Ministério da
Ciência, Tecnologia e Inovação

sid.inpe.br/mtc-m19/2011/11.15.22.02-TDI

GRADE NUMÉRICA GENERALIZADA: UM NOVO CONCEITO PARA REPRESENTAÇÃO E VISUALIZAÇÃO ANALÍTICA DE SISTEMAS DE SÉRIES TEMPORAIS

Thalita Biazzuz Veronese

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Maurício José Alves Bolzan, e Reinaldo Roberto Rosa, aprovada em 24 de novembro de 2011.

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

INPE
São José dos Campos
2011

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

Ministério da
Ciência, Tecnologia e Inovação

sid.inpe.br/mtc-m19/2011/11.15.22.02-TDI

GRADE NUMÉRICA GENERALIZADA: UM NOVO CONCEITO PARA REPRESENTAÇÃO E VISUALIZAÇÃO ANALÍTICA DE SISTEMAS DE SÉRIES TEMPORAIS

Thalita Biazzuz Veronese

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Maurício José Alves Bolzan, e Reinaldo Roberto Rosa, aprovada em 24 de novembro de 2011.

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

INPE
São José dos Campos
2011

Dados Internacionais de Catalogação na Publicação (CIP)

V559g

Veronese, Thalita Biazzuz.
Grade numérica generalizada: um novo conceito para representação e visualização analítica de sistemas de séries temporais / Thalita Biazzuz Veronese. – São José dos Campos : INPE, 2011.
xxvi + 92 p. ; (sid.inpe.br/mtc-m19/2011/11.15.22.02-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011.
Orientadores : Drs. Maurício José Alves Bolzan, e Reinaldo Roberto Rosa.

1. análise de séries temporais. 2. analítica visual. 3. grade numérica generalizada. 4. representação de dados espaçotemporais. I.Título.

CDU 519.246.8:004.62

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

Computação Aplicada

Dr. Nandamudi Lankalapalli Vijaykumar

Presidente / INPE / SJCampos - SP

Dr. Reinaldo Roberto Rosa

Orientador(a) / INPE / SJCampos - SP

Dr. Solon Venâncio de Carvalho

Membro da Banca / INPE / SJCampos - SP

Dr. Alexandre Souto Martinez

Convidado(a) / USP/ Ribeirão P. / Ribeirão Preto - SP

Dr. Arcilan Trevenzoli Assireu

Convidado(a) / UNIFEI / Itajubá - MG

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

Aluno (a): Thalita Biazzuz Veronese

São José dos Campos, 24 de novembro de 2011

"It's impossible, that's sure! So let's start working".

Philippe Petit
em *"Man on wire"*, 2008

A meus maravilhosos pais, Carlos Eduardo e Irene

AGRADECIMENTOS

Agradeço primeiramente a meus pais, pela educação, pelo amor e, acima de tudo, pelo exemplo de vida.

Agradeço a minha amada e admirada irmã, que me enche de inspiração com suas conquistas e me ampara com seu carinho sempre que preciso.

Agradeço a toda minha família, que considero um grande presente e que me garante os momentos de alegria e união sem os quais não é possível realizar qualquer conquista.

Agradeço aos meus amigos, que para minha sorte continuam ao meu lado apesar da minha ausência nos últimos anos.

Agradeço a cada um dos meus alunos, com os quais aprendo diariamente.

Agradeço a cada um dos professores que tive, aos quais devo minha interminável vontade de aprender.

Agradeço aos meus orientadores, eternos mestres em minha jornada, pelas lições e conversas, pela compreensão e pela confiança.

Agradeço ao INPE, pelas oportunidades concedidas e pelo apoio oferecido em todos os momentos desta pesquisa de doutorado, e por todas as amizades que tive o prazer de encontrar neste período.

Agradeço ao IEAv/CTA e à Anhanguera Educacional, pelas oportunidades concedidas para que eu pudesse amadurecer profissionalmente em fases fundamentais desta pesquisa, e a todos os colegas e amigos com quem tive a oportunidade de conviver durante estes períodos.

Agradeço ao CNPq (Processo nº141325/2008-9) e à FAB pelo apoio financeiro, fundamental para o desenvolvimento desta pesquisa.

Agradeço ao IFSP de Campos do Jordão, por me acolher e me apoiar neste momento de transformações e muita tensão, a que pretendo retribuir com minha intensa dedicação, e aos amigos com quem tenho a alegria de conviver nesta instituição.

Finalmente, agradeço a meu marido, a mais bela parceria que jamais imaginei que

pudesse existir, e tive a sorte de encontrar.

RESUMO

A Analítica Visual (*Visual Analytics*) é uma abordagem essencialmente computacional – incorporando aspectos multidisciplinares – dedicada à análise visual de um sistema de informação devidamente formatado para a inspeção e interpretação de dados por um especialista humano. Este trabalho está direcionado para o tratamento completo, incluindo a representação visual, de um sistema massivo de séries temporais. Este tratamento considera sua organização generalizada, a aplicação de ferramentas de análise e, por último, sua representação de forma que a visualização analítica possa ser utilizada pelo cientista na interpretação integrada dos dados para tomadas de decisão.Neste trabalho, é introduzido o conceito de grade numérica generalizada (GNG), que generaliza toda série temporal N-dimensional a partir de três tipos de parâmetros: (i) grau variacional; (ii) coeficientes de extensão; e (iii) medidas analíticas. Demonstramos neste trabalho que qualquer conjunto de medidas nos domínios do tempo e do espaço pode ser reduzido a uma GNG. A funcionalidade, praticidade e robustez de uma representação de dados baseada no conceito de GNG são investigadas a partir de exemplos de aplicações em física espacial e ambiental. A pesquisa em analítica visual desenvolvida neste projeto de doutorado em Computação Aplicada envolveu a investigação de métodos de geração de séries temporais N-dimensionais não lineares para construção de uma base canônica de dados, com o objetivo de testar a complexidade do modelo proposto. O método para análise de flutuações foi adotado como método analítico padrão para séries temporais, e o modelo de representação de GNGs foi desenvolvido em C++, XML e Java. Considerando o novo conceito proposto para grade numéricas generalizadas, como resultados principais desta pesquisa podemos destacar: proposta e apresentação de um novo modelo para representação e visualização analítica de séries temporais; e investigação de um gerador de dados canônicos para teste do modelo no contexto da sua aplicação para o monitoramento do clima espacial.

GENERALIZED NUMERICAL LATTICE: A NEW CONCEPT FOR ANALYTICAL VISUALIZATION AND REPRESENTATION OF TIME SERIES SYSTEMS

ABSTRACT

Visual Analytics is an essentially computational approach – incorporating multidisciplinary aspects – dedicated to the visual analysis of an information system to properly formatted to data inspection and interpretation by a human expert. This work is directed to the full treatment, including the visual representation, of a massive time series system. This treatment considers its generalized organization, the application of analysis tools and, finally, its representation so that the analytical visualization can be used by scientists in the integrated interpretation for decision making. In this paper, we introduce the concept of generalized numerical lattice (GNL), which generalizes every N-dimensional time series from three types of parameters: (i) variational degree; (ii) extension coefficients; and (iii) analytical measurements. We have shown in this work that any set of measures in time and space domains can be reduced to a GNL. Functionality, practicity and robustness of a data representation based on GNL concept are investigated through examples of applications in space and environmental physics. The visual analytics research developed in this doctoral program in Applied Computing involved the investigation of methods for generating nonlinear N-dimensional time series to construct a canonical data base, in order to test the proposed model complexity. The fluctuation analysis method was adopted as a standard analytical method for time series, and the GNLs representation model was developed using C++, Java and XML. Given the new concept for generalized numerical lattices, as major results of this research we can highlight: the proposal and presentation of a new model for representing and visualizing analytical time series, and the investigation of a canonical data generator for testing the model in the context of its application in space weather monitoring.

LISTA DE FIGURAS

Pág.

LISTA DE TABELAS

Pág.

LISTA DE ABREVIATURAS E SIGLAS

2D	–	Bidimensional
3D	–	Tridimensional
API	–	Application Program Interface
ATP	–	Adenosin Triphosphate
CDF	–	Common Data Format
CME	–	Coronal Mass Ejection
DFA	–	Detrended Fluctuation Analysis
DSR	–	Data System Representation
ECG	–	Eletrocardiograma
EMBRACE	–	Estudo e Monitoramento Brasileiro do Clima Espacial
FITS	–	Flexible Image Transport System
GNG	–	Grade Numérica Generalizada
GNL	–	Generalized Numerical Lattice
GOES	–	Geostationary Operational Satellites
GOIN	–	Global Observation and Information Network
GPA	–	Gradient Pattern Analysis
HDF	–	Hierarchical Data Format
ICSU	–	International Council of Scientific Unions
IDH	–	Índice de Desenvolvimento Humano
IPCC	–	Intergovernmental Panel over Climate Change
ISTP	–	International Solar-Terrestrial Physics
KPNO	–	Kitt Peak National Observatory
LAC	–	Laboratório Associado de Computação e Matemática Aplicada
MGNLA	–	Module GNL Assembly
NOAA	–	National Oceanic and Atmospheric Administration
OGC	–	Open Geospatial Consortium
OLAP	–	On-Line Analytical Processing
PSD	–	Power Spectrum Density
SAJ6	–	Solar Activity on June 6, 2000
SET	–	Série Espaçotemporal
SFU	–	Solar Flux Unit
SHD	–	Sistema Heterogêneo de Dados
SIG	–	Sistema de Informação Geográfica
SPIDR	–	Space Physics Interactive Data Resource
ST	–	Série Temporal
TRACE	–	Transition Region and Coronal Explorer
UT	–	Universal Time
WDC	–	World Data Center
XML	–	Extensible Markup Language

LISTA DE SÍMBOLOS

t	–	tempo
A	–	matriz de amplitudes
$s(x, y, z)$	–	espaço euclidiano
N	–	extensão temporal
m	–	extensão espacial em x
n	–	extensão espacial em y
l	–	extensão espacial em z
nm	–	nanômetros
$\pounds$	–	grade numérica generalizada (GNL)
κ	–	grau variacional de uma GNL
λ_ℓ	–	coeficiente de extensão em ℓ de uma GNL
ℓ	–	domínio usual em que são tomadas medidas sobre uma GNL
μ_p	–	medida analítica de índice p tomada sobre uma GNL
P	–	quantidade de medidas analíticas tomadas sobre uma GNL
ν	–	frequência
$S(\nu)$	–	espectro de potências
α	–	expoente de escala para o método DFA
β	–	expoente de escala para o método PSD
ϕ	–	fases no modelo de geração de sinais do tipo ruído *Browniano*
ε	–	constante de acoplamento
ρ	–	parâmetro de controle caótico

SUMÁRIO

Pág.

1 INTRODUÇÃO

"Imagination is more important than knowledge."

(Albert Einstein)

A interpretação de dados é parte fundamental no desenvolvimento de uma pesquisa científica. Na ciência moderna dados no formato digital são gerados a partir de experimentos numéricos e/ou a partir de instrumentos de medida que, em geral, observam variáveis sistêmicas com alta sensibilidade e alta resolução nos domínios do tempo, espaço e frequência. Na maioria das vezes, as observações envolvem múltiplas fontes de dados que podem ser provenientes de um mesmo sistema (físico, biológico, etc.), resultando em um grande conjunto heterogêneo de dados científicos que aqui denominamos *sistema heterogêneo de dados* (SHD). Como consequência, o avanço da fronteira do conhecimento na ciência moderna depende da obtenção de uma grande variedade de SHDs obtidos com alta resolução, bem como do desenvolvimento de técnicas matemáticas e algoritmos avançados para análise dos mesmos. Por isso, a gestão da informação em grandes volumes de dados científicos, em geral incorporados a ambientes multimídia distribuídos, tem sido discutida como um problema de máxima importância para a pesquisa em computação aplicada à ciência de dados (de Carvalho et al., 2006).

A maneira usual como são coletados os dados contribui significativamente para a construção deste cenário, pois na maioria das vezes esta coleta é feita na forma de registros simples, ou dados brutos, que precisam ser posteriormente contextualizados, tornando-se assim fonte de conhecimento sobre os fenômenos estudados (SANTOS, 2008). De acordo com Graves et al. (2008), a dificuldade de encontrar um ambiente capaz de integrar conjuntos de dados e ferramentas de análise leva também a uma sobrecarga nas redes de comunicação, devido à necessidade de transferência da informação da base de dados até o local de análise.

A representação da informação na forma de sistemas de bancos de dados coerentes e organizados tem se tornado, portanto, essencial para a computação científica, especialmente quando se trata de dados compostos por séries temporais geradas por observação e ou simulação de sistemas heterogêneos que envolvem, além de multiplas variáveis, múltiplos processos de geração das mesmas. Quando as múltiplas variá-

veis que representam um sistema são geradas a partir de processos não lineares que podem ou não estar associados, seu estudo sistêmico torna-se uma tarefa complexa. Nesse caso, tais sistemas são chamados aqui de *sistemas complexos*. Em geral, a observação e/ou simulação de um sistema complexo resulta na geração de um SHD.

Dessa forma, os sistemas complexos, quando se referem a sistemas reais em física, química, biologia, economia, etc., caracterizam-se por fenômenos espaçotemporais coletivos, emergindo da ação combinada de um grande número de constituintes heterogêneos, cuja dinâmica pode ser observada e analisada em detalhes (CLADIS; PALFFY-MUHORAY, 1995; BAR-YAN, 1997, p. ex.). Isso indica a existência de observações ou simulações numéricas de estruturas e processos em todas as escalas temporais, espaciais e espectrais possíveis. Devido a esta abordagem multiescala e multidimensional, a presença de informações sobre não linearidades, correlações de longo alcance e transições de fase em uma representação de dados pós-analítica[1] torna-se indispensável.

Discutir dentro de uma abordagem inédita os aspectos conceituais relacionados à identificação e à representação de um sistema heterogêneo de dados constitui o principal escopo deste trabalho. Como objetivos específicos, podemos destacar a construção de um panorama visual sobre a complexidade de um sistema e de suas interações, a geração de um sistema heterogêneo de dados canônico para teste do modelo, a aplicação do modelo sobre um sistema real e a generalização do sistema computacional desenvolvido.

O Capítulo 2 descreve o conceito de sistema de dados considerando novas abordagens da tecnologia da informação. Neste cenário, introduziremos, no Capítulo 3, uma generalização inovadora para a representação de um conjunto de séries temporais pós-analisadas que está baseada em:

(i) suas variáveis constitutivas $A(t_i, s)$. Cada variável que constitui um determinado sistema é interpretada como sendo uma medida de intensidade ou amplitude que pode variar como uma função do tempo t e do espaço euclidiano $s(x,y,z)$;

[1] Dados pós-análiticos são denominados aqui como aqueles que foram analisados por uma dada técnica matemática ou estatística e que são representados por uma medida ou informação proveniente da aplicação desta técnica. Em geral, a medida representa alguma característica intrínseca do sistema, como, por exemplo, a sua natureza determinística, estocástica ou híbrida.

(ii) suas extensões discretas nos domínios do tempo e do espaço; e

(iii) medidas de segunda ordem provenientes da análise de cada $A(t_i, s)$.

