Rosa et al. (1998).

Considerando somente os valores de G_{1A} para matrizes 3x3 é possível definir, conforme indicado em Freitas et al. (2007), os graus qualitativos de assimetria. Desta forma, quando os valores da matriz de elementos analisada tiver fases e normas muito diferentes e assimétricas o valor de G_{1A} na matriz 3x3 tenderá a ser maior que $1,222$. Esse aspecto permite utilizar a distribuição de G_{1A} na escala 3x3, para analisar a assimetria na escala fina de uma matriz.

A Figura 3.17 resume os aspectos fundamentais para o cálculo do primeiro momento do gradiente. A função implementada possui flexibilidade para definir os padrões de busca previamente definidos, considerando a indexação clássica de matrizes (linha, coluna) para definição dos pares de comparação de vetores assimétricos.

Tabela 3.1 - Tabela com os possfveis valores de G_{1A} para o tamanho de uma matriz 3x3.

G_{1A}	Nv	k_e	n_e=(3n-k_e − 3)
0	3	3	3
0.8000	5	3	9
0.6000	5	4	8
0.4000	5	5	7
1.1429	7	3	15
1.0000	7	4	14
0.8571	7	5	13
0.7143	7	6	12
0.5714	7	7	11
1.3333	9	3	21
1.2222	9	4	20
1.1111	9	5	19
1.0000	9	6	18
0.8889	9	7	17
0.7778	9	8	16

Figura 3.17 - Aspectos fundamentais para o cálculo do primeiro momento do gradiente.

3.2.2 Exemplo do cálculo de $\mathbf{G}_{1A3x3}$

Um exemplo do cálculo de G_{1A3x3} é apresentado na Figura 3.18 considerando duas matrizes geradas com os modelos simplificados para geração dos dados canônicos. As Figuras 3.18a, 3.18b e 3.18c são exemplos de diferentes tipos de padrões espaciais. Importante destacar que, o valor de G1 será o mesmo $(1,9983)$ caso as duas matrizes sejam compostas com bordas iguais e não ocorra a remoção de vetores. A remoção de vetores, resulta em diferentes distribuições de G_{1A3x3}, caracterizando assimetria em uma escala de 3x3 pixels. A utilização do histograma dos valores de G_{1A3x3} fornece uma estimativa inicial da assimetria na escala fina de 3x3 pixels. Na solução representando um processo turbulento (3.18a) os valores G_{1A3x3} concentram-se em 1 com uma distribuição simétrica ao redor desse valor. Para a solução periódica (3.18b) os valores estão concentrados em $0,8$, o que significa uma baixa assimetria do campo gradiente na escala $3x3$. Já para o caso do ruído branco (3.18c) há uma distribuição assimétrica concentrando a mediana em $1,111$, o que indica uma maior concentração de valores assimétricos na escala $3x3$. Ressalta-se que todas as soluções são normalizadas e possuem a mesma média (0) e variância (1).

A extensão para aplicação do cálculo de G_{1A} em 1-D deve seguir os procedimentos de transformação de uma série temporal em uma matriz conforme proposto por Assireu et al. (2002). Uma ilustração desse procedimento está apresentada nas Figuras 3.19 e 3.20. Note que, com a metodologia proposta, não é mais necessária a triangulação de Delaunay, mas apenas o cálculo do envoltório convexo.

3.2.3 Definição de outros operadores computacionais

Nessa seção é proposta a definição de outros operadores computacionais baseados na definição dos momentos do campo gradiente. A definição desses operadores computacionais tem como objetivo extrair outras informações do campo gradiente explorando conceitos de diversidade.

Assumindo que o valor do campo gradiente pode ser decomposto em duas matrizes distintas de normas e fases, ou em coordenadas cartesianas nas componentes x e y, pode-se definir operadores computacionais baseados no cálculo da diversidade através dessas matrizes. O cálculo de índices de diversidade é geralmente obtido através de métodos estatísticos aplicados em estudos ecológicos para a obtenção da densidade e da distribuição de populações, permitindo a quantificação da biodiversidade em diferentes habitats (SIMPSON, 1949).

Figura 3.18 - Distribuição de $G1_{A3x3}$ para a) superfície turbulenta gerada pela equação de Ginzburg-Landau; b)superfície periódica resultante do somatório de matrizes de seno e c) superfície isotrópica derivada de um ruído branco.

Neste trabalho, a Diversidade de Norma e a Diversidade de Fase são definidas a partir do índice de Simpson (SIMPSON, 1949), ambos aplicados ao campo gradiente de norma e fase definido por:

$$G_{2_D} = 1 - \sum_{i=1}^{S_r} Pr_i^2, \tag{3.20}$$

$$G_{3_D} = 1 - \sum_{i=1}^{S_\phi} P\phi_i^2, \tag{3.21}$$

onde, Pr e $P\phi$ são a proporção de normas e valores fases em campo gradiente, res-

Figura 3.19 - (a) Série temporal com tendência; (b) Matriz correspondente; (c) Campo gradiente correspondente (9 vetores assimétricos) e (d) Campo de triangulação com 16 linhas conectoras.
Fonte: Adaptada de Assireu et al. (2002).

Figura 3.20 - a) Típico padrão de variabilidade de série temporal; b) Matriz correspondente; (c) Campo gradiente correspondente (9 vetores assimétricos) e d) Campo de triangulação com 20 linhas conectoras.
Fonte: Adaptada de Assireu et al. (2002).

pectivamente. S é o total de elementos únicos considerados a partir do fatiamento do histograma. Um campo gradiente perfeitamente homogêneo em fase e norma tem o índice de diversidade 0. Se todos os elementos forem diferentes entre si, o índice tem valor 1. Computacionalmente, a implementação desse índice exige a inserção de uma componente de tolerância de norma (T_r) e fase (T_ϕ), que define quais va-

lores são considerados iguais e na prática isso é realizado fatiando o histograma de valores de norma e fase dos elementos da matriz. Para as normas isso é definido em função da dimensão da matriz em que $S_r = (T_r) \equiv N$, onde N é número de elementos do campo gradiente definido a partir da dimensão da matriz escalar. O fatiamento dos valores de fase utiliza uma tolerância a fase em graus decimais para definir o fatiamento do histograma $S_\phi = 360/(T_\phi)$, essa tolerância pode ser dada em radianos, porém é mais intuitiva no sistema de graus decimais. Um aspecto muito importante que deve ser considerado é a resolução em *bits* do sinal coletado ou analisado, isso deve ser considerado no momento da implementação computacional, principalmente no cálculo do campo gradiente e no cálculo das diversidades. Se por um lado esses operadores são tendenciosos e paramétricos, de outro lado eles são de simples implementação computacional e proporcionam uma medida de diversidade do campo gradiente que pode auxiliar na análise e interpretação de determinados tipos de dinâmicas em que há grande variabilidade em fase e/ou norma.

O outro operador computacional proposto nessa tese está relacionado a uma medida de energia da matriz de normas do campo gradiente. A norma de *Frobenius* ou traço de uma matriz quadrada, $\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2}$, é utilizada para definição do operador a partir dos valores da matriz de norma G_2 do momento gradiente. Este operador é definido por:

$$G_{2_{NF}} = \sum_{i=1}^{L}\sum_{j=1}^{C}\left|{r_{i,j}}^2\right|, \tag{3.22}$$

Em que L é o número de linhas e C o número de colunas da matriz de normas do campo gradiente. Essa medida é sempre positiva e está relacionada à energia do campo de normas do campo gradiente.

A partir da Inequalidade de Frobenius (IF)(BOTTCHER e WENZEL, 2005) definiu-se outro índice. Essa inequalidade da álgebra linear implica em dizer que para quaisquer duas matrizes reais A e B reais, o valor da inequalidade é sempre limitado pela conjectura a $\sqrt{2}$.

$$\frac{\|AB - BA\|_F}{\|A\|_F\,\|B\|_F} \leq \sqrt{2}, \tag{3.23}$$

Utilizando a decomposição do campo gradiente em coordenadas cartesianas (dx e dy) é possível associar os valores das matrizes de dx e dy dentro da inequalidade,

definido aqui como índice de Frobenius:

$$NFr_i = \frac{\|dxdy - dydx\|_F}{\|dx\|_F \|dy\|_F} \leq \sqrt{2}\ , emque,\ (\|dx\|_F > 0\ e\ \|dy\|_F > 0), \tag{3.24}$$

Esse índice pode ser utilizado como um operador computacional apto a comparar matrizes de mesma dimensão. É interessante notar que se dx e/ou dy é nulo, esse índice tende ao infinito. Em matrizes elementares 3x3 também há uma relação entre a ortogonalidade, se existem ângulos de 45° entre os vetores, formando fontes ou sumidouros (padrões convergentes ou divergentes no campo gradiente em relação a um ponto na matriz) no campo gradiente, o valor do índice tenderá a $\sqrt{2} \approx 1,4142$. Para campos gradiente completamente aleatórios o valor de NFr_i é assintótico para matrizes de dimensão maior que 20x20 elementos.

Os valores de NF_i para o exemplo apresentado na Figura 3.18 são de [0,2173; 1,4129 e 0,1002] para a solução de um processo turbulência, estrutura periódica e o ruído branco, respectivamente. O valor de $1,4129$ para a imagem (3.18b) indica a presença de estruturas ortogonais típicas de superfícies periódicas. Utilizando esse índice, pretende-se detectar transição de fase entre regimes de variabilidade que possuam estruturas ortogonais, sendo estes independentes dos valores de momentos estatísticos de média e variância. Por definição, esses últimos operadores baseados na norma de *Frobenius* estão restritos às matrizes quadráticas.

Combinando os 4 operadores propostos é possível analisar a dinâmica do padrão gradiente em uma série espaço-temporal. Essa dinâmica é decomposta nas diversidades de fase e amplitude, bem como, na medida de energia e no cálculo do índice de Frobenius para determinar a presença de estruturas ortogonais. Nas análises espaço-temporais, as mudanças no regime de variabilidade desses índices indicam a ocorrência de uma mudança na estrutura espacial.

A Figura 3.21a apresenta uma seqüência da solução da equação complexa de Ginzburg-Landau 2-D, a qual representa uma transição de fase no regime de variabilidade. A condição inicial considera um ruído branco com média 1 e desvio padrão 0. Na Figura 3.21b é possível observar a transição de fase no gráfico da série temporal de NFr_i. Nesse gráfico pode-se observar que a transição no regime de variabilidade ocorre entre os tempos de simulação 1 a 13, ocorrendo depois um relaxamento e praticamente um mesmo valor no índice. Esse resultado ilustra a utilidade desse tipo de operador para reduzir a dimensionalidade dos dados, transformando a

informação de um regime de variabilidade espaço-temporal numa media temporal.

Figura 3.21 - a) solução turbulenta da equação complexa de Ginzburg-Landau 2-D mostrando transição de fase, e b)evolução temporal do NFr_i para cada passo da solução.

A solução do mapa de Kaneko para alguns passos da simulação é mostrada na Figura 3.22. Nessa Figura é possível observar que durante a evolução da simulação há uma fragmentação das estruturas, perdendo sua correlação com as condições iniciais. Os operadores computacionais descritos nessa subseção são aplicados para cada passo da simulação e estão apresentados na Figura 3.23. Nessa Figura a série em azul representa a simulação do mapa de Kaneko para uma condição inicial do envelope de senos, a série em verde é uma condição inicial aleatória, já a série em vermelho é apenas uma matriz pseudo aleatória sem correlação espacial. Os valores de NFr_i, condição inicial do envelope de senos, indicam a presença de estruturas ortogonais identificadas pelos elevados valores nos primeiros passos da simulação. Nos passos da simulação, embora com valores muito próximos, não há uma sobreposição dos valores, indicando que ainda existe uma separação entre as duas simulações. O caso

contrário ocorre com os outros operadores mostrando que existe uma sobreposição entre a diversidade de normas e fase, bem como os valores de G_{2_N} quando ocorre a convergência dos mapas acoplados. Para todos os operadores existem passos da simulação em que os valores do campo aleatório (linha vermelha) são bem distintos da simulação do mapa acoplado, ressaltando a diferença entre as dinâmicas espaço-temporais envolvidas.

Figura 3.22 - a) Representação em imagem da solução do mapa de Kaneko (a = 2.00, e = 0.3) para duas condições iniciais: a) aleatória e b) envelope de senos.

Os operadores discutidos anteriormente estão implementados em ambiente MatLab e podem ser facilmente convertidos para softwares livres como *SciLab* e *Octave*. Além disso podem ser incorporados *à* versão do GPA em *C/C++* implementado por Strieder (2010). Na próxima seção são descritas técnicas complementares para visualização e análise de séries temporais a partir do gráfico de recorrência.

3.3 Gráfico de Recorrência Global

A última técnica implementada constitui uma combinação da técnica de gráfico de recorrência com a análise de padrões de gradiente. A técnica de gráfico de recorrência foi proposta por Eckmann et al. (1987) baseada nos conceitos de sistemas dinâmicos. O conceito formal de recorrência foi estabelecido por Henri Poincaré em 1890 (MARWAN, 2008) e sua definição está relacionada ao conceito de sistema dinâmico

Figura 3.23 - Operadores computacionais para cada passo da simulação do mapa de Kaneko: a) NFr_i ; b) Diversidade de norma G_{2_D}; c) Diversidade de fase G_{3_D} e d) Norma de Frobenius $G_{2_{NF}}$.

do espaço de fases. Esse último, de um modo simplista, pode ser definido como um espaço abstrato que representa as variáveis dinâmicas de um sistema. Se a trajetória de uma variável nesse espaço abstrato tende a passar pelos mesmos caminhos ocorre a recorrência. Esse trabalho não tem o objetivo de entrar nos aspectos fundamentais da teoria de sistema dinâmico, mas sim explorar a técnica como uma ferramenta potencial para visualização e análise de dados. Maiores informações sobre os fundamentos de sistemas dinâmicos estão disponíveis em Katok e Hasselblatt (1996) e Bergé et al. (1996).

Eckmann et al. (1987) fizeram a proposta de uma metodologia para visualizar as recorrências nos sistemas dinâmicos. Essa técnica supõe que uma variável dinâmica pode ser representada por um vetor que indica uma trajetória em um espaço matemático abstrato, o espaço de fases. Na prática a representação dessa trajetória é dada pela transformação de uma série temporal de uma medida como, temperatura, pressão, umidade ou muitas outras variáveis ambientais para o espaço de fases. No caso de exemplos clássicos de sistemas dinâmicos pode ser a representação de velocidade e posição de soluções de equações que descrevem, por exemplo, um pêndulo não-linear. A representação do gráfico de recorrência proposta por Eckmann et al.(1987) considera a construção de uma matriz de recorrência, sendo esta simétrica e composta por zeros e uns, onde o valor 1 representa uma região de recorrência e o valor 0 de não recorrência. Note que para estabelecer o que é recorrente ou não,

torna-se necessário utilizar um critério de similaridade, definido a partir de uma métrica de distância.

Formalmente o gráfico de recorrência é dado por:

$$R_{i,j}(\epsilon) = \Theta(\epsilon - \|\vec{x}_i - \vec{x}_j\|),\ i, j = 1, ..., N, \tag{3.25}$$

em que N representa o número de pontos medidos, o tamanho da série temporal analisada ou o tamanho do vetor de elementos de uma variável $vecx_i$. O componente ϵ é o limiar de distância a partir do qual será considerado recorrente. $\Theta(.)$ é a função de *Heaviside*, onde ($\Theta(x)$=0 se $x < 0$ e $\Theta(x) = 1$ caso contrário) e $\|\|.\|\|$ é a norma utilizada para comparação do que é recorrente no espaço de fases. Tipicamente utilizam-se normas tipo L_1 (norma de Manhattan) e L_2(euclidianas), entretanto a máxima norma $L\infty - norm$ é mais utilizada por ser computacionalmente mais rápida devido a utilização de uma vizinhança quadrada. Um parâmetro importante a ser estimado para definição do gráfico de recorrência é o valor do limiar ϵ. Varias metodologias são propostas para estimar o valor de ϵ, dependendo do sistema dinâmico a ser estudado (MARWAN, 2008). O aspecto mais importante do gráfico de recorrência para séries experimentais é a transformação da série temporal observada em trajetórias do espaço de fases. Essa tarefa exige certa parametrização muitas vezes difícil de ser estabelecida, considerando a necessidade de especificação das dimensões do espaço de fases e da utilização de algoritmos para a construção dessas trajetórias. Essa parametrização funciona bem em sistemas contínuos, nos quais o espaço de fases pode ser calculado através de equações diferenciais e/ou através de dimensões conhecidas do espaço de fases. Existem algoritmos propostos para transformar, por exemplo, um dado observado através de série temporal em trajetórias no espaço de fases, porém sua parametrização não é uma tarefa trivial, sendo amplamente discutida na literatura de sistemas dinâmicos (MARWAN et al., 2007).

Uma gama de derivações e extração de informação foi proposta a partir do gráfico de recorrência (MARWAN et al., 2007). Nesse trabalho apresenta-se apenas a exploração visual e geométrica do gráfico de recorrência, sem o interesse de extrair parâmetros típicos de sistemas dinâmicos como as dimensões de imersão. A representação visual do gráfico de recorrência gera padrões típicos que são chamados na literatura de tipologias. Essas tipologias caracterizam previamente o padrão de variabilidade de uma série temporal. Outro modo de explorar a geometria do gráfico

de recorrência é desconsiderar a transformação das trajetórias no espaço de fases, fazendo com que a própria série temporal, constituam a trajetória no espaço de fases. Na estatística isso é chamado de cálculo de matriz de distância (Eq. 3.26). A representação dessa matriz, quando limiarizada, exibe a mesma tipologia do gráfico de recorrência, o qual pode ser chamado de gráfico de recorrência global ou gráfico de recorrência não limiarizado (MARWAN et al., 2007).

$$D_{i,j} = \|X_i - X_j\| \tag{3.26}$$

considerando esse aspecto, o gráfico de recorrência pode ser simplificado, tornando-se de fácil implementação computacional. Essa implementação tem o objetivo de realizar um estudo qualitativo e geométrico em séries temporais. A primeira transformação calcula a matriz de distância entre os elementos da série a partir de uma série temporal ($X_{i,N}$). Para esse cálculo a distância euclidiana entre os pontos da série é utilizada. Após o cálculo dessa matriz de distância, o passo seguinte é o cálculo da matriz de recorrência, a qual é calculada a partir da limiarização da matriz de distância através da função de Heaviside. Esse limiar representa uma distância de similaridade para dizer o que pode ser recorrente. O valor pode ser baseado, por exemplo, no percentual do desvio padrão.

A partir da matriz binária, um conjunto de métricas pode ser calculado conforme proposto por (MARWAN et al., 2007). Nesse trabalho, o objetivo é implementar o método proposto para suporte e análise geométrica e visual das séries de índices de vegetação que serão descritas no próximo capítulo. Uma restrição natural desse método é a sua utilização em séries temporais extremamente curtas com N $<$ 1000, uma vez que a visualização e o cálculo da matriz se tornam custosos no caso de séries temporais compostas por muitos elementos devido *à* necessidade de combinação pareada entre os elementos da série.

A Figura 3.24 exibem os gráficos das séries temporais, a Figura 3.25 a matriz de distância e a Figura 3.26 o gráfico de recorrência. É interessante observar que a informação visualizada na matriz de distâncias (Fig. 3.25) ressalta significativamente os padrões de variabilidade da série temporal, onde valores de baixa amplitude (tons de azul escuro) significam pontos recorrentes na série temporal. Esses gráficos ressaltam as séries que possuem tendência e também os pontos atípicos. A matriz de distância indica onde há transição e quebras no regime de flutuação. Nesse sentido, o estudo e a implementação dessa técnica tiveram o objetivo de fornecer ao usuário das

séries temporais de sensoriamento remoto uma forma alternativa de visualização dos regimes de variabilidade das séries discutidas no próximo capítulo. A simples análise da matriz de distância permite ao usuário identificar pontos de mudança no regime de variabilidade mais facilmente do que analisando somente as séries temporais.

Figura 3.24 - Séries canônicas: a) série de senos; b) série de senos + ruído branco; c) Não estacionária + ruído branco; d) multi-periódica série de senos; e) E.C.G.L. - solução turbulenta; f) E.C.G.L. -intermitência; g) E.C.G.L. - turbulência em fase; h) E.C.G.L. - estruturas coerentes - pontos atípicos; i) ruído branco.

A análise geométrica dos valores limiarizados na imagem de recorrência pode indicar a presença de tendências na série, uma periodicidade ou uma maior freqüência de variabilidade. Se existem grandes áreas não preenchidas no gráfico de recorrência, isso indica a presença de tendência ou sazonalidade (Fig. 3.26). Um modo geométrico simples para verificar a presença dessas lacunas na tipologia do gráfico de recorrência é gerar uma triangulação de *Delaunay* entre os pontos limiarizados. O tamanho de cada triângulo pode ser calculado, além da estatística sobre o número total de triângulos e a área do maior triângulo encontrado. Nesse sentido, esse trabalho propõe um algoritmo para explorar a informação geométrica do gráfico de recorrência.

O algoritmo proposto consiste nos seguintes passos:

a) Transformação da série temporal numa matriz de distância;

b) Cálculo da imagem binária a partir da matriz de distância;

Figura 3.25 - Matriz de distância : a) série de senos; b) série de senos + ruído branco; c) Não estacionária + ruído branco; d) multi-periódica série de senos; e) E.C.G.L. - solução turbulenta; f) E.C.G.L. -intermitência; g) E.C.G.L. - turbulência em fase; h) E.C.G.L. - estruturas coerentes - pontos atípicos; i) ruído branco.

c) Inserção da borda triangular nos elementos da matriz triangular superior limiarizada;

d) Aplicação da triangulação de Delaunay nos pontos de recorrência;

e) Cálculo dos atributos: número total de triângulos e área do maior triângulo, números de triângulo de maior área;

f) Normalização dos atributos em função da dimensão do tamanho da série.

O algoritmo é simples de ser implementado e utiliza a biblioteca Qhull

Figura 3.26 - Gráfico de recorrência global com limiar $\epsilon de 0,05$ desvio padrão : a) série de senos; b) série de senos + ruído branco; c) Não estacionária + ruído branco; d) multi-periódica série de senos; e) E.C.G.L. - solução turbulenta; f) E.C.G.L. -intermitência; g) E.C.G.L. - turbulência em fase; h) E.C.G.L. - estruturas coerentes - pontos atípicos; i) ruído branco.

(http://www.qhull.org/) para realizar a triangulação de Delaunay conforme citado na seção anterior. A Figura 3.27 ilustra a seqüência dos passos do algoritmo proposto.

A fim de elaborar um número adimensional e válido para múltiplas escalas o número de triângulos e a área dos triângulos máximos pode ser normalizada pela dimensão da matriz de distância e número máximo de pontos de recorrência. A normalização da área do máximo triângulo é definida pela simples razão entre a área do maior triângulo dividida pela área de triangulação total, que é a metade da dimensão da

Figura 3.27 - Seqüência de passos para a triangulação do gráfico de recorrência.

matriz. Já o número total de triângulos normalizado é definido pelo número total de triângulos dividido pelo total de triângulos possíveis no domínio do gráfico de recorrência que é dado por NxN.

O número total de triângulos pode ser calculado a partir das Equações 3.17 e 3.19 em que o número total de triângulos é $n_t = 2n - 2 - k_e$, em que k_e é o total de pontos do envoltório convexo e n o total de vértices da triangulação (BERG at al., 2008). A triangulação do gráfico de recorrência deve considerar a inserção de uma borda triangular como o envoltório convexo. Nesse caso, o número de pontos do envoltório convexo é função do comprimento da série temporal N. Esse número de pontos do envoltório convexo é $k_e = 3N - 3$, considerando que o número máximo de pontos da triangulação n representa todos os possíveis pontos incluindo o envoltório convexo e todos possíveis pontos dentro desse tem-se que $n = (N^2 + N)/2$, por substituição temos que o total de possíveis triângulos (TT) é $TT = n_t = (N^2 - 2N + 1)$, sendo N o comprimento da série temporal. Com isso pode-se normalizar o número de triângulos

com o número máximo de triângulos possíveis, isso é uma medida indireta da taxa de recorrência. Outra normalização é o triângulo de máxima área que pode ser normalizado pela área máxima da triangulação. A área da máxima triangulação é a simples relação entre a dimensão da matriz de distância $area_{max} = \frac{N^2}{2}$.

O atributo total de triângulos representa quantos pontos na série temporal são recorrentes. Se há muitos pontos recorrentes o valor desse índice é alto e indica que existem valores atípicos de amplitude da série. O valor de área máxima do triângulo indica a presença de tendência na série e mudança de amplitude. Para esse comprimento de série, por exemplo, pode-se afirmar que valores muito maiores que $0,05$ estão relacionados a presença de tendência e mudanças na dinâmica de flutuação da série. Valores bem menores que $0,05$ e tendendo a zero indicam que parte da série possui alta-frequência. Se o número de triângulos de máxima área for maior que 1, indica que pode ter alguma periodicidade na série. O valor do maior triângulo gerado está relacionado com o tamanho das lacunas presente no gráfico de recorrência, ou seja, os pontos não preenchidos. A existência de grandes lacunas irregulares é um indicativo de que o regime de variabilidade da série não está confinado apenas num certo domínio de variabilidade. Quando o valor do maior triângulo encontrado é muito pequeno, é uma indicação que a série possui alta freqüência, preenchendo todas as regiões do espaço de fases. Esses operadores são concebidos apenas para gerar um indicativo em séries que possuem tendência ou variações abruptas, sendo intrinsecamente operadores derivados da geometria computacional. Os *boxplotes* da Figura 3.29 ilustram os valores de 1000 simulações para cada dimensão de uma série temporal simulada. As séries da Figura 3.29a são um ruído branco e da Figura 3.29b um passeio aleatório representando uma série com tendência. Esse exemplo ressalta a separação entre o tamanho do máximo triângulo para as séries estacionárias e não estacionárias. Esse resultado de maneira alguma substitui qualquer teste estatístico de estacionaridade, mas pode ser utilizado em séries temporais curtas (N<500) para uma separação preliminar de séries com tendência.

A Figura 3.29 ilustra a relação entre o número de triângulos e a área de cada triângulo para as nove séries canônicas utilizadas. Nas séries em que há tendência ou pontos abruptos e o número de triângulos de maior área é elevado como pode ser observado nas séries temporais c, d e h, valores acima da linha tracejada. Na série h em que existem pontos atípicos, o número de triângulos é elevado devido ao maior número de pontos de recorrência. Para a série a que possui regime periódico bem definido e sem ruídos existe um padrão de vários triângulos com mesma área o que indica a presença de tipologias periódicas também no gráfico de recorrência.

Figura 3.28 - Simulação dos valores do triângulo máximo encontrado em relação ao comprimento da série temporal.

Para as demais séries não há uma diferença significativa já que a maioria não possui tendências embora sejam visivelmente distintas.

Figura 3.29 - Relação entre o número de triângulos e a área de cada triângulo para as nove séries canônicas utilizadas: a) séries ruído branco e b)passeio aleatório.

As métricas estabelecidas não excluem a importância visual da matriz de distância e do gráfico de recorrência, que são importantes da análise visual. As tipologias da matriz de distância e do gráfico de recorrência global estão associadas ao regime de variabilidade da série. Nesse sentido essa tipologia pode ajudar a inferir o regime de variabilidade da série temporal. Marwan et al. (2007) demonstraram que linhas paralelas a diagonal estão relacionadas ao determinismo da série analisada, bem como, a presença de pontos de recorrência aleatoriamente espalhadas configuram a desordem da série, típico de um sistema sem memória. Uma variedade de métricas de quantificação do gráfico de recorrência são propostas na literatura (ZBILUT e WEBBER JR., 1992; ZBILUT e WEBBER JR. 1994, MARWAN et al. 2002).

A implementação da metodologia descrita nessa seção tem o objetivo de apresentar ao usuário das séries temporais de sensoriamento remoto uma forma alternativa de visualização dos regimes de variabilidade das séries que são discutidas no próximo capítulo. A simples análise da matriz de distância permite ao usuário não especialista identificar visualmente, por exemplo, transição entre regimes bruscos de variabilidade.

3.4 Síntese dos Operadores Computacionais Propostos

Nesse Capítulo foram apresentados os aspectos computacionais dos métodos e algoritmos utilizados. Resumidamente podem-se classificar as aplicações dos métodos propostos de acordo com os dados de entrada. A Tabela 3.2 mostra a lista de operadores computacionais estudados, suas vantagens, limitações e tipo de dado a ser tratado. Todas as implementações computacionais dos algoritmos para o cálculo dos operadores realizadas na Tese estão listadas abaixo e serão disponibilizadas como uma *Toolbox* do MatLab e também como uma publicação técnico-científica tipo Programas de Computador (PDC) do Serviço de Informação e Documentação - (SID) do INPE.

- DFA - Identificação do padrão de variabilidade da série espaço-temporal ou temporal quanto às leis de escala;
- GPA - Técnica baseada no campo gradiente para medir assimetria, diversidades e energia do campo gradiente;
- Gráfico de Recorrência - Visualização e detecção de tendências;

Tabela 3.2 - Características dos Operadores Computacionais

Operador computacional	**Observações**	**Tipo**
DFA 1-D α	Método amplamente utilizado na literatura para estudos de leis de escala na variabilidade existente em séries temporais . Restrições: Necessidade de modificações do método de acordo com o tipo de tendência a ser estudado.	1-D
DFA 2-D α	Método computacionalmente mais simples para identificar leis de escala em matrizes 2-D. Restrições: Pouca descrição e utilização do método na literatura	2-D
GPA G_{1A}	Técnica da geometria computacional que permite o cálculo de assimetria em séries curtas. Restrições: Elevado custo computacional para aplicação em séries de dimensão muito grande, dificuldade na interpretação dos resultados	1-D e 2-D
GPA G_{2_D}	Utilizado para estudar a distribuição de normas de um campo gradiente. Restrições: Depende da parametrização do que é similar em norma	2-D
GPA G_{3_D}	Permite estudar a distribuição de fases de um campo gradiente. Restrições: Depende da parametrização do que é similar em fase;	2-D
GPA G_{2_N}	Medida de energia do campo gradiente	2-D [NxN]
GPA NFr_i	Operador para detecção de campos ortogonais no campo gradiente	2-D [NxN]
Matriz de Distância e Plotes de Recorrência	Análise complementar de visualização e detecção de tendências e pontos abruptos	1-D séries curtas

4 ESTUDO DE CASO PARA DETECÇÃO DE MUDANÇA DE USO E COBERTURA DA TERRA

Look deep into nature, and then you will understand everything better.
(*Albert Einstein*)

Este capítulo apresenta um estudo de caso que utiliza os métodos computacionais apresentados no capítulo 3 e as séries temporais de índices de vegetação disponíveis no laboratório virtual de sensoriamento remoto. Esse estudo teve como objetivo avaliar os métodos computacionais propostos como ferramentas integrantes do módulo de algoritmo e análise do laboratório virtual de sensoriamento remoto.