O conceito de grade numérica generalizada (GNG) será introduzido com base em informações sobre uma sequência discreta de medidas no tempo e/ou no espaço para uma dada variável constitutiva[2]. O agrupamento dos parâmetros constituintes de uma medida dinâmica, quando escrito no formato de uma GNG, deverá fornecer uma representação alternativa da informação, integrando metadados estruturais e semânticos. Dessa forma, a integração de todas as grades numéricas obtidas a partir de um mesmo sistema de dados poderá ser utilizada para gerar um modelo de representação que permita capturar de forma visual a complexidade do sistema e de suas interações. A descrição detalhada do conceito de grade numérica generalizada será objeto de estudo do Capítulo 3.

Ferramentas de visualização em geral se baseiam na organização de dados segundo critérios específicos, e podem considerar aspectos cognitivos do sistema visual humano para tornar a análise de dados mais rápida e exploratória, possibilitando a recuperação de informações e a construção de novos conhecimentos sobre elas. Neste contexto, destacam-se algumas abordagens multidisciplinares emergentes, como a Arquitetura da Informação e a Analítica Visual. Reunindo contribuições de áreas como computação, psicologia, semiótica e *design* gráfico, a arquitetura da informação pode ser definida como a arte e a ciência de preparar a informação para ser usada mais eficiente e efetivamente por seres humanos (JACOBSON, 1999). A Analítica Visual, em particular, tem como objetivo principal a integração de ferramentas de análise e de visualização para permitir aos tomadores de decisão evidenciar suas capacidades cognitivas e perceptuais sobre o processo analítico e, ao mesmo tempo, aplicar habilidades computacionais avançadas para melhorar o processo de descoberta (KEIM et al., 2006). Os fundamentos destas duas abordagens, bem como trabalhos envolvendo a aplicação de técnicas de visualização a conjuntos de dados espaçotemporais, são descritos no Capítulo 4.

Motivado por estes fatores, o modelo de representação baseado em grades numéricas generalizadas e sua utilização como uma ferramenta de visualização e análise

[2]Uma variável constitutiva corresponde a qualquer variável fundamental gerada por um processo que é constituinte dinâmico do sistema, como, por exemplo, um gradiente de temperatura que causa alteração no valor da velocidade de deslocamento de um fluido. Nesse exemplo tanto a temperatura como a velocidade são variáveis constitutivas.

exploratória de todo e qualquer sistema heterogêneo de dados baseado em medidas espaçotemporais é apresentado no Capítulo 5.

O desenvolvimento deste modelo envolveu o estudo e a avaliação de um conjunto de ferramentas de análise diversas, bem como sua implementação computacional, através de estudos de caso envolvendo dados sintéticos e reais. Aspectos particulares de cada método de análise, como, por exemplo, sua robustez em relação à quantidade de medidas disponíveis, tornam sua aplicabilidade dependente das características dos dados analisados. Por esta razão, o sistema computacional desenvolvido neste trabalho baseia-se em um conjunto aberto de parâmetros pós-analíticos, possibilitando ao analista decidir quais métodos devem ser incorporados ao seu esquema de representação.

Os métodos de análise explorados e implementados neste trabalho são apresentados no Capítulo 6, que traz também os resultados da aplicação do modelo sobre um estudo de caso em dados sintéticos, acompanhado de uma descrição de uma nova abordagem para geração de dados espaçotemporais sintéticos, e em dados reais de Física Espacial. Aplicações em outros domínios, como, por exemplo, dados limnológicos e meteorológicos, são discutidas e avaliadas no contexto do desempenho computacional das ferramentas implementadas. As considerações finais e perspectivas futuras são apresentadas no Capítulo 7.

2 SISTEMAS DE DADOS COMPLEXOS

"Quando considero (...) o pequeno espaço que ocupo e que vejo,
abismado na imensidade dos espaços que ignoro e que me ignoram,
assusto-me e pasmo de me ver aqui e não além, porque não há razão para
ser aqui e não além, para ser agora e não depois."

(Blaise Pascal)

Em cada área da ciência existem diversas formas de coletar dados em sistemas naturais a partir dos quais sejam extraídas informações estruturais e realizados diferentes tipos de análise. Por essa razão, pesquisadores geralmente acumulam grandes conjuntos de dados e resultados de forma independente, ocupando extenso espaço de armazenamento, que aumenta com a precisão das medidas proporcionada pelos avanços tecnológicos.

Uma tendência da ciência moderna consiste na busca por ferramentas que capacitem a integração de dados e resultados com o objetivo de obter novas soluções e *insights* sobre os problemas da atualidade, em especial aqueles intrinsicamente interdisciplinares, como os estudos sobre clima espacial e meio ambiente. Nesse sentido, encontrar novos modelos de representação aplicáveis ao maior número de áreas possível, uniformizando os sistemas de dados provenientes de processos de formação de padrões espaçotemporais, tem se apresentado como um desafio para a ciência da informação. Usualmente tais dados, embora possam representar fenômenos de diferentes áreas e processos, estão associados a medidas de variáveis que são funções do espaço e do tempo. Com base nestas variáveis em comum, é introduzida neste capítulo a definição de dados espaçotemporais no contexto desta tese, bem como o conceito de representação e reconhecimento de padrões espaçotemporais.

2.1 Dados de Séries Temporais

Dados de séries temporais são aqueles gerados de forma sequencial no tempo[1]. Uma série temporal é, portanto, um conjunto de observações $A(t_i)$, cada uma obtida em um instante de tempo t_i. Em uma série temporal discreta, as medidas são tomadas

[1]Do ponto de vista de organização de dados, múltiplas dimensões no tempo são possíveis se um sistema acomoda múltiplas linhas de tempo para um evento, tais como tempo válido, tempo de transição ou tempo de decisão (RODDICK; SPILIOPOULOU, 1999)

em instantes de tempo t_i regular ou irregularmente[2] espaçados (MORETTIN; TOLOI, 1987).

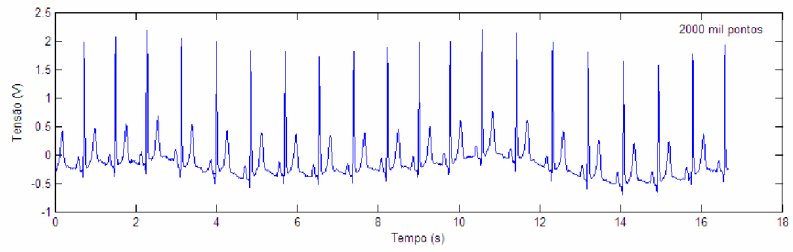
O registro do eletrocardiograma (ECG) de um paciente e a posição de um projétil em movimento são exemplos de variáveis cujas medidas podem gerar séries temporais contínuas, enquanto a média de manchas solares observadas anualmente ou o valor anual do índice de desenvolvimento humano (IDH) de um país se apresentam na forma de séries temporais discretas. Tais exemplos, como mostra a Figura 2.1, ilustram a ampla aplicabilidade dos conceitos e ferramentas de análise de séries temporais às mais diversas áreas da ciência, como física, química, biologia, meio ambiente, medicina, sociologia e economia.

A quantidade de informação presente em uma série temporal não é necessariamente proporcional ao número de pontos que a compõem. Duplicar o número de observações, reduzindo pela metade o intervalo entre os valores pode não fornecer conhecimento adicional, especialmente se, por exemplo, as observações sucessivas da série forem altamente correlacionadas (KENDALL; ORD, 1990), o que ilustra a influência da estrutura interna da série sobre o cenário de análise do sistema em observação, atuando de maneira complementar aos dados descritivos ou metadados.

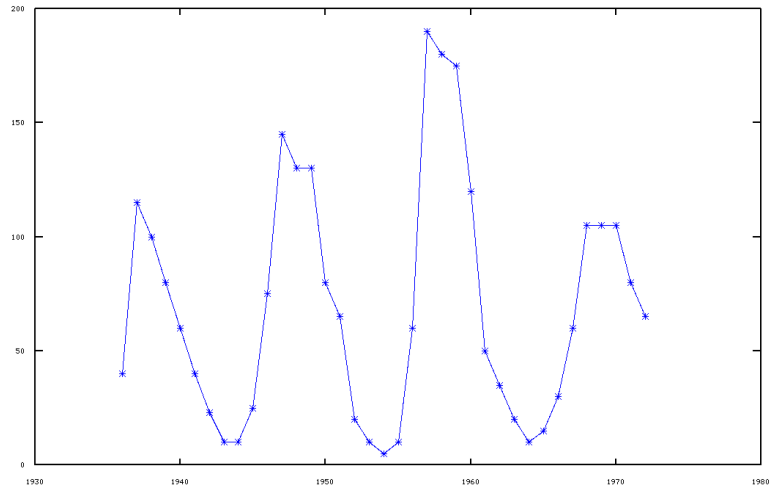
Cada unidade que compõe um conjunto de dados relacionais é chamada de registro, que é descrito através de uma coleção de atributos associados a ele. Computacionalmente, um conjunto de dados pode ser representado na forma de um arquivo no qual as linhas identificam os diferentes objetos (registros) e cada coluna corresponde a um atributo. A representação de uma coleção de séries temporais na forma de um conjunto de dados pode ser realizada através de diferentes estratégias, de acordo com o objetivo da aplicação. Na forma de dados temporais ou sequenciais, cada instante de tempo pode ser associado a um registro. Neste caso, os atributos armazenam valores associados àquele tempo. Esta estratégia é usualmente utilizada em aplicações de previsão de tempo (WITTEN; FRANK, 2005). Alternativamente, cada atributo pode representar um instante de tempo, e observações sucessivas de cada série temporal são associadas a atributos sucessivos no conjunto de dados.

Os dados de séries temporais podem, ainda, ser vistos como um tipo especial de dados sequenciais, em que cada registro corresponde a uma série temporal (TAN

[2]Nos casos em que as observações aparecem de forma irregular, por ausência de determinados valores ou espaçamento desigual, interpolação, reamostragem ou abordagens alternativas devem ser consideradas (BROCKWELL; DAVIS, 1991; KENDALL; ORD, 1990).



(a)



(b)

Figura 2.1 - (a) Série temporal contínua de um eletrocardiograma real discretizado por registro digital das medidas. (b) Série temporal discreta interpolada do número relativo de manchas solares entre 1936 e 1972.
Fonte: (a) Guerra (2009); (b) Andrews e Herzberg (1985).

et al., 2009). No presente trabalho, este conceito será generalizado para abranger também dados espaçotemporais, ou seja, dados gerados por observações medidas no tempo e no espaço tridimensional. Esta abordagem será descrita na seção 2.3.

Os objetivos específicos da análise de séries temporais podem variar de acordo com diversos fatores, como a natureza do sistema observado e a área de aplicação. Em termos gerais, porém, os tipos mais comuns de investigação estão relacionados à caracterização e à modelagem da série, com o objetivo de descrevê-la matematicamente, realizar previsões[3] do seu comportamento futuro, estimativas sobre pontos relativos a instantes de tempo passados ou desenvolver um sistema de controle ba-

[3]usando técnicas de *forecasting*, *backcasting* ou *hindcasting*, etc.

seado no modelo gerado. A escolha do método de análise a ser utilizado também depende da complexidade da série. Métodos lineares aplicam-se a sistemas cuja dinâmica intrínseca é governada pelo paradigma linear, ou seja, de que pequenas causas levam a pequenos efeitos, e todo comportamento irregular do sistema se deve a alguma entrada randômica externa ao sistema (KANTZ; SCHREIBER, 2003). Uma vez que na natureza a maior parte dos sistemas apresentam comportamento mais complexo, torna-se necessário considerar a possibilidade de que a série temporal tenha sido gerada por um processo dinâmico não linear. Neste caso, devem ser utilizados métodos de análise não linear. A abordagem de análise adotada neste trabalho, que leva em consideração todas as componentes de variabilidade que podem descrever uma série temporal, será detalhada no Capítulo 6.

2.2 Dados Espaciais

Conceitualmente, o termo "dados espaciais" pode ser controverso se tentarmos generalizá-lo para todos os domínios da ciência. A própria definição de espaço, por exemplo, ocupa uma das posições de maior diversidade de significados linguísticos. Assim, enquanto para a Astronomia o espaço corresponde à região que abrange a atmosfera terrestre e o espaço cósmico (HOUAISS; VILLAR, 2001), em sistemas de informação geográfica (SIG) o termo "espacial" – ou "geoespacial" – refere-se à localização geográfica dos eventos sobre a superfície da Terra (CASTLE; CROOKS, 2006). Esta distinção conceitual é refletida nas formas de representação, visualização e análise dos dados espaciais relativos às diferentes áreas de aplicação, e diversos trabalhos vêm sendo desenvolvidos para atender às demandas de cada uma delas, como pode ser visto, por exemplo, em Verhein (2009), Wehrend e Lewis (1990), Graves et al. (2008), Roddick e Spiliopoulou (1999).

Há, no entanto, um interesse cada vez maior na identificação de características e interesses comuns a diferentes setores da ciência para o desenvolvimento de métodos e ferramentas de maior generalidade. Um tipo de informação espacial presente em quase todos os sistemas da natureza se refere aos processos de formação de padrões que podem se estender em direções mutuamente ortogonais (MECKE; STOYAN, 2000). Padrões espaciais são representados visualmente na forma de imagens. Portanto, neste contexto, qualquer imagem isolada representa um padrão espacial.

Nesta tese, o termo "espacial" refere-se ao espaço euclidiano composto por três dimensões espaciais, usualmente representadas pelo sistema cartesiano de coordenadas

x, y, z. Portanto, referimo-nos à definição mais geral de "espaço" utilizada pela Física, que está incorporada ao domínio físico espaçotemporal[4].

2.3 Dados Espaçotemporais

Padrões espaciais medidos sequencialmente, ou seja, à medida que o tempo evolui, geram conjuntos de dados denominados espaçotemporais, integrando os conceitos de tempo e espaço definidos nas seções anteriores. Um exemplo de série espaçotemporal em física é aquela gerada a partir do imageamento estroboscópico de uma camada granular oscilatória (UMBANHOWAR et al., 1996). Milhares de pequenas esferas (que possuem diâmetros variando de 0.5 a 1 mm, por exemplo) são dispostas de forma aleatória sobre uma plataforma formando uma camada granular. Quando à plataforma é imposto um movimento vibratório com uma determinada frequência, o sistema, formado pelas pequenas esferas interagindo por forças de contato, tende a convergir para um padrão regular. Este processo, na física, é conhecido como relaxação espaçotemporal (ROSA; RAMOS, 2003), representado parcialmente pela sequência de três imagens mostradas na Figura 2.2. No último quadro (imagem mais à direita), é destacado um subdomínio da grade, visualizado em três dimensões. Note que, apesar da natureza tridimensional (3D) das estruturas, o dado gerado para análise apresenta apenas duas dimensões espaciais. No contexto do arquivo de dados, a altura da camada granular (uma medida de amplitude) está distribuída em uma estrutura matricial $m \times n$. Portanto, a terceira dimensão é a variável $A(t_i, m, n)$. Nesse caso, cada medida ou elemento de amplitude que compõe a matriz que gera a imagem é denominado píxel.