Entender a dinâmica de cobertura da terra de forma transdisciplinar é um processo essencial na Ciência do Sistema Terrestre. Essa dinâmica é resultado da ocorrência simultânea de múltiplos processos biológicos, físicos, geológicos, meteorológicos, sociológicos, econômicos, dentre outros que ocorrem simultaneamente no espaço e no tempo. O cálculo das emissões de gases do efeito estufa é um exemplo clássico que depende do conhecimento sobre o uso e cobertura da terra. No geral, estes estudos são extensamente apoiados por dados e técnicas de sensoriamento remoto e de geoprocessamento. Em síntese existem duas linhas de abordagem comumente utilizadas, a detecção de mudanças e a modelagem dos processos de mudança.

A detecção de mudanças tem como objetivo quantificar o tipo e o período da mudança, sendo muito utilizada para o monitoramento ambiental e a avaliação de impactos ambientais com enfoque ecológico/econômico, voltado para quantificar o que e quanto mudou, bem como, os efeitos destas mudanças nos ecossistemas envolvidos. Um exemplo clássico é o monitoramento do desflorestamento na floresta amazônica como os programas PRODES e DETER. A abordagem da modelagem dos processos de mudança tem o objetivo de gerar cenários futuros de uso e cobertura da terra para, de alguma forma, fornecer informações para os tomadores de decisão nas políticas de gestão territorial e ambiental. Ambas as linhas de pesquisa são intensamente apoiadas pela computação aplicada. Na primeira linha de pesquisa, uma gama de métodos de reconhecimento de padrões e processamento de imagens são tradicionalmente utilizados (COPPIN et al., 2004).

Na segunda linha de pesquisa, modelos matemáticos, empíricos ou híbridos são integrados aos sistemas de informações geográficas. A partir de dados de sensoriamento remoto e/ou outros tipos de dados espaciais esses modelos são utilizados na geração de cenários futuros do uso e cobertura da terra. Um exemplo dessa aplicação são

os ambientes de modelagem como o TerraME e seus vários módulos (CARNEIRO, 2003). Na linha de detecção de mudanças de uso e cobertura da terra, as técnicas de classificação multitemporal são amplamente discutidas na literatura de sensoriamento remoto. Lu et al. (2004) e Coppin et al.(2004) fazem uma ampla revisão desses métodos. Esses autores classificam de forma qualitativa, as vantagens e as desvantagens, bem como o tipo de algoritmo, a escala temporal e o tipo de dado de entrada. De maneira geral, a maioria das metodologias é baseada em distâncias ou classificações estatísticas e direcionadas a alguns tipos de uso e cobertura da terra. Lu et al. (2004) propõem seis categorias de métodos de detecção de mudanças baseados em: (1) álgebras, (2) transformações matemáticas e radiométricas, (3) classificação, (4) modelos avançados, que incluem a transformação de dados radiométricos em medidas biofísicas também utilizando dados de campo, (5) sistemas de informações geográficas, e (6) análise visual de imagens.

A abordagem desse trabalho é diferente e complementar aos métodos tradicionalmente descritos na literatura tanto para a detecção quanto para a modelagem e simulação. Essa abordagem consiste em dois tipos de análise: *(i)* aplicação dos métodos DFA-1D, Análise de Padrões Gradiente e gráfico de recorrência para suporte a visualização e a caracterização da dinâmica de uso e cobertura da terra, e; *(ii)* aplicação de métodos computacionais em séries espaço-temporais para uma indicação inicial das áreas onde possam ter ocorrido mudanças no uso ou cobertura da terra, bem como para estudos localizados da variabilidade dinâmica da vegetação. Na próxima seção são descritas as aplicações para análise das séries e na seção posterior para análise de séries espaço-temporais.

4.1 Detecção de Mudanças de Uso e Cobertura da Terra com Séries Temporais

A metodologia utilizada consiste em avaliar os operadores computacionais 1-D propostos e implementados no Capítulo 3 como integrantes do módulo de algoritmo e análise do laboratório virtual apresentado no Capitulo 2.

Com o objetivo de ilustrar os resultados, selecionou-se três áreas com e sem mudanças no uso e cobertura da terra em diferentes biomas. As séries temporais de EVI2 são apresentadas na Figura 4.1, representando respectivamente: a) um processo de desflorestamento na região amazônica; b) um desflorestamento no cerrado, e; c) uma área de floresta semi-decidual sem alterações. Os operadores computacionais utilizaram a série temporal filtrada com a transformada de *wavelets* como padrão, mas isso não exclui a possibilidade de utilização da outra série sem a filtragem

wavelets.

Figura 4.1 - Séries temporais EVI2 a) um processo de desflorestamento na região amazônica; b) um desflorestamento no cerrado, e; c) uma área de floresta semi-decidual sem alterações.

O primeiro operador computacional utilizado é o expoente α do método DFA-1D. Um procedimento adotado para eliminar o efeito da tendência sazonal é o cálculo da anomalia, conforme proposto por Telesca e Lasaponara (2005) que analisaram séries temporais de índices de vegetação com o DFA 1-D, removendo a tendência sazonal. A técnica de calcular a anomalia é amplamente descrita na literatura e consiste em subtrair de cada valor da série a sua correspondente média mensal decadal e dividir pelo desvio padrão também mensal. Ou seja, a nova série de $EVI2_{an} = (EVI2 - < EVI2 >_m)/\sigma_{EVI2_m}$ onde $< EVI2 >_m$ e σ_{EVI2_m} são, respectivamente, a média e desvio padrão mensal, ou seja, um vetor de 12 elementos cada, representando a média de todos os janeiro(s), fevereiro(s), e assim por diante. A transformação em anomalia normaliza a série e diminui os efeitos de sazonalidade.

Analisando os valores de α após a remoção da tendência temos os valores de 1,15,

0,80 e 0,70 para o desmatamento na floresta, desmatamento no cerrado e área de floresta semi-decidual, respectivamente. Esse resultado do expoente α é apresentado na Figura 4.2.

Figura 4.2 - Expoente α calculado para: a) um processo de desflorestamento na região amazônica; b) um desflorestamento no cerrado; e c) uma área de floresta semi-decidual sem alterações.

Com objetivo de realizar uma análise exploratória inicial sobre os métodos propostos foram selecionadas 30 amostras por classe para diferentes tipos de uso como áreas de floresta, pivô central e áreas desflorestadas em floresta e em cerrado. Pode-se observar no gráfico *boxplot* da Figura 4.3 que existe uma tendência de que desflorestamentos em área de floresta apresentem valores maiores de *alpha*.

Figura 4.3 - Boxplot dos valores α para as séries temporais amostradas.

Esse comportamento é esperado uma vez que, após o desflorestamento, a série modifica sua tendência com o forte decaimento nos valores do índice. O desflorestamento nas áreas de cerrado apresenta uma intersecção com as áreas de floresta, já para as áreas de intensa agricultura ocorre maior dispersão, mas no geral os valores são mais baixos devido ao maior número de ciclos na série-temporal. Do ponto de vista da dinâmica da vegetação o resultado demonstra que o processo após o desflorestamento na floresta deve ser diferente do que ocorre no cerrado. Uma possível explicação para isso é que geralmente no cerrado, para essa região, o tipo de uso subseqüente é a agricultura, enquanto que na floresta são observados três tipos de uso, a pastagem, a agricultura ou a regeneração da própria floresta. O algoritmo do DFA-1D está disponível para novas análises e testes. No processo de visualização, valores de α bem maiores que 1 em áreas de floresta podem indicar a ocorrência de um processo de desflorestamento dado pela não estacionaridade da série temporal de EVI2.

O outro operador computacional testado para séries unidimensionais é a distribuição das assimetrias dos valores de $G1_{A3x3}$. Esse operador é aplicado às anomalias das séries temporais de EVI2 conforme metodologia proposta por Assireu et al. (2002) e Freitas et al. (2007). Um passo inicial é transformar a série temporal em submatrizes 3x3 e calcular os valores de $G1_{A3x3}$ em cada sub-escala, conforme descrito no Capítulo 3. Na Figura 4.4 o histograma ilustra a distribuição das assimetrias dos quatro tipos de uso da terra e um dado sintético que é um ruído branco de mesmo tamanho amostral. A distribuição de assimetrias utilizando os valores de $G1_{A3x3}$ é difícil de ser analisada, inviabilizando a utilização dos valores de $G1_{A3x3}$ para esse tipo de análise.

O exemplo da Figura 4.5 mostra a comparação de duas séries de EVI2, filtrada (Fig. 4.5a) e não filtrada (Fig. 4.5d). No gráfico da Figura 4.5b e 4.5d é possível visualizar a distribuição das assimetrias ao longo da série. No histograma apresentado nas Figuras 4.5c e 4.5e é evidenciado um incremento nas assimetrias mais altas (1,222), isso é esperado uma vez que a série não filtrada possui ruído de alta freqüência. Nesse sentido, a utilização do $G1_{A3x3}$ na escala mais fina pode servir como um instrumento de comparação geométrica e qualitativa para avaliar a presença de ruídos principalmente em séries curtas em que não exista uma estatística robusta.

A última análise unidimensional implementada é uma ferramenta para suporte na visualização das séries temporais EVI2, utilizando a combinação entre o gráfico de recorrência global e a matriz de distância para ressaltar visualmente às alterações

Figura 4.4 - Histograma dos valores de $G1_{A3x3}$ para as séries temporais amostradas por classe de uso.

no uso e cobertura da terra. Os exemplos são apresentados para as séries da Figura 4.7 conforme indicado: Figura 4.6a) matriz de distância; Figura 4.6b) limiarização utilizando o parâmetro de 1% do desvio padrão, e; Figura 4.6c) série temporal correspondente na cor de linha verde. É importante notar que quando há uma grande alteração do uso da terra isso é visualmente ressaltado na matriz de distância (FIG-4.6a). Na Figura (Fig. 4.6a) há uma maior regularidade na variabilidade da série temporal, os pontos em vermelho na matriz de distância representam a distância entre os pontos extremos máximos e mínimos e, para essa matriz, as distâncias máximas entre as séries ocorrem em 2004 e 2005 com 2010. No gráfico de recorrência (Fig. 4.6b) há uma maior regularidade com ausência de grandes lacunas.

Em uma análise visual do primeiro caso torna-se bem evidente que o regime de variabilidade mudou após o ano de 2007 tendo uma variabilidade periódica em 2009 e 2010. Os valores na matriz de distância significam a diferença do EVI2 entre os pontos da série verde (Fig. 4.7c). No gráfico de recorrência global existem áreas grandes e irregulares não preenchidas, indicando que a série não é estacionária. O gráfico de área na série temporal é apenas auxiliar e, como mencionado no Cap. 2, permite identificar os períodos de seca e de chuva. Para a Figura 4.8a a matriz de distância indica novamente uma área em que ocorreu a mudança no regime de variabilidade e, após o ano de 2006, o regime de variabilidade tornou-se periódico

Figura 4.5 - Espectro de assimetria com valores de $G1_{A3x3}$ para as séries temporais: a) Série EVI2; b)espectro de assimetrias na escala *3x3* para série (a); c) histograma das assimetrias para série (a); d)Série EVI2 sem filtragem por *wavelets*; e) espectro de assimetrias na escala *3x3* para série (d); c) histograma das assimetrias para série (d).

indicando o uso agrícola. A elevação dos valores de distância máximos maior que $0,8$ em um curto período de tempo indica a presença de área agrícola, e pode ser confirmada pelo rápido incremento no valor do índice. A grande concentração de pontos entre 2000 e 2006 no gráfico de recorrência (Figura 4.8b) indica que há pouca variação de amplitude no regime de variabilidade do EVI2 nesse período. Esse processo é semelhante ao ocorrido no exemplo da Figura 4.7 entre os anos de

2000 e 2007.

Figura 4.6 - Gráfico de recorrência para uma série temporal de floresta. a)Matriz de distância; b) gráfico de recorrência e c)série temporal.

Utilizando os operadores computacionais derivados do gráfico de recorrência global pode-se ter um indicativo se a série é não estacionária. A Figura 4.9 ilustra o gráfico boxplot das duas métricas extraídas do gráfico de recorrência: o número de triângulos e a área do máximo triângulo, ambos normalizados, representam respectivamente: as amostras de floresta, desflorestamento em floresta, desflorestamento em cerrado, agricultura intensa e um ruído branco. O gráfico de dispersão para essas amostras é mostrado na Figura 4.9c. Para a área do máximo triângulo os valores de desflorestamento em floresta são maiores, indicando a presença de grandes lacunas no gráfico de recorrência. Isso ocorre porque quando há um desflorestamento, o tempo de limpeza da área é bem maior numa região de floresta densa do que numa região de cerrado, fazendo com que os valores de índice permaneçam baixos durante um longo período, muitas vezes mais que anos. Essa diferença de amplitude dos valores de EVI2 entre o período de floresta e após o desflorestamento com um longo período com pouca cobertura vegetal é visível no gráfico de recorrência (Figura. 4.9b).

Figura 4.7 - Gráfico de recorrência para uma série temporal de desflorestamento em floresta. a)Matriz de distância; b) gráfico de recorrência e c)série temporal.

Os resultados dos operadores também podem contribuir no processo de uma possível classificação das séries, indicando as áreas com alterações no uso com mudança no regime de variabilidade da série temporal de EVI2. Para as demais classes de uso os valores de tamanho do máximo triângulo são bem próximos e com pouca variância. O total de triângulos está relacionado com a taxa de recorrência, em que as séries de desflorestamento em cerrado apresentam os maiores valores. Esse número é um indicativo de que existem pontos extremos na série por um curto período de tempo, fazendo com que uma maior região do gráfico de recorrência esteja preenchida. Para as áreas de agricultura intensa ocorre um número menor de triângulos definido principalmente pela maior variabilidade dos dados em amplitude, isso faz com que menos pontos do gráfico de recorrência estejam preenchidos em função da maior variabilidade da distância entre os valores de EVI2 da série temporal.

Todas as classes de uso estão bem distantes do regime de variabilidade de um ruído branco, o qual serve apenas como uma referencia ilustrativa. O gráfico de dispersão indica que o incremento no total de triângulos ou na área do máximo triângulo pode ser um indicativo de mudanças. A grande dificuldade dessa metodologia está em

Figura 4.8 - Gráfico de recorrência para uma série temporal de desflorestamento em cerrado. a)Matriz de distância; b) gráfico de recorrência e c)série temporal.

identificar regimes de mudança baseado na variabilidade do índice quando não há muitas alterações nessas métricas. A Figura 4.10 ilustra um caso desses em que uma área de pivô com agricultura intensa foi convertida para áreas de cana-de-açúcar. A Figura 4.10 mostra que existe uma mudança no regime de variabilidade dos índices a partir do ano de 2007. Os valores das métricas não permitem concluir se ocorreu ou não uma grande variabilidade no regime, para esse tipo de análise são necessárias métricas que explorem o domínio da freqüência. Ressalta-se que para esse tipo de série (Figura 4.10c), com alta variabilidade nos valores de EVI2 devido ao intenso cultivo agrícola, os parâmetros da filtragem por *wavelets* necessitam ser alterados. Desta forma, recomenda-se que sejam utilizadas as séries temporais de EVI2 com resolução temporal de 8 dias para os estudos em área agrícolas. Estas séries temporais em breve também estarão disponíveis para toda América do Sul no laboratório virtual.

As técnicas e operadores computacionais ilustrados nessa seção buscam contribuir para o entendimento e o processo de visualização das mudanças de uso e cobertura da terra. Em síntese, o método do GPA G_{1A} pouco contribuiu nas etapas de inferência

Figura 4.9 - Distribuição por amostra dos valores de a)número de triângulos; b) área do máximo triângulo são normalizados e c)gráfico de dispersão entre o número de triângulos e área do máximo triângulo são normalizados

sobre o uso da terra. A maior utilidade está na avaliação da presença de ruídos nas séries, pois em teoria, séries com maior presença de assimetrias na escala mais fina podem estar associadas a um incremento de ruídos no sinal observado.

No contexto da filtragem das séries temporais de EVI2, isso pode ser um parâmetro para verificar se a série filtrada diminuiu as assimetrias numa escala mais fina. Pelo lado da visualização, a utilização do gráfico de recorrência e principalmente da matriz de distância é um apoio para análise visual das séries temporais. Utilizando as métricas derivadas do gráfico de recorrência, pode-se estabelecer parâmetros qualitativos que indiquem preliminarmente quais áreas possuem tendências de mudança. Esse procedimento em um primeiro momento pode servir para uma primeira estimativa das áreas em que ocorreram mudanças drásticas na vegetação, bastando para isso utilizar limiares rígidos e conservadores durante um processo de classificação, quando for de interesse.

4.2 Detecção de Mudanças de Uso e Cobertura da Terra com Séries Espaço-Temporais

Esta seção explora a variabilidade espaço-temporal nos dados disponíveis no laboratório virtual. Na área de análise de sinais e modelagem, Bennet (1979) propôs

Figura 4.10 - Gráfico de recorrência para uma série temporal de agricultura com mudança do tipo de cultura. a)Matriz de distância; b) gráfico de recorrência e c)série temporal

uma abordagem completa sobre os objetivos, métodos, tipos e análises de séries espaço-temporais. Esse autor destaca que um dos objetivos de estudar os sistemas espaço-temporais é tentar entender e compreender a sua variabilidade. Em uma visão sistêmica isso é realizado para propor modelos que descrevam ou analisem essa variabilidade espaço-temporal.

O emprego do termo espaço-temporal é amplamente utilizado na literatura de sensoriamento remoto e banco de dados geográficos. Em bancos de dados geográficos e sistemas de informações geográficas este termo refere-se a como representar, modelar ou analisar a variabilidade do espaço geográfico no tempo. Na área de sensoriamento remoto o mesmo é amplamente utilizado quando uma técnica é aplicada em um conjunto de imagens ou qualquer outra medida temporal. Uma análise espaço-temporal deveria tratar simultaneamente o tempo e o espaço, entretanto, em geral isto não ocorre, uma análise espaço-temporal resume-se a dois tipos de abordagens, técnicas estritamente espaciais que são aplicadas ao longo do tempo em dados espaciais ou o contrário, técnicas temporais unidimensionais aplicadas num espaço. De qualquer

forma caracteriza-se como um dado espaço-temporal um conjunto de medidas que tenha pelo menos duas dimensões, sendo uma o tempo pelo menos uma dimensão no espaço. Uma análise em um conjunto de dados com essas dimensões constitui uma análise espaço-temporal seja em física, modelos matemáticos, geografia, biologia ou qualquer outra disciplina. Uma forma de organização para esse tipo de dado é proposta por Veronese (2011).

Nas ciências aplicadas que utilizam dados reais ou experimentais o termo espaço-temporal tornou-se recorrente devido ao desenvolvimento e a popularização de novos sensores, principalmente de dispositivos eletrônicos, por isso a coleta e o tratamento dos dados espaço-temporais torna-se cada vez mais comum. Uma analogia comparativa com a mecânica dos fluidos permitiria identificar dois tipos de coletas de dados com abordagens diferenciadas: i)abordagem Lagrangeana: coleta dados espaciais ao longo do tempo por um sensor variando a posição no espaço, e; ii) abordagem Euleriana: observação de um mesmo espaço ao longo do tempo. Os sistemas móveis de coletas de dados ilustram uma abordagem Lagrangeana, um exemplo típico são os derivadores oceânicos que coletam vários parâmetros físicos e oceanográficos ao longo do tempo em coordenadas geográficas diferentes. As imagens de sensoriamento remoto são um exemplo típico de abordagem Euleriana, onde imagens observando uma mesma região geográfica (espaço) são adquiridas ao longo do tempo.

Nesse trabalho utilizou-se a abordagem Euleriana para estudar o uso e cobertura da terra. O pressuposto básico consiste em considerar que as variações espaciais de uso e cobertura da terra alteram a vegetação que, por sua vez, pode ser observada numa seqüência temporal de imagens de sensoriamento remoto. Esse trabalho considera que se existe mudanças drásticas no regime de variabilidade da vegetação ao longo do tempo pode haver alterações de uso e cobertura da terra.

Em uma associação com a física-estatística pode-se considerar que uma determinada área com mudança no seu uso e cobertura modifica o seu regime de variabilidade espaço-temporal, como num processo de transição de fase. Em uma analogia pode-se dizer que, se uma região tradicionalmente coberta por áreas de pastagem modificou seu uso para áreas de agricultura, haveria mudança no regime de variabilidade espaço-temporal da vegetação, alterando o ecossistema, a economia, bem como diversos outros fatores. O problema principal consiste em verificar se é possível identificar essas áreas com mudanças diretamente em imagens de sensoriamento remoto. Do ponto de vista dos estudos de mudanças de uso e cobertura da terra isso poderia contribuir para regiões pouco estudadas com a carência de mapas e outras

informações sobre o uso da terra. Na parte dos sistemas de detecção de mudanças serve como um alerta inicial indicando os *hotspots* de mudança de uso e cobertura da terra, para um posterior detalhamento, utilizando o método mais apropriado como proposto por Lu et al. (2004). Nesse sentido é apresentada a seguir a metodologia e os resultados de aplicação dos operadores computacionais em dados espaço-temporais descritos no Capítulo 2.

A metodologia básica é dividida em dois passos: i) extração de uma métrica a partir das séries espaço-temporais, esse procedimento é realizado com a aplicação dos operadores computacionais descritos no Capítulo 3, e; ii) elaboração de métricas para indicar as regiões com grandes alterações no uso e cobertura da terra. Inicialmente, o algoritmo deve ser capaz de recuperar, recortar e elaborar séries espaço-temporal a partir das séries temporais disponíveis no laboratório virtual. Isso é realizado pelo módulo de gerenciamento de dados que permite recuperar um conjunto de séries temporais por limites geográficos e intervalo de datas previamente estabelecidos. Em um passo seguinte, o conjunto de operadores computacionais descritos no capítulo 3 é aplicado no domínio espacial para cada data da série espaço-temporal, ou seja, em cada recorte de imagem. Além dos operadores computacionais DFA-2D e GPA os momentos estatísticos média, variância, curtose e assimetria são calculados. Os momentos estatísticos são fundamentais para comparação e uma avaliação preliminar e estimativa de parâmetros estatísticos básicos das séries espaço-temporais.

O resultado desse primeiro passo são séries temporais dos índices α do DFA-2-D, os operadores G_{2_D} , G_{3_D}, G_{2_N} e NFr_i e os momentos estatísticos média, variância, curtose e assimetria para cada recorte geográfico previamente definido. Essa generalização da metodologia permite a sua aplicação para estudos localizados onde se deseja entender um pouco da variabilidade espaço-temporal da vegetação antes de qualquer procedimento de classificação, seleção e delimitação de áreas de estudo, associação com parâmetros biofísicos medidos em campo entre outros. Alguns exemplos são descritos no final dessa seção.

Outra aplicação da metodologia descrita anteriormente é a geração de índices de mudança espacialmente localizados. A metodologia básica consiste em obter as séries temporais dos índices utilizados para diversas amostras distribuídas espacialmente numa dada região. Essa amostragem pode ser realizada tanto de modo aleatório como sistemático e são representadas por limites geográficos de dimensão previamente estabelecidos. Para cada uma das amostras é aplicada a metodologia descrita anteriormente, gerando assim as séries temporais dos operadores propostos. Como

os valores dos índices utilizados possuem sazonalidade e amplitudes diferentes é necessário um procedimento de normalização que é realizado calculando a anomalia para cada série temporal do operador. O procedimento para o cálculo da anomalia consiste na simples operação de subtração da série original pela média mensal dividida pelo desvio padrão do período correspondente. Após cada série temporal ser transformada em anomalia é calculado o ajuste linear através do algoritmo de mínimos quadrados. Os valores de *p-value* são utilizados para identificar os ajustes significativos a $p < 0,01$.

A partir do cálculo da inclinação da reta β_1 e do limiar de significância *p-value* é elaborada uma categorização dos pontos em que ocorreram mudança. Para essa categorização de cada variável utilizou-se o *boxplot* para selecionar os pontos considerados "$outliers$" ou pontos atípicos. Esses valores atípicos do coeficiente angular da reta (β_1) indicam que ocorreu alguma mudança nos valores de EVI2 das séries espaço-temporais, logo uma mudança espacial na dinâmica da vegetação. Para definição dos valores atípicos utiliza-se a amplitude inter-quartis $IQR = q3 - q1$, composta pela diferença entre o quartil superior ($q3$) e o inferior ($q1$) da distribuição dos valores do coeficiente de inclinação da reta β_1. Os valores atípicos positivos são definidos a partir dos quartis superiores $q3 + 1,5 * IQR$, enquanto os valores inferiores foram definidos a partir do quartil inferior $q1 - 1,5 * IQR$. Esse valor de 1,5 define a posição dos *whiskers* que são os limites do *boxplot*, valores fora desse limite são considerados atípicos. O passo seguinte é realizar entre as 9 variáveis as operações de união e intersecção entre todos os conjuntos de valores atípicos de β_1. Utilizando o somatório do número de intersecções é possível identificar áreas espaciais em que ocorreram mudanças ao longo do tempo. A síntese dessa metodologia está apresentada na Figura 4.11.

O modelo de tendência é o ajuste linear simples dado por:

$$Y_t = \beta_0 + \beta_1 t + a_t, \tag{4.1}$$

onde Y_t são os valores da série temporal, no caso representando a série temporal de anomalia para cada variável a ser estudada, β_0 e β_1 os coeficientes a serem estimados pelo método de mínimos quadrados e a_t são os resíduos.

Minimizando a soma do quadrado dos resíduos :

$$f(\beta_0, \beta_1) = \sum_{t=1}^{N} (\beta_0 + \beta_1 t + at)^2. \tag{4.2}$$

Obtendo as equações normais por mínimos quadrados :

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{t=1}^{N} t = \sum_{t=1}^{N} Y_t, \tag{4.3}$$

$$\hat{\beta}_0 \sum_{t=1}^{N} Y_t + \hat{\beta}_1 \sum_{t=1}^{N} t = \sum_{t=1}^{N} tY_t. \tag{4.4}$$

As soluções são:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \bar{t} \hat{\beta}_1 = \frac{\sum_{t=1}^{N} tY_t - \left[(\sum_{t=1}^{N} t)(\sum_{t=1}^{N} Y_t) \right] /N}{\sum_{t=1}^{N} t^2 - (\sum_{t=1}^{N} t)^2/N}, \tag{4.5}$$

onde,

$$\overline{Y} = \frac{1}{N} \sum_{t=1}^{N} Y_t, \tag{4.6}$$

$$\bar{t} = \frac{1}{N} \sum_{t=1}^{N} t. \tag{4.7}$$

Já os momentos estatísticos são definidos a partir do recorte da matriz A e são calculados para cada recorte da matriz em cada data da imagem t.

Momento de primeira ordem média:

$$\mu_{1t} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} A_t(i, j). \tag{4.8}$$

Momento de segunda ordem variância:

$$\mu_{2t} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (A_t(i,j) - \mu_{1t})^2. \tag{4.9}$$

Momento de terceira ordem assimetria:

$$\mu_{3t} = \frac{1}{(\mu_{2t})^2} \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (A_t(i,j) - \mu_{1t})^3. \tag{4.10}$$

Momento de quarta ordem curtose:

$$\mu_{4t} = \frac{1}{(\mu_{2t})^4} \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (A_t(i,j) - \mu_{1t})^4. \tag{4.11}$$

em que, N e M são as dimensões do recorte de matriz, i e j a posição em linha e coluna dos elementos da matriz, no caso de um recorte de imagens a posição dos pixels.

Figura 4.11 - Síntese da metodologia para espacialização de pontos de mudança.

Para exemplificar o trabalho são selecionadas todas as imagens da porção centro-sul da América do Sul. O primeiro passo para aplicação da metodologia é a distribuição de amostras sistemáticas de 32x32 km ao longo de toda área de estudo. Para cada amostra são elaboradas séries espaço-temporais de 32x32 km representadas por uma imagem de dimensão 128x128 pixels. Cada série espaço-temporal é composta por 217 imagens, adquiridas no intervalo de tempo entre agosto de 2000 e dezembro de 2009. Para cada imagem da série espaço-temporal são aplicados os operadores computacionais DFA-2D, GPA e calculados os momentos estatísticos. A partir da transformação das séries espaço-temporais em 9 séries temporais representantes dos padrões de variabilidade espacial é realizado, para cada série, o cálculo de anomalia e o ajuste linear representado pelo coeficiente angular e pelo *p-value*. Os valores do coeficiente angular (β_1) são utilizados como um índice para identificar as séries com tendências. O conjunto da distribuição de tendências β_1 por variável é apresentado na Figura 4.12. Nessa Figura os pontos em vermelho são os valores identificados como atípicos representando as amostras com tendência em alguma dessas variáveis.

Figura 4.12 - Distribuição dos valores de $\beta 1$ para os valores de anomalia.

O passo seguinte consiste na espacialização dos pontos atípicos, indicando onde estão localizados esses pontos. Considerando que cada amostra representa uma região geográfica isso pode ser representado em um mapa de proporções, utilizando a coordenada geográfica central do recorte de imagem amostrado.