Um conjunto ordenado de píxeis onde em cada um há uma medida de amplitude é comumente chamado de matriz ou imagem. Entretanto, como no exemplo da Figura 2.3, existem casos onde a amplitude é medida em x, y e z. Nesse caso, cada medida ou elemento de amplitude que compõe o arquivo que gera a imagem 3D é denominado vóxel. Um conjunto ordenado de vóxeis onde em cada um há uma medida de amplitude é comumente chamado de hipercubo. Cada quadro de uma sequência temporal de matrizes (2D) ou hipercubos (3D) será denominado aqui por *snapshot*. De forma geral, todos os casos espaçotemporais (2D e 3D) podem ser denotados pelo termo simplificado "série temporal", identificando a ordem da dimensão espacial quando

[4] O domínio físico espacial admite representar uma variável em outros sistemas de coordenadas: polares, cilíndricas e esféricas, as quais não serão consideradas explicitamente na conceituação de espaço deste trabalho.

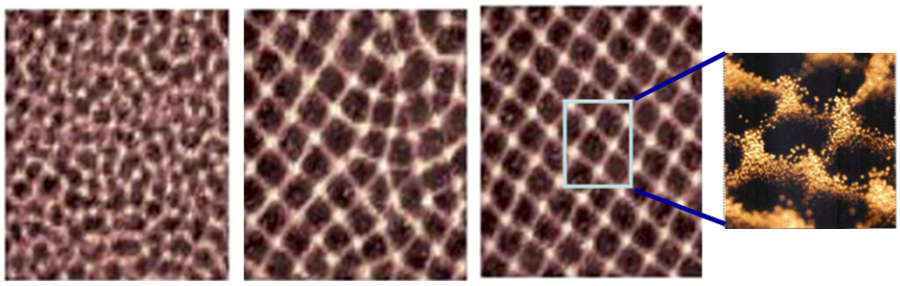


Figura 2.2 - Exemplo de formação de padrões espaçotemporais: sequência de três imagens mostrando o processo relaxação espaçotemporal em uma camada granular.
Fonte: Rosa et al. (2003)

ela estiver presente.

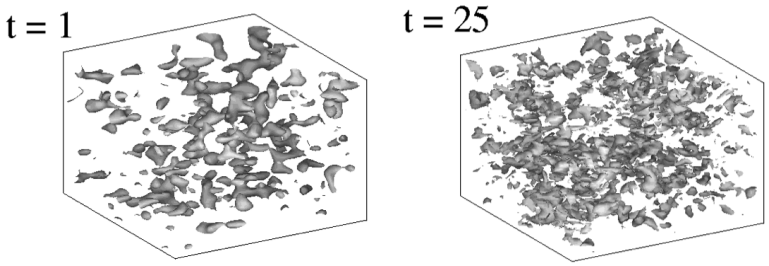


Figura 2.3 - Sequência de duas imagens (hipercubos) mostrando o processo de fragmentação de estruturas em um fluido com vorticidade turbulenta. A simulação foi realizada através do código ZEUS, considerando uma caixa $128 \times 128 \times 128$ e 25 *snapshots*.
Fonte: Rosa et al. (2011)

Na Figura 2.4 são apresentados três exemplos de séries temporais obtidos através de diferentes processos de medidas nas áreas: nanotecnologia, oceanografia e física solar. Em cada sequência horizontal composta por três imagens (superior, intermediária e inferior) são mostrado três instantâneos (*snapshots*) representando a evolução de um processo físico cuja medição resulta em uma sequência de imagens digitais. A sequência superior mostra a formação de porosidade em uma amostra de silício com dimensões da ordem de 100 $nm \times 100$ nm, observada através de um microscópio de força atômica. As imagens foram obtidas a cada 2 segundos e digitalizadas como matrizes de tamanho 64×64 (SILVA et al., 2000). Na sequência intermediária, é

visualizada, por fotogrametria aérea da NOAA, a evolução da mancha de óleo que vazou do navio Exxon Valdez, em 1989, sobre a península do Alaska (NOAA, 1992). A sequência inferior mostra a evolução da configuração do campo magnético durante a explosão solar observada em 6 de junho de 2000 através do telescópio a bordo do satélite TRACE (ROSA et al., 2008).

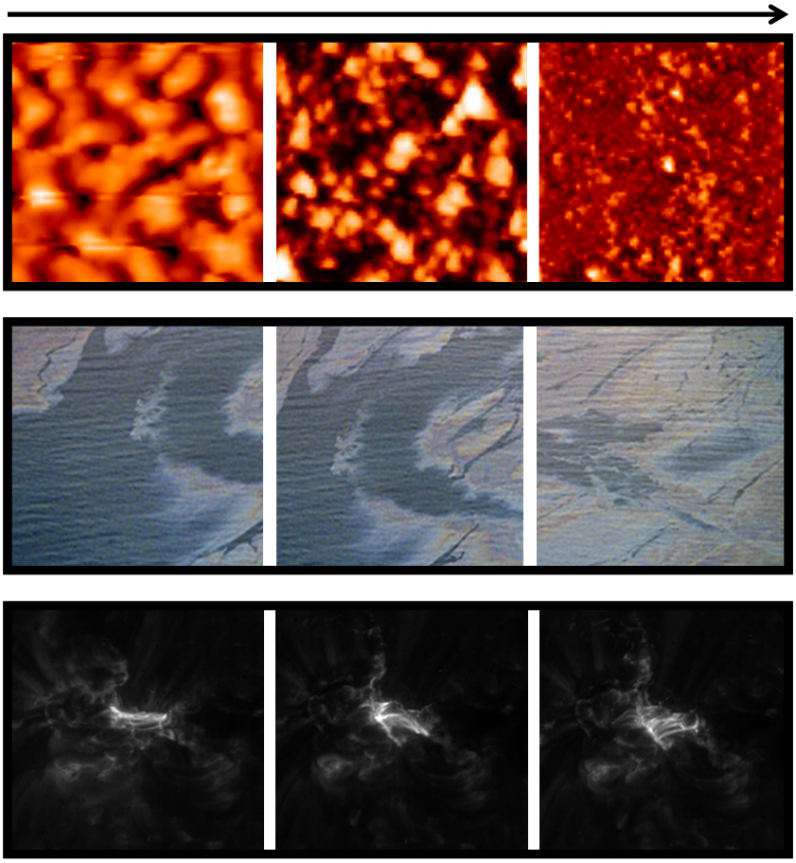


Figura 2.4 - Exemplos de séries temporais observadas na forma de imagens digitais. A seta indica a evolução temporal da esquerda para a direita. A sequência superior está relacionada com a observação da formação de porosidade em uma amostra de silício. A sequência intemediária mostra a evolução do grande vazamento de petróleo do navio Exxon Valdez em 1989. Na sequência inferior é apresentada a evolução de uma região solar ativa observada na faixa de ultravioleta (171A). Fonte: Adaptado de Silva et al. (2000), NOAA (1992) e Rosa et al. (2008).

Na tentativa de entender, modelar e analisar os processos naturais de formação de padrões espaçotemporais, métodos computacionais são incorporados a diferentes campos de aplicação, estabelecendo novas áreas interdisciplinares. A mineração de

dados, que consiste na aplicação de técnicas estatísticas e matemáticas sobre grandes conjuntos de dados a fim de descobrir correlações, tendências e padrões significativos, apresenta importância cada vez maior neste contexto. O processamento digital de imagens, por exemplo, tornou-se uma ferramenta chave na análise de dados espaço-temporais de todas as áreas da ciência natural e no estudo de fenômenos científicos complexos (JÄHNE, 1993).

O surgimento de novas aplicações nas ciências ambientais e da saúde, como o monitoramento dos níveis regionais de ozônio e o mapeamenteo de doenças, vem aumentando nos últimos anos a popularidade dos modelos espaçotemporais. Outra razão para este aumento está no recente avanço na eficiência dos recursos computacionais, devido ao grande volume frequentemente ocupado pelos conjuntos de dados espaçotemporais (STROUD et al., 2001). Novas informações são incorporadas aos bancos de dados em astrofísica, por exemplo, muitas vezes antes que os astrônomos tenham tempo hábil para analisar as informações disponibilizadas previamente. Observa-se a mesma tendência em outras áreas da ciência espacial[5], como o estudo das relações solar-terrestres, associado à recente área denominada Clima Espacial[6], que investiga e monitora a interferência da atividade solar sobre o nosso planeta.

A seguir é apresentado um conjunto de dados relacionados ao clima espacial, utilizado como estudo de caso no contexto desta tese.

2.4 Um Exemplo de Sistema Heterogêneo de Séries Temporais

A ciência espacial em geral requer análises sistêmicas a fim de obter informações para previsão do clima espacial. Esta análise é necessária para compreender os processos físicos estudados através da identificação das relações entre diferentes parâmetros, ou para usar os dados disponíveis na construção de um modelo dos processos solares geoefetivos. Neste contexto, a atividade solar é uma das principais causas da variabilidade do clima espacial, responsável pelos fenômenos observados no meio interplanetário, no campo geomagnético e na ionosfera (Figura 2.5).

Nesta seção é apresentado um conjunto heterogêneo de dados provenientes de três fontes: (i) *Space Physics Interactive Data Resource* (SPIDR, 2011), (ii) *The NASA International Solar-Terrestrial Physics* (ISTP, 2011) e (iii) *The Ondrejov Solar Radio*

[5]Neste caso, "espaço" se refere ao espaço sideral, no contexto da Astronomia.

[6]Programa operacional e de pesquisa do INPE, iniciado em 2008 (EMBRACE – Estudo e Monitoramento Brasileiro do Clima Espacial – `http://www.inpe.br/climaespacial/`).

Data Archive (JIRICKA et al., 1993).

Tabela 2.1 - Dados selecionados para o evento SAJ6. N, m, n e l indicam respectivamente, as extensões temporal em t e espaciais em x, y e z.

	Dados	Instrumento (Fonte)	N	m	n	l
ST2	Explosão em Rádio (3GHz)	Ondrejov (Sol)	2988	-	-	-
ST3	Fluxo de raio-X	GOES (Sol)	5760	-	-	-
ST4	Densidade iônica	ACE (IMF)	5600	-	-	-
SET1	Dst	Kyoto (MAG)	8736	-	-	-
SET2	Imagem UV	TRACE (Sol)	3	256	256	-
SET3	Imagem WL	TRACE (Sol)	9	512	512	-
SET4	Mapa	Simulação	1000	64	64	64

Os metadados selecionados para ilustrar esta aplicação são exibidos na Tabela 2.1, todos relacionados à atividade de clima espacial observada no período de 5 a 8 de junho de 2000. Nesta tabela, as extensões temporal e espacial indicam, respectivamente, as quantidades de medidas tomadas no tempo e no espaço que compõem cada série temporal.

2.4.1 A atividade solar (evento SAJ6)

Em 6 de junho de 2000, a ejeção de massa coronal foi acompanhada por dois dos mais intensos *flares* solares já observados. A Figura 2.5 mostra o magnetograma do disco solar completo observado pelo Observatório Nacional Kitt Peak (KPNO) às 17:42 UT e a imagem de raio-X correspondente observada pelo satélite Yohkoh às 18:55 UT. A região ativa (no visível) e o *flare* (em raio-X) são identificados em cada imagem.

A Figura 2.6 mostra a série temporal selecionada para: (i) a explosão solar observada pelo observatório Ondrejov desde 16:35 UT; (ii) o fluxo de raio-X observado pelo satélite NOAA GOES-8[7] desde 00:00 UT de 5 de junho; (iii) a densidade iônica interplanetária observada pelo satélite ACE desde 00:00 UT de 5 de junho; e (iv) o índice Dst observado pelo Centro de Dados Kyoto em 4 e 5 de junho. Observe que de 6 a 8 de junho há assinaturas especiais do evento SAJ6 observadas desde o sol até a magnetosfera.

[7] *Geoestationary Operational Satellites* da NASA.

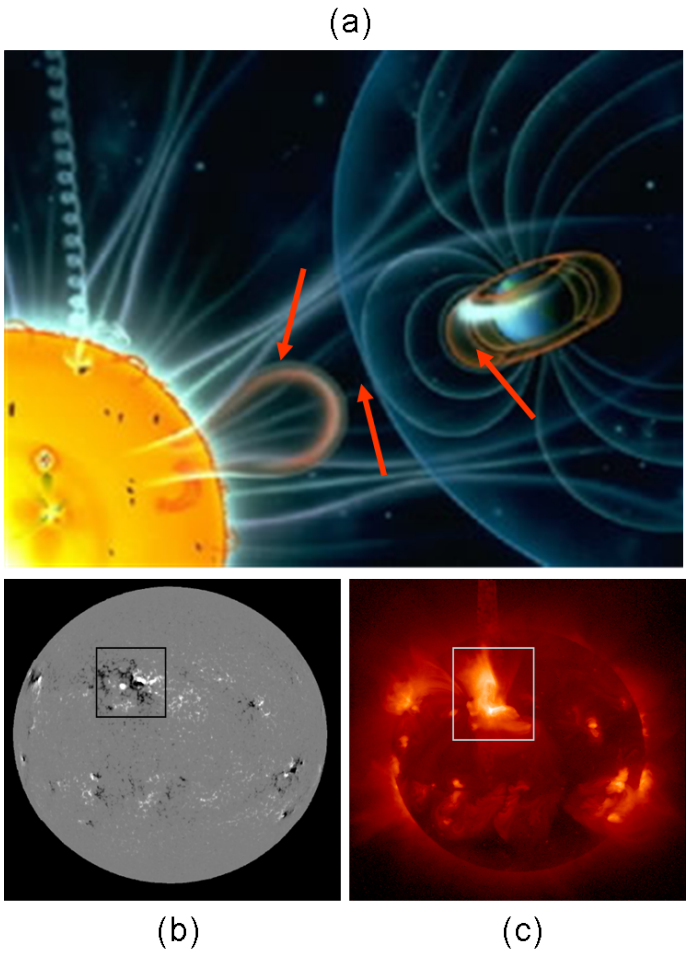


Figura 2.5 - (a) O cenário físico solar-terrestre: as setas indicam, da esquerda para a direita, três regiões nas quais os dados exibidos pela Tabela 2.1 são observados – coroa solar, campo magnético interplanetário e magnetosfera. (b) Uma imagem completa do sol observado pelo TRACE em 6 de junho de 2000. (c) Uma imagem completa do sol observado pelo YOHKOH em 6 de junho de 2000.