A Figura 4.13 exibe a espacialização desses pontos indicando áreas com grande mudança na dinâmica da vegetação. A Figura 4.13 representa a operação de intersecção entre os valores atípicos, onde o número 7 indica que 7 variáveis foram consideradas

pontos atípicos.

Figura 4.13 - Mapa da distribuição de *hotspots* de pontos de mudança.

Na Figura 4.13 pode-se observar uma grande concentração de pontos no estado de Mato Grosso e Minas Gerais. No Mato Grosso esta concentração é consequência do processo de desflorestamento da floresta amazônica onde os maiores índices de

supressão vegetal concentram-se na região chamada de arco do desflorestamento. No estado de Minas Gerais, a grande maioria dos pontos está relacionada ao reflorestamento ocorrido na porção nordeste do estado, o qual apresentou um grande contraste com a vegetação de cerrado. Essas grandes mudanças estão também relacionadas a instalação de novos reservatórios hidrelétricos como é o caso da divisa entre Goiás e Tocantins. Os pontos no estado de São Paulo estão muito relacionados a expansão do cultivo de cana-de-açúcar. Para chegar a essas conclusões é necessário utilizar a interpretação das séries temporais de EVI2 disponíveis no laboratório virtual conforme descrito no Capítulo 2. Nesse sentido as séries das 9 variáveis também estão integradas ao laboratório virtual para regiões previamente fixadas no espaço. Isso permite a análise conjunta entre as séries temporais e o espaço temporal de EVI2. Esses índices podem servir aos tomadores de decisão, pesquisadores e à população em geral como um indicativo em pequenas e grandes escalas de onde estão ocorrendo mudanças na dinâmica da cobertura vegetal associadas ao uso ou cobertura da terra. Nos exemplos a seguir são exploradas a variabilidade dos índices espaciais para alguns tipos de aplicações de uso e cobertura da terra.

O índice de Frobenius, por exemplo, está relacionado a mudança de estruturas ortogonais e tipo fonte e sumidouro [1] no regime de variabilidade espacial. Já as diversidades de fase e norma são medidas do grau de desorganização do campo gradiente. A norma de Frobenius do conjunto de normas é uma medida de energia do campo gradiente e ressalta o número e a intensidade de transição entre dois segmentos homogêneos na imagem, ou seja, as intensidades de borda de uma imagem. O expoente α do DFA-2D é uma medida do grau de fragmentação da imagem, onde valores próximos a 1 indicam regiões bem fragmentadas e homogêneas. Os valores bem maiores que 1 são relacionados a diminuição da fragmentação e ao surgimento de segmentos homogêneos maiores na imagem. Os momentos estatísticos são medidas no domínio escalar da imagem e estão relacionados à amplitude e ao tipo de distribuição estatística da imagem.

O exemplo da Figura 4.14 e 4.15 mostra a variabilidade dinâmica de uma área localizada no estado do Rio Grande do Sul com intenso uso agrícola. Os padrões de variabilidade das séries indicam um regime de alta frequência, refletindo o padrão de variabilidade da vegetação ao longo do ano. As áreas de agricultura intensa possuem dois ciclos de cultivo anual, modificando intensamente a cobertura vegetal no decorrer de um ano. A flutuação de α próximo a 1 indica uma área bem fragmentada com

[1] Padrões no campo gradiente indicando áreas de convergência e divergência em um determinado ponto.

estruturas espaciais pequenas, esse é um padrão típico das regiões que possuem um intenso cultivo em áreas pequenas. As séries temporais dos momentos estatísticos em geral ressaltam as mesmas variações das métricas propostas.

O maior valor no índice de vegetação em 2006 estava relacionado ao retardo da safra agrícola de 2006/2007 que ocorreu devido a um período de seca observado em 2005. Este fenômeno pode ser observado como pontos atípicos na série de diversidade de norma e na série de curtose no ano de 2005. Para esse último tipo de interpretação é necessário um conhecimento adicional sobre a região de estudo, aspecto fundamental para qualquer tipo de análise de mudança de uso e cobertura da terra.

Figura 4.14 - Sequencia de imagens para uma área agrícola.

Na Figura 4.16 e 4.17 é apresentado um exemplo de mudança drástica no regime de variabilidade de seis séries temporais, ocorrida em função da instalação do reservatório hidrelétrico que ocorreu entre 2005 e 2006. Somente com a análise da média, variância e diversidade de fase não seria possível identificar tal mudança no padrão de variabilidade. O decaimento da flutuação do índice *alpha* é menor uma vez que, após o enchimento do reservatório, as amplitudes de variabilidade da vegetação tornam-se menores dada a menor proporção de cobertura vegetal.

A Figura 4.18 e 4.19 apresenta um exemplo de uma área de reflorestamento, em que uma grande área de eucalipto foi plantada numa região com uma cobertura vegetal menos densa. Isso provoca um contraste entre a vegetação natural e as áreas reflorestadas. Essas áreas de reflorestamento são caracterizadas por grande biomassa

Figura 4.15 - Distribuição das 9 variáveis para área de agricultura.

e um ciclo fenológico mais longo alterando o padrão de variabilidade espacial da vegetação devido ao surgimento de estruturas grandes e homogêneas. Avaliando as

Figura 4.16 - Sequencia de imagens para uma área de instalação de reservatório hidrelétrico.

séries temporais dos operadores da Figura 4.19 há tendências na maioria da variabilidade temporal dos operadores. Nos valores de variância e assimetria a mudança de variabilidade é mais evidente a partir do ano de 2005. Os valores de α aumentam ao longo do tempo indicando a presença de estruturas maiores e mais homogêneas na imagem.

O caso de uma área de desflorestamento, apresentado na Figura 4.20 corresponde a um intenso desflorestamento que ocorreu nos últimos 11 anos, alterando o regime de variabilidade das séries temporais a partir do ano de 2004. Nesse exemplo, a diminuição nos valores da média (Figura 4.21c), bem como a mudança no regime de variabilidade observado na diversidade de fase e de norma indicaram a ocorrência de mudança.

Além dos exemplos apresentados, todas as séries temporais de variabilidade do índice de vegetação estão integradas ao globo virtual permitindo o cálculo por coordenada geográfica e período analisado. Medidas simplificadas como estatística descritivas anuais das 9 métricas calculadas podem auxiliar nas inferências sobre as mudanças e diferenças numa escala temporal anual. Isso abre a possibilidade de analisar, comparar estas séries com outras variáveis coletadas em campo e verificar se existe alguma correlação com a variabilidade dinâmica da vegetação ou do uso e cobertura da terra.

Um exemplo da Figura 4.22 ilustra o cálculo das médias de 11 anos de dados,

Figura 4.17 - Distribuição das 9 variáveis da área de instalação de reservatório hidrelétrico.

de julho de 2000 a julho de 2011, para os valores de α do DFA-2-D e média de EVI2 para uma escala de 32x32 km. Os valores médios de α (Fig. 4.22a) indicam

Figura 4.18 - Sequencia de imagens para uma área de reflorestamento.

que o comportamento dinâmico espaço-temporal da vegetação segue diferentes leis de escala de acordo com o bioma e com a variabilidade espacial e temporal da vegetação. Os valores α mostram que existe uma similaridade entre os valores no nordeste brasileiro e na região da frontreira entre Bolívia, Argentina e Paraguai. Tal similaridade pode estar relacionada ao tipo de vegetação e principalmente aos efeitos da sazonalidade sobre a fenologia da vegetação.

Os valores de α podem auxiliar na regionalização de modelos, indicando áreas que a vegetação possui um comportamento espaço-temporal similar. Isso também pode ser de grande valia para o entendimento dos processos dinâmicos da vegetação e como isso ocorre em escalas espaciais e temporais. Um grande desafio para trabalhos futuros torna-se definir caminhos para correta interpretação dos valores de α e sua correlação com as demais variáveis. Os valores médios de EVI2 (4.22b) estão relacionados a quantidade média de biomassa, ressaltando a grande variabilidade nos padrões da vegetação na América do Sul. Ressalta-se que valores elevados de EVI2 não estão correlacionados linearmente com a biomassa e sim com a vegetação fotosinteticamente ativa.

Do lado computacional um conjunto de ferramentas de reconhecimento de padrões e mineração de dados pode ser utilizado para a procura de um determinado padrão de variabilidade no conjunto de múltiplas séries. A principal vantagem dos métodos apresentados refere-se a possibilidade de análise visual e exploratória dos regimes de variabilidade da dinâmica da vegetação. Por outro lado, a necessidade de um

Figura 4.19 - Distribuição das 9 variáveis da área de reflorestamento.

conhecimento prévio não intuitivo, pode ser visto como uma desvantagem para a interpretação e análise das séries temporais de EVI2 apresentadas no Capítulo 2.

Figura 4.20 - Sequencia de imagens para uma área de desflorestamento.

Ressalta-se que essa mesma metodologia pode ser aplicada para o entendimento da variabilidade de outros tipos de séries temporais e espaço-temporais, utilizando os algoritmos já implementados e disponíveis. A síntese das implementações computacionais realizadas está descrita no Apêndice D.

Figura 4.21 - Distribuição das 9 variáveis da área de desflorestamento.

Figura 4.22 - Distribuição dos valores médios de 2000 a 2010 na escala de 32x32km para os valores de: a) α do DFA-2-D e b) média de EVI2.

5 CONCLUSÕES

When you have only one way of expressing yourself, you have limits that you dont appreciate. When you get a new way to express yourself, it teaches you that there could be a third or a fourth way. It opens up your eyes to a much broader universe.
(*David Donoho*)

A pesquisa de doutorado em computação aplicada realizada nesse trabalho estudou a combinação de técnicas de análise de séries temporais, originalmente introduzidas pela física computacional, para suporte *à* visualização e caracterização do uso e cobertura da terra utilizando dados de sensoriamento remoto.

A definição e implementação dos algoritmos a partir do método DFA, análise de padrões de gradiente e gráfico de recorrência permitiu discutir alguns aspectos computacionais técnicos das metodologias. Alguns desses aspectos computacionais até então não foram descritos na literatura. Quanto a análise de padrões de gradiente, especificamente no cálculo do primeiro momento gradiente, recomenda-se que trabalhos futuros concentrem-se na definição de simetria e na utilização de meta-heurísticas para busca e eliminação de vetores simétricos. Um bom começo seria explorar grupos de simetria no plano que são utilizados para geração de padrões espaciais repetitivos. Entre eles destacam-se os 17 grupos de simetria no plano, grupos de *Rosette* e *Frieze*, e para três dimensões os grupos cristalográficos (MAINZER, 2005ab).

A implementação computacional de dinâmicas simplificadas para geração de padrões permitiu a construção de um conjunto de dados canônicos. Esses além de servirem de suporte para validação dos operadores computacionais propostos, em trabalhos futuros permitirão estudos específicos utilizando as características de cada modelo implementado. O estudo e a implementação dos operadores computacionais podem ser utilizados não somente na área de sensoriamento remoto, mas nas aplicações do grupo de pesquisa do LAC na área de Física Computacional de Sistemas Complexos: Aplicações Espaciais e Geofísicas do INPE ($http://dgp.cnpq.br/buscaoperacional/detalhegrupo.jsp?grupo = 008710519IU3BKidentificacao$).

O conceito de Laboratório Virtual de Sensoriamento Remoto para Visualização e Análise da Dinâmica do Uso e Cobertura da Terra, uma contribuição inédita, devido ao potencial de popularizar a utilização dos dados de sensoriamento remoto. A integração das séries temporais com o globo virtual abre caminhos para novos estudos explorarem tanto os detalhes da escala espacial em um regime estático dado

pela imagem de alta resolução disponível no globo virtual, quanto à dinâmica dessa área representada pela série temporal dos índices de vegetação. No contexto do INPE, o Laboratório Virtual constitui uma ferramenta complementar as atividades de pesquisa em sensoriamento remoto que pode ser útil na validação de mapas de desflorestamento, classificações do uso e cobertura da terra e nos estudos em agricultura e floresta.

A utilização do Laboratório Virtual de Sensoriamento Remoto para Visualização e Análise da Dinâmica do Uso e Cobertura da Terra não está restrita aos dados MODIS. Existe uma continuidade nos programas de observação da terra com sensores de moderada resolução espacial, um exemplo é o programa de novos satélites orbitais de órbita polar da NASA-DoD-NOAA como o NPOESS - *National Polar-orbiting Operational Environmental Satellite System* que prevê uma série de sensores similares ao MODIS chamado de *Visible Infrared Imaging Radiometer Suite* VIIRS. Esse sensor já está em fase pré-operacional com o satélite NPP (*NPOESS Preparatory Project*) renomeado para *Suomi National Polar-orbiting Partnership*.

O estudo de caso utilizando os operadores computacionais e os dados do laboratório virtual proporcionam uma nova abordagem nos estudos da caracterização do uso e cobertura da terra. Nesse sentido, o gráfico de recorrência representa uma ferramenta potencial para visualização das séries temporais de sensoriamento, permitindo uma visualização complementar das séries temporais. Os operadores computacionais derivados desse método podem ser utilizados previamente para identificação de áreas com tendências e pontos atípicos. O método DFA 2-D é uma abordagem para analisar imagens de sensoriamento remoto permitindo a detecção de áreas fragmentadas diretamente nas imagens. O conceito de *hotspots* de mudanças podem auxiliar na identificação das regiões com mudanças drásticas na estrutura espacial da vegetação. Essa análise contribuirá para as regiões em que há escassez de mapas de uso e cobertura da terra. A mesma análise poderá indicar em escala local se existem mudanças nos regimes de variabilidade da vegetação, logo essas podem estar associadas a mudança do uso e cobertura da Terra. Além disso, pode ser aplicada para outras variáveis como temperatura de superfície e outros índices de vegetação.

Entre as perspectivas futuras que serão desenvolvidas a partir desta pesquisa de doutorado, destacam-se as seguintes:

a) Estudo do gráfico de recorrência utilizando medidas quantitativas (ZBILUT e WEBBER JR., 1992; WEBBER JR. e ZBILUT, 1994, MARWAN

et al. 2002) bem com a extensão para outras dimensões como proposto por Marwan (2007).

b) Uso de segmentação multitemporal das imagens disponíveis no laboratório virtual para definição de segmentos homogêneos e generalização das séries temporais para esses segmentos.

c) Extensão geográfica do Laboratório Virtual para outros continentes, prioritariamente para África onde há uma escassez de dados geográficos e uma grande demanda por estudos do uso e cobertura da terra. Inclusão de outras séries temporais como temperatura da superfície, focos e cicatrizes de queimadas e outros índices de vegetação.

d) Avaliação do Médodo DFA-2D e GPA na caracterização espacial de outros dados de sensoriamento remoto, como imagens de RADAR e LIDAR.

e) Escolha e elaboração de um índice equivalente ao EVI2 para estudar turbidez e clorofila em lagos e reservatórios.

Trabalhos em desenvolvimento:

a) Metodologias para validação de mapeamentos da cana-de-açúcar pelo projeto CanaSat baseadas nas séries temporais disponibilizadas pelo globo virtual.

b) Estudos de métodos pedagógicos e recursos computacionais de computação gráfica e simbologia para o treinamento interativo de usuários não especialistas para interpretação e utilização das séries temporais disponíveis no laboratório virtual.

c) Aprofundamento e definição do conceito de pixels com grande amplitude em um curto período de tempo. Esse conceito será fundamental para o mapeamento e monitoramento de algumas culturas agrícolas bem como o mapeamento de alguns tipos de vegetação sazonalmente inundadas nas regiões alagáveis.

d) Extensão dos conceitos utilizados para oceanografia disponibilizando as séries processadas de variáveis fundamentais com temperatura da superfície do mar, clorofila, campos de velocidade, precipitação e altimetria dos oceanos.

Por fim, essa tese ofereceu a oportunidade de pesquisar, estudar e implementar diversas ferramentas para o estudo de séries temporais e espaço-temporais. Num contexto geral os métodos computacionais da física estatística computacional possuem grande potencial de utilização nas ciências aplicadas, sua maior vantagem é estarem relacionados a um processo físico passível de interpretação e associações com fenômenos universais. Por outro lado o desconhecimento do usuário comum de todos os conceitos teóricos e formais da física e matemática dificulta muitas vezes a interpretação e a sua ampla utilização.

Outro aspecto a ressaltar é a construção de elo muito importante entre a computação aplicada e o sensoriamento remoto aplicado. Essa integração de conhecimento e pesquisa permitiu uma forma inédita de disponibilização de dados e ferramentas computacionais que são úteis nas mais diversas áreas do conhecimento como na agropecuária, ciências florestais, engenharia ambiental enfim em muitos estudos da nova ciência do sistema terrestre.

Ainda é importante ressaltar que o trabalho de tese está inserido no contexto de ambas as linhas de pesquisa do curso de pós-graduação em computação aplicada, a modelagem computacional e a tecnologia da informação e extração de informações. Na primeira linha através do estudo de métodos computacionais pesquisados no contexto dos sistemas dinâmicos complexos. Na segunda linha com a combinação de conhecimento para a manipulação, processamento e disponibilização, e extração de informações de dados no contexto espacial.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRY, P.; SELLAN, F. The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation, **Appl. and Comp. Harmonic Anal.**, v.3, n.4, p. 377-383. 1996.

ADDISON P. S. **Fractals and chaos**: an illustrated course. CRC Press, 1997. 256p.

AFSARMANESH, H.; BENABDELKADER, A.; KALETAS, E.C.; GARITA, C. HERTZBERGER, L.O. Towards a multi-layer architecture for scientific virtual laboratories. **HPCN** v. 1823, p.163-176. 2000.

AFSARMANESH, H. A reference architecture for scientific virtual laboratories. **Future Generation Computer Systems**, v17, no. 8: p.999-1008. 2001.

AGUIRRE, L. A. **Introdução a identificação de sistemas**: técnicas lineares e não-lineares aplicadas a sistemas reais. 3 ed. Belo Horizonte: UFMG, 2007. 730p.

ALVAREZ-RAMIREZ J.; RODRIGUEZ E.; ECHEVERRIA J. C. Detrending fluctuation analysis based on moving average filtering, **Physica A: Statistical Mechanics and its Applications**. v. 354,n. 15 P. 199-219. August 2005.

ALVAREZ-RAMIREZ, J.; RODRIGUEZ, E.; CERVANTES, I.; ECHEVERRIA, J.C. Scaling properties of image textures: a detrending fluctuation analysis approach, **Physica A: Statistical Mechanics and its Applications**, v. 361, n. 2, p. 677-698 , mar. 2006.

ANDERSON E.; BAI Z.; BISCHOF C.; BLACKFORD S.; DEMMEL J.; DONGARRA J.; DU CROZ J.; GREENBAUM A.; HAMMARLING S.; MCKENNEY A.; SORENSEN D.**LAPACK users' guide**. 3. ed. Philadelphia: SIAM, 1999.

ANDRADE, A. P. A.; RIBEIRO, A. L. B.; ROSA, REINALDO ROBERTO. Gradient pattern analysis of cosmic structure formation: Norm and phase statistics. **Physica D - Nonlinear Phenomena**, v. 223, n. 2, p. 139-145, 2006.

ARANSON, I.S.; KRAMER, L. The world of the complex Ginzburg-Landau equation. **Rev. Mod. Phys.** 74:99-143, 2002.

ARNEODO, A. ; BACRY, E.; GRAVES, P.V.; MUZY, J.F. Characterizing long-

range correlations in DNA sequances from wavelets analysis. **Phys. Rev. Lett.** v.74, n.16, p. 3293-3296, 1995.

ASSIREU, A. T.; ROSA, R.R.; VIJAYKUMAR, N. L.; LORENZZETTI, J.A.; REMPEL, E. L.; RAMOS, F. M.; ABREU, L. D.; BOLZAN, M. J. A.; ZANANDREA, A. Gradiente pattern analysis of short nonstationary time series: an application to lagrangian data from satellite tracked drifters. **Physical D: Nonlinear Phenomena**, v. 168-169, p. 397-403, Aug. 2002. Publicado como: INPE-9012-PRE/4690.

AUSLOOS M.; IVANOVA K. Introducing false EUR and false EUR exchange rates. **Physica A: Statistical Mechanics and its Applications,**v. 286, n. 1-2, , p. 353-366,October 2000.

BAK P.; TANG C.; WEISENFELD W. Self-organized criticality: an explanation of 1/f noise, **Phys. Rev. Lett.**, v. 59 p.381-384, 1987.

BALLAGH, L. M.; RAUP, B. H.; DUERR, R. E.; KHALSA, S. J. S.; HELM, C.; FOWLER, D.; GUPTE, A. Representing scientific data sets in KML: methods and challenges,**Computers Geosciences**, v. 37, 1, Virtual Globes in Science, p. 57-64. 2011 ISSN 0098-3004, DOI: 10.1016/j.cageo.2010.05.004.

BARBER, C.B.; DOBKIN, D.P.; HUHDANPAA, H.T. The Quickhull algorithm for convex hulls, **ACM Trans. on Mathematical Software**, 22(4):469-483, Dec 1996. Disponível em: http://www.qhull.org

BARBOSA, EDUARDO BATISTA DE MORAES. **Estudos da variabilidade de séries temporais ambientais através de espectros de assimetria implementados em software cientifico livre.** 2006-05-26. 96 p. (INPE–TDI/). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2006.

BARDET, J.-M.; LANG, G.; OPPENHEIN, G.; PHILIPPE, A.; STOEV, S.; TAQQU, M.S. Generators of long-range dependence processes: a survey, **Theory and applications of long-range dependence**, Birkhäuser, p. 579-623. 2003

BARDET J.M.; KAMMOUN, I. Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes. **IEEE Transactions on Information Theory**, v. 54, n. 5, may 2008.

BARNES, W.L.; PAGANO, T.S.; SALOMONSON, V.V. Pre-launch characteris-

tics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. **IEEE Transactions on Geoscience and Remote Sensing** , v.36, n.4 , July, p.1088-1100, 1999.

BARONI, M. P. M. A.; ROSA, R. R.; SILVA, A. F.; PEPE, I. ; ROMAN, L. S.; RAMOS, F. M.; AHUJA, R.; PERSSON, C.; VEJE, E. Modeling and gradient pattern analysis of irregular SFM structures of porous silicon . **Microelectronics Journal**, v. 37, n. 4, p. 290-294, 2006.

BASHAN, A.; BARTSCH, R.; KANTELHARDT, J. W.; HAVLIN, S. Comparison of detrending methods for fluctuation analysis, **Physica A: Statistical Mechanics and its Applications**, v. 387, n. 21, p. 5080-5090,Sep. 2008.

BENNETT, R. J. **Spatial time series.** London, Pion. 1979. 674p.

BERG MARK , CHEONG OTFRIED , KREVELD MARC VAN , OVERMARS MARK . **Computational geometry**: algorithms and applications, (3rd revised ed.), Berlin, Springer-Verlag. 2008. 386p.

BERGÉ, PIERRE, YVES POMEAU, MONIQUE DUBOIS-GANCE. **Dos ritmos ao caos.** Tradução: Roberto Leal Ferreira. São Paulo: Editora UNESP, 1996. 301p.

BOASHASH, B., **Time-frequency signal analysis and processing:** a comprehensive reference, Oxford, Elsevier Science. 2003, 743p.

BONNEY, R.; COOPER, C.B.; DICKINSON, J.; KELLING, S.; PHILLIPS, T.; ROSENBERG, K.V.; SHIRK, J. Citizen science: a developing tool for expanding science knowledge and scientific literacy. **Bioscience** v.59, p.977-98, 2009.

BOX, G. E. P.; JENKINS, G. M.; REISENSEL, G. **Time series analysis : forecasting and control**. 03. ed., Englewood Cliffs, Prentice Hall. 1994. 575 p.

BRETHERTON, F. P.; SINGLEY, P.T., Metadata: a user's view. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, (VLDB'94), 20., 1994, Santiago de Chile, Chile. **Proceedings. . .** Santiago: Morgan Kaufmann, 1994. ISBN 1-55860-153-8.

BRITO, R. C. **Ambiente gráfico flyby para análise de padrões gradientes de silício poroso.** 2004-05-14. 120 p. (INPE-12894-TDI/1009). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004.

BUTLER, D., Virtual globes: the web-wide world. **Nature**, v.439 (7078), p. 776-778. 2006.

CARBONE A.; CASTELLI G.; STANLEY H.E. Time dependent Hurst exponent in financial time series, **Physica A**, v.344, p267-271, 2004.

CARNEIRO, T. G. S. **Nested-ca**: a foundation for multiscale modelling of land use and land cover change. 2006. 114 p. (INPE-14702-TDI/1227). Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006.

CHATÉ, H. Spatiotemporal intermittency regimes of the one-dimensional complex Ginzburg-Landau equation, **Nonlinearity** v.7 p.185-204, 1994.

CHATFIELD C., **The analysis of time series:** an introduction. 6ed., New York, Chapman Hall, 2004. 283 pp

CHEN, X.; LIN, G. ; FU, Z. Long-range correlations in daily relative humidity fluctuations: A new index to characterize the climate regions over China, **Geophys. Res. Lett.**,v. 34, L07804, 2007.

CHIANCA, C.V.; TICONA, A.; PENNA, T.J.P. Fourier-detrended fluctuation analysis, **Physica A: Statistical Mechanics and its Applications** v. 357, n. 3-4, , p. 447-454, nov. 2005.

CHIANG, G.; TOBY O.H.; DOVE, M. T.; BOVOLO, C. I.; EWEN, JOHN. Geovisualization Fortran library. **Computers Geosciences**, v. 37, 1, Virtual Globes in Science, p. 65-74, 2011. DOI: 10.1016/j.cageo.2010.04.012.

CHRISTOFOLETTI, A. **Modelagem de sistemas ambientais.** São Paulo, Edgar Blücher, 2000. 236p.

COOLEY, J. W; TUKEY,J. W. An algorithm for the machine computation of the complex fourier series,**Mathematics of Computation**, Vol. 19, p. 297-301, 1965.

COPPIN P.; JONCKHEERE I.; NACKAERTS K.; MUYS B.; LAMBIN E. Digital change detection methods, In Ecosystem Monitoring: a review. **International Journal of Remote Sensing.** Taylor e Francis Ltd., n. 9, p. 1565-1596, 10 mai. 2004

CORONADO, A.V.; Carpena P. Size effects in correlation measures. **Journal of**

Biological Physics, n.31 ,p. 121-133, 2005.

COSTA JUNIOR, R. A. ; ROSA, R. R. ; MATTEDI, A. P. ; RAMOS, F. M. . Gradient pattern analysis of extended convection-diffusion. **Physica. A**, v. 344, p. 447-455, 2004.

COSTA JÚNIOR, R. A. **Desenvolvimento e aplicações de um ambiente computacional para a análise de padrões - gradientes**. 2004. 202 p. (INPE-14618-TDI/1197). Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004.

COX, S.M.; MATTHEWS, P.C. Exponential time differencing for stiff systems. **J. Comp. Phys.**, v.176 p.430-455, 2002.

DANTAS, M. S. ; ROSA, R. R. ; SANT'ANNA, N. ; CEREJA JR, M. G. ; VERONESE, T. B. ; BIANCHI, S. ; ROSA, J. C. ; ALEXIEV, K. M. ; J. D. S. DA SILVA. The VLADA white paper: building an active virtual lab for advanced data analysis. **Journal of Computational Interdisciplinary**. Sciences, v. 2, p. 47-56, 2011.

DAUBECHIES, I. **Ten lectures on wavelets.** CBMS-NSF Regional Conference Series in Applied Mathematics 61, Philadelphia, PA: Soc. Ind. Appl. Math, 1992, 377 p.

DEFRIES, R.S.; HOUGHTON, R.A.; HANSEN, M.C.; FIELD, C.B.; SKOLE, D. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. **Proceedings of the National Academy of Sciences of the United States of America** v. 99, n. 22, p. 14256-14261, Out. 2002.

DEFRIES, R., ASNER, G., HOUGHTON, R. (*Ecosystems and land use change*. Geophysical Monograph Series. Vol. 153. Washington, DC. American Geophysical Union. 2004. 309 p.

DEFRIES, R.S., Terrestrial vegetation in the coupled human-earth system: Contributions of remote sensing. **Annual Review of Environment and Resources.** v.33 p369-390. 2008.

DOMINGUES, M. O.; MENDES JR, O.; COSTA, A. M. .On wavelet techniques in atmospheric sciences. **Advances in space research**, v. 35, n. 5, p. 831-842, 2005. Publicado como: INPE-13145-PRE/8404.

DUDA, R.O., HART, P.E. **Pattern classification and scene analysis**, New York, John Wiley Sons, Inc. 1973. 512 p.

DUHAMEL, P. ; VETTERLI, M. Fast fourier transforms: a tutorial review and a state of the art, **Signal Processing**, Vol. 19, , p. 259-299, 1990.

DUMOUCHEL, W. H.; O'BRIEN, F. L. . Integrating a robust option into a multiple regression computing environment. **Computer Science and Statistics:**Proceedings of the 21st Symposium on the Interface. Alexandria, VA: American Statistical Association, 1989.

DUTRA, L.V.**Classificação de texturas usando modelos ARMA e distâncias da função de autocorrelação.** 130p. Tese (Doutorado em Computação Aplicada). PI(INPE-5067-TDL-406).Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1990.

ECKMANN J. P.; KAMPHORST S. O.; RUELLE D. Recurrence plots of dynamical systems. **Europhysics Letters** v.5 (9), p.973-977, 1987.

EDUARDO RODRIGUEZ, E.; ECHEVERRIA, J.C.; ALVAREZ-RAMIREZ,J. Detrended fluctuation analysis of heart intrabeat dynamics, **Physica A: Statistical Mechanics and its Applications**, v. 384, n. 2, p. 429-438, 2007.

EFRON, B. **The jackknife, the bootstrap, and other resampling plans.** Society of Industrial and Applied Mathematics CBMS-NSF Monographs, v.38. 1982. 100p.

EFRON, B.; R. J. TIBSHIRANI, **An introduction to the bootstrap**, Washington DC, Chapman Hall, 1993. 456p.

ERMENTROUT G.B.Stable small-amplitude solutions in reaction-diffusion systems, **Quart. Appl. Math.** n39. p.61-86, 1981.

EVA, H.; BELWARD, A; EVARISTO, M.; DI BELLA, C.;, GOND, V.; JONES, S.; SGRENZAROLI, M.; FRITZ, S. A land cover map of South America, **Global Change Biology**, v.. 10, p. 731-744. 2004.

FRAEDRICH, KLAUS; BLENDER, RICHARD. Scaling of atmosphere and ocean temperature correlations in observations and climate models. **Phys. Rev. Lett.** v.90,n. 10, p.108501-108504 ,2003.

FRANZKE, C.L.E. ; GRAVES, T. ; WATKINS, N.W.; GRAMACY, R.B.; HUGHES. C. Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity. (2012) **Philosophical Transactions of the Royal Society A**, 370(1962), p. 12501267; 2012, arXiv:1101.501

FREITAS, R. M. **Técnias de análise de séries temporais aplicadas {a** detecção de desflorestamento em tempo real 194 p. (INPE-15248-TDI/1335). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2007.

FREITAS R. M.; SHIMABUKURO, Y. E., Combining wavelets and linear spectral mixture model for MODIS satellite sensor time series analysis, **JCIS - Journal of Computational Interdisciplinary Sciences**, v. 1, p. 51-56. 2008.

FREITAS, R. M.; ROSA, R. R.; SHIMABUKURO, Y. E. Using gradient pattern analysis for land use and land cover change detection. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2010, Honolulu. **Proceedings...** Honolulu, Hawaii, USA: IEEE, 2010. v. 1. p. 3648-3651.

FREITAS, R. M.; ARAI, E.; ADAMI, M.; SOUZA, A. F.; SATO, F. Y.; SHIMABUKURO, Y. E.; ROSA, R. R.; ANDERSON, L. O.; RUDORFF, B. F. T. Virtual laboratory of remote sensing time series: visualization of MODIS EVI2 data set over south america. **Journal of Computational Interdisciplinary Sciences** v.2(1), p.57-68. 2011. DOI: 10.6062/jcis.2011.02.01.0032. http://www.epacis.net/jcis

FRIEDL, M.A.; MCIVER, D.K.; HODGES, J.C.F.; ZHANG, X.Y.; MUCHONEY, D.; STRAHLER, A.H.; WOODCOCK, C.E.; GOPAL, S.; SCHNEIDER, A.; COOPER, A.; BACCINI, A.; GAO, F.; SCHAAF, C. Global land cover mapping from MODIS: algorithms and early results. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 287-302, Nov. 2002.

FRIGO, M. ; JOHNSON, S. G. FFTW: An Adaptive software architecture for the FFT, **Proceedings...** of the International Conference on Acoustics, Speech, and Signal Processing, Vol. 3,p. 1381-1384. 1998.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SALDANHA, T. C. B.; VISANI, V.; PIMENTEL, M. F., Estudo comparativo sobre filtragem de sinais instrumentais usando transformadas de fourier e wavelet. **Química Nova**, v.24: p.874-884, 2001.

GEARY, R. C. The contiguity ratio and statistical mapping. **The Incorporated**

Statistician ,v5, n.3,p. 115-145. 1954.

GEWEKE, J.; PORTER-HUDAK S.,The estimation and application of long memory time series. **J. Time Ser. Anal.** v.4 p221. 1983.

GHIL, M.; R. M. ALLEN; M. D. DETTINGER; K. IDE, D. KONDRASHOV. Advanced spectral methods for climatic time series, **Rev. Geophys.**, v. 40, n. 1, p. 1-41, 2002.

GOODCHILD, M.F., Citizens as voluntary sensors: spatial data infrastructure in the World of Web 2.0. **Int. J. Spat. Data Infrastructures Res**. v.2, p.24-32. 2007.

GOVINDAN, R.B.; WILSON, J.D.; PREISS, H.; ESWARAN H.; CAMPBELL, J.Q.; LOWERY, C.L. Detrended fluctuation analysis of short datasets: An application to fetal cardiac data, **Physica D: Nonlinear Phenomena**, v. 226, n. 1, ,p. 23-31, feb. 2007.

GU, G.; ZHOU W. Detrended fluctuation analysis for fractals and multifractals in higher dimensions. **Phys. Rev. E**, v. 74, n. 061104, 2006.

HARALICK, R. M.; SHANMUGAN, K.; DISNSTEIN, I. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, vol. SMC-3, n. 6, November. 1973.

HARTIGAN, J.A. Statistical theory in clustering, **Journal of Classification**, v.2, p.63-76. 1985.

HENTSCHEL H.G.E.; PROCACCIA I. The infinite number of generalized dimensions of fractals and strange attractors, **Physica D**: Nonlinear Phenomena, Volume 8, Issue 3,p. 435-444, Sep. 1983. 0167-2789, 10.1016/0167-2789(83)90235-X.

HUETE, A. R.; DIDIAN K.; SHIMABUKURO Y. E.; RATANA P.; SALESKA S. R.; HUTYRA L. R.; YANG W.; NEMANI R. R.; MYNEMI R. Amazon rainforests green-up with sunlight in dry season, **Geophys. Res. Lett.**, 33, 2006.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.P;GAO, X;FERREIRA, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v.83, p.195-213. 2002.

HUFFMAN, G.J.; ADLER R.F.; BOLVIN, D.T.; GU, G., NELKIN, E.J.; BOW-

MAN, K.P.; HONG, Y.; STOCKER, E.F. ; WOLFF , D.B. The TRMM multi-satellite precipitation analysis: quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. **J. Hydrometeor.**, v.8:p.38-55. 2007.

HURST, H.R. Long-term storage in reservoirs. **Trans. Amer. Soc. Civil Eng.** v.116, p.770-799. 1951.

HUYBERS PETER; CURRY WILLIAM. Links between annual, milankovitch and continuum temperature variability **Nature** 441, p.329-332, 2006.

JIANG, Z.; HUETE, A. R.; DIDAN, K.; MIURA, T., Development of a two-band enhanced vegetation index without a blue band, **Remote Sensing of Environment**, v. 112(10), p. 3833-3845.2008.

JÖNSSON, P.; EKLUNDH, L. TIMESAT-A program for analyzing time series of satellite sensor data. **Computers Geosciences**, v.30, p.833-845. 2004.

JUSTICE, C. O.; TOWNSHEND, J.R.G.; VERMOTE, E.F.; MASUOKA, E.; WOLFE, R.E.; SALEOUS, N.; ROY, D.P.; MORISETTE, J.T. An overview of MODIS land data processing and product status. **Remote Sensing of Environment**, v. 83, p. 3-15, 2002.

KAI, SHI; CHUN-QIONGA,LIU; NAN-SHANA, AI; XIAO-HONG, ZHANG. Using three methods to investigate time-scaling properties in air pollution indexes time series. **Nonlinear Analysis: RealWorld Applications**, v.9 p.693-707, 2008.

KANEKO K. Pattern dynamics in spatio-temporal chaos, **Physica D**, v.34 1989.

KANTELHARDT J. W. ;KOSCIELNY-BUNDE EVA HENIO; REGO, H. A.;HAVLIN, SHLOMO; BUNDE, ARMIN. Detecting long-range correlations with detrended fluctuation analysis, **Physica A: Statistical Mechanics and its Applications**, v.295, n. 3-4, ,p.441-454, June 2001.

KANTELHARDT, J. W.; KOSCIELNY-BUNDE, E.; RYBSKI, D.; BRAUN, P.; BUNDE, A., HAVLIN, S. Long-term persistence and multifractality of precipitation and river runoff records, **J. Geophys. Res.**, 111, D01106, 2006.

KANTELHARDT, J. W.; STEPHAN, A.; ZSCHIEGNER, E.; KOSCIELNY-BUNDE, E.; HAVLIN, S.; BUNDE, A., STANLEY, E. H. Multifractal detrended fluctuation analysis of nonstationary time series, **Physica A: Statistical Mechanics and its Applications**, v 316, n.1-4, p. 87-114, Dec. 2002.

KANTZ, H. E SCHREIBER, T. **Nonlinear time series analysis**. Cambridge, England, University Press, 1997. 365p.

KAPLAN, L. M. Extended fractal analysis for texture classification and segmentation. **IEEE Trans. on Image Processing**, v. 8, n. 11, p. 1572-1585, Nov. 1999.

KATOK, A. ; HASSELBLATT, B. **Introduction to the modern theory of dynamical systems**. Cambridge, University Press, 1995. 793p.

KING, M.; KAUFMAN, Y; MENZEL, W. AND TANRÉ, D. Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS). **IEEE Transactions Geoscience and Remote Sensing**, v.30, n.1, Jan., 2002.

KOOPMANS, L.H. **The spectral analysis of time series**, probability and mathematical statistics 22, San Diego: Academic Press. 1995. 359p.

LAMBIN, E.F.; LINDERMAN, M. Time series of remote sensing data for land change science. **IEEE Transactions on Geoscience and Remote Sensing**, v. 44, n.7, p. 1926-1928, July 2006.

LAPOLA, D. M.; SCHALDACH, R.; ALCAMO, J.; BONDEAU, A.; KOCH, J.; KOELKING, C.; PRIESS, J. A., Indirect land-use changes can overcome carbon savings from biofuels in Brazil. **Proceedings of the National Academy of Sciences** , v. 107, n. 8, p.3388-3393. 2010.

LE PENNEC, E.; MALLAT, S. Sparse geometric image representation with bandelets, **IEEE Trans. on Image Processing**, vol 14, no. 4, p. 423-438. 2005.

LI Z.; KAFATOS M. Interannual variability of vegetation in the united states and its relation to el niño/southern oscillation. **Remote Sensing of Environment**, v. 71, n. 3, p. 239-247, 2000.

LONGLEY, P. To what extent are the fundamental spatial concepts that lie behind GIS relevant in design? In **Spatial Concepts in GIS and Design**. Santa Barbara, CA: UCSB. 2008

LONGLEY, P.; GOODCHILD, M.F; MAGUIRE, D.J.; RHIND, D.W. **Geographical information systems and science**, Nova Jersey, EUA, John Wiley Sons Inc. 2005. 472p.

LOPES R.;BETRUNI N. Fractal and multifractal analysis: A review. **Medical Image Analysis**, v.13, n.4, p634-649, 2009.

LU, D.; MAUSEL, P.; BRONDIZIO, E.; MORAN, E. Change detection techniques. **International Journal of Remote Sensing**. Taylor e Francis Ltd., n. 25, p. 2365-2407, 20 jun. 2004.

MAINZER, K. **Symmetry and Complexity.** The spirit and beauty of nonlinear science; world scientific series on nonlinear science series A, Singapore, 2005a.

MAINZER, K. Symmetry and complexity in dynamical systems. **European Review** , 2005b, 13, Supplement 2, 29-48.

MALLAT, S. A theory for multi resolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.11, p. 674-693. 1989.

MALLAT, S.**A wavelet tour of signal processing**, 3rd edition, Academic Press. 1999. 805p

MALAMUD BRUCE D., TURCOTTE DONALD L., **Self-affine time series:** I. generation and analyses, In: Renata Dmowska and Barry Saltzman, Editor(s), Advances in Geophysics, Elsevier, 1999, Volume 40, p. 1-90, ISSN 0065-2687, ISBN 9780120188406, 10.1016/S0065-2687(08)60293-9.

MANDELBROT, B. **The fractal geometry of nature.** New York: W.H. Freeman and Company. 1983. 420p.

MARTINES, Mônica de Cássia Siqueira. **Algumas observações sobre a características de Euler** : uma introdução de elementos da história da matemática no ensino médio. Dissertação de Mestrado. UNESP Rio Claro, 118p. 2009.

MARWAN N.; WESSEL N.; MEYERFELDT U.; SCHIRDEWAN A.; KURTHS J. Recurrence plot based measures of complexity and its application to heart rate variability data, **Physical Review E**, 66(2), p.026702-026710. 2002 DOI:10.1103/PhysRevE.66.026702

MARWAN N.; KURTHS J.; SAPARIN P. Generalised recurrence plot analysis for spatial data, **Physics Letters A**, 360(4-5), p.545-551, 2007a. DOI:10.1016/j.physleta.2006.08.058.

MARWAN N.; ROMANO M. C.; THIEL M.; KURTHS J. Recurrence plots for the analysis of complex systems, **Physics Reports**, 438(5-6),p.237-329. 2007b. DOI:10.1016/j.physrep.2006.11.001

MATHER, P.M. **Computer processing of remotely-sensed images**. Chichester, UK, John Wiley Sons, 1999, 292p.

MATSOUKAS, C.; ISLAM, S. ;RODRIGUEZ-ITURBE, I. Detrended fluctuation analysis of rainfall and streamflow time series, **J. Geophys. Res.**, 105(D23), 29,p165-172, 2000.

MARAUN, D.; RUST, H. W.; TIMMER J. Tempting long-memory - on the interpretation of DFA results. **Nonlin. Processes Geophys.**, v.11, p.495-503, 2004.

MEYER Y. **Wavelets and operators**, Cambridge studies in advanced math., vol. 37, Cambridge, Univ. Press, 1992, 223 p.

MICHALSKI, S. Blocks adjustment–reduction of bias and variance of detrended fluctuation analysis using Monte Carlo simulation, **Physica A: Statistical Mechanics and its Applications**, n.387, n.1, , p. 217-242, 2008.

MIELNICZUK, J.; WOJDYLLO, P. Estimation of hurst exponent revisited. **Computational Statistics e Data Analysis**. v.51 p.4510-4525, 2007.

MONTEIRO, L. H. A. **Sistemas dinâmicos** (2ed, revisada e ampliada). 2. ed. São Paulo, Editora Livraria da Física, 2006. v. 1. 625 p.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, v.37, n.1, p. 17-23. 1950.

MORAN, E.F. **Developing the amazon.** Bloomington. Indiana, EUA: Indiana University Press. 1981.

MORETTIN, P. A. **Ondas e ondaletas:** da analise de Fourier a analise de ondaletas. São Paulo, Editora da USP-SP., 1999. 272p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo, Edgard Blucher, 2004. 534p.

MUZY, J.F.,BACRY, E., ARNEODO, A. Wavelets and multifractal formalism for singular signals: Application to turbulence data. **Phys. Rev. Lett.** v.67 n.3515-

3518, 1991.

MYNENI ET AL. Large seasonal changes in leaf area of amazon rainforests. **Proc. Natl. Acad. Sci.**, v.104 p4820-4823, 2007 doi:10.1073/pnas.0611338104.

NAGARAJAN R., Reliable scaling exponent estimation of long-range correlated noise in the presence of random spikes. **Physica A: Statistical Mechanics and its Applications.** v. 366, p.1-17, 2006.

NAGARAJAN, R.; KAVASSERI,R. G. Minimizing the effect of trends on detrended fluctuation analysis of long-range correlated noise. **Physica A: Statistical Mechanics and its Applications.** v. 354, p.182-198, 2005.

NEWELL„A.C. ; WHITEHEAD J.A. Finite bandwidth, finite amplitude convection, **J. Fluid Mech.**, v.38, p.279-303, 1969.

NEWMAN, G.; GRAHAM, J.; CRALL, A.; MELINDA, L.. The art and science of multi-scale citizen science support.**Ecological Informatics**, v.6, p.217-227, 2011.

NIELSON, G. M., Visualization in scientific and engineering computation. **IEEE Computer**, n. 24 , v. 9, p. 58-66. 1991.

OSBORNE, A.; PROVENZALE, A. Finite correlation dimension for stochastic systems with power-law spectra. **Physica D**, n. 35, p. 357-381, 1989.

OKABE, A.; BOOTS, B.; SUGIHARA, K. **Spatial tessellations: concepts and applications of voronoi diagrams**. New York, Wiley, 1992.

OSWIECIMKA, P.; KWAPIEN, J.;DROZDZ, S. Wavelet versus detrended fluctuation analysis of multifractal structures. **Phys. Rev. E** v.74, n.016103, 2006.

PENG, C. -K.; BULDYREV,S. V.; Goldberger, A. L.; Havlin, S.; Mantegna, R. N.; Simons, M.; Stanley, H. E. Statistical properties of DNA sequences, **Physica A: Statistical and Theoretical Physics**, v. 221, n. 1-3, p. 180-192, 1995.

PENG, C.K.; BULDYREV, S.V.; Goldberger, A.L.; Havlin, S.; Sciortino, F.; Simons, M.; Stanley, H.E. Long-range correlations in nucleotide sequences. **Nature**, v.356, n.6365, p. 168-170, 1992.

PENG, C.K.; BULDYREV, S.V.; HAVLIN, S.; SIMONS, M.; STANLEY, H.E.; GOLDBERGER, A.L. Mosaic organization of DNA nucleotides. **Phys. Rev. E**,

v.49, p.1685-1689. 1994.

PERCIVAL, D. B.; WALDEN A. T. **Spectral analysis for physical applications**: multitaper and conventional univariate techniques. Cambridge: Cambridge University Press, 1993. 583p.

PIN, ENRICO DEL; CARNIEL, ROBERTO;TARRAGA, MARTA. Event recognition by detrended fluctuation analysis: an application to Teide-Pico Viejo volcanic complex, Tenerife, Spain, **Chaos, Solitons e Fractals**, v. 36, n. 5, ,p. 1173-1180. 2008.

PRIGOGINE, I. **Order out of chaos**,Toronto, Bantam, 1984. 349p.

RAMBALDI, G., KYEM, P.A.K., MCCALL. P, M. M.; WEINER D., Participatory spatial information management and communication in developing countries. **EJISDC. Electron. J. Inf. Syst. Dev. Countries**, p.1-9 2006.

RAMOS, F.M.; ROSA, R.R.; NETO, C.R.; ZANANDREA, A., Generalized complex entropic form for gradient pattern analysis of spatio-temporal dynamics, **Physica A**, n. 283 v.1-2: p.171-174, Aug. 2000.

ROBERTS, D.A.; KELLER, M.; SOARES, J.V. Studies of land-cover, land-use, and biophysical properties of vegetation in the large Scale Biosphere Atmosphere experiment in Amazônia. **Remote Sensing of Environment**, v. 87, p.377-388, 2003.

ROSA, R.R.; NETO, C.R.; RAMOS, F.M.; SHARMA, A.S.; VALDIVIA, J.A. Computational operator for dynamical complex pattern recognition. in: European Physical Society (Ed.), **EPS-modelling Collective Phenomena in Complex Systems**, 22F, p.304-305, 1998.

ROSA, R. R.; SHARMA, A. S.; VALDIVIA, J. A. Characterization of asymmetric fragmentation patterns in spatially extended systems. **International Journal of Modern Physics C**, v. 10, p. 147-163, 1999.

ROSA, R. R. ; PONTES, J. ; CHRISTOV, C. ; RAMOS, F. M. ; RODRIGUES NETO, C. ; REMPEL, E. L. ; WALGRAEF, D. . Gradient pattern analysis of swift-hohenberg dynamics: Phase Disorder Characterization. **Physica A**, v. 283, n. 1-2, p. 156-159, 2000.

ROSA, R. R.; RAMOS, F. M. **Análise de padrões gradientes e a física estatís-**

tica da formação de estruturas espaço-temporais. In: . Tendências da física estatística no Brasil. São Paulo: Livraria da Física, 2003.

ROSA, R. R.; CAMPOS, M. R.; VIJAYKUMAR, N. L.; FUJIWARA, S.; SATO, O. T. Gradient pattern analysis of structural dynamics: application to molecular system relaxation. **Brazilian journal of physics**, v. 33, n. 3, p. 605-610, Sep. 2003.

ROSA, R. R.; BARONI, M. P. M. A.; ZANIBONI, G. T.; da Silva, A. F.; ROMAN, L. S.; PONTES, J.; BOLZAN, M. J. A. Structural complexity of disordered surfaces: analyzing the porous silicon sfm patterns. **Physica A**, v. 386, n. 2, p. 666-673, 2007.

ROSA, R.R.; KARLICKY, M.; VERONESE, T.B.;VIJAYKUMAR, N.L.; SAWANT, H.S.; BORGAZZI,A.I.; DANTAS, M.S.; BARBOSA, E.B.M., SYCH, R.A.; MENDES, O. Gradient pattern analysis of short solar radio bursts, **Advances in Space Research**, v.42, p844-851. 2008.

ROSA, R. R. **Visualização e análise computacional de séries temporais.** São José dos Campos: INPE, 2007. Disciplina CAP-400. Notas de Aula.

ROSA, R. R. **Análise de séries temporais**, em Computação e Matemática Aplicada às Ciências e Tecnologias Espaciais. 1 ed. São José dos Campos: INPE„ v. 1, 2008, p. 200-220.

RUDORFF, B.F.T.; AGUIAR, D.A.; SILVA, W.F.; SUGAWARA, L.M.; ADAMI, M.; MOREIRA, M.A. Studies on the rapid expansion of sugarcane for ethanol production in são paulo state (Brazil) using landsat data. **Remote Sensing**, v. 2, p. 1057-1076. 2010.

RUDORFF B. F. T., ADAMI M., AGUIAR D.A.. The soy moratorium in the amazon biome monitored by remote sensing images, **Remote Sensing** v3, n1., p.185-202. 2011.

RUNNING, S. W.; JUSTICE, C.; SALOMONSON, V.V.; HALL, D.; BARKER, J.; KAUFMAN, Y.; STRAHLER, A.; HUETE, A.; MULLER, J.P.; VANDERBILT, V.; WAN, Z.M.; TEILLET, P.; CARNEGGIE, D. Terrestrial remote sensing science and algorithms planned for EOS/MODIS. **International Journal of Remote Sensing.** n.15: p. 3587-3620, 1994.

RZEPA, H, TONGE, A. TongeV. Chemlab: A virtual chemistry laboratory. **Journal of Chemical Information and Computer Science**, v38, p.1048-1053, 1998.

SAKAMOTO T.; YOKOZAWA M.; TORITANI H.; SHIBAYAMA M.; ISHITSUKA N. ; OHNO H. A crop phenology detection method using time-series MODIS data. **Remote Sensing of Environment**, v, 96, n. 3-4,p. 366-374, 2005.

SALINAS, SÍLVIO R. A.. **Introdução à física estatística.** EdUSP, 2ed., São Paulo, 1997, 472p.

SCHOWENGERDT, R. **Remote sensing**: models and methods for image processing. 2.ed. San Diego, USA: Academic Press, 1997. 522p.

SEGEL L.A. Distant side-walls cause slow amplitude modulation of cellular convection, **J. Fluid Mech.** v.38, p.203-224, 1969.

SELLAN, F. Synthese de mouvements browniens fractionnaires a l'aide de la transformation par ondelettes, **C. R. Acad. Sci.**, vol. t, no. I, p.351 - 358 , 1995.

SHANG, P.; LU, Y.;KAMAE,S.; Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis, **Chaos, Solitons Fractals**, v. 36, n.1,p. 82-90, 2008.

SHIMABUKURO, Y.S.; SMITH, J.A. The least-squares mixing models to generate fraction images derived from remote sensing multiespectral data. **IEEE Trans. Geosci. Remote Sens.**, v. 29, n. 1, p. 16-20, 1991.

SIMPSON, E. H. Measurement of diversity. **Nature**, v163, p.688. 1949.

SKOLE, D.; TUCKER, C. Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. **Science**, v. 260, p. 1905- 1910, 1993.

STANLEY, H.; E. SCALING, Universality, and renormalization: three pillars of modern critical phenomena, **Rev. Mod. Phys.** v.71, p. 358-366, 1999.

STANLEY,H. E. ; AMARAL, L. A. N.; GOLDBERGER, A. L.; HAVLIN, S.; IVANOV P. CH.; PENG, C.-K. Statistical physics and physiology: monofractal and multifractal approaches, **Physica A** v.270, p. 309-324, 1999.

STANLEY,H. E. ; AMARAL, L. A. N.; GOPIKRISHNANA, P.; IVANOV, P. CH.; KEITTB, T.H.; PLEROUA, V. Scale invariance and universality: organizing principles in complex systems. **Physica A**, v281, p60-68, 2000.

STAUDACHER, M.; TELSER, S.; AMANN, A.; HINTERHUBER, H.; RITSCH-

MARTE, M. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep, **Physica A** v.349, p.582, 2005.

STOICA, P. ; MOSES, R.L. **Introduction to spectral analysis**, Prentice-Hall, 1997, p. 319 p. STREET, J. O.; R. J. CARROLL; AND D. RUPPERT. A note on computing robust regression estimates via iteratively reweighted least squares. **The American Statistician.** Vol. 42, , p. 152-154. 1988.

STRIEDER, C. **Cálculo do coeficiente de assimetria gradiente em C/C++**: exemplos de aplicações em astrofísica e cosmologia. 2010. 103 p. (sid.inpe.br/mtc-m19@80/2010/08.03.18.34-TDI). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010. Disponível em: <http://urlib.net/8JMKD3MGP7W/3827KLH>. Acesso em: 18 fev. 2012.

STUART J.T. On the non-linear mechanics of wave disturbances in stable and unstable parallel flows, **J. Fluid. Mech**. v.9 p.353-370, 1960.

TANRÉ D.; DEROO C.; DUHAUT P.; HERMAN M.; MORCRETTE J.J.; PERBOS J.; DESCHAMPS P.Y.; Description of a computer code to Simulate the Satellite Signal in the Solar Spectrum. **International Journal of Remote Sensesing**, v.11, n. 4, p. 659-668, 1990.

TAQQU, M.S.; TEVEROVSKY, V.; WILLINGER, W. Estimators for long-range dependence: an empirical study, **Fractals**, v. 3, n. 4, p.785-798, 1995.

TELESCA, L.; LAPENNA, V. Measuring multifractality in seismic sequences. **Tectonophysics.**, v.423, p. 115-123, 2006.

TELESCA, L.; LAPENNA, V.; MACCHIATO, M. Spatial variability of the time-correlated behaviour in Italian seismicity. **Earth and Planetary Science Letters**, v.212, n.3, p. 279-290, 2003.

TELESCA, L.; LASAPONARA, R. Discriminating dynamical patterns in burned and unburned vegetational covers by using SPOT-VGT NDVI data, **Geophys. Res. Lett.**, v.32, L21401, 2005.

TELESCA, L.; LASAPONARA, R. Pre- and post-fire behavioral trends revealed in satellite NDVI time series, **Geophys. Res. Lett.**, v. 33, L14401, 2006a.

TELESCA, L.; LASAPONARA, R. Vegetational patterns in burned and unburned areas investigated by using the detrended fluctuation analysis, **Physica A: Statis-**

tical Mechanics and its Applications v. 368, n. 2, p. 531-535, 2006b.

TELESCA, L.; LASAPONARA, R., Investigating fire-induced behavioural trends in vegetation covers, **Communications in Nonlinear Science and Numerical Simulation**, v. 13, n. 9, p. 2018-2023, 2008.

THAYN, J.B.; PRICE, K.P., Julian dates and introduced temporal error in remote sensing vegetation phenology studies. **International Journal of Remote Sensing**, v.29, p. 6045-6049. 2008.

THEODORIDIS, S.; KOUTROUMBAS , K., **Pattern recognition**, 4th Edition, New Jersey, Academic Press. 984p.

TLACZALA, W., GRAJNER, G. ZAREMBA, M. Virtual laboratory with simulated nuclear experiments. **IEEE Transactions on Instrumentation and Measurement**, v. 57, n. 8, p. 1766-1770. August 2008.

TURCOTTE D. L. **Fractals and chaos in geology and geophysics**. 2 Edition, Cambridge University Press, Cambridge,UK, 1997, 398p.

VALENCIA, M.; ARTIEDA, J.; ALEGRE, M.; MAZA, D. Influence of filters in the detrended fluctuation analysis of digital electroencephalographic data. **Journal of Neuroscience Methods.** v.170, p.310-316, 2008.

VALÉRIO, A.M. ; KAMPEL, M. ; ASSIREU, A.T. ; STECH, J.L. . The asymmetric fragmentation operator applied to meteo-limnological time series in a tropical reservoir. **Ecological Informatics**, v. 8, p. 29-36, 2012.

VANDEWALLE, N.; AUSLOOS M., BOVEROUX,P. The moving averages demystified. **Physica A**, v. 269, n. 170,p. 170-176, 1999.

VASCONCELOS, D. B. **Correlação espacial em redes de mapas acoplados** 107 p. Tese. (doutorado em física). Universidade Federal do Parana, Curitiba, 2005.

VERMOTE, E.F.; TANRÉ, D.; DEUZÉ, J.L.; HERMAN, M.; MORCRETTE, J. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. **IEEE Transactions on Geoscience and Remote Sensing**. v.35, n.3, p. 675-686, May. 1997.

VERONESE, T. B. **Grade numérica generalizada**: um novo conceito para representação e visualização analítica de sistemas de séries temporais. versão: 2012-01-23.

Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011.

VERONESE, T. B.; ROSA, R. R.; BOLZAN, M. J. A.; FERNANDES, F. C. R.; SAWANT, H. S.; KARLICKY, M. Fluctuation analysis of solar radio bursts associated with. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 73, n. 11-12, p. 1311-1316, 2011.

WAN Z.; ZHANG Y.; ZHANG Q.; LI Z.-L. Validation of the land-surface temperature products retrieved from Terra Mode rate Resolution Imaging Spectroradiometer data, **Remote Sens. Environ.**, 83, 163-180, 2002.

WEBBER, C.L. JR.; ZBILUT, J.P. Dynamical assessment of physiological systems and states using recurrence plot strategies. **J. Appl. Physiol.** v.76: p.965-973.1994

WEBER, R.O.; TALKNER, P. Spectra and correlations of climate data from days to decades. **Journal of Geophysical Reseach,** v.106, n.D17, p.131-144, SEP., 2001

WERON R., Estimating long-range dependence: finite sample properties and confidence intervals, **Physica A: Statistical Mechanics and its Applications**, v.312, n. 1-2, p. 285-299, 2002.

WINTERBOTTOM, D.M., Simulating the complex Ginzburg-Landau equation in 1D using pseudo-spectral code and ETD2 exponential time-stepping. **The Complex Ginzburg-Landau equation**. University of Nottingham, 2005. Disponível em: *http : //codeinthehole.com/tutorials/cgl/references.html*. Acesso em 10 de dezembro de 2008.