A Figura 2.7 (a) mostra três *snapshots* selecionados, observados pelo TRACE[8], telescópio espacial da NASA, desde 15:08 UT, representando a evolução da estrutura do *loop* magnético mútuo complexo do qual o *flare* solar é desenvolvido. A Figura 2.7 (b) mostra uma composição de nove imagens de luz branca obtidas pelo TRACE do complexo *spot* central na região ativa 1926 que produziu três *flares* de classe X em 6 e 7 de junho de 2000. A Figura 2.7 (c) mostra um mapa da radiação 3D simulado da coroa solar no qual a coloração verde mostra a radiação em rádio de um choque guiado por CME[9], enquanto as colorações amarela e vermelha revelam

[8] *Transitional Region and Coronal Explorer*
[9] *Coronal Mass Ejection*

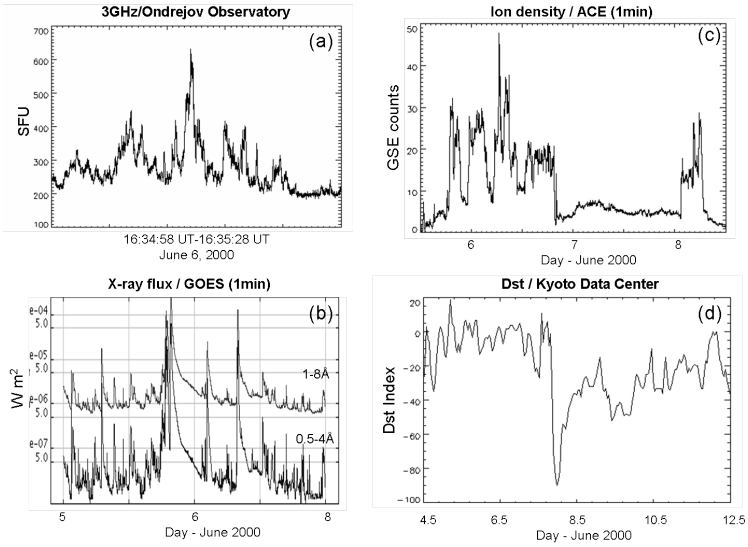


Figura 2.6 - Conjunto de séries temporais selecionado para SAJ6: (a) explosão solar Ondrejov observada em 3GHz; (b) fluxo de raio-X solar GOES; (c) densidade iônica do meio interplanetário ACE; (d) índice Dst geomagnético do centro de dados Kyoto.

a estrutura denominada *stream* próxima ao sol (COHEN; HOLLIS, 2011).

2.5 Generalização das Séries Temporais de um SHD

Como vimos no exemplo anterior, um SHD compreende séries temporais de medidas que podem ou não ser tomadas também nas diferentes coordenadas do domínio espacial $s(x, y, z)$. Portanto, o primeiro objetivo no desenvolvimento de uma representação robusta de um conjunto de medidas no tempo é discutir uma possível generalização destas medidas em relação aos domínios espaciais.

Note que qualquer série temporal, tanto no caso de uma medida de amplitude apenas no domínio do tempo, $A(t)$, como nos domínios espacial e temporal, $A(t, x)$, $A(t, x, y)$ e $A(t, x, y, z)$, pode ser interpretada, na sua forma digitalizada (discreta), como um conjunto de números (reais ou inteiros) ordenados no domínio do tempo, e, quando for o caso, também ordenados nos domínios x, y e z do espaço euclidiano. A ordenação dos números que representam medidas distribuídas regularmente no tempo e no espaço pode ser interpretada como sendo uma "grade" de medidas. Considere, por exemplo, a série temporal mostrada na Figura 2.8 composta por 5 medidas de temperatura tomadas apenas no tempo a cada intervalo de duas horas. A mesma série pode ser vista na forma de uma grade abstrata a partir de duas

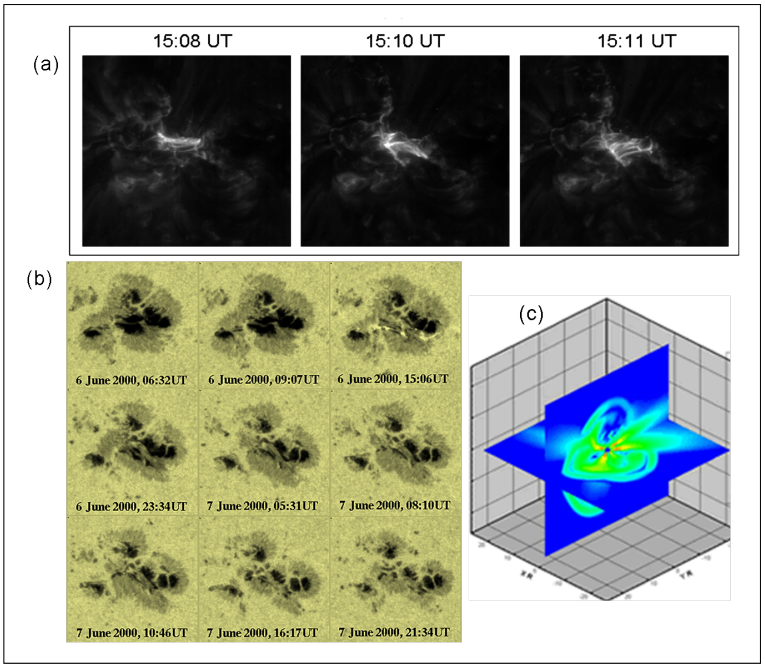


Figura 2.7 - Dados espaçotemporais selecionados para SAJ6: (a) sequência de imagens locais observadas pelo TRACE em 171$\dot{A}$; (b) Sequência de imagens de luz branca locais tomadas pelo TRACE; (c) Exemplo de um possível mapa de radiação simulado da coroa solar por Schmidt and Gopalswamy
Fonte: (c)Cohen e Hollis (2011).

representações, como mostram as Figuras 2.8 (b) e 2.8 (c). Entretanto, o conceito de grade aqui adotado considera cada valor numérico de uma dada amplitude medida $A(t)$ como sendo um elemento de uma grade espacial tridimensional. Quando a medida de $A(t)$ está sendo tomada em um único ponto do domínio $s(x, y, z)$, os elementos de grade estão ordenados apenas no tempo como no exemplo da Figura 2.8. O caso mais geral, como ilustrado na Figura 2.9, é tomar a medida de amplitude em cada instante t em vários elementos da grade $s(x, y, z)$. Na Figura 2.9 (b) é mostrado um exemplo de medida tomada apenas sobre nove elementos da grade definida pelo domínio $s(x, y)$.

Com base no que foi discutido acima, no próximo capítulo será introduzido o conceito de grade numérica, generalização que provê uma representação mais robusta para um determinado SHD. À luz deste novo conceito veremos que cada uma das séries espaçotemporais apresentadas anteriormente pode ser interpretada como uma grade numérica. Uma das novidades nesta nova abordagem para generalização de

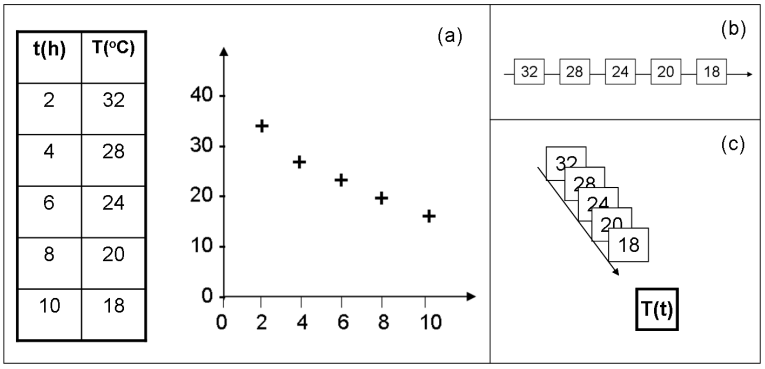


Figura 2.8 - Exemplo de uma série temporal composta por cinco valores de temperatura medidos a cada intervalo de duas horas a partir do meio-dia. (b) Uma possível representação alternativa para o conjunto de medidas, mostrando apenas os valores numéricos ordenados no tempo. (c) A mesma representação destacando que cada medida $T(t)$ equivale a um elemento da "grade numérica" temporal.

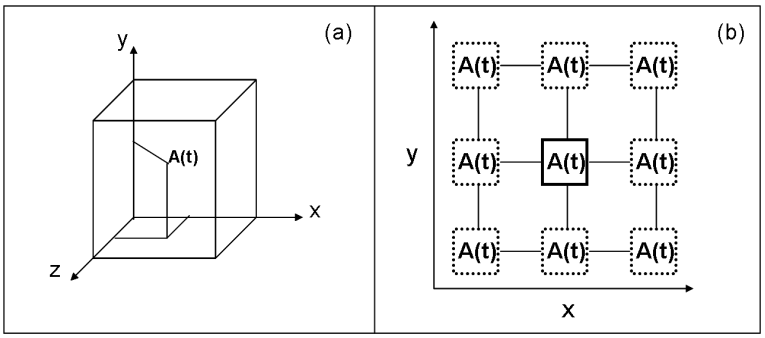


Figura 2.9 - (a) Qualquer medida de amplitude $A(t)$ pode também ser tomada em elementos de uma grade 3D distribuídos ao longo das coordenadas espaciais (x, y, z). (b) Um exemplo de medidas tomadas numa grade com nove elementos distribuídos apenas em duas dimensões espaciais.

uma série temporal é a inclusão de informação que provenha da sua análise (estatística, por exemplo) prévia. Note que, caso todos os dados descritos na Tabela 2.1 já estivessem previamente analisados por técnicas matemáticas de interesse (o espectro de potências, por exemplo), esta medida analítica poderia constar na visualização. A inclusão de uma medida analítica poderia auxiliar a interpretação comparativa entre todas as séries em busca de uma informação integrada que pudesse, por exemplo, auxiliar no monitoramento das séries para previsão dos fenômenos físicos associados à ocorrência da explosão solar.

3 GRADE NUMÉRICA GENERALIZADA

"Tanto o rigor como a liberdade têm suas vantagens. Por que não mesclar os dois?"

(Jan V. White)

Conceitualmente, uma grade pode ser definida como um traçado em quadrículas usado para apresentar visualmente um conjunto de informações (LAROUSSE DO BRASIL, 2007), que, reunidas em um pequeno espaço, podem facilitar a compreensão do processo representado (WHITE, 2006).

No contexto deste trabalho, uma grade numérica corresponde a qualquer distribuição discreta de quantidades numéricas estruturadas em um espaço cartesiano delimitado por dimensões espaciais lineares. A partir desta definição, é estabelecido um novo formalismo para uma generalização matemática de conjuntos de dados obtidos por múltiplas medidas sobre um único sistema, com base no conceito de grade numérica generalizada (VERONESE et al., 2009).

Na Figura 3.1 são apresentados os exemplos mais triviais de grades numéricas representativas de medidas que podem variar no espaço euclidiano. Cada ponto de uma grade regular é identificado de acordo com suas dimensões espaciais (x, y, z) e associado a uma medida representada por um valor numérico (Figura 3.1 (a)). Quando os valores associados a estes pontos podem variar em função do tempo, a dinâmica espaçotemporal da grade assume uma representação matemática generalizada, como é ilustrado em (b).

Formalmente, uma GNG é definida como uma representação estrutural de dados na forma de uma grade ou de um elemento de grade[1], $\pounds$, na qual uma dada série espaçotemporal é representada por três tipos de coeficientes, os dois primeiros representando metadados associados a informações estruturais (domínios de função e extensão, respectivamente), e o terceiro tipo relacionado a propriedades pós-analíticas, como momentos estatísticos, índice de espectro de potências, quantidades morfométricas, etc. Assim, a estrutura da grade $\pounds$ pode ser escrita como uma equação do tipo:

[1] O símbolo $\pounds$ foi escolhido em alusão à letra "L", da qual deriva e que corresponde à inicial da palavra *lattice*, termo em Inglês para o substantivo "grade".

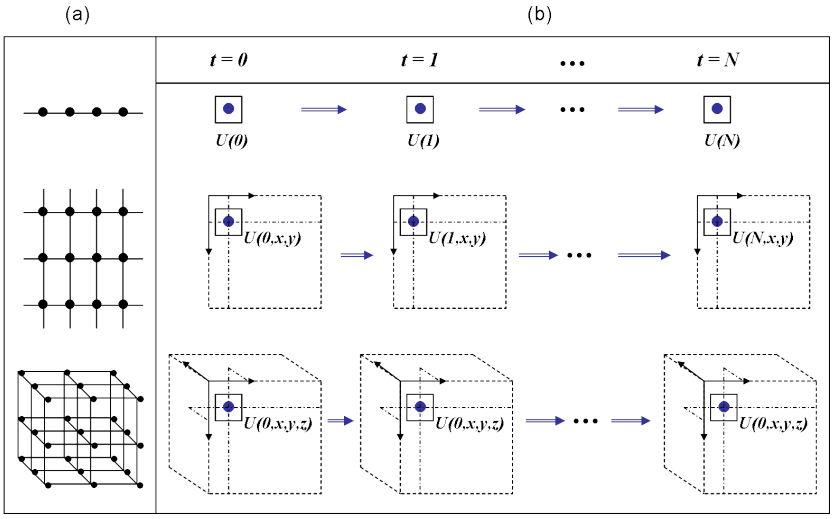


Figura 3.1 - (a) Exemplos de grades regulares que são interpretadas como grades numéricas quando há em cada ponto da grade um valor numérico associado. (b) Exemplos de grades numéricas dinâmicas que admitem uma generalização.

$$\pounds = f^* \left(\kappa, \lambda_\ell, \mu_p \right) \quad , \tag{3.1}$$

onde κ representa o grau variacional, que é a quantidade de variáveis de estado dos domínios fundamentais – tempo e espaço euclidiano tridimensional – e dimensões extras ou novas variáveis acopladas; λ_ℓ indicam os coeficientes de extensão, dados pela quantidade de medidas discretas em cada domínio usual ℓ; e μ_p é o conjunto de propriedades pós-analíticas caracterizando o comportamento estático e/ou dinâmico da variável constitutiva A. A notação f^* assume aqui o significado de uma aplicação discreta que respeita uma determinada regra de organização e representação de dados.

3.1 Grau variacional

O grau variacional κ depende de quantos dos possíveis tipos de variáveis constitutivas do sistema real estão disponíveis no sistema de dados. No contexto deste trabalho, em uma GNG todas as medidas são consideradas variantes pelo menos no tempo. Em um outro escopo, poderíamos adotar a frequência como domínio fundamental[2]. Como mostra a Tabela 3.1, adota-se, para $\kappa = 1$, uma variável constitutiva, $A(t)$.

[2]Esta abordagem alternativa será descrita no Capítulo 6 para uso em bancos de dados de astronomia.

Então podemos ter, para uma sequência de dados discreta, um observável temporal $A(t_i)$, correspondendo a $\kappa = 2$. Se tivermos informação espaçotemporal medida em uma, duas ou três dimensões, respectivamente, o grau variacional aumenta para 3, 4 ou 5. Para uma série espaçotemporal (p. ex., uma sequência de N imagens, cada uma delas composta por $m \times n$ píxeis) temos $\kappa = 4$. Quando os dados representam um hipercubo dinâmico composto por $l \times m \times n$ vóxeis, temos uma $\pounds$ com $\kappa = 5$. Observe que grades numéricas generalizadas compostas por dimensões extra e novas variáveis acopladas (funcionais) correspondem a $\kappa \geq 6$.

Tabela 3.1 - As variáveis constitutivas como uma função do grau variacional κ.

κ	Variáveis Constitutivas
1	$A_0 = t$
2	$A_0 = t; A_1 = f(t)$
3	$A_0 = t; A_1 = x; A_2 = f(x, t)$
4	$A_0 = t; A_1 = x; A_2 = y; A_3 = f(x, y, t)$
5	$A_0 = t; A_1 = x; A_2 = y; A_3 = z; A_4 = f(x, y, z, t)$
6	$A_0 = t; A_1 = x; A_2 = y; A_3 = z; A_4 = g(A_i, i \leq 3); A_5 = f(x, y, z, A_4, t)$
$\vdots$	$\vdots$

GNG usuais para dados experimentais correspondem a séries temporais convencionais ($\pounds^2$) e séries espaçotemporais, como sequências de imagens ($\pounds^4$). As grades $\pounds^3$ e $\pounds^5$ usualmente são observáveis a partir de simulações numéricas. Em geral, $\pounds^{\kappa \geq 6}$ compostas por dimensões extra e novas variáveis acopladas (funcionais) podem ser importantes, por exemplo, em genômica, sistemas de informações geográficas e cosmologia experimental. Como já dito anteriormente, também é possível definir os funcionais como variáveis espectrais quando o domínio da frequência está explicitamente envolvido. O estudo de caso descrito no Capítulo 6 é baseado em exemplos de GNG usuais relacionadas ao acoplamento magnético solar-terrestre, nos quais apenas $\pounds^2$ e $\pounds^4$ são consideradas.