WOLFE, R.E.; NISHIHAMA, M.; FLEIG, A.J.; KUYPER, J.A.; ROY, D.P.; STOREY, J.C.; PATT, F.S. Achieving sub-pixel geolocation accuracy in support of MODIS land science. **Remote Sensing of Environment**, v. 83, p. 31-49, 2002.

XIAO, X.; HAGEN, S.; ZHANG, Q.; KELLER, M.; MOORE, B.; Detecting leaf phenology of seasonally moist tropical forests in South America with multi-temporal MODIS images, **Remote Sensing of Environment**, p.465-473. 2006.

XU, LIMEI; IVANOV, P. CH.;HU, KUN;CHEN, ZH; CARBONE, A.; STANLEY, H. E. Quantifying signals with power-law correlations: A comparative study of detrended fluctuation analysis and detrended moving average techniques. **Phys. Rev. E** 71, 051101, 2005.

XU NA; SHANG, PENGJIAN; KAMAE, SANTI; Minimizing the effect of exponential trends in detrended fluctuation analysis, **Chaos, Solitons Fractals**, v.41, n.1, p.311-316, 2009.

YARON, D. KARABINOS, M. LANGE, D. GREENO, J. G. LEINHARDT G. The chemcollective: virtual labs and online activities for introductory chemistry courses, **Science**, Vol. 328, No. 5978. April, 2010.

YANG, T. Computational verb decision trees. **International Journal of Computational Cognition** (Yang's Scientific Press) v. 4, p.34-46. 2006

YUAN, Y.; SHAW M.J., Induction of fuzzy decision trees. **Fuzzy Sets and Systems**, v.69, p.125-139.1995.

ZBILUT, J.P.; WEBBER, C.L.JR. Embeddings and delays as derived from quantification of recurrence plots. **Phys. Lett . A** v.171: p.199-203. 1992.

ZHAN, H. Scaling in global ocean chlorophyll fluctuations, **Geophys. Res. Lett.**, n. 35, L01606, 2008.

ZHENG, H.; SONG, W.; WANG J. Detrended fluctuation analysis of forest fires and related weather parameters, **Physica A: Statistical Mechanics and its Applications**, v. 387, n.9, p. 2091-2099, 2008.

ZHU, X.; FRAEDRICH,K.; BLENDER, R. Variability regimes of simulated Atlantic MOC, **Geophys. Res. Lett.**, v.33, L21603, 2006.

APÊNDICE A - ANÁLISE DE SÉRIES TEMPORAIS E MÉTODO DFA

A.1 Revisão Bibliográfica do Método DFA

Neste apêndice é apresentada uma revisão bibliográfica sobre os principais fundamentos e as aplicações do método DFA. A primeira subseção descreve um panorama geral sobre a análise de séries temporais e as aplicações em sensoriamento remoto. A segunda subseção apresenta uma breve descrição dos fundamentos, variantes, propostas e críticas ao método DFA. Na segunda subseção relacionam-se os trabalhos de aplicação do DFA com ênfase na análise em sistemas ambientais. Na última subseção os principais trabalhos de extensão do método DFA para análise bi-dimensional são apresentados.

A.1.1 Análise de Séries Temporais em Sensoriamento Remoto

Nas últimas décadas, o lançamento de novas plataformas orbitais e missões multi-sensores proporcionaram um aumento progressivo do volume de dados de sensoriamento remoto, abrangendo diferentes resoluções espaciais, espectrais e temporais. Anteriormente, o conjunto de séries temporais era limitado a sensores de baixa resolução espacial como o Advanced Very High Resolution Radiometer (AVHRR) a bordo dos satélites da série NOAA e o sensor Vegetation a bordo do satélite SPOT. Entretanto, o sensor MODIS, a bordo dos satélites Terra (EOS-AM1, lançado em 1999) e Aqua(EOS-PM1, lançado em 2002) da NASA , representou uma revolução no fornecimento de imagens de satélite para estudos de séries temporais. Desde o ano de 2000, o processamento e a disponibilização das imagens deste sensor foram padronizados e validados através de uma parceria entre a NASA e um consórcio de pesquisadores, garantindo melhorias na qualidade geométrica, na resolução espacial e radiométrica destes dados (BARNES, 1998; JUSTICE et al., 2002).

A análise de séries temporais de dados de sensoriamento remoto, antes restrita a estudos globais em oceanografia e atmosfera, hoje constitui uma realidade no sensoriamento da superfície terrestre, principalmente relacionada à detecção e o monitoramento de mudanças no uso e cobertura da terra (LAMBIN e LINDERMAN, 2006). No campo ecológico, a análise de séries temporais tem permitido novas especulações sobre padrões fenológicos da vegetação, como por exemplo, o crescimento da floresta amazônica no período seco (HUETE et al., 2006; XIAO et al., 2006; MYNENI et al., 2007). Novas abordagens utilizando ferramentas como a transformada *wavelets* permitiram estudos da variabilidade inter-anual da vegetação e sua relação com anos de El-Niño (LI e KAFATOS, 2002) e determinação de padrões fenológi-

cos em culturas agrícolas (SAKAMOTO et al., 2005). Uma variedade de técnicas de análise de séries temporais tem sido empregada na área de observação da Terra. Na literatura da área de sensoriamento remoto a nomenclatura de análise de séries temporais aplica-se desde a visualização à classificação multi-temporal de imagens até o emprego de técnicas de dinâmica não linear em séries históricas de dados.

No campo da climatologia, comumente utilizado pelas comunidades de sensoriamento remoto da atmosfera e oceanografia, uma revisão das principais técnicas de análise de séries temporais é apresentada por Ghil et al. (2002). Estes autores dividem sua revisão em três classes de métodos, os quais utilizam realce da relação sinal/ruído, análise espectral e análise multivariada. Nos métodos que realçam a relação sinal/ruído destacam-se a análise do espectro de singularidades *singular spectrum analysis* (SSA), usando as funções ortogonais empíricas *empirical orthogonal functions* (EOFs) e as transformadas *wavelets*. Nos métodos de análise espectral os autores ressaltam que tanto a abordagem determinística quanto a estocástica depende de uma boa estimação do espectro de potência e, nesse contexto, os métodos tradicionais de estimativa do espectro de potência são apresentados por esses autores. Destacam-se o método do periodograma, correlograma e estimadores para reduzir tendências e a variância como a janela de Barllet e modelo Auto Regressivo (AR). Outros métodos como máxima entropia e *multitaper* também são apresentados, destacando-se nesse último a análise harmônica. Na classe dos modelos multivariados os autores descrevem o *Principal Oscillation Patterns* (POPs), extensão do método de máxima entropia utilizando modelos autoregressivos, e a análise do espectro de singularidades multicanal. Estes mesmos autores fazem uma relação entre as séries temporais e a dinâmica não linear, ressaltando a dificuldade no tratamento dos dados uma vez que na maioria dos casos as séries climatológicas são curtas e contaminadas por ruídos, porém inúmeros avanços foram destacados como o cálculo da dimensão de correlação, estimação de atratores e análise do espaço de fase.

No que tange o processo de classificação multi-temporal, outras técnicas estatísticas e de reconhecimento de padrões são comumente observadas na análise de séries temporais em sensoriamento remoto. Dentre elas destacam-se: redes neurais artificiais, árvores de decisão, análise de componentes principais, máquina de vetores de suporte, análise de agrupamentos, dentre outros (SCHOWENGERD, 1997; MATHER, 1999). A tese foi desenvolvida utilizando uma abordagem complementar aos métodos citados anteriormente, integrando técnicas intrinsecamente computacionais que surgiram através de grupos de pesquisa em sistemas complexos e física computacional com o principal objetivo de analisar séries temporais e espaço-temporais. Nesse

trabalho a definição de séries temporais é utilizada unicamente para representação unidimensional enquanto séries espaço-temporais representam a evolução de um padrão bi-dimensional no tempo conforme será descrito nas seções posteriores.

Ressalta-se que no âmbito da análise estatística de séries temporais, não abordada neste trabalho, como mencionado na introdução, destaca-se uma vasta gama de modelos estocásticos para análise e previsão como os *AutoRegressive model* (AR), *AutoRegressive Moving Average* (ARMA), *Autoregressive Conditional Heteroskedasticity* (ARCH), *AutoRegressive Integrated Moving Average* (ARIMA), *Fractional AutoRegressive Integrated Moving Average* (FARIMA), *Seasonal AutoregRessive Integrated Moving Average* (SARIMA), dentre outros (BOX et al., 1994; MORETTIN e TOLOI, 2004; AGUIRRE, 2007).

A.1.2 Espectro de Potência e Correlações de Longo Alcance

A análise de uma série temporal consiste na busca de periodicidades, tendências, pontos influentes e não-linearidades (MORETTIN e TOLOI, 2004), exigindo como conhecimento prévio a freqüência de amostragem, as escalas características e as propriedades de periodicidade da série temporal em estudo. No contexto da física estatística é importante conhecer se existem correlações entre um valor observado num instante *(t)* com valores observados nos instantes *(t+*Δt*)* ou *(t-*Δt*)*. Desta forma, é possível estudar correlações em várias escalas, comumente chamadas (*lags*) ou passos na análise estatística de séries temporais. Na estatística, por exemplo, isso é tratado como análise de auto-correlação de uma série. Nesta análise calculam-se as covariâncias entre os elementos da série para cada escala. A razão entre covariâncias nos diversos passos do tempo gera a denominada função coeficiente de autocorrelação-FAC conforme apresentado no Capítulo 3, Eq. 2.1 (MORETTIN e TOLOI, 2004).

O exemplo da Figura A.1a apresenta o cálculo da função de auto-correlação para uma série com correlação de longo alcance. Note que o decaimento da função é exponencial (Figura A.1b). A Figura A.1c apresenta como exemplo um ruído branco, ou seja, um dado estacionário e sem qualquer tipo de memória o que é confirmado pela função de autocorrelação (Figura A.1d).

Uma variedade de métodos para estudar o comportamento da função de autocorrelação está disponível na literatura. Dentre eles pode-se citar a análise de semivariogramas, a Análise R/S *Rescaled Range Statistic*, análise de média de valores extremos, técnicas espectrais como o espectro de potência e o método de transformada *wavelets* (MALAMUD e TURCOTTE, 1999). No campo da análise de sinais

Figura A.1 - Função de autocorrelação (c e d) para a série temporal sem memória (a) e uma série temporal com memória de longo alcance (c).

e na estatística de séries temporais é um tema recorrente na especificação e identificação de modelos (BOX et al., 1994; MORETTIN e TOLOI, 2004, CHATIFIELD, 2004). Este trabalho não tem o objetivo de descrever ou entrar em métodos formais da estimação de tipos de ruídos, tendência em séries temporais, dentre outros aspectos relacionados às séries temporais do ponto de vista estatístico. Para maiores detalhes sobre esses métodos recomenda-se a leitura de Chatfield (2004) que aborda de forma introdutória os principais aspectos da análise estatística de séries temporais.

A análise de periodicidades e/ou padrões cíclicos em séries temporais tem um ramo bastante definido chamado análise espectral. De uma forma bem geral, a análise espectral descreve uma série temporal ou sinal, como uma sobreposição de componentes periódicas. Seja um função contínua f(t) periódica, pode-se representar essa função domínio da frequência através da combinação linear de frequências. Essa operação é conhecida como *transformada de Fourier*. Uma extensa literatura está disponível sobre a análise de Fourier e os vários tipos de análise espectral (PERCIVAL

e WALDEN, 1993; KOOPMANS, 1995; STOICA e MOSES, 1997; MORETTIN, 1999). De uma maneira simplificada pode-se assumir que a transformada de Fourier (*G(f)*) de uma função contínua f(t) variando no tempo (t) é dada por:

$$G(f) = \int_{-\infty}^{\infty} f(t)e^{2\pi ift}dt, \tag{A.1}$$

em que $e^{-2\pi ift} = cos(2\pi ift) + isen(2\pi ift)$, f é a frequência, t o tempo, e i denota a unidade imaginária em que $i^2 = -1$.

O espectro de potência *S(f)* é a representação de todas as frequências de *G(f)* e é dado por:

$$S(f) = |G(f)|^2 . \tag{A.2}$$

A representação gráfica de $S(f)$ em função de f é conhecida como periodograma. O cálculo de $S(f)$ não é muito trivial em sinais discretos uma vez que quando discretizados os valores de $S(f)$ serão estimativas. Dentre outros métodos de estimativa de $S(f)$ podemos citar, periodograma, *Mulitapper*, método das covariâncias, Welch, dentre outros. Em geral, a grande maioria desses métodos estão baseados no cálculo da transformada de Fourier. Em um domínio discreto o cálculo da transformada de Fourier é realizado geralmente com os algoritmo chamado *Fast Fourier Transform* (FFT) (COOLEY e TUKEY, 1965). Atualmente derivações desse algoritmo estão disponíveis nas bibliotecas matemáticas dos vários softwares, uma delas é o *Fastest Fourier Transform in the West* (FFTW) (http://www.fftw.org/) proposta por Frigo e Johnson (1998). Uma revisão completa sobre o algoritmo FFT e seus derivados pode ser encontrada em Duhamel e Vetterli (1990). Neste trabalho, não tem o objetivo de descrever tais métodos, mas, sim ilustrar como é a relação entre *S(f)* e *f*.

Se essa relação segue um lei de potência pode-se dizer que,

$$S(f) \propto f^{-\beta} , \tag{A.3}$$

Como existe uma relação entre *S(f)* e a função de autocorrelação F.A.C.(Eq. 2.1, Capítulo 3) dada pelo teorema de Wiener-Khinchin, não demonstrado aqui, os ex-

poentes β e γ apresentam uma relação linear e podem ser classificados indicando a presença de persistência, não correlação ou anti-persistência, bem como se uma série é estacionária ou não. Uma classificação desse tipo é proposta por Malamud e Turcotte (1999) e mostrada na tabela a seguir. Essa relação funciona bem para séries com muitos pontos N>1000 e quando se tem um bom estimador de β. Os exemplos da Figura A.2 ilustram os valores de β estimados para dois tipos de séries temporais, um ruído branco (A.2a e A.2c) e um passeio aleatório tipo movimento browniano (A.2b e A.2d). A estimativa do espectro de potência foi realizada pelo método do periodograma que utiliza a *FFT*.

Tabela A.1 - Classificação do expoente β. FONTE: Adaptado de Malamud e Turcotte (1999).

intervalo de β	tipo de persistência	estacionário
$\beta > 1$	persistência forte	não
$1 > \beta > 0$	persistência fraca	sim
$\beta = 0$	não correlacionado	sim
$-1 < \beta < 0$	anti-persistência	sim

Figura A.2 - Série estacionária com respectivo espectro de potência (a,b) e série não estacionária (c, d).
Fonte: Adaptado de Malamud e Turcotte (1999).

O movimento tipo browniano $\beta = 2$ pode ser obtido pelo simples somatório de um ruído branco, e é definido por:

$$mB = \sum_{v=1}^{n} \epsilon_v, \quad em\ que,\ (\ mB \in \Re,\ v \in N), \tag{A.4}$$

em que ϵ_v é um ruído branco e mB é uma série com variância $(N\sigma^2)$, e n o total de elementos da série. Para simulação do movimento fracionário Browniano (mfB) $\beta > 1$, não abordada nesse trabalho, existe uma gama de algoritmos distintos para esse fim. Entre eles pode-se citar o modelo proposto por Osborne e Provenzale (1989) que utiliza o espectro de potência e o algoritmo proposto por Sellan (1995) e implementado por Albry e Sellan (1996) que utiliza a transformada *wavelets* para sintetizar o mfB. Uma revisão dos algoritmos geradores de mfB pode ser encontrada em Barded et al. (2003), nesse trabalho de revisão os autores avaliam mais de 10 tipos de geradores sintéticos do mfB. A importância de estudar esse tipo de movimento está relacionado a sua presença nos processos dinâmicos da natureza. A popularização do mfB foi realizada principalmente com os trabalhos do matemático francês Benoît B. Mandelbrot que mostrou sua relação com os sistemas dinâmicos e geometria dos fractais e principalmente sua presença na natureza (MANDELBROT, 1983; ADDISON, 1997). O movimento browniano fracionário é geralmente parametrizado via expoente de Hurst (H) que é equivalente ao α do DFA, a relação entre H e β é dada linearmente por:

$$\beta = 2H - 1,\ para\ -1\ <\ \beta\ <\ 3, \tag{A.5}$$

cuja demonstração completa pode ser encontrada em Taqqu et al. (1995) e Malamud e Turcotte (1999). Ao extrair o valor de α ou β de uma série, seja ela temporal ou espaço temporal pode-se inferir sobre as propriedades estatísticas, indicando o tipo de persistência presente no dado bem como se ele é ou não estacionário. Por outro lado essa estimativa não é trivial principalmente em dados experimentais e dependentes do tipo de técnica utilizada para estimativa dos expoentes.

Geralmente a não estacionaridade em dados experimentais pode ser gerada por sazonalidades e/ou pela adição de ruídos não estacionários durante o processo de aquisição. A estacionaridade de uma série temporal pode ser entendida como a não modificação sistemática da média ao longo do tempo, ou seja, uma série sem nenhuma tendência, nenhuma mudança na variância e também com a remoção de

componentes periódicos (CHATFIELD, 2004). De maneira geral, as técnicas tradicionais que exploram a caracterização de classes de universalidade via lei de potência[1] falham quando o dado apresenta tendências (STANLEY et al., 1999). Um exemplo clássico dessa falha é mostrado pelo espectro de potência. A Figura A.3 apresenta a soma (Fig. A.3a) e a concatenação de duas séries senoidais (Fig. A.3c), respectivamente. Neste exemplo, as correlações de longo alcance não podem ser diferenciadas pelo espectro de potência (Fig. A.3c e A.3d). Embora com amplitudes diferentes, os dois picos de freqüência característicos não podem ser discriminados no espaço. Uma técnica alternativa para solucionar esse problema é a transformada *wavelets*. Em resumo esta técnica busca analisar o sinal tanto no domínio da freqüência quanto no domínio da escala (tempo). Maiores detalhes estão disponíveis em Meyer (1992), Morettin (1999) e Domingues et al. (2005) que descrevem os fundamentos e aplicações da transformada *wavelets*. Uma análise mais simples e alternativa à análise por *wavelets* é o método DFA que possui a implementação mais simples e já está amplamente consolidado na literatura, os exemplos apresentados na próxima seção ilustram a aplicação do DFA.

A.1.3 Avaliação e Variantes do DFA

O método DFA foi originalmente estabelecido por Peng et al. (1994) para análise de correlações de longo alcance em mosaicos de nucleotídeos de DNA. Derivado da análise de flutuação (PENG et al.,1992), o DFA teve como seu antecessor o método de determinação do expoente de Hurst (Análise R/S Rescaled range Statistic). O expoente de Hurst (H) fundamenta-se nas medidas do engenheiro inglês Harold Edwin Hurst que correlacionou dados ambientais com padrões aleatórios. Através de observações de dados hidrológicos, Hurst elaborou uma relação entre escalas atuantes em uma série temporal que permite identificar suas relações de dependência ou independência (HURST 1951; ADDISON, 1997, MIELNICZUK e WOJDYLLO, 2007). Em geral a estimativa do expoente de *Hurst* H é dada por:

$$\log \frac{R}{S} = H * logN + logk, \tag{A.6}$$

em que, R é a amplitude da série, S o desvio padrão, *log k* é uma constante que significa o ponto onde a reta intercepta o eixo Y, num gráfico *log-log* e N é o número de pontos da série. Portanto, no gráfico *log-log* de R/S a inclinação da reta será H, representando o mesmo que o expoente α. Se a série se comportasse como um

[1] coeficiente angular em escala *loglog* invariantes *à* escala(Stanley, 2000).

Figura A.3 - Série periódica com respectivo espectro de potência série série estacionária (a,b) e não estacionária (c, d).
Fonte: Adaptado de Stanley et al. (1999)

ruído branco, H seria igual a 0,50. A série dos desvios acumulados aumentaria com a raiz quadrada dos valores de N, isto significa que quando$H$ difere de 0,50 as observações não são independentes. Cada observação carrega uma memória de todos os eventos que a precederam, esta memória não é de curta duração, mas sim de longa duração, teoricamente possui duração infinita (ADDISON, 1997). Nesse contexto, eventos recentes têm maior impacto que os eventos mais distantes, mas sempre com a propagação da influência residual.

O problema da análise do R/S de Hurst e também do FA são dados espúrios, comumente observados nos dados experimentais. Ambos os métodos são sensíveis a dados espúrios (TAQQU et al., 1995, BASHAN et al., 2008). Nesse contexto, o método DFA tornou-se popular na área de análise de séries temporais e sinais, tendo como principal objetivo a detecção de correlações de longo alcance e suas relações com leis de potência. Segundo Bashan et al. (2008) já foram publicados mais de 500

artigos relacionados com o DFA incluindo aplicações, avaliações do método e novas variantes do DFA para aplicações específicas.

Entre os principais variantes do DFA destacam-se mais de 10 diferentes abordagens do método. O Fourier DFA (CHIANCA et al., 2005), por exemplo, tem o objetivo de remover tendências em séries periódicas que apresentem baixas freqüências de modulação. Esta variante do método pode ser útil, por exemplo, para análise de sinais periódicos contaminados por ruídos. Outras variantes baseadas em filtros (RODRIGUEZ at al., 2007), médias móveis (CARBONE et al., 2004), médias centradas (ALVAREZ-RAMIREZ et al., 2005) e funções contínuas (STAUDACHER et al., 2005) têm demonstrados sua aplicabilidade em casos específicos. Por exemplo, Nagarajam e Kavasseri (2005) e Nagarajan (2006) utilizam filtros baseados na decomposição de valor singular-SVD para minimizar os efeitos de saltos (pontos atípicos) aleatórios presentes em um sinal.

A avaliação dos diferentes métodos proposta por Bashan et al. (2008) indicam o uso do DFA original como o melhor método para dados cujas tendências são desconhecidas. Através de séries geradas artificialmente Xu et al. (2005) fizeram a comparação entre o DFA original e variantes baseados em médias móveis e ponderadas para vários tipos de correlação (valores de α). Os autores apresentaram uma representação de mapas de probabilidade em que mostram, para diversos valores de α e em diferentes comprimentos de série, uma probabilidade associada. Os resultados indicaram o melhor método para cada faixa de valores de α simulados, em geral, o DFA original possui o melhor desempenho global.

Uma generalização do DFA para aplicações multifractais foi proposta por Kantelhardt et al. (2002). O*Multifractal Detrended Fluctuation Analysis* (MDFA) tem se mostrado como uma ferramenta alternativa à análise *wavelets*, propriamente ao método *Wavelet Transform Modulus Maxima* (WTMM) (MUZY et al., 1991; ARNEODO et al. , 1995). Os autores do MDFA ressaltam a vantagem em relação ao WTMM quanto ao custo computacional, além dos melhores resultados obtidos para séries curtas. O mesmo resultado foi encontrado por Wiecimka et al. (2006) que mostraram o melhor desempenho do MFDFA para dados simulados e séries reais com características multifractais.

Weron (2002) propôs um método que permite calcular empiricamente um intervalo de confiança para os valores de α estimados pelo DFA. Este autor compara os resultados do DFA com a análise do R/S de Hurst e com o método GPH proposto por GewekePorter-Hudak (GPH)(1983). Os resultados indicaram que o DFA possui

o menor desvio padrão dentre os três métodos comparados. Grau-Carles (2006) fez uma análise com a técnica *bootstrap* para estimar os valores de α em séries curtas. Este autor conclui que sua metodologia permite diferenciar falsas correlações de longo alcance em séries muito curtas (n $<$ 200), sendo sua metodologia de grande utilidade na análise de séries temporais do mercado financeiro.

Mielniczuk e Wojdyllo (2007) constroem uma metodologia para correção do método de análise do R/S de Hurst. Segundo esses autores, os resultados desse novo estimador são melhor que o DFA para séries simuladas com alfa entre ($0{,}5 < \alpha < 0{,}9$). Os autores ressaltam que o tempo computacional do algoritmo é bem menor que o DFA e a análise do R/S tradicional.

Michalski (2008) testou o DFA através de uma extensiva simulação Monte Carlo utilizando sete diferentes geradores de séries sintéticas com memórias de curto e longo alcance. Segundo esse autor o ajuste das escalas durante a estimação do valor de α é fundamental para o melhor desempenho do método. Por exemplo, este autor demonstra que a utilização de escalas grandes para o cálculo da função de flutuação F(s) pode influenciar muito nos resultados da estimativa de α.

Bardet e Kammoun (2008) descreveram analiticamente que o DFA apresenta problemas de convergência para correlações de longo alcance. Segundo esses autores o valor máximo de α estimado pelo DFA é restrito a 2 e em casos de uma tendência constante o valor de α está restrito a 1,5. Os autores reforçam que o DFA não é robusto para tudo e não deve ser aplicado em processos com tendência. Embora, sem qualquer tipo de comparação com o DFA, os autores indicam o método baseado em *wavelets* como o melhor estimador dos valores de α, além disso, segundo esses autores, utilizando o algoritmo de Mallat o custo computacional é menor que o DFA. Os resultados encontrados por esses autores contrariam os resultados encontrados por Taqqu et al. (1995) que demonstraram que o DFA não apresenta tendências na estimação de α. Por outro lado, Bardet e Kammoun (2008) corroboram com os trabalhos realizados no desenvolvimento de variantes do DFA que buscam atender geralmente tipos de tendências específicas.

Xu et al. (2009) apresentaram um algoritmo baseado em filtragem pela transformada discreta de Laplace para minimizar efeitos de tendências exponenciais na estimativa do valor de α no gráfico *loglog* para o método Multifractal DFA. Os autores testaram o método com dados reais de velocidade de tráfego em uma estrada em Beijing na China e complementaram o trabalho de Shang et al. (2008) que apresentaram uma análise detalhada da velocidade de tráfego nessa rodovia através do método

Multifractal DFA.

A.1.4 Exemplos de Aplicações

No campo da análise de sinais na área de ciências médicas e biológicas, o DFA tem sido extensivamente aplicado à análise de sinais cardíacos (STANLEY et al., 1999; GOVINDAN et al.; 2007), eletroencefalogramas (Valencia et al.; 2008), cadeias de nucleotídeos de DNA (PENG et al, 1994; PENG et al, 1995), dentre outros. A área de economia e finanças é um outro segmento que vem utilizando o DFA constantemente, dentre vários trabalhos podemos citar Vandewalle at al. (1999) e Grau-Carles (2006) que analisam a variabilidade de índices cambiais. Como o objetivo desta revisão é descrever aplicações do DFA na área ambiental, as aplicações em outras áreas como ciências médicas, biológicas, dos materiais, econômicas e sociais, não serão abordadas. A seguir serão descritos exemplos de aplicações ambientais nas áreas de meteorologia, oceanografia, ecologia, hidrologia e geofísica.

Um exemplo clássico em climatologia é descrito por Weber e Talkner(2001) que analisam séries históricas de amplitude térmica diurna, temperatura, precipitação, pressão e umidade do ar. Os dados foram derivados do registro de estações meteorológicas localizadas na Europa e Estados Unidos. Os autores classificaram estas estações de acordo com sua localização geográfica e elevação, e para cada variável calcularam o expoente de Hurst utilizando o DFA, a análise R/S de Hurst e análise espectral. Os resultados indicaram comportamentos diferenciados para cada grupo (montanha, marítima, continental e baixa elevação). A análise através do método DFA sugere que há um comportamento multifractal para alguns grupos e variáveis, além da presença de correlações de longo alcance. Segundo os autores a comparação com os outros métodos permitiu maior embasamento na discussão e conclusões.

Chen et al. (2007) realizaram uma análise climatológica de 50 anos da umidade relativa diária em 73 estações meteorológicas distribuídas sobre o território Chinês. Estes autores propõem um novo índice baseado no DFA para classificação da variabilidade da umidade relativa na china. O índice denominado χ é simplesmente a multiplicação de α obtido do DFA pelo desvio padrão de cada série de anomalia de umidade relativa, multiplicado por um fator 100 para visualização dos dados. A Figura A.4 apresenta a distribuição geográfica das estações meteorológicas com os valores de α e os valores do índice χ separados por regiões geográficas. Os autores enfatizam que esse simples índice permite a classificação das regiões geográficas em função da variabilidade da umidade relativa, instrumento que pode ser de grande utilidade para o governo chinês e cientistas no entendimento da variabilidade climática

e suas correlações de longo alcance.

Figura A.4 - a) Valores de α para umidade relativa. b) Valores do índice χ para umidade relativa separados por região geográfica. Fonte: Chen et al. (2007).

Kai et al.(2008) realizaram uma análise de correlações de longo alcance em séries de índices de poluição da cidade de Shanghai, China. Nessa análise foram utilizados 4 índices de poluição sobre os quais os autores aplicaram os métodos para verificar a presença de correlações de longo alcance no contexto de sistemas complexos bem como na idéia de Self-Organized Criticality (SOC)[2]. Com o uso do DFA, a análise R/S de Hurst e análise espectral encontraram correlações de longo alcance e sugerindo que a poluição atmosférica pode ser invariante à escala, porém ressaltaram que essa hipótese ainda deve ser provada ou negada. Segundo esses mesmos autores alguns comportamentos não lineares da série devem ser descritos por um método multifractal [3].

No campo das aplicações em oceanografia, Ausloos e Ivanova (2001) analisaram as correlações de longo alcance do índice *Southern Oscillation Index* (SOI) que é a função básica que caracteriza o fenômeno de El Niño. Utilizando o método DFA os autores encontraram a presença de anti-correlações em escalas menores que 70 meses e correlações de longo alcance em escalas maiores. Os autores ressaltam que os resultados concordam com os demais trabalhos encontrados na literatura, que indicam as oscilações em baixa freqüência do SOI como fonte de erros em modelos

[2]Conceito utilizado em física-estatística para descrever fenômenos críticos que apresentam auto-organização e transição de fase (BAK et al. 1987)

[3]Uma análise multifractal é uma generalização de um fractal e assume que existem mais de um expoente para descrever múltiplas dimensões fracionárias. Uma revisão dos métodos de análise de fractais e multifractais é descrita por LOPES e BETRUNI (2009).

estatísticos de previsão.