3.2 Coeficientes de extensão

Os coeficientes de extensão λ_ℓ se referem ao comprimento do conjunto de dados em cada domínio. Então, λ_1 se refere ao número de pontos N que compõem o vetor $A(t)$ (sua extensão no domínio do tempo); λ_2 é a extensão dos dados no domínio do espaço euclidiano x; λ_3 é o tamanho da próxima dimensão discreta y, e assim

por diante. Então, uma $\pounds^{2,\lambda_1}$ representa uma série temporal $A(t)$ composta por λ_1 pontos, enquanto uma $\pounds^{4,\lambda_1,\lambda_2,\lambda_3}$ representa uma série espaçotemporal composta por λ_1 imagens de tamanho $\lambda_2 \times \lambda_3$ com uma medida de intensidade $A(t,x,y)$ em cada píxel correspondente (x,y). Como exemplos, uma $\pounds^{2,10^4}$ representa uma série temporal $A(t)$ composta por 10.000 pontos, enquanto uma $\pounds^{4,10^2,64,64}$ representa uma sequência dinâmica de 100 imagens de tamanho 64×64. Uma dada GNG $\pounds^{5,10^2,64,64,64}$ representa uma sequência dinâmica de 100 hipercubos de tamanho $64 \times 64 \times 64$ com uma medida de intensidade $A(t,x,y,z)$ em cada voxel correspondente (x,y,z). Note que, para um dado grau variacional κ, temos κ variáveis, sendo A_0 correspondente à variável fundamental, que, para séries dinâmicas, é o tempo. Portanto, utilizando esta generalização, vemos que qualquer série de medidas no tempo e no espaço pode ser reduzida à representação de uma grade numérica $\pounds^{\kappa,\lambda_1,\lambda_2,\lambda_3,\lambda_4}$. A partir daqui, o termo "série temporal" será utilizado neste trabalho para denotar todos os casos usuais $\kappa = 2$, $\kappa = 4$ e $\kappa = 5$, uma vez que o tempo é a variável dinâmica fundamental.

3.3 Medidas analíticas

Há diversos parâmetros e métricas μ_p relevantes para caracterização de séries temporais. Quando $\kappa = 2$, a autocorrelação de $A(t)$ pode ser a primeira propriedade a ser considerada. Então, μ_1, numa dada abordagem, poderia ser a correlação cruzada de $A(t)$ com si mesma, uma medida com valores no intervalo: $-1 \leq \mu_1 \leq 1$, que caracteriza intensidades repetidas, tais como a presença de um sinal periódico corrompido por ruído, ou a frequência fundamental em $A(t)$ imposta por suas frequências harmônicas (DUNN, 2005). Quando o padrão de variabilidade de $A(t)$ é um ruído branco perfeito temos $\mu_1 \approx 0$ (não correlacionado, apresentando distribuição de probabilidade normal). Desta forma, μ_1 como autocorrelação é capaz de detectar não aleatoriedade em dados e pode ser usada para identificar um modelo de série temporal apropriado se $A(t)$ tem algum componente determinístico (KANTZ; SCHREIBER, 2003). Consequentemente, um segundo tipo de medida analítica seria a caracterização do ruído $1/f^{\mu_2}$ do espectro de potência de $A(t)$ (KESHNER, 1982). Neste caso, o índice do espectro de potências μ_2 é usado para identificar as escalas nas quais a grade apresenta mais forte correlação. Há muitas outras propriedades possíveis, como variância, assimetria e curtose, a divergência Kullback-Leibler (BURNHAM; R., 2002), dimensões de tipo fractal (PEITGEN et al., 2004), entropia Kolmogorov-Sinai (PEITGEN et al., 2004), índice de singularidade espectral (BOLZAN et al., 2009), etc., que podem ser associadas a GNG com $\kappa = 2$. Todas essas medidas podem, de alguma

forma, ser generalizadas para aplicação em grades $\pounds^4$ e $\pounds^5$.

Em particular, para $\pounds^4$, as medidas analíticas podem resultar de propriedades morfométricas e/ou de processamento de imagens, como, por exemplo, funções de correlação espacial, funcionais de Minkowski (MECKE; STOYAN, 2000) e coeficientes de assimetria da Análise de Padrões Gradientes (GPA) (ROSA et al., 1999; ROSA et al., 2003; ROSA et al., 2007), que caracterizam aspectos estruturais do padrão $A(t, x, y)$ observado no domínio espaçotemporal. Desta forma, as possíveis medidas analíticas constituem um conjunto aberto $\{\mu_1, \mu_j, ..., \mu_p\}$, no qual a escolha dos elementos considerados torna-se dependente da natureza dos dados e da aplicação.

Portanto, no contexto GNG, quando são consideradas medidas analíticas já realizadas, utilizaremos a notação geral: $\pounds^{\kappa,\lambda_1,...,\lambda_\kappa-1}_{\mu_1,\cdots,\mu_P}$, com μ_1 representando, por exemplo, o índice do espectro de potências, que pode ser calculado para $A(t)$, $A(t, x, y)$ e $A(t, x, y, z)$. Tal parâmetro é aqui exemplificado como o mais simples exemplo para μ_1. Assim, uma GNG definida por $\pounds^{2,10^4}_{-1.66}$, portanto, representaria uma série temporal $A(t)$ composta por 10.000 pontos, com índice de espectro de potências igual a -1.66. Tal valor é utilizado para diagnosticar, por exemplo, comportamento do tipo turbulento e, portanto, pode sugerir um modelo ou processo adequado para a série temporal $A(t)$. Neste exemplo, μ_1 é uma propriedades analítica explicitamente incorporada e representada em uma dada GNG.

3.4 Definição Formal de GNG

Uma vez introduzida a noção intuitiva de uma grade numérica generalizada, seu conceito formal, considerando apenas os casos até $\pounds^5$, pode ser apresentado considerando as seguintes definições:

Definição 1 (Coeficientes de Extensão): Seja $D_U : \{t, x, y, z\}$ o domínio espaçotemporal usual em coordenadas cartesianas definido pelos subdomínios t para o tempo, x, y e z para o espaço cartesiano. Se D_U é discreto e composto por λ_1 instantes ordenados de tempo, λ_2 posições ordenadas em x, λ_3 posições ordenadas em y e λ_4 posições ordenadas em z, então o elemento genérico $\lambda_\ell \in \mathbb{Z}$ é o coeficiente de extensão em cada sub-domínio de D_U, $\ell = 1$ para t, $\ell = 2$ para x, $\ell = 3$ para y e $\ell = 4$ para z.

Definição 2 (Grau Variacional κ): Seja $\{A_{\kappa-1}\}$ um conjunto de medidas numéricas, de grau κ, tomadas em D_U. Se $\kappa = 1$, então $A_0 = t$, que é a variável do domínio

fundamental; se $\kappa = 2$, então $A_1 = A(t)$; se $\kappa = 3$, então $A_2 = A(t, x)$; se $\kappa = 4$, então $A_3 = A(t, x, y)$; se $\kappa = 5$, então $A_4 = A(t, x, y, z)$.

Definição 3 (Grade Numérica Generalizada): Seja $A(t)$ uma série de medidas numéricas tomadas em D_U previamente analisada através dos parâmetros μ_1, μ_2,$\cdots$, μ_P. Se são conhecidos o seu grau variacional e seus coeficientes de extensão, então $A(t)$ pode ser escrita na forma $\pounds^{\kappa,\lambda_1,\lambda_2,\lambda_3,\lambda_4,\mu_1,\cdots,\mu_P}$, definida como Grade Numérica Generalizada.

Para este trabalho, foram implementadas duas medidas analíticas, baseadas nas seguintes técnicas: a análise não tendencial de flutuações (PENG et al., 1994), aplicada a séries temporais sem variabilidade espacial, ou seja, grades $\pounds^2$; e a análise de padrões gradientes (ROSA et al., 2003), aplicada a séries temporais com variabilidade no espaço euclidiano bidimensional (x, y), ou seja, grades $\pounds^4$. No próximo capítulo serão discutidas as medidas analíticas decorrentes da aplicação destas ferramentas, bem como os critérios envolvidos na escolha das mesmas. A explicação detalhada das técnicas é apresentada nos Apêndices C e D.

3.5 O Estudo de Caso em Clima Espacial e Processos Ambientais

Com base no conceito de grade numérica generalizada, é possível reduzir uma série temporal analisada a um simples elemento de informação que chamaremos, sem perda de generalidades, de GNG. No exemplo da Figura 3.2, uma série temporal (figura superior) composta por 3600 medidas de temperatura da floresta amazônica no nível da copa das árvores tomadas em um intervalo de 1 hora foi analisada através do seu espectro de potências (figura inferior). O valor do índice espectral funciona como um caracterizador do processo responsável pela geração do sinal[3]. Portanto, a série temporal acrescida da sua saída analítica pode ser representada e visualizada pela GNG $\pounds_{\mu_1}^{\kappa,\lambda_1} = \pounds_{-1.78}^{2,3600}$. Note que, como para $\kappa = 2$ há apenas um coeficiente de extensão, a mesma representação pode ser escrita, na forma reduzida, como $\pounds_{-1.78}^{3600}$. O fato de haver apenas um coeficiente de extensão na notação da GNG implica que se trata de uma variável medida apenas no domínio temporal.

Dependendo da análise desejada, o parâmetro analítico pode ser determinado pela natureza da série temporal. Para caracterizar a lei de escalas das amplitudes de séries

[3]Neste exemplo, trata-se de um processo estocástico do tipo turbulento: não gaussiano e com autocorrelação significativa entre muitas escalas, que definem um regime inercial com lei de potência $1/v^\beta$, com inclinação do espectro de potências $\beta \approx -5/3$.

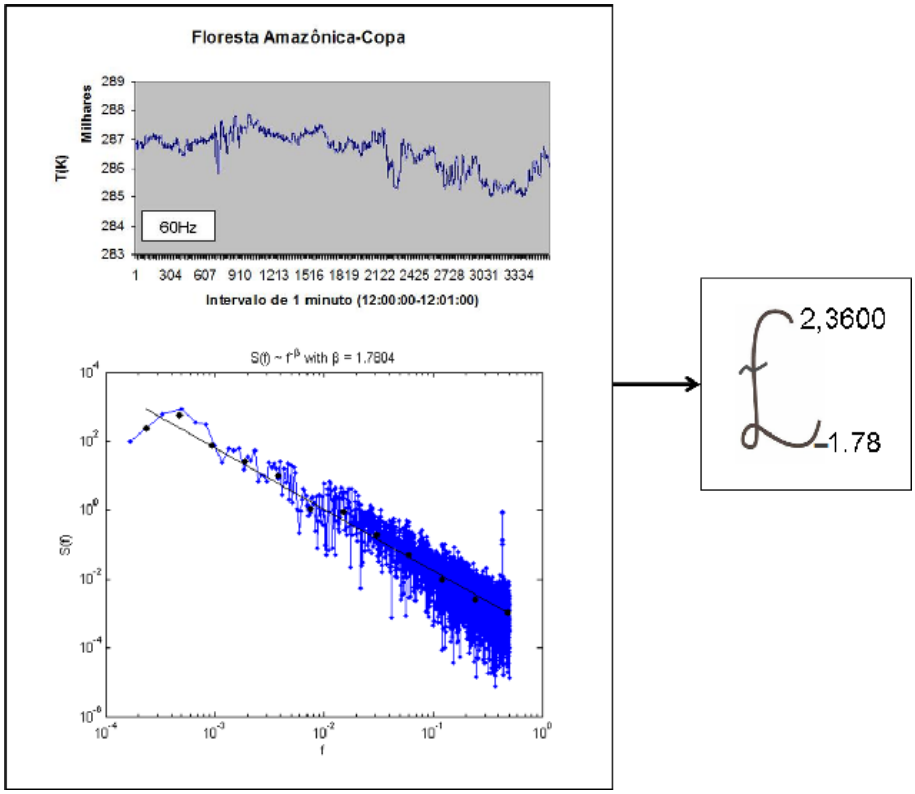


Figura 3.2 - Série temporal de medidas de temperatura na floresta amazônica e seu respectivo espectro de potências, representada como uma GNG com grau variacional $\kappa = 2$, coeficiente de extensão $\lambda_1 = 3600$ e medida analítica $\mu_1 = -1.78$.

curtas é recomendado, ao invés do espectro de potências tradicional, o uso da técnica DFA (*Detrended Fluctuation Analysis*). Essa alternativa será discutida com maior detalhe no Capítulo 6. Esta técnica de análise fornece o expoente de escala da série que caracteriza o seu nível de persistência ou anti-persistência durante a evolução do processo (ver Apêndice C). Será a técnica adotada para a análise padrão de GNG neste trabalho, por ser muito mais robusta do que o espectro de potências para análise de séries curtas, comuns tanto em medidas em física ambiental como em física espacial (VERONESE et al., 2011b).

Um exemplo típico de representação por GNG em física espacial é mostrado nas Figuras 3.3 e 3.4. Na Figura 3.3, uma série temporal proveniente de 2956 registros de explosões solares observadas na frequência de 3 GHz, analisada a partir da técnica DFA, é representada pela sua respectiva GNG (já em sua forma reduzida, isto é, sem necessidade de escrever o grau variacional).

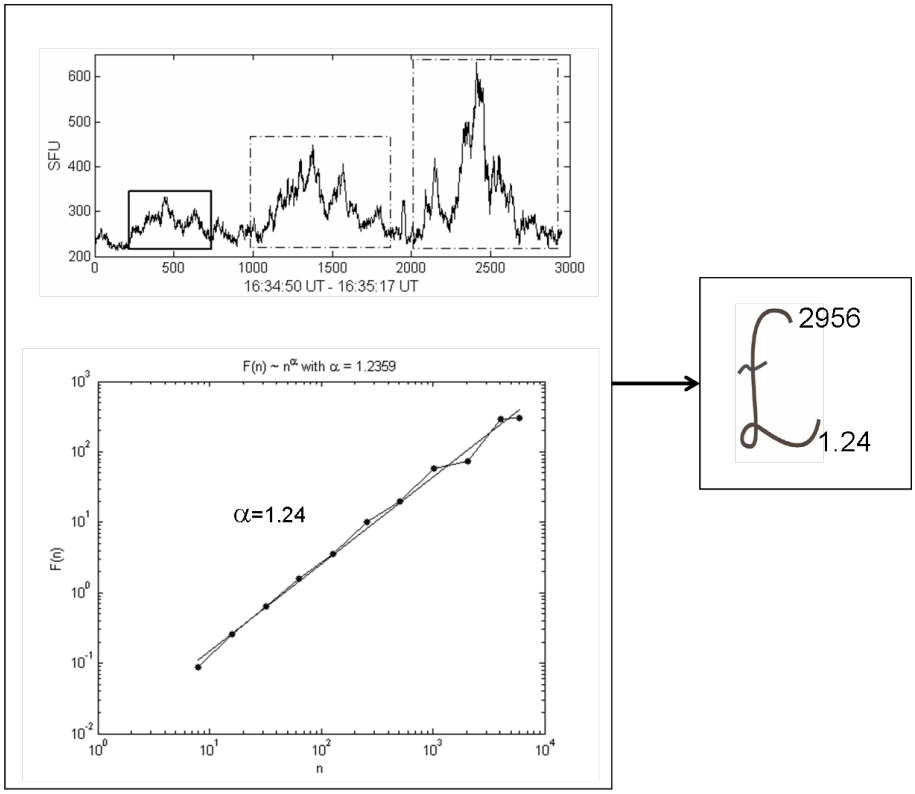


Figura 3.3 - Série temporal de medidas de SFU (*Solar Flux Unit*), representada como uma GNG com coeficiente de extensão $\lambda_1 = 2956$ e parâmetro analítico $\mu_1 = 1.24$. Medidas obtidas a partir do radiotelescópio de Ondrejov da Republica Tcheca.

Na Figura 3.4, uma série espaçotemporal representada por duas imagens de tamanho 64×64 (*snapshots*), observadas pelo radio-interferômetro de Nobeyama no Japão (SAWANT et al., 2011) e analisadas através da técnica GPA (*Gradient Pattern Analysis* – ver Apêndice D), é representada pela sua respectiva GNG.

Tomando como base o último exemplo (Figuras 3.3 e 3.4), cabe perguntar se é possível compor em um mesmo modelo de representação um sistema de GNG; por exemplo, como representar numa mesma estrutura coerente as duas grades mostradas nessas figuras. No próximo capítulo, discutiremos este desafio dentro de um novo paradigma denominado "Analítica Visual", que permitirá construir um tipo complementar de representação de séries temporais denotadas na forma de GNG.

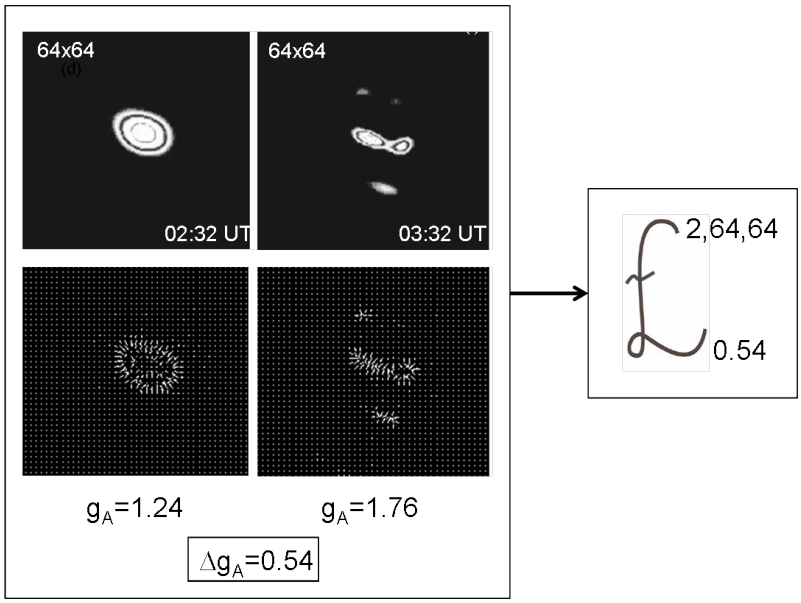


Figura 3.4 - Na parte superior, uma sequência de duas imagens de SFU (*Solar Flux Unit*) medidas em coordenadas espaciais x e y na frequência de 17 GHz a partir do rádio-interferômetro de Nobeyama no Japão. Os campos gradientes de cada imagem, a partir dos quais são calculados os coeficientes de assimetria, são mostrados na parte inferior. À direita é exibida a respectiva GNG com coeficientes de extensão: temporal ($\lambda_1 = 2$); espaciais ($\lambda_2 = \lambda_3 = 64$); e o parâmetro analítico espacial ($\mu_2 = 0.54$), dado pela diferença entre os coeficientes de assimetria de cada imagem.

4 VISUALIZAÇÃO DE DADOS CIENTÍFICOS

"Computer science is no more about computers than astronomy is about telescopes."

(Edsger Dijkstra)

Como mencionado no capítulo anterior, há uma diferença aparentemente intransponível entre a crescente velocidade em que são gerados novos dados e a lenta evolução na capacidade de analisá-los eficientemente (FAYYAD et al., 1996). Combinar técnicas sofisticadas e tradicionais na análise de dados complexos e heterogêneos, característica da mineração de dados, pode auxiliar a descoberta de informações que levem a uma melhor compreensão da natureza dos conjuntos de dados científicos, em geral de grande volume e dimensionalidade variável (GABER, 2010; ROBERTSON, 1991).

Historicamente, o progresso científico quase sempre esteve vinculado à análise visual dos processos naturais (EARNSHAW; WISEMAN, 1992). Segundo Keim et al. (2006), os três objetivos principais da visualização são a apresentação, a análise confirmatória e a análise exploratória. A apresentação de informações multidimensionais graficamente sobre uma tela bidimensional, o que de maneira geral caracteriza as técnicas de visualização, possibilita atingir resultados de forma mais rápida e eficiente. Isso é importante, por exemplo, em um processo operacional de monitoramento de informações para tomada de decisão. Assim, áreas a princípio independentes, como computação gráfica, ciência cognitiva, processamento de imagens, processamento de sinais e *design*, convergem através de métodos análogos que proporcionam à visualização científica ferramentas adicionais para a pesquisa e a investigação de relações entre os dados. Neste contexto, a representação visual de um conjunto de dados científicos, em particular do tipo GNG, introduzido no capítulo anterior, desempenha um papel cada vez mais importante na tentativa de compreender e monitorar o fenômeno a partir do qual séries temporais são geradas.

Embora a mineração de dados enfatize em geral abordagens algorítmicas ou matemáticas, a incorporação de técnicas de visualização, ou a mineração visual de dados[1], pode ser uma ferramenta chave na análise automática de conjuntos massivos de da-

[1]Detalhes sobre métodos de visualização em mineração de dados podem ser encontrados em Keim et al. (2005), Tan et al. (2009) e Keim (1996).

dos (TAN et al., 2009). Projetados para lidar com bases de dados estáticas[2], métodos tradicionais de mineração de dados vêm sendo explorados na busca por representações mais eficientes de dados sequenciais (LAST et al., 2004), como é o caso das séries temporais, aqui tratadas como GNG.

Frequentemente apenas parte dos dados disponíveis possuem qualidade suficiente para ser submetidos, por exemplo, a tarefas de treinamento em métodos de reconhecimento de padrões[3] (WITTEN; FRANK, 2005). Neste caso, ferramentas de visualização podem facilitar o trabalho de um especialista humano, muitas vezes necessário para filtrar informações que eventualmente possam comprometer o resultado final da análise. Isso acontece porque nós, seres humanos, somos dotados de habilidades que nos permitem reconhecer intuitivamente padrões visuais que poderiam nos escapar se apresentados algébrica ou textualmente (LAROSE, 2006). Outra motivação usual para a utilização de técnicas de visualização é a tentativa de integrar conhecimento do domínio[4] à análise, uma tarefa muitas vezes difícil ou até mesmo impossível quando restrita a ferramentas puramente estatísticas ou algorítmicas. Informações descritivas sobre os dados (metadados), por exemplo, são um tipo de conhecimento *a priori* capaz de tornar mais poderosa qualquer estratégia de representação de dados, porém incorporá-las de forma útil à análise continua a ser um desafio (WITTEN; FRANK, 2005).

4.1 Dados descritivos ou metadados

Uma das principais motivações para a incorporação de atributos descritivos a modelos de visualização é que muitas vezes as informações contidas nos mesmos estão relacionadas, e a identificação destas relações pode ser importante na análise dos dados. Em geral, há pouca dificuldade técnica em incluir conhecimento sobre essas relações nos algoritmos de análise, porém ainda é um desafio encontrar a melhor forma de expressar estes metadados para que eles possam ser visualizados pelo analista humano a partir do algoritmo (WITTEN; FRANK, 2005).

Com este objetivo, constantemente novos formatos de arquivos são desenvolvidos

[2]Bases de dados estáticas caracterizam-se pela irrelevância da ordenação dos registros ou objetos da base de dados sobre os padrões de interesse (LAST et al., 2004).

[3]O reconhecimento de padrões visa à classificação de objetos dentro de um número de categorias ou classes (THEODORIDIS; KOUTROUMBAS, 2003).

[4]O conhecimento do domínio (*domain knowledge*) refere-se ao conhecimento – adquirido ou intuitivo – inerente a um especialista humano sobre os dados analisados (HJØRLAND; ALBRECHTSEN, 1995; WITTEN; FRANK, 2005; TAN et al., 2009).

na tentativa de uniformizar o armazenamento e a transmissão de informações. No contexto de dados científicos, podemos citar os formatos HDF, CDF e FITS (ver Apêndice A). Outras iniciativas voltadas para tipos específicos de dados podem ser encontradas, como o padrão OGC (Open Geospatial Consortium), aplicado a Sistemas de Informações Geográficas (DOYLE; REED, 2001). A escolha da abordagem usada para representar um determinado conjunto de dados depende inicialmente do contexto da aplicação. Diferentes representações podem revelar características bem distintas, e por isso é preciso identificar com precisão o tipo de informação em que estamos interessados (ROBERTSON, 1991). Na próxima seção, é apresentado o estado da arte em aplicações de ferramentas de representação e visualização de dados espaçotemporais.

4.2 Estado da Arte

A busca por novas técnicas capazes de auxiliar especialistas humanos na análise massiva e imediata de dados complexos tem levado ao surgimento de novas áreas baseadas, em geral, na integração de disciplinas como visualização científica, análise estatística, representação do conhecimento e ciências cognitivas (THOMAS; COOK, 2006). Questões relacionadas a segurança, saúde e meio ambiente são exemplos de aplicações que podem gerar uma sobrecarga de dados e exigir tomadas de decisão robustas e urgentes. Assim, mais do que a análise automática, o campo emergente da Analítica Visual[5] (*Visual Analytics*) permite aos tomadores de decisão evidenciar suas capacidades cognitivas e perceptuais sobre o processo analítico e, ao mesmo tempo, aplicar habilidades computacionais avançadas para melhorar o processo de descoberta (KEIM et al., 2006).

Na literatura encontra-se uma ampla diversidade de trabalhos documentando pesquisas em visualização e/ou representação de dados espaçotemporais em diferentes áreas da ciência (PAULOVICH, 2008; de Pinho, 2009; GRÉGIO et al., 2009; HEIJMANS, 2002). Uma taxonomia extensível, capaz de descrever e classificar técnicas de visualização analisando os mecanismos e fundamentos de percepção visual envolvidos, pode ser encontrada em Jr. et al. (2006).

Desenvolvidos essencialmente para sistemas de dados corporativos, as aplicações OLAP (*On-Line Analytical Processing*) englobam técnicas para a exploração de ma-

[5]Analítica Visual é a tradução do termo em Inglês, no qual *Analytics* refere-se à ciência da análise humana (CHINCHOR et al., 2010).

trizes multidimensionais de valores através da criação de tabelas de resumo a partir de dados de várias dimensões ou compostos por vários valores de atributos. Focados na análise interativa, tais sistemas normalmente fornecem capacidades extensivas para a visualização dos dados e geração de resumos estatísticos (TAN et al., 2009). O potencial das aplicações OLAP sobre dados científicos vem sendo explorado em alguns trabalhos recentes. Por exemplo, integrando ao OLAP outras ferramentas de bancos de dados, visualização e modelagem, Cao et al. (2010) buscam representar e organizar um sistema virtual de dados de bacias hidrográficas dos Estados Unidos da América, como mostra a Figura 4.1.

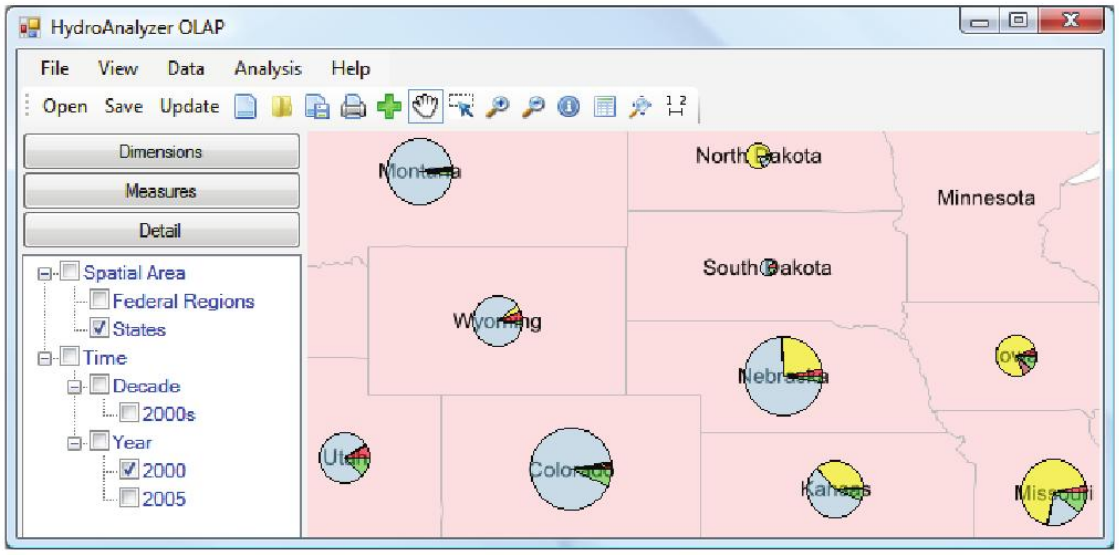


Figura 4.1 - Exemplo de representação de dados multidimensionais através do conceito de OLAP.
Fonte: Cao et al. (2010).

Na área de física espacial, a base de dados distribuída SPIDR (*Space Physics Interactive Data Resource*), originalmente desenvolvida em 1995 como uma demonstração para o projeto internacional GOIN (*Global Observation and Information Network*), é descrita em Zhizhin et al. (2008). O sistema, que permite selecionar, visualizar e modelar dados históricos de clima espacial[6] distribuídos pela Internet, é agora uma

[6]O termo "clima espacial" se refere às condições variáveis temporalmente no ambiente espacial capazes de afetar sistemas tecnológicos implantados na Terra ou no espaço (HANSLMEIER, 2007). Neste contexto, espaço corresponde à região que abrange a atmosfera terrestre e o espaço cósmico (HOUAISS; VILLAR, 2001).

fonte de dados padrão para a Física solar-terrestre, compondo o Centro Mundial de Dados (WDC – *World Data Center*) criado pelo Conselho Internacional de Uniões Científicas (ICSU – *International Council of Scientific Unions*).