Fraedrich e Blender (2003) utilizaram o DFA para obtenção dos valores de α de séries de temperatura de superfície do mar e do ar. A análise foi realizada com 100 anos de observações de estações meteorológicas e de 1000 anos de dados simulados por modelos climatológicos que acoplam a interação oceano-atmosfera. A análise permitiu a espacialização dos valores de α para os dois casos estudados. A Figura A.5 apresenta os resultados encontrados por estes autores. Os resultados indicam que, em geral, a temperatura sobre os continentes apresenta menor memória que sobre os oceanos. Cabe ressaltar que os dados foram simulados, logo devem ser consideradas as restrições e escala dos modelos para uma conclusão sobre os fenômenos físicos envolvidos.

Figura A.5 - a) Valores de α para 100 anos de dados observados de temperatura. b) Valores de α para os dados simulados 1000 anos.Fonte: Fraedrich e Blender (2003).

Zhu et al. (2006) investigaram os regimes de variabilidade da circulação meridional no oceano atlântico. O DFA foi aplicado em séries de 500 anos de simulação mensal de dois modelos acoplados de interação oceano-atmosfera utilizados pelo *International Panel for Climate Change* (IPCC). Os resultados indicam a concordância entre os valores de α entre os dois modelos. Os maiores valores de correlação de longo alcance foram encontrados no atlântico norte, o que segundo os autores, pode ser um indicativo de que ações antropogênicas provocam alterações no clima.

Huybers e Curry (2006) utilizam o DFA para encontrar as leis de potência na temperatura de superfície no planeta Terra. Utilizando dados de re-análise obtidos pelo NCEP-NCAR e datação espectral de gelo e de rochas os autores investigaram as relações entre a variabilidade da temperatura anual, continua e do ciclo de Milanko-

vitch[4]. Os resultados indicaram que a variabilidade anual, contínua e de Milankovitch representam uma resposta determinística a insolação. Os autores encontram uma coerência entre as estimativas do método espectral (*Thomsons multitaper method*) com o método DFA.

Uma análise de séries temporais de clorofila, no oceano, derivado de imagens do sensor orbital *SeaWifs* foi realizada por Zhan (2008). Utilizando 8 anos de dados foram construídas 26079 séries de 414 elementos, cada série representa a média geométrica dos valores de clorofila contidos em 1 grau de resolução geográfica. O autor encontrou correlações de longo alcance, sendo que para todo oceano a média do valore de α foi 0,83 com o mínimo de 0,51 e máximo de 1,26. A análise do comportamento em diferentes escalas permitiu o autor inferir sobre possíveis efeitos de fenômenos de meso-escalas na concentração de clorofila.

No campo da hidrologia, destaca-se o trabalho de Matsoukas et al. (2000) que analisou os dados de precipitação e escoamento superficial com o DFA. Os dados utilizados compreendem medidas em alta resolução temporal (15 min.) de 9 localidades dos Estados Unidos. Os resultados indicaram a presença de dois regimes: entre 1 hora e 5 dias correlações de curto alcance, e entre 10 dias a poucos anos a presença para correlações de longo alcance. Resultados similares foram encontrados para o escoamento superficial, dado a sua correlação com os dados de precipitação.

Uma análise multifractal da precipitação e vazão de rios foi realizada por Kantelhardt et al. (2006). Nessa análise os autores utilizaram o Multifractal DFA para caracterização de 99 estações meteorológicas e 42 estações hidrológicas distribuídas pelos 4 continentes. Diferente dos resultados encontrados por Matsoukas et al. (2000), os autores não encontraram correlações de longo alcance nos dados de precipitação, enquanto que os dados de vazão apresentaram correlação de longo alcance. Os resultados também indicaram persistência na vazão não relacionada com a precipitação. Os autores sugerem que os expoentes multifractais sejam utilizados para caracterização das estações pluviométricas e hidrológicas. Segundo os autores, essa caracterização pode auxiliar na parametrização de modelos de escoamento superficial.

No campo do sensoriamento remoto aplicado a ecologia de florestas destaca-se os trabalhos de Telesca e Lasaponara (2005), Telesca e Lasaponara (2006a), Telesca e Lasaponara (2006b) e Telesca e Lasaponara (2008). Utilizando dados do sensor

[4]Ciclos de variabilidade do clima de acordo com o movimentos orbitais do planeta Terra

Vegetation a bordo do satélite SPOT os autores investigaram o estado da vegetação após uma ocorrência de queimada. Basicamente, os autores construíram séries temporais a partir da observação de amostras (pixels) nas séries de imagens de NDVI[5] para áreas de vegetação sadia e queimada. Aplicando o DFA nessas séries os autores caracterizaram estas áreas com o expoente α, mostrando que após a queimada os valores de α tendem aumentar. Do ponto de vista ecológico e florestal isso pode ser explicado pela perda e recuperação de biomassa com a queimada. Nesse contexto, a correlação entre os valores de NDVI e a precipitação ou até mesmo com a variabilidade climática é maior, ou seja, a vegetação modifica sua estrutura física e fisiológica e tende a crescer (ganhar biomassa) em pouco espaço de tempo, o que provocará uma tendência na série de NDVI, conforme mostrado na Figura A.6a. Os valores α para áreas queimadas e não queimadas são apresentados na Figura A.6b. Os autores ressaltam que o método proposto pode ser utilizado na caracterização de outros tipos de vegetação. Cabe ressaltar que essa metodologia pode ser avaliada com outros índices de vegetação e até mesmo utilizando outras regiões espectrais com o objetivo de investigar a variabilidade do comportamento espectral da vegetação. No contexto de classificação de imagens, o expoente α pode ser incluído como um atributo a ser utilizado pelo classificador.

Figura A.6 - a) Série temporal de NDVI após uma queimada. Região de Ventimiglia, norte da Itália. Fonte: Telesca e Lasaponara (2006). b) Valores do expoente α para vegetação queimada e não queimada. Fonte: Telesca e Lasaponara (2005)

Zheng et al. (2008) realizaram análise que compara as condições meteorológicas com a incidência de incêndios florestais no Japão. Utilizando o DFA os autores encontra-

[5]Índice gerado a partir da região espectral do vermelho e infravermelho próximo que está relacionado com o vigor da vegetação

ram valores de α para os dados de área queimada, temperatura, umidade relativa e precipitação. Os dados meteorológicos utilizados têm a resolução temporal horária, já os dados de queimada possuem resolução temporal de duração em minutos. Os autores encontraram dois regimes para α, o que pode indicar um comportamento multifractal. Os resultados apontam que os valores de α e as escalas de correlação de longo alcance para as quatro variáveis são próximas. Os valores são 3,9 ; 5,20 ; 4,67 e 5,20 dias para as variáveis tempo de queimada em floresta, temperatura, umidade relativa e precipitação, respectivamente. Segundo os autores as correlações entre as variáveis podem ser úteis para entender a dinâmica do fogo nas áreas de floresta. Cabe ressaltar que a análise foi realizada no domínio unidimensional, sendo que as distribuições espaciais das variáveis meteorológicas bem como os tipos de vegetação podem influenciar na dinâmica do fogo em regiões florestais.

Na área de geofísica, a utilização mais comum do DFA é na área de sísmica. Telesca et al. (2003) analisaram a atividade sísmica na Itália de 1986 a 2001 e encontraram correlação entre as atividades sísmicas principalmente na área central da Itália, embora os autores reconheçam que uma maior quantidade de dados é necessária para melhorar a análise. Esta mesma área central da Itália foi investigada por Telesca e Lapenna (2006) com a abordagem multifractal DFA. Os autores concluíram que após um violento evento (terremoto) em 1997 ocorreu a diminuição nos parâmetros que caracterizam o comportamento multifractal da série. Ou seja, as atividades sísmicas modificaram a dinâmica de um comportamento de escalas heterogêneas para escalas homogêneas.

Ainda no campo da sísmica, Pin et al. (2008) propõem uma ferramenta baseada no DFA para detecção automática de falsos alarmes na atividade sísmica do complexo vulcânico de Teide-Pico, nas ilhas Canárias. Basicamente, o sistema analisa, para curtos períodos de tempo (20 min.), o desvio padrão dado pela curva de ajuste do valor de α. Segundo os autores, um desvio padrão maior está associado à presença de atividades irregulares. Para cada dia, a relação entre o desvio padrão de séries espaçadas a cada 20 min é calculada. O desvio padrão de todas as janelas é estabelecido como um limiar, e valores acima desse desvio indicam a presença de falsos alarmes. Embora os autores não apresentem as curvas analisadas, a variabilidade entre os valores de α ao longo de 24 horas possui uma forte modulação entre o início e a metade do dia, conforme apresentado na Figura A.7. Tal modulação por sua vez pode estar associada às atividades antropogênicas relatadas pelo autor como fonte de ruído nos dados.

Figura A.7 - Valores de α ao longo de um dia. Fonte: Pin et al. (2008)

A.1.5 Extensões para outras dimensões

Os exemplos de aplicações apresentados na seção anterior demonstram a utilização do DFA nas comunidades de meteorologia, oceanografia, ecologia e geofísica. Do ponto de vista da coleta e tratamento dos dados a abordagem unidimensional é mais utilizada. Investigar as correlações de longo alcance no domínio maior que o unidimensional não é uma tarefa trivial, uma vez que geralmente as correlações em um campo maior que o 1-D podem ser heterogêneas ao longo de determinadas direções. A capacidade de analisar imagens, campos escalares ou vetoriais tridimensionais, pode auxiliar em muito a análise de um dado fenômeno físico. Porém, técnicas robustas, computacionalmente de fácil implementação e que descrevam as relações de dependência entre escalas e em diferentes dimensões ainda estão em desenvolvimento. Segundo Gu e Zhou (2006), a transformada *wavelet* é um exemplo típico de uma técnica robusta, porém, com dificuldades na implementação computacional e na escolha da melhor parametrização.

No contexto de imagens, a análise de correlações está diretamente ligada à textura. A definição de textura está relacionada à forma da estrutura do campo bi-dimensional e a sua caracterização pode ser realizada por meio da análise de rugosidade (KAPLAN, 1999). Na área de reconhecimento de padrões e processamento de imagens a análise de textura é fundamental para entender, caracterizar e classificar uma imagem. Isto tem várias aplicações na área de sensoriamento remoto e medicina (KAPLAN, 1999). Segundo esse mesmo autor a análise de texturas naturais como

nuvens, modelos de terrenos, fumaça entre outros pode ser caracterizada pelo expoente de Hurst. Este expoente controla a rugosidade visual de um processo em todas as escalas. Um exemplo típico desse processo é o movimento fracionário Browniano (fBm) popularizado por Mandelbrot (1983).

A caracterização de texturas em imagens utilizando o expoente de Hurst derivado do DFA é descrita em Alvarez-Ramirez et al. (2006). A metodologia proposta por esse autor é baseada em uma função descritora de textura proposta por Kaplan (1999). Basicamente, é uma análise 1-D em várias orientações da imagem. O algoritmo proposto por Alvarez-Ramirez et al. (2006) transforma a imagem bi-dimensional, numa dada orientação θ, em duas seqüências ortogonais, linhas e colunas. Para cada conjunto de séries calcula-se a função de flutuação de acordo com o DFA. Os autores avaliaram o método proposto em 4 diferentes imagens. O exemplo de dois resultados encontrados pelos autores é apresentado na Figura A.8 que representa a granulação solar (Figura A.8a e A.8b) devido à atividade magnetohidrodinâmica e emulsão de água e óleo (Figura A.8c e A.8d) observada por um microscópio eletrônico. Observando o comportamento da função α_θ ao longo das escalas podemos inferir sobre a anisotropia da imagem bem como o *lag* que não se possui mais correlação espacial, em que α_θ => 0,5. Como a série de α_θ apresenta uma variação, ao longo de n, podemos concluir que a textura apresenta diferentes rugosidades nas várias escalas. O decaimento α_θ ao longo do aumento nas escalas pode ser utilizado tanto para análise de dependência espacial ao longo de diferentes orientações, bem como para parametrização de amostragens estatísticas na imagem. Ressalta-se que há na literatura abordagens tradicionais de processamento de imagens para descritores de textura (DUTRA, 1990; SCHOWENGERD, 1997, MATHER, 1999).

A metodologia proposta por Alvarez-Ramirez et al. (2006) foi pioneira ao utilizar o método DFA para abordagem 2-D, porém, isso não representa a generalização do método para dimensões maiores que 1. Uma generalização similar para análise da anisotropia é apresentada no Anexo 1. A generalização para espaços bi-dimensionais foi proposta por Gu e Zhou (2006) como mostrado no capítulo 3.

Figura A.8 - a) Imagem da granulação solar devido à atividade magnetohidrodinâmica c) emulsão de água e óleo observada por um microscópio eletrônico. b e d) comportamento da função ${}_{a}vr_{\theta}(n)$ e α_{θ} ao longo das escalas em pixels (n). Fonte: Alvarez-Ramirez et al. (2006)

APÊNDICE B - MODELOS NUMÉRICOS UTILIZADOS PARA GERAÇÃO DE DADOS SINTÉTICOS

Nesse apêndice são apresentados modelos simplificados de sistemas físicos que são capazes de gerar dados canônicos representando os mais diversos padrões de variabilidade típicos de sistemas complexos. Alguns desses modelos são relativamente simples com poucas variáveis, mas apresentam conjuntos de soluções complexas. Toda implementação será disponibilizada como uma ToolBox do MatLab e também como uma publicação técnico-científica tipo Programas de Computador (PDC) do Serviço de Informação e Documentação - (SID) do INPE.

B.1 Envelope de senos

O modelo de envelopes de senos é o modelo simples que tem o objetivo de representar padrões de periodicidades nos dados. O modelo estudado incorpora ruído gaussiano aditivo e é representado computacionalmente por:

Para o caso 1-D:

$$X_t = \sum_{j=1}^{n} (As_j sen(\omega_j t + \phi_j) + \epsilon_{j,t}), \tag{B.1}$$

em que, X é a série unidimensional de senos, t é a posição no tempo ou espaço unidimensional variando de $[1..N]$, onde As_j é a amplitude em cada componente j, $\omega = 2\pi ft$ é frequência, e ϕ o ângulo de fase, n o número de componentes periódicas com os *j-ésimos* parâmetros e $\epsilon_{j,t}$ é o um ruído gaussiano de média 0, variância σ^2 e dimensão *1xN*.

Para o caso 2-D:

$$Ms = \sum_{k=1}^{m} (sin(B * X_k) + sin(B * X_k)' + \epsilon_{m,k}), \tag{B.2}$$

em que Ms é a matriz de envelopes senoidais, X é uma série multi-periódica definida pela Equação B-1 de dimensão *1xN*, **B** é um vetor unitário de dimensão *1xN* e m é o número de matrizes periódicas a serem somadas a partir da *k-ésima* série multi-periódica gerada. $\epsilon_{m,k}$ é uma matriz de ruídos gaussiano de média μ e variância σ^2 e dimensão *NxN*.

Com este simples modelo é possível simular a periodicidade em séries temporais e envelopes de amplitude em séries espaço-temporais, os exemplos nas Figuras B.1 e

B.2 mostram o envelope de senos 1-D e 2-D simulados a partir dos modelos descritos anteriormente. Um exemplo clássico dessa periodicidade em sensoriamento remoto é no comportamento espectro temporal da vegetação. Padrões de periodicidade como apresentado na Figura B-1 são típicos em séries temporais de índices de vegetação, uma vez que a fenologia da vegetação está correlacionada com variação sazonal do fluxo de radiação solar.

Figura B.1 - a)Série simulada com 8 períodos, N=300, $\sigma^2 = 0.1$; b)Série simulada com 10 períodos, N=100, μ =0, $\sigma^2 = 0.3$

B.2 Grade de Mapas Acoplados - Mapa de Kaneko

Uma grade de mapas acoplados é um sistema dinâmico que modela o comportamento de sistemas não lineares. Em geral, eles são utilizados nos estudos de dinâmica caótica em sistemas com dimensão maior que 1. Uma característica das grades de mapas acoplados é sua discretização que ocorre tanto no tempo quanto no espaço. Uma grade de mapa acoplado também apresenta a característica de irreversibilidade, ou seja, a partir de um tempo discreto t não é possível derivar a solução t-n. A aplicação das grades de mapas acoplados é ampla, variando desde simulações da dinâmica de

Figura B.2 - a)Envelope de senos simulados 2-D com 3,5 períodos, N=100, $\sigma^2 = 0$; ;b)Envelope de senos simulados 2-D com 3,5 períodos, N=100, $\sigma^2 = 0.3$;c) Projeção do eixo Z no plano xy do envelope de senos 2-D, com 0.5 períodos, N=100, $\sigma^2 = 0$; d) Projeção do eixo Z no plano xy do envelope de senos 2-D, com 0,5 períodos, N=100 e $\sigma^2 = 0.3$.

populações, reações químicas, processos convectivos, escoamentos de fluidos até a simulação de falhas em grades de computadores (KANEKO, 1989).

Uma grade de mapa acoplado é um mapeamento interativo, em que cada nó de uma grade está relacionado com seu vizinho através de um fator de acoplamento. A grade de mapa acoplado escolhida para este trabalho foi proposta por Kaneko (1989) e tornou-se um exemplo clássico nos estudos de fenomenologia espaço-temporal em sistemas complexos. A grade de mapas acoplados é dada por:

$$u_s^{t+1} = (1-\epsilon_k)f(u_s^t) + \frac{\epsilon_k}{2}f(u_{s+1}^t) + f(u_s^t) \quad t \in \mathbb{N},\ \epsilon_k \in [0,1], \tag{B.3}$$

em que,

u_s^t é a grade de mapas acoplados no instante t e $f(x)$ é dado pelo mapa logístico.

As condições iniciais da grade variam desde ondas, superfícies gaussianas até ruídos aleatórios. Já as condições de contorno podem ser livres, fixas ou periódicas. A função f(x) escolhida para essa grade foi o mapa logístico, por representar um sistema dinâmico simples, porém com comportamento complexo. O mapa logístico é representado por mapeamento polinomial de grau 2 dado por:

$$x_{n+1} = \mu_k x_n (1 - x_n) \mu_k \in (0, 4], \tag{B.4}$$

em que ε é o fator de acoplamento e μ_l é o parâmetro de não linearidade, variando os parâmetros ϵ_k e μ_k a grade acoplada exibe uma rica variedade de fenômenos, incluindo a formação de domínios com evolução caótica e regular, duplicação de períodos, etc. Uma completa fenomenologia da dinâmica espaço-temporal desta equação é descrita por Vasconcelos (2005). Este autor descreve para cada variação dos parâmetros de controle ϵ_k e μ_k o tipo de estrutura e padrão formado ao longo da grade de mapas acoplados.

A Figura B.3 apresenta soluções da grade de mapas acoplados com fator de acoplamento $\epsilon_k = 0,3$ e $\mu_k = 1,75$ o que segundo Vasconcelos (2005) isso representa a fenomenologia de intermitência espaço-temporal. A condição inicial adotada foi uma distribuição aleatória com média 0 e variância 1.

Figura B.3 - Solução da grade de mapas acoplados mostrando o surgimento de padrões intermitentes na seqüência da solução

B.3 Equação de Ginzburg-Landau

A Equação de Ginzburg-Landau é uma das equações mais estudadas na comunidade física (ARANSON e KRAMER, 2002). Em sua forma complexa é dada por:

$$A_t = \mu_{gl} A + (1 + i\alpha_{gl}) A_{xx} - (1 + i\beta_{gl})|A|^2 A. \tag{B.5}$$

Esta equação descreve qualitativamente e muitas vezes quantitativamente vários fenômenos físicos, incluindo a propagação de ondas não lineares, transições de fase de segunda ordem, convecção de Rayleigh-Bérnard e supercondutividade. A Equação Complexa de Ginzburg-Landau (ECGL) descreve a evolução das amplitudes em um sistema fora do equilíbrio. Esta função é uma equação diferencial parcial muito estudada na física e é conhecida como uma equação de "Modulação", "Amplitude" ou "Envelope". O termo A que representa as amplitudes, é chamado de parâmetro de ordem, e é definido como uma função complexa de escala no domínio do tempo (t) e do espaço (x); a, μ_{gl} e β_{gl} são parâmetros fenomenológicos. Embora apresente termo complexo a ECGL é considerada uma equação reativa-difusiva, em que $(1 + ia)$ é a constante de difusão e $(1 + i\alpha_{gl})A_{xx}$ é o termo Difusivo) e $\mu_{gl}A - (1 + i\beta_{gl})|A|^2 A$ é o termo reativo.

A ECGL criou diversos conceitos chaves que foram desenvolvidos anos antes de se tornar bem conhecida, particularmente no contexto do estudo de formação de padrões. Nesse contexto histórico, Landau em 1937 foi o primeiro a propor a dependência espacial lenta no contexto do espalhamento de raio X em cristais. Em 1944 Landau foi um dos pioneiros ao estudar os padrões de relaxação temporal lenta em sistemas turbulentos. Entretanto, o conceito de transição de fases tornou-se bem conhecido com o sucesso de sua colaboração com Ginzburg em 1950 que formularam a teoria que hoje é chamada de Ginzburg-Landau para a supercondutividade (ARANSON e KRAMER, 2002). As equações reais do Ginzburg-Landau foram derivadas primeiramente como equações da amplitude de onda-longa por Newell e Whitehead (1969) e Segel (1969) no contexto de convecção em misturas binárias perto do início da instabilidade. As equações complexas do Ginzburg-Landau foram derivadas primeiramente por Stuart (1971) que estudaram o fluxo de Poiseuille, e por Ermentrout (1981) com a investigação de sistemas reativos-difusivos. As principais aplicações da equação de Ginzburg-Landau concentram-se na área de: convecção de Rayleigh-Bernárd, Escoamento de Taylor-Couette , chamas estabilizadas, ondas hidrotermais, eletroconvecção em cristais líquidos, modelagem de transições de fase

de segunda ordem.

A Equação de Ginzburg-Landau possui solução analítica para um conjunto restrito de problemas, uma vez que os termos não lineares A_{xx} e de segunda ordem não podem ser resolvidos analiticamente. Soluções numéricas da parte real podem ser desenvolvidas a partir dos métodos Espectrais, Pseudo-espectrais, Euler diferenças finitas, e Runge-Kutta. Em geral, os métodos numéricos mais utilizados são os métodos espectrais e solvers de EDO como Runge-Kutta pois, apresentam menos problemas de convergência da solução. A ECGL apresenta uma vasta fenomenologia de padrões espaço-temporais, Chaté (1994) parametrizou esses padrões fenomenológicos de acordo com os parâmetros α_{gl}, μ_{gl} e β_{gl}. De acordo com esse autor é possível gerar soluções que representam processo como intermitência e turbulência, chaos e transição de fase. Neste trabalho será adotado o código aberto proposto por Winterbottom (2005) que utiliza método pseudoespectral (COX e MATTHEWS, 2002) para solução da ECGL em 1-D e 2-D. Um exemplo de solução 1-D turbulenta é apresentado na Figura B.4. A condição inicial foi um pulso. A Figura B.5 apresenta uma solução de estruturas coerentes para o caso 2-D com condições iniciais uma matriz de ruído branco.

Figura B.4 - Solução da ECGL 1-D pelo código proposto por Winterbottom(2005), $\alpha_{gl} = 2,5$, $\mu_{gl} = 1$ e $\beta_{gl} = -1,97$, condições de contorno periódica.

Figura B.5 - Solução da ECGL 2-D pelo código proposto por Winterbottom(2005), $\alpha_{gl} = 0$, $\mu_{gl} = 1$ e $\beta_{gl} = -1,5$, condições de contorno periódica.

APÊNDICE C - ELABORAÇÃO DAS IMAGENS MODIS

A sequência de passos para elaboração da série temporal interpolada é apresentada nas Figuras a seguir (C-1, 2, 3 e 4). O processo de formação de uma imagem composição de 16 dias é descrito na Figura C.1. A Figura C.2 mostra um exemplo de uma série antes da interpolação linear e a Figura C.3 os valores depois da interpolação linear. A Figura C.4 ilustra os valores de distância entre os dias interpolados para esse exemplo. Essas Figuras buscam ilustrar como é elaborada a métrica qualitativa de *gaps* que tem o objetivo de medir o percentual de pixels interpolados em cada série temporal.

Figura C.1 - Passo 1

Figura C.2 - Passo 2

Figura C.3 - Passo 3

Figura C.4 - Passo 4

APÊNDICE D - FLUXOGRAMA DAS IMPLEMENTAÇÕES COMPUTACIONAIS

A estrutura computacional desenvolvida na tese é apresentada na Figura D.1. Essa estrutura consiste em módulos especificamente elaborados para o desenvolvimento da tese. O primeiro módulo implementado é o conjunto de funções para geração dos modelos sintéticos e métodos de análise, este é discutido no Capítulo 2 e Apêndice B. Esses modelos foram implementados em ambiente MatLab o que permite a generalização para outras áreas do conhecimento, não limitando as aplicações ao sensoriamento remoto. Essa estrutura de modelos sintéticos é fundamental para testar os métodos implementados em modelos amplamente conhecidos no domínio da física computacional e que representam os mais diversos regimes de variabilidade complexa.

Figura D.1 - Fluxograma simplificado da estrutura de implementação computacional.

As funções para pré-processamento, armazenamento, gerenciamento e visualização das séries temporais e espaço-temporais foram implementadas em ambiente MatLab e ANSI C e são discutidas no capítulo 3. Entre outras funções desses módulos pode-

se citar a geração dos índices de vegetação, conversão de formatos, interpolação e filtragem pelo método de wavelets, funções de atualização dos sistemas de arquivos, re-processamento de novos conjuntos de dados e visualização acoplada ao globo virtual. O Módulo de Detecção de Mudanças é descrito no capítulo 4 e utiliza as ferramentas descritas no capítulo 3 e os dados descritos no capítulo 4. Todas as funções desse módulo de detecção de mudança também estão implementadas em ambiente MatLab.

APÊNDICE E - EXEMPLOS DE VISUALIZAÇÃO E INTERPRETAÇÃO DAS SÉRIES EVI2

Neste apêndice apresenta-se um conjunto de séries temporais ilustrando a integração com o globo virtual e diferentes tipos de mudança do uso e cobertura.

As Figuras disponíveis nas próximas páginas são compostas pela imagem em perspectiva do Google Earth, imagem do googleMaps e séries temporais de precipitação (gráfico de área em azul) e séries EVI2 i) filtrada com *wavelets*(linha em verde) e 2)filtrada sem *wavelets* (linha em vermelho). A referência do *eixo x*, linha tracejada, é o dia 1 de janeiro e os *eixos y* indicam a amplitude da precipitação acumulada mensal (mm) e valores de EVI2 (eixo y da direita). O retângulo em branco nas imagens indica a localização aproximada do pixel MODIS, e não necessariamente representam o comportamento espectral somente do seu interior dada as diferenças de escala e desconhecimento do georeferenciamento das imagens disponíveis no Google Earth e Google Maps. Todas as figuras foram geradas pelas funções implementadas em ambiente MatLab usando recursos da Google Earth e Google Maps API. Os exemplos apresentados nas próximas páginas buscam ilustrar os diferentes tipos de uso e cobertura da terra. Os exemplos ilustram como a interpretação das séries gera uma informação dinâmica sobre as imagens de alta-resolução disponíveis no globo virtual.

Figura E.1 - Série temporal com desmatamento em cerrado, a partir de 2008 transformação em área agrícola.

Figura E.2 - Série temporal com desmatamento em cerrado, a partir de 2007 transformação em área agrícola.

Figura E.3 - Série temporal para uma área de floresta.

Figura E.4 - Série temporal para uma área de agricultura intensa, pivô central.

Figura E.5 - Série temporal para uma área de agricultura intensa, pivô central.

Figura E.6 - Série temporal para uma área de cana-de-açúcar.

Figura E.7 - Série temporal com desflorestamento em floresta em 2007 e uso agrícola a partir de 2009.

Figura E.8 - Série temporal com desflorestamento em floresta em 2004 e conversão para agrícola em 2007.

Figura E.9 - Série temporal com desflorestamento em floresta em 2007 com posterior regeneração.

Figura E.10 - Série temporal com desflorestamento em floresta em 2007 com uso agrícola a partir de 2009.

Figura E.11 - Série temporal para uma área convertida de caatinga para culturas permanentes.

ANEXO A - Exemplo de Aplicação de Método DFA anisotrópico para monitoramento ambiental

Análise de Anisotropia de Imagens Utilizando o Método DFA

Ramon Morais de Freitas[1], Reinaldo Roberto Rosa[1], Yosio Edemir Shimabukuro[1]
Patricia Carneiro Genovez[2], Maria Cristina Bentz[2]

Instituto Nacional de Pesquisas Espaciais – INPE
Caixa Postal 515 – 12245-970 – São José dos Campos - SP, Brasil
[2] PETROBRAS CENPES
Av. Jequitibá, 950 - Ilha do Fundão - 21949-900 - Rio de Janeiro - RJ, Brasil
{ramon,yosio}@dsr.inpe.br
{reinaldo}@lac.inpe.br
{cris,genovez.sampling}@petrobras.br

abstract This work presents the Detrended Fluctuation Analysis (DFA) method for anisotropic analysis in remote sensing data. Radar images derived from Radarsat-1 satellite were used for this analysis. The objective was to improve understanding about DFA operational capabilities for anomalies detection in remote sensing data applied to oceanography and petroleum exploration and production systems. The study area is Campos Basin located offshore of Rio de Janeiro, Brazil. The proposed method based on eulerian approach that combines geocoding oil platform points over near real time radarsat images. Using onnidirectional sampling over radar images, we applied DFA method for each radial direction time series. The DFA anisotropic shows anomalies direction and ranging over plataforms geocoding points. Therefore the preliminary results indicate that the proposed method can be used for anisotropic analysis of the surface. abstract

resumo Este trabalho apresenta o método de análise de flutuações destendenciadas - Detrended Fluctuation Analysis (DFA) para análise de anisotropia de imagens de sensoriamento remoto. Imagens Radar do sensor Radarsat-1 foram utilizadas nesta análise. O objetivo do trabalho é avaliar o DFA como uma métrica para detecção de anomalia nas imagens de radar com aplicação na área de oceanografia e exploração e produção de petróleo. A área de estudo está localizada na bacia de Campos localizada na área off-shore do estado do Rio de Janeiro. A metodologia proposta é baseada numa abordagem euleriana que combina um ponto georeferenciado na imagem, ex. plataforma de petróloe, em imagens de radar adquiridas em tempo próximo ao real. Utilizando amostragem onimidirecional nas imagens de radar aplicou-se o DFA em

cada direção e em diferentes distâncias. Os resultados preliminares mostram que o DFA pode ser utilizado como uma métrica para detectar a anomalia, indicando sua direção e distância de um ponto georeferenciado na imagem.