Wong e Thomas (2004) definem a *Analítica Visual* como uma abordagem contemporânea que combina a arte da intuição humana e a ciência da dedução matemática para diretamente reconhecer padrões e derivar conhecimento e discernimento sobre eles. Esta abordagem deu origem nos últimos anos a diversos trabalhos (GREEN et al., 2008; CHEN et al., 2009) e, mais recentemente, a uma nova área: a *Analítica Multimídia*, que incorpora a análise multimídia à Analítica Visual (CHINCHOR et al., 2010).

O modelo de representação de dados desenvolvido neste trabalho tem por objetivo constituir um novo formalismo para generalização matemática de conjuntos de dados espaçotemporais gerados por múltiplas medidas tomadas sobre um único sistema físico cujos processos apresentam diferentes padrões de variabilidade no tempo, como é o caso dos dados do programa EMBRACE do INPE. Por meio da integração de metadados, parâmetros pós-analíticos e ferramentas de visualização, um modelo complementar deve facilitar a análise exploratória de sistemas de dados complexos baseados em medidas automáticas sobre um determinado conjunto de GNG.

4.3 Analítica Visual Aplicada a GNG Científicas

No contexto de dados científicos de séries temporais, as ferramentas apresentadas nas seções anteriores ilustram a dificuldade em desenvolver um modelo de representação que generalize conjuntos de séries temporais ao mesmo tempo em que as analisa e apresenta de forma automática suas propriedades de variabilidade no tempo e no espaço. Em geral, cientistas utilizam estas ferramentas de maneira complementar e dependente do tipo de análise e/ou tomada de decisão desejada.

Muitas vezes, gráficos e imagens fornecem indícios ou *insights* que devem ser seguidos de técnicas de análise sistêmica confirmatórias ou exploratórias para que se atinjam conclusões e/ou decisões confiáveis sobre todo o sistema de dados. Assim, uma visualização prévia de metadados e medidas analíticas tomadas sobre todo o conjunto de dados representativo de um mesmo sistema de séries temporais pode, por outro lado, auxiliar o cientista na tarefa de realizar uma seleção prévia, eliminando a necessidade de gerar modelos visuais sobre informações irrelevantes em

função de seus atributos, ou destacando a importância de determinados dados em relação ao objetivo da análise integrada.

É importante ressaltar que a importância da aplicação prévia ou posterior desta visualização analítica é, a princípio, diretamente proporcional à quantidade de grades numéricas que compõem o conjunto de dados considerado. Portanto, uma representação alternativa, reunindo em um cenário simples informações descritivas e analíticas sobre um sistema heterogêneo de dados, pode atuar como uma ferramenta auxiliar em pesquisas envolvendo análise, representação e visualização de séries temporais nas mais diversas áreas do conhecimento. Com este objetivo, neste trabalho desenvolvemos um modelo integrado de representação e visualização analítica de séries temporais, denominado *modelo de representação baseado em grades numéricas generalizadas*.

Como descrito em (VERONESE et al., 2009), este modelo foi inicialmente projetado para ser desenvolvido em módulos, como mostra a Figura 4.2.

De acordo com o projeto inicial, o módulo MGNLA (*Module GNL Assembly*) seria responsável pela construção das grades numéricas generalizadas a partir da leitura dos arquivos de dados "$Data_i$". O termo GNL corresponde à sigla para *Generalized Numerical Lattice*, enquanto DSR se refere a *Data System Representation*. No segundo módulo seria construído o sistema de representação de dados baseado em grades numéricas generalizadas, em função dos resultados do módulo de análise sistêmica de dados de segunda ordem, gerando uma estrutura matricial. As variáveis C_i indicam as quantidades de GNG para cada valor de $i = 1, \cdots, \kappa - 1$. Desta forma, a variável L corresponde ao número total de GNG compondo o sistema representado.

A configuração final do modelo resultante da implementação computacional destes módulos será descrita no próximo capítulo.

Como vimos neste capítulo, neste trabalho será introduzido um modelo de representação mais simples do que aqueles existentes, incorporando porém ingredientes que são de interesse direto da comunidade científica, como é o caso da análise de séries temporais disponibilizada diretamente no modelo de representação.

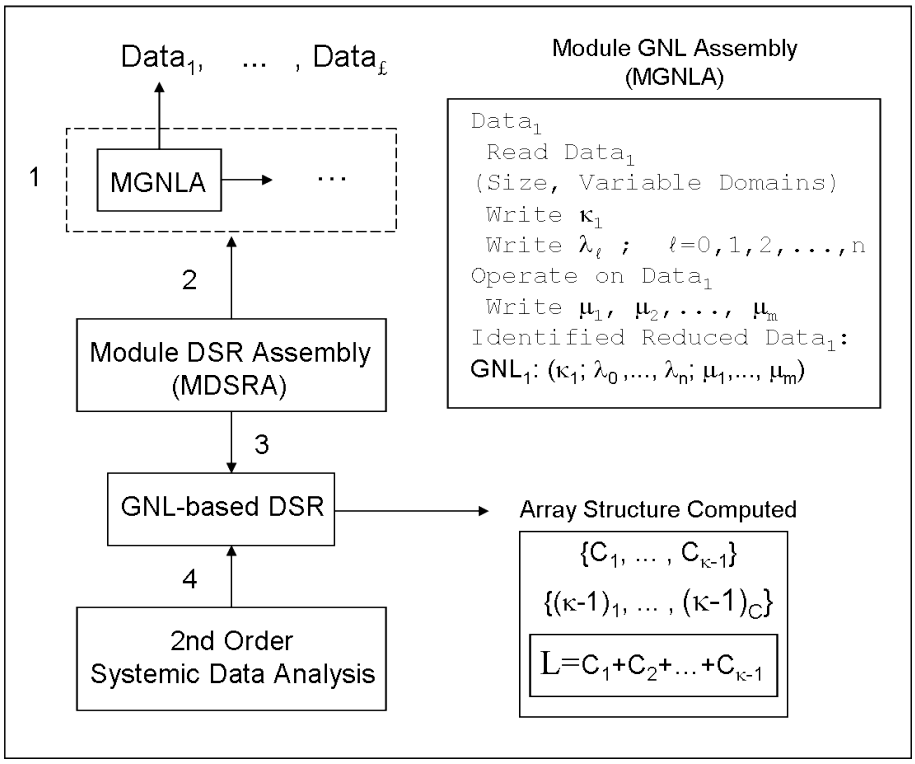


Figura 4.2 - Projeto inicial do modelo de representação de dados baseado em grades numéricas generalizadas.
Fonte: Veronese et al. (2009).

5 MODELO DE REPRESENTAÇÃO BASEADO EM GRADES NUMÉRICAS GENERALIZADAS

"Man is still the most extraordinary computer of all."

(John F. Kennedy)

5.1 Modelo conceitual

Baseado em um conjunto de grades numéricas generalizadas representando uma coleção de medidas experimentais de um mesmo sistema observado, o modelo de análise e representação de séries temporais desenvolvido neste trabalho consiste em uma estrutura organizada no formato de uma tabela contendo todas as GNG relativas a um sistema de dados particular. É intuitivo notar que a forma como introduzimos o conceito de GNG permite construir uma estrutura na qual os elementos de cada linha podem ser ordenados por λ_1 (a extensão temporal dos dados), e os elementos de cada coluna, por κ, ambos em ordem ascendente, como na estrutura ilustrada na Figura 5.1. Uma propriedade importante desta representação é que a coluna da extrema direita apresenta os totais marginais τ_κ, ou seja, em cada linha é exibido o número total de GNG com o mesmo κ. Quando o sistema de dados é representado por todos os valores de κ, o valor da coluna mais à direita corresponde ao total de séries temporais $A(t)$ (primeira linha), séries espaçotemporais em uma dimensão espacial $A(t, x)$ (segunda linha), imagens $A(t, x, y)$ (terceira linha), hipercubos $A(t, x, y, z)$ (quarta linha), etc. Quando não há dados para algum κ, a representação deve considerar o próximo κ automaticamente. Por exemplo, quando não há $\pounds_3$, o total de séries espaçotemporais $\pounds_4$ é exibido na célula mais à direita da segunda linha.

Este modelo baseado em GNG pode ser útil em muitas áreas e aplicações, incluindo Mineração de Dados baseada em propriedades estruturais de séries temporais, modelagem de dados multidimensionais, sistemas de informação multivariada, especialmente aqueles obtidos em física espacial e ambiental, genômica, neurociência, medicina e outros bancos de dados espaçotemporais em ciências gerais. A tarefa de modelagem do sistema de dados em uma representação organizada de GNG de forma estrutural e semântica é obtida a partir da execução de uma sequência de passos pré-estabelecida. Inicialmente, as observações do sistema real são realizadas usando di-

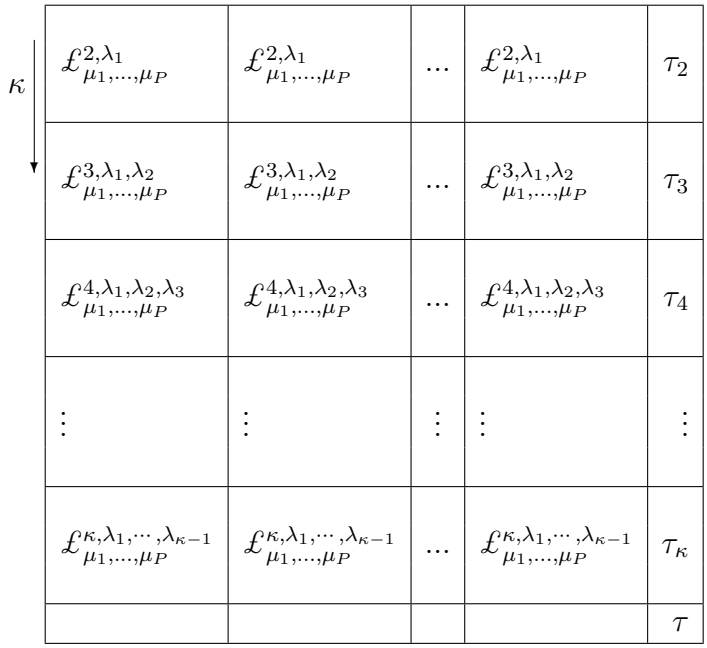


Figura 5.1 - Modelo de Representação de Dados baseado em GNG.

ferentes instrumentos e/ou experimentos numéricos. As medidas resultantes podem ser organizadas como metadados e arquivos de dados. Então, o processamento de dados é realizado através de uma classe de programas que organizam e manipulam os dados, geralmente grandes quantidades de dados numéricos. É possível adquirir medidas sobre um sistema observável em tantas variáveis quantas se desejem, e então realizar uma parametrização e analisar a informação a fim de incorporá-la à estrutura de GNG aqui proposta. O modelo final é obtido através do agrupamento e uniformização de todas as informações geradas a partir dos arquivos – um para cada tipo de dado – em um único Modelo de Representação de Dados, a partir do qual a análise sistêmica pode ser realizada para identificar, por exemplo, atributos da variabilidade solar responsáveis por anomalias em satélites geoestacionáios. Este processo é resumido na Figura 5.2.

Na próxima seção, é descrita a estratégia adotada para resolver computacionalmente as etapas do processo de construção do modelo de representação baseado em grades numéricas generalizadas.

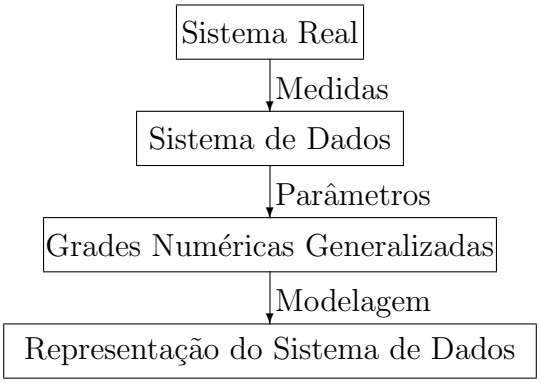


Figura 5.2 - Representação de um sistema de dados a partir da observação de um sistema real.

5.2 Implementação Computacional

O ambiente computacional é composto por dois módulos complementares, como mostra a Figura 5.3. O primeiro módulo, desenvolvido em linguagem C++, é responsável por identificar, processar e armazenar de forma organizada em um único arquivo XML as informações referentes a todos os arquivos de entrada (grades numéricas) associados ao sistema real representado. Os arquivos de entrada devem estar em um formato homogêneo, que será descrito na próxima seção. A seguir, o módulo para construção do esquema visual de representação de dados, implementado em linguagem Java, é aplicado sobre o arquivo XML, produzindo uma janela de visualização contendo a representação formal do sistema de grades numéricas generalizadas resultante.

Ao conjunto de módulos que formam este sistema computacional, demos o nome de *DLattice++*, em referência à palavra "grade" (em Inglês, *lattice*), base conceitual, e à linguagem C++, base computacional do sistema.

5.2.1 Encapsulamento dos dados de entrada

Computacionalmente, Sistemas Heterogêneos de Dados constituem uma coleção de arquivos que podem estar em diferentes formatos. Muitas vezes, como discutido no Capítulo 4, novos formatos são especialmente desenvolvidos para facilitar a organização dos dados em um determinado tipo de aplicação. Para prover maior liberdade ao usuário, é necessário, portanto, que o sistema computacional de construção do modelo de representação baseado em GNG não esteja restrito a dados de entrada em formatos específicos. Assim, as informações numéricas representativas das medi-

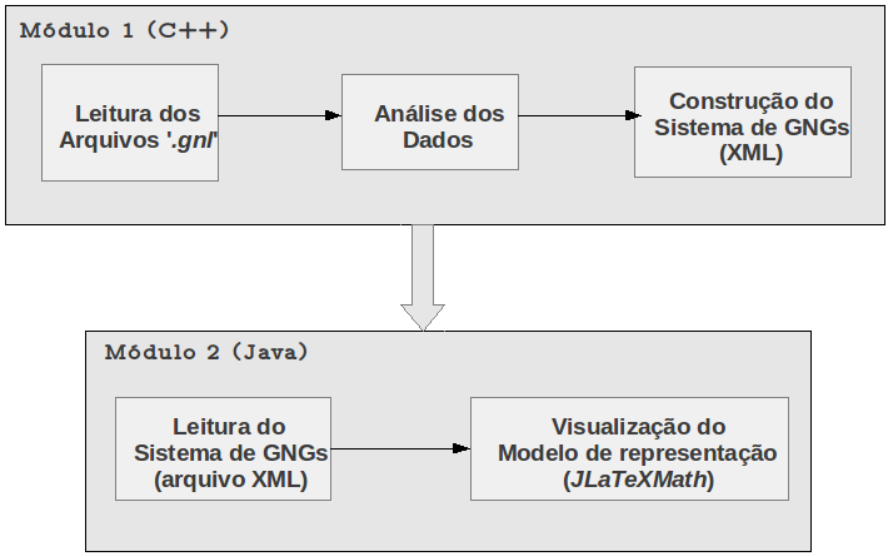


Figura 5.3 - Diagrama do programa *DLattice++*: módulos computacionais da implementação computacional do modelo de representação baseado em GNG.

das que compõem cada grade devem ser encapsuladas em um arquivo de extensão '.gnl', obedecendo a uma regra de organização para converter cada arquivo em uma representação homogênea.