A.1 Introdução

O descobrimento de novos campos de exploração de petróleo off-shore tem levado a indústria do óleo e órgãos governamentais a ter cada vez mais atenção nas questões ambientais que envolvem as atividades de exploração e produção de petróleo no mar. Considerando a presença significativa de plataformas e navios sondas pela costa brasileira, o monitoramento de possíveis ocorrências envolvendo o derramamento de óleo e descontroles operacionais[1] deve ser realizado no contexto da prevenção, caracterização e investigação. O monitoramento sistemático de áreas oceânicas é importante para detectar prontamente possíveis acidentes envolvendo derramamento de óleo, fornecendo informações estratégicas para a tomada de decisão durante as ações de combate ao óleo no mar ou nas áreas costeiras.

Nesse contexto, um sistema de monitoramento orbital utilizando imagens de RADAR é fundamental, pois garante o imageamento sob quaisquer condições meteorológicas, dia ou noite, além de uma grande área de recobrimento (Fingas e Brown, 2005). Com as imagens de RADAR é possível detectar áreas de baixo retroespalhamento do sinal SAR que podem estar associadas a alvos escuros na superfície do mar. Em geral, feições de alvos escuros (baixo retroespalhamento) podem estar associadas à presença de óleo, concentração de algas, baixa intensidade de vento, óleos biológicos dentre outros fenômenos oceanográficos(Brekke e Solberg, 2005; Bentz, 2006).

A interpretação e classificação automática de alvos escuros verificados em imagens de RADAR ainda é um problema em aberto devido à grande quantidade de feições que apresentam resposta semelhante ao óleo nestas imagens. Várias metodologias de seleção de atributos e classificação de imagens RADAR têm sido propostas (Brekke e Solberg, 2005) todavia, o trabalho de intérpretes experientes é indispensável para a elucidação das feições encontradas nas imagens. Imagens meteo-oceanográficas, fornecendo informações sobre concentração da clorofila na superfície do mar, temperatura na superfície do mar e intensidade e direção de vento auxiliam na identificação destas feições nas imagens de RADAR, embora, inspeções de campo abordo de aeronaves são essenciais para comprovar a natureza destas feições (Bentz, 2006). Dessa forma, o problema na interpretação das imagens está relacionado com o tempo de

[1]Nomenclatura dada aos vazamentos devido perfuração de novos poços e mistura de água e óleo durante o processo de extração

análise, edição e principalmente subjetividade da interpretação (Solberg et al., 2005). Nesse contexto, os sistemas automáticos de detecção de ocorrências podem auxiliar em muito o fotointérprete, tanto na questão de tempo quanto na análise dos diferentes tipos de ocorrências, inclusive nas relacionadas a derramamento de óleo (Brekke e Solberg, 2005; Bentz, 2006; Keramitsoglo et al. 2006).

Do ponto de vista de um alvo fixo no mar, que apresente riscos de vazamentos (ex. plataformas de petróleo), torna-se interessante caracterizar a direção e distância aproximada de uma anomalia que ocorra no entorno de uma plataforma, seja um derramamento de óleo, presença de algas, baixa intensidade do vento, óleos biológicos, dentre outros. Com auxílio de técnicas de sensoriamento remoto e computação esse procedimento pode ser automatizado e realizado em tempo próximo do real. O pré-requisito básico é o conhecimento da posição geográfica do alvo fixo e do georeferenciamento da imagem. Esse trabalho tem como hipótese a afirmação de que a análise da anisotropia das imagens RADAR da superfície do mar pode auxiliar o fotointérprete na caracterização da variabilidade espacial em torno de um alvo fixo. Nesse contexto, o objetivo deste trabalho é propor e avaliar uma metodologia que utiliza o método DFA para análise da anisotropia de imagens com aplicações na área de oceanografia e indústria de exploração de petróleo.

A.2 Método DFA

O método *Detrended Fluctuation Analysis* foi proposto por Peng et al. (1994) e tem se estabelecido como uma importante ferramenta para detecção de correlações de longo alcance em séries temporais não estacionárias. Originalmente o método DFA, baseado na teoria de caminhos aleatórios (*random walk theory*) (Mandelbrot, 1983; Addison, 1997), é um aperfeiçoamento do método *Fluctuation Analysis* FA (Peng et al., 1992). Inicialmente proposto para aplicações em análise seqüencial de DNA, o método DFA tem aplicações na análise de séries temporais nas mais diversas áreas (Kantelhardt et al., 2001; Bashan et al., 2008). Nessa seção, será apresentada a descrição matemática do método, bem como os passos para sua implementação computacional e alguns exemplos de aplicações em geociências.

A abordagem do DFA é fundamentada na análise de flutuações de uma série temporal em diferentes escalas. Na prática o DFA permite eliminar a tendência de uma ST em diferentes escalas, analisando as flutuações intrínsecas do dado. Entenda-se como flutuações, a medida de variabilidade do sinal associada a variância de cada segmento da série, em diferentes escalas. O algoritmo do DFA é composto por 6 passos básicos.

(1) Dado uma série temporal $\{X_{i,N}\}$ com $i = 1, ..., N$ eqüidistantes espaçados no tempo, o primeiro passo é o cálculo da *função perfil*. Esta função é obtida pela soma cumulativa da série temporal. Nesse passo já podemos subtrair a média da série temporal <X>, conforme descrito a seguir.

$$Y(i) = \sum_{i=1}^{N} X_i - < X > \tag{A.1}$$

(2) subdivide-se o perfil Y(i) em Ns segmentos não sobrepostos de mesma escala s, em que

$$N_s \equiv [N/s] \tag{A.2}$$

(3) calcula-se a tendência local de cada segmento (ν) através do ajuste polinomial de mínimos quadrados. Para cada segmento (ν) calcula-se a série "destendenciada" através da diferença entre o perfil Y(i) e o polinômio $p_\nu(i)$. Em geral, usa-se o polinômio de grau linear (p=1), obtendo-se o DFA1 de primeira ordem, que retira tendências lineares da função perfil Y(i). Polinômios de maior grau, como por exemplo quadráticos (p=2), podem ser utilizados nesse passo para retirar tendências de ordens maiores obtendo-se, para cada grau de polinômio p, um valor da função Ys(i) e conseqüentemente um valor de DFAp associado ao grau do polinômio.

$$Y_s(i) = Y(i) - p_\nu(i) \tag{A.3}$$

(4) calcula-se a variância para cada segmento (ν) dada por:

$$F_s^2(\nu) =< Y_s^2(i)Y(i) >= \frac{1}{s}\sum_{i=1}^{s} Y_s^2[(\nu - 1)s + i] \tag{A.4}$$

(5) calcula-se a função de flutuação DFA para cada segmento s, conforme descrito a seguir:

$$F(s) = \left[\frac{1}{N_s}\sum_{\nu=1}^{N_s} F_s^2(\nu)\right]^{\frac{1}{2}} \tag{A.5}$$

(6) Se a série original apresenta correlações de longo alcance, o valores de $F^{(n)}(s)$ seguem uma lei de potência:

$$F^n(s) \propto s^\alpha \tag{A.6}$$

Para grandes valores de s vale a relação com o expoente da autocorrelação γ, onde α=1-$\gamma/2$ com $0 < \gamma < 1$ (Taqqu et al., 1995). Onde n é a ordem do polinômio utilizado para tirar a tendência da série. Na prática isso permite o cálculo do expoente α através do ajuste linear em escala logarítmica de s em função de $\mathrm{F}^n(s)$.

O expoente de flutuação pode ser classificado de acordo com uma faixa dinâmica de valores (Kantelhardt et al., 2001; Bashan et al., 2008; Zheng et al., 2008):

$\alpha < \frac{1}{2}$: anti-correlacionado, sinal anti-persistente.
$\alpha \cong \frac{1}{2}$: sem correlação, ruído branco, não apresenta memória.
$\alpha > \frac{1}{2}$: apresenta correlações de longo alcance.
$\alpha \cong 1$: 1/f noise ou ruído rosa.
$\alpha > 1$: random walk, não estacionário.
$\alpha \cong \frac{3}{2}$: ruído browniano, integração do ruído branco.

A.3 Procedimento Proposto

O procedimento básico proposto e utilizado nesse trabalho é a aplicação do método DFA conforme descrito na seção anterior para análise de anisotropia de campos bi-dimensionais. Porém, antes da aplicação do DFA devem-se calcular as coordenadas do raio de aplicação. O conceito básico é a transformação de coordenadas polares para cartesianas e vice-versa. A generalização apresentada aqui considera uma abordagem euleriana, ou seja, temos um ponto fixo de observação, no qual queremos analisar a presença de anisotropia de um campo 2-D ou 3-D de uma variável escalar ou vetorial qualquer.

Para o caso bi-dimensional, a partir de um ponto fixo *P(x,y)* de uma matriz escalar ou vetorial M de dimensões [L,C], são construídos n segmentos radiais de raio Rs igualmente distribuídos em um vetor real θ_n dado por:

$$\theta_n = [0 : \frac{(2\pi)}{n} : (2\pi)] \tag{A.7}$$

Dada as restrições: (x-Rs) $\geq$ 1, (x+Rs) $\leq$ C, (y-Rs) $\geq$ 1 e (y+Rs)$\leq$ L, para cada posição do vetor $\{\theta_n\}$,variando entre 0 a 2π, constrói-se uma série unidimensional a partir do ponto fixo $P(x, y) \in M$, dada pelas coordenadas cartesianas da seqüência de pares $[(X_{\theta n}), (Y_{\theta n})]$:

$$X_{\theta n} = (x + ([1 : Rs] * cos(\theta_n)) \tag{A.8}$$

$$Y_{\theta n} = (y - ([1 : Rs] * sin(\theta_n))) \tag{A.9}$$

Note que cada vetor $X_{\theta n}$, $Y_{\theta n}$ deve ser arredondado para o valor inteiro mais próximo, uma vez que os valores de coordenadas da matriz M são inteiros. Esse procedimento pode ser realizado pela simples operação numérica de arredondamento descrita por:

$$Y = \lfloor (Z + (\rceil Z \lceil - \lfloor Z \rfloor)/2) \rfloor \tag{A.10}$$

Onde, Z representa o vetor a ser arredondado, no caso $X_{\theta n}$ e $Y_{\theta n}$, construindo n séries para cada segmento S_θ dado por,

$$S_\theta = M(X_{\theta n}, Y_{\theta n}) \tag{A.11}$$

onde, M representa uma matriz 2-D a ser amostrada, e ($X_{\theta n}$,$Y_{\theta n}$) é o par ordenado para cada segmento radial de raio Rs e direção θ gerados a partir do ponto fixo *P(x,y)*.

Após a amostragem da matriz 2-D, o passo seguinte é aplicar o método DFA em cada segmento S_θ obtendo assim n valores de α_θ. O método DFA é aplicado conforme descrito na seção anterior. A visualização dos valores α_θ é realizada em coordenadas polares, ou seja, θ_n representa a fase e o valor do respectivo expoente α é exibido como o raio nessa coordenada.

A.4 Experimentos

A.4.1 Dados

No presente trabalho foram utilizados recortes de 4 imagens RADAR de diferentes datas provenientes do satélite RADARSAT-1, banda C, polarização HH, possuindo resolução radiométrica de 8 bits, resolução espacial de 50m x 50m e no formato compactado MrSid.

A.4.2 Metodologia

O método descrito na seção 3 foi utilizado para avaliação da anisotropia de imagens RADAR [2] em plataformas de petróleo localizadas na região da bacia de Campos, RJ. A Figura 1 apresenta respectivamente, a) uma fotografia de uma plataforma de petróleo, b) sua localização na imagem de RADAR, c) recorte amostrado nesta imagem com um raio de 5km, d) amostragem radial a partir do centro desse recorte. Os pontos claros (na escala de cinza) na imagem RADARSAT são gerados pelo sinal retroespalhado desses alvos devido a reflexão de canto. Com o objetivo de exemplificar e avaliar o método proposto foram analisados inicialmente mais de 50 campos de anisotropia para as 4 datas. As análises apresentadas neste trabalho se restringiram em selecionar apenas alguns exemplos de imagens que representem campos anisotrópicos e isotrópicos.

A.4.3 Resultados e Discussão

No exemplo a seguir (Figura 2) observa-se uma imagem de uma plataforma de petróleo cujos valores de nível digital normalizado - ND[3] , em torno da plataforma são isotrópicos. Para facilitar a visualização a escala de cinza foi invertida, assim os maiores valores de ND são escuros e os menores são claros. A Figura 2-a é a representação da imagem, na Figura 2-b é a representação da superfície, em que os picos representam os alvos com alto retroespalhamento, como plataformas e embarcações. Na Figura 2-c é apresentado a imagem amostrada gerada a partir do raio Rs para 360 valores de θ. Analisando o gráfico em coordenadas polares (Figura 2-d) para toda a variação da fase θ os valores do raio, α, flutuam em torno da média sem apresentar direções preferenciais, ou seja, há isotropia na escala Rs de 6,75 km. Porém, quando aumenta-se o valor da escala, Rs para 13,5 km (Figura 3), percebe-se que o campo se torna anisotrópico da direção aproximada entre 120° e 150° devido à presença de uma mancha escura(Figura 3-d).

Um outro exemplo (Figura 4) consiste em uma região de baixa intensidade de vento em que os valores de α_θ são próximos a um ruído branco (Figura 4-d). Para as áreas anisotrópicas os valores apresentam correlação de longo alcance, uma vez que há transição entre o baixo retroespalhamento do baixo vento para um maior retroespalhamento da superfície do mar que é mais rugosa. Esse aumento no retroespalha-

[2]O autor agradece a Dra. Cristina Maria Bentz e a Petrobrás SA pelo fornecimento das imagens utilizadas nesse trabalho.

[3]normalização aplicada para transformação da escala de 8 bits para valores entre 0 e 1, dado pela divisão dos valores dos pixels por 255.

Figura A.1 - a) Fotografia de uma plataforma de petróleo (Fonte: Petrobrás); b)Localização desta plataforma na imagem RADARSAT-1 c); recorte de imagem centrado na plataforma; e d) amostragem radial a partir do centro da imagem (plataforma)

Figura A.2 - a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando que a superfície é isotrópica num raio entre 6,5 km

Figura A.3 - a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando que a superfície se torna anisotrópica num raio entre 6,5 a 13 km

mento é resultado da reflexão difusa dos pulsos emitidos pelo RADAR nessa região. Um campo totalmente isotrópico pode ser observado na Figura 5, em que os valores de α_θ oscilam em torno de um ruído do tipo 1/f noise. Esses valores indicam as correlações de longo alcance ao longo de todas as direções seguem esse tipo de ruído. Para uma região de resíduo oleoso o campo anisotrópico é observado na Figura 6, entre as direções 240° e 270° o campo apresenta mesma variabilidade, pois as amostras estão dentro do resíduo oleoso. Contudo, os valores do expoente α_θ são maiores que para baixa intensidade de vento (Figura 4). Isso indica que a textura deve ser avaliada na discriminação de alvos escuros. Nas outras direções 0°, 50°, 210° e 275° o valor de α_θ é maior, dado a região de transição entre o resíduo oleoso e a água do mar.

Figura A.4 - a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando a anisotropia num região de baixa intensidade de vento

É interessante ressaltar que os valores de α_θ, além de reduzirem a dimensionalidade dos dados para análise, possuem um significado físico, indicando que existem corre-

Figura A.5 - a) Recorte de imagem centralizado na plataforma b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando uma superfície totalmente isotrópica

Figura A.6 - a) Recorte de imagem centralizado na plataforma; b) Visualização em superfície do recorte da imagem; c) Amostragem radial a partir da plataforma; e d) Valores de α_θ em coordenadas polares, mostrando a anisotropia devido ocorrência de resíduo oleoso

lações de longo alcance ao longo das direções. Isto está relacionado com a interação do pulso emitido pelo RADAR com a superfície do oceano. Em futuros trabalhos essa análise será explorada correlacionando a rugosidade do oceano observada pelo expoente α_θ com dados de correntes, ventos de superfície e modelos hidrodinâmicos de interação de dois fluidos.

A.5 Conclusões

Este trabalho apresentou uma nova abordagem para aplicação do método DFA para análise da anisotropia de campo 2-D. Um exemplo para demonstrar o potencial da metodologia na área de meio ambiente foi aplicado à indústria de exploração e produção de petróleo. Os resultados preliminares indicam a possibilidade da utilização do método DFA para caracterizar a presença de anisotropia ao redor das plataformas e/ou embarcações em imagens de RADAR. A detecção de anisotropia ao redor destes pontos pode vir a constituir informação auxiliar estratégica para o processo de interpretação visual das imagens, tendo potencial de indicar não apenas a presença

de anomalias visíveis ao redor das plataformas, como também, anomalias imperceptíveis ao olho humano, como orientações do campos de vento. O método pode ser utilizado como uma métrica capaz de reduzir a dimensionalidade dos dados, uma vez que, para cada alvo de interesse georeferenciado na imagem, pode-se criar uma métrica que descreva a anisotropia ao entorno do alvo na imagem. Neste mesmo contexto, a caracterização de séries espaço temporais em uma abordagem euleriana representa também outra aplicação potencial para o método DFA.

A metodologia DFA apresenta como limitação a utilização de raios muito pequenos para as análises, produzindo segmentos com poucos pixels, restritos a séries muito curtas (n<64). Esta limitação abre caminhos para o teste de novas metodologias para avaliação da anisotropia em imagens. Todavia, ressalta-se o potencial do método DFA para aplicações em áreas do sensoriamento remoto como, por exemplo, analisar a presença de anisotropia em alvos homogêneos em imagens ópticas derivadas de sensores hiperespectrais, proporcionando um estudo da anisotropia de um alvo em várias regiões do espectro eletromagnético.

Quanto ao método proposto cabe ressaltar que este pode ser generalizado também para 3 dimensões, bastando a utilização de geometria esférica. A combinação da metodologia com técnicas de classificação, oriundas do reconhecimento de padrões e inteligência artificial abre caminho para novas aplicações nos sistemas ambientais. Fora do domínio ambiental existe um potencial de emprego da metodologia no reconhecimento automático de alvos, como por exemplo, nos sistemas de navegação em tempo real.

Referências Bibliográficas

Addison P. S.(1997). **Fractals and Chaos**: An Illustrated Course. CRC Press, 256p.

Bashan, A.; Bartsch, R.; Kantelhardt, J. W.; Havlin, S.(2008). Comparison of detrending methods for fluctuation analysis, **Physica A: Statistical Mechanics and its Applications**, v. 387, n. 21, p. 5080-5090,Sep.

Bentz, C. M. (2006). **Reconhecimento Automático de Eventos Ambientais Costeiros e Oceânicos em Imagens de Radares Orbitais**. Tese de Doutorado. Rio de Janeiro, XIII, 115 p., COPPE/UFRJ, Engenharia Civil.

Brekke, C.; Solberg, A.H.S. (2005). Oil spill detection by satellite remote sensing, **Remote Sensing of Environment**, v. 95, n. 1,p. 1-13, mar.

Fingas, M.F.; Brown, C.E. (2005). Review Of Oil Spill Remote Sensing Presented at the Eighth. **Proceedings...** International Conference on Remote Sensing for Marine and Coastal Environments, Halifax, Nova Scotia, May.

Kantelhardt, J. W.; Stephan, A.; Zschiegner, E.; Koscielny-Bunde, E.; Havlin, S.; Bunde, A., Stanley, E. H. (2002). Multifractal detrended fluctuation analysis of nonstationary time series, **Physica A: Statistical Mechanics and its Applications**, v 316, n.1-4, p. 87-114, Dec.

Keramitsoglou I., Cartalis C., Kiranoudis C. (2006). Automatic Identification of oil spills on satellite images. **Environmenthal Modelling Software**, v. 21, p. 640-652.

Mandelbrot, B. (1983). **The Fractal Geometry of Nature.** New York: W.H. Freeman and Company. 420p.

Peng, C.K.; Buldyrev, S.V.; Goldberger, A.L.; Havlin, S.; Sciortino, F.; Simons, M.; Stanley, H.E. (1992).Long-range correlations in nucleotide sequences. **Nature**, v.356, n.6365, p. 168-170.

Peng, C.K.; Buldyrev, S.V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. (1994).Mosaic organization of DNA nucleotides. **Phys. Rev. E**, v.49, p.1685-1689.

Solberg, A.; Clayton, P.; Indregard, M. (2005). D2 - Report on benchmarking oil spill recognition approaches and best practice. **Kongsberg Satellite Services - Norway Archive** n.: 04-10225-A-Doc, 2.1.

Taqqu, M.S.; Teverovsky, V.; Willinger, W. (1995). Estimators for long-range dependence: an empirical study, **Fractals**, v. 3, n. 4, p.785-798.

Zheng, H.; Song, W.; Wang J. (2008).Detrended fluctuation analysis of forest fires and related weather parameters, **Physica A: Statistical Mechanics and its Applications**, v. 387, n.9, p. 2091-2099.

ANEXO B - Publicações Relacionadas ao Trabalho de Tese

Nas próximas páginas estão inclusas as publicações decorrentes dessa pesquisa, conforme regimento do curso de Computação Aplicada.

As demais publicações realizadas durante a pesquisa de doutorado de (2007-2012) e em projetos de colaboração podem ser acessadas em:

http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=C283835

Journal of Computational Interdisciplinary Sciences (2011) 2(1): 57-68
© 2011 Pan-American Association of Computational Interdisciplinary Sciences
Printed version ISSN 1983-8409 / Online version ISSN 2177-8833
http://epacis.org

Virtual laboratory of remote sensing time series: visualization of MODIS EVI2 data set over South America

Ramon Morais de Freitas[1], Egidio Arai[1], Marcos Adami[1], Arley Souza Ferreira[1], Fernando Yuzo Sato[1], Yosio Edemir Shimabukuro[1], Reinaldo Roberto Rosa[1], Liana Oighenstein Anderson[1,2] and Bernardo Friedrich Theodor Rudorff[1]

Manuscript received on February 4, 2011 / accepted on March 24, 2011

ABSTRACT

Over the last ten years millions of gigabytes of MODIS (Moderate Resolution Imaging Spectroradiometer) data have been generated which is forcing the remote sensing users community to a new paradigm in data processing for image analysis and visualization of these time series. In this context this paper aims to present the development of a tool to integrate the 10 years time series of MODIS images into a virtual globe to support LULC change studies. Initially the development of a tool for instantaneous visualization of remote sensing time series within the concept of a virtual laboratory framework is described. The virtual laboratory is composed by a data set with more than 500 million EVI2 (Enhanced Vegetation Index 2) time series derived from MODIS 16-day composite data. The EVI2 time series were filtered with sensor ancillary data and Daubechies (Db8) orthogonal Discrete Wavelets Transform. Then EVI2 time series were integrated into the virtual globe using Google Maps and Google Visualization Application Programming Interface functionalities. The Land Use Land Cover changes for forestry and agricultural applications are presented using the proposed time series visualization tool. The tool demonstrated to be useful for rapid LULC change analysis, at the pixel level, over large regions. Next steps are to further develop the Virtual Laboratory of Remote Sensing Time Series Framework by extending this work for other geographical regions, incorporating new computational algorithms, testing data from other sensors and updating the MODIS time series.

Keywords: MODIS, EVI2, wavelets transform, time series analysis, virtual globe, land use and land cover changes, forest, agriculture, South America, instantaneous visualization.

1 INTRODUCTION

Over the last few decades, multi-temporal images of Earth observation satellites have turned into a paramount source of information for monitoring the planet Earth, particularly to study the land use and land cover changes (LULC) [1]. Such studies are gaining more attention not only by scientists but also by policy makers and media, since terrestrial ecosystems exert major influence on climate change and climate variability [2].

Remote sensing sensors such as: the AVHRR (Advanced Very High Resolution Radiometer) on board of the NOAA (National Oceanic and Atmospheric Administration) satellites; the Vegetation on board of the SPOT (Satellite Pour l'Observation de la Terre) satellites; and the MODIS (Moderate Resolution Imaging

Correspondence to: Ramon Morais de Freitas – E-mail: ramon@dsr.inpe.br
[1]National Institute for Space Research – INPE, São José dos Campos, SP, Brazil.
[2]University of Oxford, Environmental Change Institute – ECI, Oxford, UK.

Spectroradiometer) on board of the Terra (EOS-AM1) and Aqua (EOS-PM1) satellites have been responsible for the construction of long term time series dataset. All these sensors acquire images on an almost daily basis which is an important characteristic for optical sensors designed to observe LULC changes. For the MODIS sensor a significant advancement was achieved by improving both spatial and spectral resolutions. Furthermore, an international consortium of scientists has focused on providing validated MODIS data with high radiometric and geometric quality, since the launch of the Terra satellite [3]. Over the last ten years millions of gigabytes of MODIS data have been generated which is forcing the remote sensing users community to a new paradigm in data processing for image analysis and visualization of these long term time series.

On the other hand the development of geobrowser tools, based on virtual globes, has provided free access to high spatial resolution images and geographical maps derived from remote sensing satellites. The development of these virtual globes allowed researchers and general public to visualize geospatial data, to understand multi-scale geography, to process data and to publish information [4, 5, 6]. The visualization of long term remote sensing data sets for scientific purposes has great potential for better understanding the complex space-time dynamics of terrestrial ecosystems. This tool is useful for scientists to understand more efficiently the different phenomena embedded in a large volume of data [7]. However, the pre-processing and extraction of information from these data sets require specific software and advanced technical knowledge to put them available to the end-users in a friendly and accessible way. The integration of time series for studies on LULC changes using virtual globes such as Google Maps (http://maps.google.com/) Google Earth (http://earth.google.com/) and Microsoft Virtual Earth (http://www.microsoft.com/maps/) are not yet easily accessible to users due to constraints in data storage and the lack of a specific computational architecture for integration and visualization of this time-series. In this context, this paper aims to present the development of a tool to integrate the 10 years time series of MODIS images into a virtual globe to support LULC change studies.

2 THE VIRTUAL LABORATORY OF REMOTE SENSING TIME SERIES

The macro framework of the *Virtual Laboratory of Remote Sensing Time Series* is divided into five components that are presented in Figure 1. The Data set component includes the hardware and software structures to storage the remote sensing time series data. The *Dataset manager* establishes the connections among all the laboratory components. The *Algorithm module* and the *Analysis module* provide the basis for the time series analysis and visualization. The *Visualization module* establishes the interface between the laboratory and the end-user using virtual globe facilities. The *Visualization module* is the main purpose of the present work and is described using the 10 years of MODIS time series data set.

2.1 Modis data set

The present work was developed for the South America continent that comprises about 18 million km^2, representing 12% of the Earth land surface. The region is characterized by different biomes such as tropical and seasonal forest, caatinga forest, grassland, savanna and others (Global Change Biology South America map [8]).

The MODIS images were acquired at the portal Warehouse Inventory Search Tool WIST NASA (https://wist.echo.nasa.gov). The selected product was the MOD13Q1 (collection 5) which is the composition of 16 days at spatial resolution of 250 m. The time frame of data acquisition is July 2000 to December 2010. The study area is divided in 29 MODIS tiles with 1,200 $\times$ 1,200 km each as illustrated in Figure 2. A total of 6,293 tiles were acquired corresponding to 3.5 TB of raw data. The images are in sinusoidal projection (WGS84 datum). All bands were reprojected to the geographic coordinate system with the same datum and converted from HDF (Hierarchical Data Format) to GeoTIFF format to ensure data portability among software. The total storage capacity is approximately 295 GB for each time series per spectral band. The vegetation index chosen for the present study is the EVI2 (Enhanced Vegetation Index 2; [9]) which highlights the land cover variations. It is computed using the surface reflectance of the *Red* and *NIR* (near infrared) bands available in the MOD13Q1 product (Equation 1):

$$EVI2 = 2.5 * \frac{NIR - Red}{(NIR + 2.4 * Red + 1)} \quad (1)$$

In addition, the view zenith angle band and the blue band (surface reflectance) were used to pre-filter the MODIS time series as explained in the next section.

2.2 Filtering Procedures

Optical remote sensing data are frequently affected by cloud cover and sensor noise that interfere in the ability to characterize spatial-temporal land cover dynamics. In order to construct a

Figure 1 – Virtual Laboratory of Remote Sensing Time Series framework.

continuous and consistent time series it is necessary to filter the data. The filtering procedures were performed in two steps generating two EVI2 time series: 1) without wavelet transform; and 2) with wavelet transform. The first filtering procedure applied rules that were designed following the methodologies proposed by [10, 11] in which data were eliminated from the original time series if the reflectance in the blue band is greater than 10% or if the sensor view zenith angle is greater than 32.5°. These threshold procedures eliminate clouds contaminated and off-nadir pixels. The filtered data were then linear interpolated based on the date of the pixel of the image composition to provide equally spaced time series.

Further filtering was applies in the second filtering procedure using the wavelet transform following the methodology proposed by [12]. The signal decomposition by wavelets eliminates the high frequencies typically associated with the presence of noise. The wavelet transform is given by [13, 14]:

$$W(a,b) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{|a|}} \psi_{a,b} * \left(\frac{t-b}{a} \right) dt \quad (2)$$

$$\psi_{a,b}(t) \equiv \frac{1}{\sqrt{a}} \psi \frac{(t-b)}{a}, a > 0, \ -\infty < b < +\infty \quad (3)$$

where a is the scale parameter, b is the translation parameter, $f(t)$ is the function to be transformed and ψ^* is the mother wavelet function complex conjugate. The function (Equations 2 and 3) is not continuous; therefore, it needs to be discretized by using discrete values of $(a, b) = (2^m, 2^n, k)$, where m, n and k are integer values and limited by the length of the time series. This allows the expansion of the mother wavelet to other scales. The *Discrete Wavelet Transform* (DWT) decomposes a discrete signal at different resolution levels. A DWT, defined in

Figure 2 – MODIS tiles coverage for South America.

Equation 4, is a mapping function of a discrete digital signal into a sequence of coefficients

$$DWT(m,k) = \frac{1}{\sqrt{a_0^m}} \sum_n^k x(n)\psi\left(\frac{k - nb_0a_0^m}{a_0^m}\right) \quad (4)$$

where ψ is the mother wavelet and the scaling parameters a and b are translation functions of the integer parameter m in which $a = a_0^m$, $b = nb_0a_0^m$ and k is an integer variable that refers to the number of samples of a given input signal (x) that depends on the signal length. The parameters a and b allow expanding the mother wavelet into several daughter wavelets.

The implementation of the discrete function (Equations 5, 6 and 7) is performed through recursive algorithms of low pass and high pass frequency filters known as pyramid algorithm [15]. For each decomposed signal in the jth scale signal there is a ratio coefficient of details (D) given by high-pass filters, and approximations (A) given by low-pass filters associated with the mother wavelet,

$$f_m(t) = A_m(t) + \sum_{j=1}^{m} D_j(t)\,, \quad (5)$$

$$D_M(t) = \sum_{k=-\infty}^{\infty} W_{m,k}\,\psi_{m,k}(t)\,, \quad (6)$$

$$A_m(t) = \sum_{k=-\infty}^{\infty} V_{m,k}\,\varphi_{m,k}(t)\,, \quad (7)$$

where $W_{m,k}$ are the wavelets coefficients, $\varphi_{m,k}$ is a scaled and translated basis function called scaling function and $V_{m,k}$ are the scaling coefficients.

The Daubechies wavelet Db8 was used as mother function (Fig. 3). This function is orthogonal ensuring that the decomposed signal is reconstructed without the presence of residues

Figure 3 – Wavelets functions: a) Db8 mother wavelet (ψ) and b) Scale function Db8 (φ).

due to asymmetries of the wavelet mother function. In this procedure, each set of pixels corresponds to a time series profile of the stacked images. The vector t was broken into 8 different scales and the time series was reconstructed using the five largest scales corresponding to lower frequencies. The higher frequencies were eliminated because they are usually associated with the presence of sensor noise and spectral responses contaminated by clouds and shadows.

The names given to the two EVI2 time series generated by the filtering procedures were:

1) *without wavelet*, and
2) *with wavelet*.

2.3 Data Integration and the Web Tool

The generated *Data set* (Fig. 1) is composed by over 500 million EVI2 time series filtered with and without wavelet transform for the entire South America continent. In order to construct this data set a significant computational effort was carried out involving more than 60 days of processing time using three personal computers (PC) with Linux OS. All computational procedures used Matlab and Ansi C platforms.

The EVI2 time series were integrated into the virtual globe (GoogleMaps) using the *Dataset manager* (Fig. 1) that was specifically developed for this purpose. To visualize the EVI2 time series in the virtual globe a website was built, which is available at http://www.dsr.inpe.br/laf/series.html, based on JavaScript and PHP platforms using Google Maps and Google Visualization Application Programming Interface functionalities. For each call of a geographic coordinate from the virtual globe the two EVI2 time series are instantaneously recovered. The information of the time series recovered by the call refers to one MODIS pixel. The integration with the virtual globe shows static geographical space using high spatial resolution satellite images provided by Google Maps server (Fig. 4a). However, caution should be taken for analyzing these time series due to the different spatial resolution of the images. Each MODIS pixel represents roughly an area of 6.25 ha (250 $\times$ 250 m) while the high spatial resolution image provided by Google Maps is only used to locate the MODIS pixel. In addition, a tool was build to assess the elevation anisotropy around the selected point (red balloon). This tool uses the elevation model information available in the Google maps API. The anisotropy visualization is a simple polar plot of elevation around two sample circles, allowing a rapid view of the topography around the selected point (Fig. 4c). This tool allows interactivity and provides a range of distance between the center of the selected coordinate and the sampled circles. Figure 4d shows the 10 years EVI2 time series using interactive plot provided by Google charts API functionalities. The red and blue lines represent the time series filtered *with wavelet* and *without wavelet*, respectively.

3 APPLICATIONS

The intense anthropogenic pressure forces the LULC change processes in South America by converting natural forest and savanna areas to pasture and agriculture. At sub-tropical region

Figure 4 – Website display components: a) Google earth virtual globe used to select the geographic coordinate of the area of interest; b) list of selected points; c) polar plot of elevation around two sampled circles showed in Google image; and d) EVI time series for the selected plot (red line filtered *with wavelet* and blue line filtered *without wavelet*).

there is the intensification of agriculture related to food and biofuels production [16, 17]. In this context some application examples are provided using the visualization module described in the present work for LULC change analysis.

3.1 Land Use and Land Cover Change Over Forest

Figure 5a shows a forested region with several deforested areas in the municipality of Feliz Natal in Mato Grosso state, Brazil. The elevation around the selected point (11°55′S; 54°10′W) varies from 354 to 360 m indicating a reasonable flat terrain (Fig. 5b). The MODIS EVI2 time series are presented in Figure 5c. Analyzing the time series a decrease in the EVI2 values can be observed in 2004 indicating a significant biomass loss due to the deforestation process. From 2005 to 2007 there was almost no vegetation regrowth as indicated by the low EVI2 profile during this period. For the years of 2007 and 2008 a typical spectral response for agricultural areas can be observed, characterized by a rapid increase followed by a rapid decrease of the vegetation index values indicating the well defined and short growth cycles of agricultural annual crops.

Figure 6 shows the seasonality of a selected point from an area that was deforested in 2004 in the National Park of Xingu, Mato Grosso state, Brazil. The visual analysis of the time series indicates a land conversion from forest to pastures. It can also be observed that the deforestation process started in the first quarter of 2004 and ended in the last quarter of the same year. It is interesting to note that the two areas observed in Figures 5 and 6 present different types of land use change that can be easily observed by analyzing the 10 years MODIS EVI2 time series. The double arrows illustrate the land use or land cover in the period.

Figure 5 – a) GoogleMaps image; b) elevation polar plot; and c) EVI2 time series plot for the selected coordinated point.

It is also interesting to note that the two differently filtered EVI2 time series generated similar curves for the selected plots of Figures 5c and 6c. A smoother curve is observed for the EVI2 *with wavelet* (red curve), and but no significant difference is observed for the *without wavelet* (blue curve) in terms of LULC change analysis. On the other hand the smoothed filtered EVI2 time series can be used in the data mining and other classification procedures when the high frequency signal is not interesting.

3.2 Land Use and Land Cover Change Over Sugarcane

Figure 7a shows a region with intense sugarcane cultivation in the municipality of Brotas, São Paulo state, Brazil. Figure 7c shows the dynamic of nine sugarcane crop years. With some knowledge on sugarcane cultivation several information can be extracted from the time series curve. A brief description of the sugarcane cultivation practices in this area can be given by the analyst in the following form. At the very beginning of the time series the low EVI2 values indicate bare soil over which the sugarcane was planted in the beginning of 2001. This sugarcane plant grew for about 18 months when it was harvested for the first time around July 2002. After the first cut the sugarcane ratoons were harvested once a year, around July, from 2003 to 2007. The sugarcane field was renovated in late 2007 when it was rotated with an annual summer crop followed by new sugarcane planted in late 2008 and harvested in mid 2009. More information about this plot can be obtained at http://www.dsr.inpe.br/laf/canasat/ [17].

The above description indicated that with a minimum of technical knowledge about sugarcane agricultural practices it is possible to the analyst recovering the 10 years history of specific plots and fields. This can be of great interest to certifiers that need to know the LULC change history.

3.3 Land Use and Land Cover Change Over Savanna

Figure 8 shows an agricultural region at the frontier of the savanna located in western Bahia state, Brazil. The region was originally covered by savanna and has been gradually converted

Figure 6 – a) GoogleMaps image; b) elevation polar plot; and c) EVI2 time series plot for selected coordinated point.

to intense agricultural land use. This region is characterized by large soybeans, corn, cotton and coffee plantations. Three time series over typical plots can be observed in Figures 8c, 8d and 8e. Figure 8c shows the typical behavior of a savanna with relative low values EVI2 and small amplitudes. Figure 8d presents the curve of a savanna area which was converted to agriculture during the 2007/2008 crop season. Figure 8e shows the conversion of savanna to agriculture after 2002.

4 GENERAL REMARKS AND FUTURE WORKS

This work presented a visualization module for the virtual laboratory using the ten years history of MODIS EVI2 time series for the entire South America continent. The visualization module demonstrated to be useful for rapid land use and land cover change analysis, at the pixel level, over large regions. The smoothed filtered EVI2 time series using wavelet transformation could be used for checking dates of land use change such as date of deforestation, planting date of agricultural crops, seasonal effect on pasture land and others. Regarding to the integration with virtual globes such as Googlemaps, this work showed an innovation because it allows the public access and instantaneous visualization. This work can be extended for any geographical region since MODIS data are available for the entire globe.

Future works will be focused on providing quality parameters associated with each time-series, through an under-development validation procedure. Due to the area coverage, a near-real time open access techniques will be evaluated for allowing land cover types to be detected by any analyst with a GPS. Table 1 describes the main algorithms and tools for data analysis to be implemented in future virtual laboratory versions. A user friendely interface will be implemented to connect the Algorithm module and the Analysis module facilities with virtual laboratory users.

Next versions of the *Virtual Laboratory of Remote Sensing Time Series* can be implemented with other data sets such as temperature, rainfall, fraction components of linear mixture model and

Figure 7 – a) GoogleMaps image; b) elevation polar plot; and c) EVI2 sugarcane time series plot for selected coordinated point.

Table 1 – Main algorithms and tools for data analysis to be implemented in the Virtual Laboratory of Remote Sensing Time Series.

Algorithms and Analysis Tools	Applications
Time-Frequency Analysis Tools (Fourier, Wavelets, Chiplets transforms) [18, 19, 20]	Understand time series variability based on spectral and temporal analysis tools
Gradiente Pattern Analysis [21, 22, 23]	Non linear advanced time series analysis based on symmetric and structural concepts
Detrended Fluctuation Analysis [24, 25]	Non-linear advanced time series analysis based on scale invariance concepts
Cluster Analysis Tools [26, 27]	Dataset analysis and classifications
Support Vector Machines and Decision Tree Classifiers [28, 29, 30]	Temporal analysis and Land Use and Land Cover Change classifications
GIS file formats integration tools [31, 32]	Supplementary analysis
Statistical regions analysis tools [33]	Regional and local focus analysis
Dataset subset and download facilities	Peopleware and users facilities
Metafile and Metadata usage [33]	Ancillary data for image interpretation and Land Use and Land Cover Analysis
Seasonality tools [18]	Compute the pos-processing products such as: annual, monthly, regional anomalies and phenology cycles.

Figure 8 – a) GoogleMaps image; b) elevation polar plot; c) EVI2 time series plot for selected coordinated point 1, savanna area; d) EVI2 time series plot for selected coordinated point 2, deforestation area in 2007; and e) EVI2 time series plot for selected coordinated point 3, deforestation area in 2002.

vegetation indices. The concept, protocols, hardware support and other software engineering specifications have been carried out to make the instantaneous visualization and analysis of time series a reality for remote sensing and general GIS users community.

ACKNOWLEDGMENTS

The authors thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) agencies for partial financial support to the work, and INPE (Earth Observation – OBT, Remote Sensing Division, DSR and Laboratory of Agriculture and Forest – LAF) for infrastructure and computing support. The MODIS data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (lpdaac.usgs.gov). The author thanks the referees for their helpful suggestions concerning the presentation of this paper.

REFERENCES

[1] LAMBIN EF & LINDERMAN M. 2006. Time series of remote sensing data for land change science. IEEE Transactions on Geoscience and Remote Sensing, 44(7): 1926–1928.

[2] DEFRIES RS, ASNER GP & HOUGHTON RA. 2004. Ecosystems and Land Use Change. American Geophysical Union, Washington, DC.

[3] JUSTICE CO, TOWNSHEND JRG, VERMOTE EF, MASUOKA E, WOLFE RE, SALEOUS, N, ROY DP & MORISETTE JT. 2002. An overview of MODIS Land data processing and product status. Remote Sensing of Environment, 83: 3–15.

[4] BUTLER D. 2006. Virtual globes: the web-wide world. Nature, 439(7078): 776–778.

[5] BALLAGH LM, RAUP BH, DUERR RE, KHALSA SJS, HELM C, FOWLER D & GUPTE A. 2011. Representing scientific data sets in KML: Methods and challenges, Computers & Geosciences, 37(1), Virtual Globes in Science, p. 57–64. ISSN 0098-3004, DOI: 10.1016/j.cageo.2010.05.004.

[6] CHIANG G, TOBY OH, DOVE MT, BOVOLO CI & EWEN J. 2011. Geo-visualization Fortran library, Computers & Geosciences, 37(1), Virtual Globes in Science, p. 65–74, ISSN 0098-3004, DOI: 10.1016/j.cageo.2010.04.012.

[7] NIELSON GM. 1991. Visualization in Scientific and Engineering Computation. IEEE Computer, 24(9): 58–66.

[8] EVA H, BELWARD A, EVARISTO M, DI BELLA C, GOND V, JONES S, SGRENZAROLI M & FRITZ S. 2004. A land cover map of South America, Global Change Biology, 10: 731–744.

[9] JIANG Z, HUETE AR, DIDAN K & MIURA T. 2008. Development of a two-band Enhanced Vegetation Index without a blue band. Remote Sensing of Environment, 112(10): 3833–3845.

[10] SAKAMOTO T, YOKOZAWA M, TORITANI H, SHIBAYAMA M, ISHITSUKA N & OHNO H. 2005. A crop phenology detection method using time series MODIS data. Remote Sensing of Environment, 96(3-4): 366–374.

[11] THAYN JB & PRICE KP. 2008. Julian dates and introduced temporal error in remote sensing vegetation phenology studies. International Journal of Remote Sensing, 29: 6045–6049.

[12] FREITAS RM & SHIMABUKURO YE. 2008. Combining wavelets and linear spectral mixture model for MODIS satellite sensor time series analysis. JCIS – Journal of Computational Interdisciplinary Sciences, 1: 51–56.

[13] DAUBECHIES I. 1992. Ten lectures on wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics 61, Philadelphia, PA. Soc. Ind. Appl. Math, 377 pp.

[14] MEYER Y. 1992. Wavelets and operators, Cambridge Studies in Advanced Math., vol. 37, Cambridge Univ. Press, Cambridge, 223 p.

[15] MALLAT S. 1989. A theory for multi resolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11: 674–693.

[16] LAPOLA DM, SCHALDACH R, ALCAMO J, BONDEAU A, KOCH J, KOELKING C & PRIESS JA. 2010. Indirect land-use changes can overcome carbon savings from biofuels in Brazil. Proceedings of the National Academy of Sciences, 107(8): 3388–3393.

[17] RUDORFF BFT, AGUIAR DA, SILVA WF, SUGAWARA LM, ADAMI M & MOREIRA MA. 2010. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. Remote Sensing, 2: 1057–1076.

[18] MALLAT S. 1999. A wavelet tour of signal processing, 2nd Edition, Academic Press.

[19] LE PENNEC E & MALLAT S. 2005. Sparse Geometric Image Representation with Bandelets, IEEE Trans. on Image Processing, 14(4): 423–438.

[20] BOASHASH B. 2003. Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, Oxford: Elsevier Science.

[21] ROSA RR, PONTES J, CHRISTOV CI, RAMOS FM, RODRIGUES NETO C, REMPEL EL & WALGRAEF D. 2000. Physica A, 283: 156.

[22] ASSIREU AT, ROSA RR, VIJAYKUMAR NL & LORENZZETTI JA. 2002. Gradient pattern analysis of short nonstationary time series: an application to Lagrangian data from satellite tracked drifters. Physica D, Elsevier, 169c: 397–403.

[23] FREITAS RM, ROSA RR & SHIMABUKURO YE. 2010. Using Gradient Pattern Analysis for land use and land cover change detection. In: International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, Honolulu, 1: 3648–3651.

[24] PENG CK et al. 1994. Mosaic organization of DNA nucleotides. Phys Rev E, 49(2): 1685–1689.

[25] KANTELHARDT JW et al. 2001. Detecting long-range correlations with detrended fluctuation analysis. Phys A, 295(3-4): 441–454.

[26] DUDA RO, HART PE. 1973. Pattern Classification and Scene, Analysis, New York: John Wiley & Sons, Inc.

[27] HARTIGAN JA. 1985. Statistical Theory in Clustering. Journal of Classification, 2: 63–76.

[28] THEODORIDIS S, KOUTROUMBAS K. 2009. Pattern Recognition, 4th Edition, Academic Press.

[29] YANG T. 2006. Computational Verb Decision Trees. International Journal of Computational Cognition (Yang's Scientific Press), 4(4): 34–46.

[30] YUAN Y, SHAW MJ. 1995. Induction of fuzzy decision trees. Fuzzy Sets and Systems, 69: 125–139.

[31] LONGLEY P, GOODCHILD MF, MAGUIRE DJ, RHIND DW. 2005. Geographical information systems and science: John Wiley & Sons Inc.

[32] LONGLEY P. 2008. To what extent are the fundamental spatial concepts that lie behind GIS relevant in design? In Spatial Concepts in GIS and Design. Santa Barbara, CA: UCSB.

[33] BRETHERTON FP, SINGLEY PT. 1994. Metadata: A User's View, Proceedings of the International Conference on Very Large Data Bases (VLDB). pp. 1091–1094.

USING GRADIENT PATTERN ANALYSIS FOR LAND USE AND LAND COVER CHANGE DETECTION

Ramon Morais de Freitas, Reinaldo Roberto Rosa, Yosio E. Shimabukuro

Instituto Nacional de Pesquisas Espaciais (INPE)
Av. dos Astronautas, 1758 - São José dos Campos, SP, 12227-010, Brazil
E-mail: {ramon, yosio}@dsr.inpe.br; reinaldo@lac.inpe.br

ABSTRACT

In this work, the computational operation based on Gradient Pattern Analysis - GPA was applied for the first time in MODIS spatial-temporal images over the Amazon region. The study area is located in the Pará State, eastern Brazilian Amazonia. Using MOD09 8-day composite product from 2000 to 2009 was elaborated the EVI2 spatial-temporal series of the study area. For each pixel we performed smooth time-series applying wavelets transform method for noise reduction. The GPA objective was characterizing small symmetry breaking, amplitude and phase disorder due to spatial-temporal fluctuations driven by the deforestation and flooded changes detected by MODIS images. For the characterization of spatial-temporal series the Gradient Pattern Analysis showed a new approach to understand LULC changes in the remote sensing images.

***Index Terms*—** Amazonia, LULC, Gradient Pattern Analysis, deforestation.

1. INTRODUCTION

The land use and land cover (LULC) change monitoring using remote sensing data is an important instrument for Government policy and surveillance of preservation areas. A near real-time change detection and characterization method would allow control of the increase of new anthropic and natural environmental changes. In general, the biodiversity is reduced by LULC change [1].

Remote sensing and geospatial analysis are substantive tools of land change studies because these technologies facilitate observations across large extends of Earth surface. In global scales, the Global Land Cover Facility (GLCF, glcf.umiacs.umd.edu) provides remote sensing data set and products for quantifying the land cover change around the world. The GLCF investigates the land cover dynamics and remote sensing derived products to explain the LULC changes. The multi-temporal MODIS higher-level quality data sets are used to understand this dynamics. In regional scale, Brazilian Government policy has been monitoring Amazonia deforestation process by remote sensing tools as DETER and PRODES INPE's programs (www.inpe.br).

Maps and measurements of LULC changes can be directly derived from remote sensing data using image processing, statistical pattern recognition methods and human interpretation. In general, this processing transforms the remote sensing data to thematic maps or classification images applying supervised or unsupervised procedures. Sophisticated computational methodologies using object image segmentation and neural networking classifiers aren't reality on the LULC change operational programs because the computational cost. This is a rule set for multi-temporal monitoring on the global and regional scales operational programs. The feature selection and original dataset dimension reduction is an alternative procedure for this problem. In this context, this work has the objective to propose a low computational cost methodology to detect and to monitor LULC changes using spatial temporal remote sensing data. Traditional methods of LULC changes detection is based on pattern recognition and statistical measurements, while the proposed method encompasses statistical physics and non linear dynamics concepts [2] to understand the LULC detection.

2. METHODOLOGY

The proposed methodology uses Gradient Pattern Analysis [3,4] a modern technique for analyzing spatially extended dynamics[1]. The measurements obtained from GPA are based on the spatial-temporal correlations between large and small amplitude fluctuations of the structure represented as a dynamical gradient pattern. Based on a scalar elementary field (Fig. 1-a) which can be represented by pixels values of an image or subset image in remote sensing, the first moment is represented by gradient field (Fig. 1-b), the second (Fig. 1-c) and third moments (Fig. 1-d) is a norm and phase representation, respectively, for each element of the

[1] Statistical physics concept that is related to 2-D spatial pattern marked by gradual changes through a series of states.

first moment. The fourth moment is the Euler's formula for complex representation of the first gradient moment which shows the relationship between the trigonometric and the complex exponential functions. By means of four gradient moments is possible to quantify the relative fluctuations and scaling coherence at a dynamical numerical lattice and this is a set of proper measures of the pattern complexity and equilibrium [3, 4].

Fig. 1. Gradient Pattern Analysis - The gradient moments.

Using gradient moments we can define computational operators as LULC change metric. The second and third gradient moment represents the norm and phase of gradient field, respectively. Using these moments we could measure the diversity of norm and phase of gradient field applying the Simpson's Index of diversity over the second and third gradient moments. These two computational operators are defining by:

$$D_{g_2} = 1 - \sum_{i=1}^{S} Pr_i^2 \quad (1)$$

$$D_{g_3} = 1 - \sum_{i=1}^{S} P\emptyset_i^2 \quad (2)$$

where, Pr and PØ are the proportion of norms and phases values on gradient field, respectively. S is total of unique elements in the gradient field base on histogram bins. A perfectly homogeneous gradient field in norm and phase would have a diversity index score of 0. A perfectly heterogeneous gradient element would have a diversity index score of 1. To measure the energy of gradient field, we define a Frobenius metric that relate the total of energy of gradient field:

$$||N_{g_2}||_F = \sum_{i=1}^{m} \sum_{j=1}^{n} |r_{i,j}|^2 \quad (3)$$

being, $||N_{g_2}||_F$ the Frobenius norm of g_2 matrix, m and is row and columns, respectively. The gradient field represented by Cartesian coordinates x and y directions we can be used the Frobenius inequality [5] as a metric.

$$||N_I||_F = \frac{||dxdy - dydx||_F}{||dx||_F\,||dy||_F} \leq \sqrt{2} \quad (4)$$

where, $||N_I||_F$ is the Frobenius index defined by Frobenius inequality. dx an dy are gradient field represented by Cartesian coordinates. This $||N_I||_F$ value can be used to compare the gradient fields in the same size.

In this work, the computational operators' base on GPA was applied for the first time in MODIS spatial-temporal images over the Amazon region. The GPA objective was to characterize small symmetry breaking, amplitude and phase disorder due to spatial-temporal fluctuations driven by the deforestation and flooded changes detected by MODIS images. Combining the diversity index and Frobenius metrics we can establish a dynamics pattern of land cover changes.

3. EXPERIMENTS

3.1. Remote Sensing Data and Processing

The study area is located in the Pará State, eastern Brazilian Amazonia (Figure 2). The region has well-defined dry and wet seasons with yearly rain about 2,100 mm a dry season occurring from June to October. The test site encompasses several landscape types as tropical forest, regrowth, deforested areas, croplands and pasture. It presented high land cover changes rates in the last years. In this region, recent agricultural expansion and flooding episodes occurrences carried out a constant land cover changes. The samples of land cover classes were collected to photo interpretation to demonstrate the method.

Using MOD09 8-day composite product from 2000 to 2009 was elaborated the spatial-temporal series of the study area. The time-series of MODIS/Terra surface reflectance images are from collection 5 of 8-day L3 Global product at 250 m resolution (MOD09A1), acquired from 2000 to 2009. 414 images were used in this study. This MODIS product is an estimate of the surface spectral reflectance for each band, as it would have been measured at ground level with no atmospheric scattering or absorption effects, generated by applying the atmospheric correction algorithm. The MOD09A1 product includes the RED and NIR bands with originally 250 m spatial resolution data.

The next step we calculate the Enhanced Vegetation Indices 2 - EVI2 [6] using MOD09 RED and NIR infrared spectral bands for each composite. The EVI2 has best similarity with traditional EVI. The EVI2 index was selected due to enhance the vegetation signal with improved sensitivity in high biomass regions as Amazonia forest regions.

Fig. 2. Location of study area with selected plots. The image is an EVI2 filtered 2002 wet season image.

For each pixel we performed smooth time-series using wavelets transform method for noise reduction [7]. The procedure was necessary due to high noise data. For each image we compute norm, phase diversity index and Frobenius norm metrics in the 50x50 pixel window. Utilizing field data and photo interpreter experience, 4-four regions with 50x50 pixels were selected with land use and land cover changes (i.e. deforestation process and seasonally flooded forest region) for demonstrating the method. Using scatter plot of this two metrics we show the dynamics of these areas. The Dg3 versus $||N_{g_2}||_F$ scatter plot represents the space phase of change. In this context, the temporal dynamics changes could be visualized by this scatter plot. The $||N_I||_F$ Frobenius index used for data dimensionality reduction represent the spatial-temporal behavior in the 1-D time-series.

4. RESULTS

Fig. 3 shows the multi-temporal images of June for 2000 to 2009 time period. Each image has 50x50 MODIS 250 meters pixels with a total area of 156 km^2. Fig. 4 shows the scatter plot of phase diversity and norm gradient variables from spatial- temporal series. Seasonally flooded forest (Site 1) region has variability in phase data due to permanent edge contrast between the river banks and the forest. High dispersion region is related to the edge presence due to deforestation and natural contrast regions (i.e. deforestation areas and river banks and). The forest without deforestation activity shows the high diversity index and low norm values, that's represent small variation of gradient field. This small variation is related with sensor noise and forest canopy structure. The deforestation region shows permanent areas of non photosynthetic vegetation, this cluster is compact because no land cover change occurs during period. The forest with deforestation activity shows a transition phase from forest cluster region to new cluster of deforestation. The increase of Frobenius norm ($||N_{g_2}||_F$) is related with the biomass losses driven by deforestation activity in June of 2004. For large deforestation regions the cluster dispersion of Frobenius norm axis tends to increase.

Figure 5 shows the multi-temporal $||N_I||_F$ behavior for 4 study sites. The white noise bounds are represented by point lines. This area represents a random gradient field, in this context, forest areas have a similar behavior with white noise. The deforestation process (Figure 4, site 4) is detected by $||N_I||_F$ metric (Figure 5, Cyan line) with rapid increase of metric values. A transition phase in the flooded areas (Red line) is observed in 2005 year probably related to strong drought occurrence.

Fig. 3. Multi-temporal EVI2 images for 4 study sites. Site 1 – – seasonally flooded forest region.; Site 2 – forest without deforestation activity; Site 3 – deforestation region and Site 4- forest with deforestation activity .

Fig. 4. Scatter plot of Phase Diversity Index and Frobenius norm of g_2. Red - Site 1; Blue - Site 2; Green - Site 2; and Cyan - Site 4 ;

Fig. 5. Multi-temporal $||N_I||_F$ behavior for 4 study sites.

5. CONCLUSION

This work showed the preliminary results of characterization of spatial-temporal remote sensing series. The Gradient Pattern Analysis showed a new approach to visualize land cover change directly in the remote sensing images. The temporal behavior of these three metrics series indicated a new potential metric to understand the land cover and land use dynamics applied directly to remote sensing images. Future work will concentrate on the development of procedures for selecting appropriate thresholds for phase diversity versus norm scatter plot.

6. ACKNOWLEDGMENTS

This study is part of the Large-Scale Biosphere Atmosphere Experiment in Amazonia (LBA) and was partially supported by the National Institute for Space Research (INPE) and FAPESP grant number 2008/05268-3.

7. REFERENCES

[1] DeFries, R. S. , G. P. Asner, and R. A. Houghton, *Ecosystems and Land Use Change.* American Geophysical Union, Washington, DC. 2004.

[2] H.E. Stanley, "Phase Transitions, Power laws and universality". *Nature* , 378 , 554. 1995.

[3] R.R. Rosa, A.S. Sharma, J.A. Valdivia, "Characterization Of Asymmetric Fragmentation Patterns In Spatially Extended Systems" *Int. J. Mod. Phys.C*, 10(1), p.147. 1999.

[4] F.M. Ramos, R.R. Rosa, C. Rodrigues Neto, A. Zanandrea. "Complex Entropic Forms for Gradient Pattern Analysis of Spatio-temporal Dynamics", *Physica A*, p.283, 171 2000.

[5] A. Böttcher, D.Wenzel, "How big can the commutator of two matrices be and how big is it typically?", *Linear Algebra Appl.* 403. p. 216–228, 2005.

[6] Z. Jiang, A. R. Huete, K. Didan,, T Miura. "Development of a two-band Enhanced Vegetation Index without a blue band", *Remote Sensing of Environment*, 112(10), p 3833-3845. 2008.

[7] R. M Freitas, Y. E. Shimabukuro . "Combining wavelets and linear spectral mixture model for MODIS satellite sensor time-series analysis", *JCIS - Journal of Computacional Interdisciplinary Sciences*, v. 1, p. 51-56, 2008.

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.