Para gerar um arquivo '.gnl', é incorporado à série temporal um cabeçalho contendo os valores dos coeficientes de extensão, λ_i, com $i = \kappa - 1, ..., 0$. O corpo do arquivo deve conter, portanto, apenas os valores numéricos que representam as medidas tomadas sobre o sistema, dispostos sequencialmente nas linhas do arquivo, como um vetor coluna de $\prod_{i=1}^{\kappa-1} \lambda_i$ elementos.

Todos os dados que devem compor o modelo de representação devem estar localizados na mesma pasta de arquivos. Desta forma, cada arquivo de dados é lido e analisado sequencialmente. Após a conclusão da análise de todas as grades numéricas, o módulo de visualização, descrito na Seção 5.2.3, realiza a integração de todas as informações.

5.2.2 Módulo de análise

A leitura dos arquivos '.gnl' é realizada por uma rotina implementada em linguagem C++. Em função do conteúdo do cabeçalho do arquivo, o módulo de análise

armazena os metadados e medidas espaçotemporais da grade numérica representada pelo arquivo de entrada em estruturas de dados dinâmicas, baseadas em objetos definidos em Press et al. (2007)[1], e seleciona as ferramentas de análise adequadas de acordo com a variabilidade espaçotemporal da grade. As ferramentas de análise implementadas neste módulo são descritas no próximo capítulo.

A etapa final deste módulo é responsável por salvar em um arquivo XML os metadados e resultados analíticos referentes a todas as grades numéricas representativas da entrada.

5.2.3 Módulo de visualização

Implementado em linguagem Java, o módulo de visualização realiza a leitura do arquivo XML, coletando os resultados obtidos pelo módulo de análise e metadados de todas as grades numéricas representativas do sistema real a ser modelado. Para a codificação das notações matemáticas estabelecidas nos Capítulos 3 e 5, foi utilizada a API *JLaTeXMath* (DENIZET; LEDRU, 2011), que permite reproduzir fórmulas matemáticas escritas em Latex sobre interfaces visuais geradas em Java.

Conforme o modelo de representação incorpora uma maior quantidade de GNG, o desempenho da sua visualização analítica torna-se mais dependente do tamanho da tela onde o modelo será visualizado. Para fins de monitoramento utilizando o modelo, são necessários monitores panorâmicos, comumente utilizados em salas de rastreio e controle de satélites, como aquelas utilizadas nos institutos e agências espaciais.

A Figura 5.4 mostra um exemplo de controle em tempo real da atitute de um satélite da NOAA. Na figura inferior, uma grade digital composta por uma rede de monitores ilustra um exemplo de tecnologia que poderá ser útil para o monitoramento do Clima Espacial (no contexto do programa brasileiro EMBRACE) baseado em GNG.

O módulo de análise de dados está configurado de forma a permitir que o usuário insira a sua rotina/biblioteca para cálculo de medidas analíticas mais conveniente. Todos os dados de entrada devem estar no formato '.gnl' descrito anteriormente.

[1] A opção pelo uso das estruturas de dados implementadas em Press et al. (2007) foi baseada na compatibilidade destas estruturas com as funções disponibilizadas por sua biblioteca, uma vez que algumas destas funções foram incorporadas à implementação das ferramentas de análise utilizadas neste trabalho.

Figura 5.4 - Acima, um exemplo de sala de monitoramento para aplicações espaciais. Note que o monitor mais à esquerda apresenta uma tabela, que poderia ser a visualização de um Modelo de Representação Analítica baseada em GNG. Abaixo, um exemplo de grades de monitores que também pode ser adaptada para o mesmo propósito descrito anteriormente.
Fonte: NOAA e *World Press*.

5.2.4 Flexibilidade dos módulos

O módulo de visualização apresentado na seção anterior está em adaptação para proporcionar maior flexibilidade em relação a todos os atributos visualizados na grade. Na configuração atual, o usuário pode alterar com facilidade o código fonte de forma a determinar como cada propriedade da grade será visualizada. Por exemplo, o uso do símbolo $£$ é opcional, os parâmetros κ e λ podem ser posicionados na parte superior ou inferior de cada GNG, e a precisão das medidas analíticas pode ser adaptada de acordo com a aplicação.

O programa *DLattice++* será disponibilizado através da ferramenta *SourceForge* (http://sourceforge.net/).

6 TESTE DO MODELO E APLICAÇÕES

"O objetivo da computação é gerar insight, não números."

(Richard Hamming)

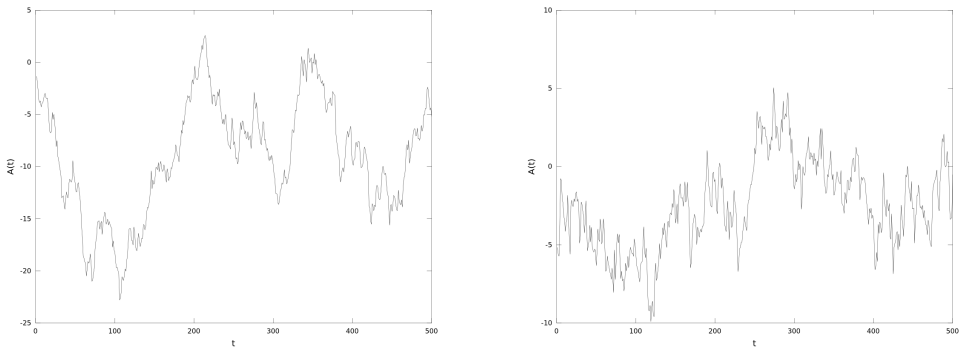
Para um determinado conjunto de GNG, é recomendado, antes da aplicação do modelo de representação analítica, realizar um teste do seu desempenho, uma vez que o custo computacional dos algoritmos utilizados no processo de construção do modelo depende das características dos dados e das ferramentas de análise escolhidas. Neste capítulo, portanto, é apresentado um sistema de GNG como referência para os casos mais usuais ($\kappa = 2$ e $\kappa = 4$), permitindo a realização de um teste de desempenho básico anterior à aplicação em casos reais, como, por exemplo, o estudo do Clima Espacial. A complexidade das operações matemáticas envolvidas nas ferramentas de análise pode também comprometer a aplicabilidade do modelo a um determinado sistema de dados. Portanto, torna-se necessário avaliar o desempenho do modelo computacional com base em algumas ferramentas de análise usuais. Finalmente, após executados os testes sobre sistemas canônicos, a aplicação do modelo será discutida no contexto de dois estudos de caso reais.

6.1 Sistemas Canônicos

Com o objetivo de demonstrar a aplicabilidade do modelo a qualquer conjunto heterogêneo de séries temporais, uma coleção de dados gerados sinteticamente foi utilizada, simulando um sistema complexo representado por medidas que variam no espaço bidimensional e no tempo. Uma vez que constituem modelos simplificados de sistemas físicos, tais dados serão denominados "sistemas canônicos"[1]. Os processos escolhidos para construção das grades numéricas generalizadas canônicas para os graus variacionais 2 e 4 são baseados no padrão de variabilidade temporal conhecido como *Movimento Browniano* (Apêndice A). Nesta seção serão descritos os algoritmos que implementam este processo para $\pounds^2$ e $\pounds^4$.

Não encontramos na literatura abordagens matemáticas para a geração de ruídos

[1] O termo "canônico" é utilizado aqui como uma tradução livre para o termo em Inglês *canonical*, um adjetivo que significa "reduzido à forma mais simples e mais significante sem perda de generalidade". No contexto da Ciência da Computação, o termo se refere a uma maneira usual ou padrão de organização (BLOECHLE et al., 2006).



(a) (b)

Figura 6.1 - Exemplos de sinais do tipo ruído *Browniano* gerados a partir do algoritmo baseado na Equação 6.1, para os seguintes valores de β: (a) $\beta \approx 1.5$; (b) $\beta \approx 2$.

do tipo *Browniano* na forma de séries espaçotemporais com grau variacional 4 ($\pounds^4$). Para a geração das séries espaçotemporais canônicas do tipo $\pounds^4$, foi desenvolvida, portanto, uma adaptação inédita da equação do mapa acoplado, incorporando a equação do movimento *Browniano* (VERONESE et al., 2011a). Esta aplicação do mapa acoplado, apresentada na Seção 6.1.2, constitui uma abordagem inovadora para sintetização de $\pounds^4$, que pode ser utilizada também como conjunto de amostras de teste e validação para ferramentas de análise de dados espaçotemporais provenientes de processos não lineares em outras abordagens.

6.1.1 GNG canônicas do tipo $\pounds^2$

O modelo utilizado aqui para gerar sinais do tipo ruído *Browniano* foi apresentado por Osborne e Provenzale (1989). Cada elemento $A(t_i)$ de uma série com N pontos, considerando $t_i = i\Delta t$ com $i = 1, \cdots, N$, distribuídos de acordo com o espectro de potências $S(\nu) \propto \nu_j^{-\beta}$, onde β corresponde ao expoente de escala, é calculado a partir da seguinte fórmula:

$$A(t_i) = \sum_{j=1}^{N/2} [S(\nu_j)\Delta\nu]^{1/2} \cos(\nu_j t_i + \phi_j), \quad i = 1, \cdots, N, \tag{6.1}$$

onde $\nu_j = j\Delta\nu$, com $j = 1, \cdots, N/2$ e $\Delta\nu = 2\pi/N\Delta t$, e ϕ_j são as fases, definidas randomicamente.

A Figura 6.1 mostra dois exemplos de sinais do tipo ruído *Browniano* gerados a partir da Equação 6.1, com o expoente de escala correspondendo a 1.5 (a) e 2 (b).

6.1.2 GNG canônicas do tipo $\pounds^4$

A variabilidade obtida pelo modelo *Browniano* descrito na seção anterior foi incorporada, de forma inédita, a cada ponto (m, n) do modelo de formação de padrões espaciais denominado mapa acoplado, onde m e n correspondem, respectivamente, às extensões espaciais em x e y, originando um novo modelo de geração de séries espaçotemporais ($\kappa = 4$) canônicas, aqui denominado mapa acoplado *Browniano*.

Considere um elemento arbitrário $a^t_{m,n}$ de uma matriz de amplitudes $A(t, x, y) = \{a^t_{m,n}\}$. A forma explícita do mapa acoplado global é dada por:

$$a^{t_{i+1}}_{m,n} = (1-\varepsilon)a^{t_i}_{m,n} + \frac{\varepsilon}{4}(f(a^{t_i}_{m+1,n}) + f(a^{t_i}_{m,n+1}) + f(a^{t_i}_{m-1,n}) + f(a^{t_i}_{m,n-1})), \quad (6.2)$$

onde ε é a constante de acoplamento e $f(a)$ é uma equação genérica, que, nos modelos originais para o estudo de caos espaçotemporal, corresponde à equação logística, $f(a) = \rho a(1-a)$, em que ρ é uma constante real que representa o parâmetro de controle caótico (RAMOS et al., 2000). Neste trabalho, devido à natureza dos dados de clima espacial, substituímos a equação logística $f(a)$ pela equação do movimento *Browniano* descrita na seção anterior. Portanto, de forma inédita, a equação do mapa acoplado *Browniano* é aqui representada por:

$$a^{t_i}_{m,n} = (1-\varepsilon_i)f_B(a^{t_i}_{m,n}) + \frac{\varepsilon_i}{4}(f_B(a^{t_i}_{m+1,n}) + f_B(a^{t_i}_{m,n+1}) + f_B(a^{t_i}_{m-1,n}) + f_B(a^{t_i}_{m,n-1}), \quad (6.3)$$

onde $\varepsilon_i \in [0, 1]$ e f_B corresponde ao termo $A(t_i)$ na Equação 6.1.

A Figura 6.2 mostra dois exemplos de séries temporais $\pounds^4$ geradas pela Equação 6.3 com expoente de escala $\beta = 2$. A série temporal superior foi gerada considerando o fator de acoplamento constante $\varepsilon = 0.5$. Na série temporal inferior, o valor do fator de acoplamento sofreu variação ao longo do tempo, assumindo valores crescentes dentro do intervalo $[0, 1]$.

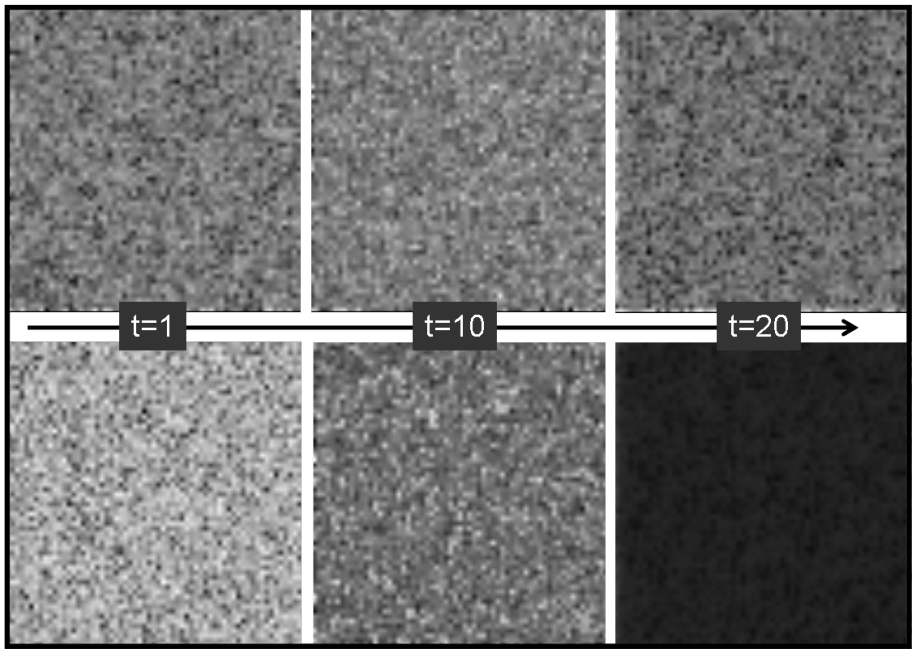


Figura 6.2 - *Snapshots* de duas séries temporais $\pounds^4$ geradas pela equação do Mapa Acoplado *Browniano*. Os três *snapshots* superiores foram gerados com fator de acoplamento constante $\varepsilon = 0.5$, enquanto os *snapshots* da série inferior foram gerados com fatores de acoplamento $\varepsilon = 0$, $\varepsilon = 0.5$ e $\varepsilon = 1$. Em ambas as séries foi utilizado o expoente de escala $\beta = 2$.

Dessa forma, o modelo canônico *Browniano* composto pelas Equações 6.1 e 6.3 constitui o simulador de grades $\pounds^2$ e $\pounds^4$ para testes do modeo descrito no capítulo anterior.

6.2 Escolha dos Métodos de Análise

Como foi dito no Capítulo 3, a escolha das medidas analíticas para aplicação sobre as grades numéricas generalizadas que compõem o sistema de dados a ser modelado depende essencialmente dos objetivos da visualização analítica, das características particulares do sistema real modelado e das limitações impostas por fatores qualitativos e/ou quantitativos dos dados disponíveis.

Para verificação e validação do modelo desenvolvido neste trabalho, é necessário, porém, selecionar medidas analíticas significativas para o contexto dos estudos de caso explorados. Nesta seção serão descritas as técnicas de análise utilizadas para implementação destas medidas analíticas, bem como os critérios de escolha envolvidos neste processo.

6.2.1 Grau variacional 2

O conceito de lei de potência, associado a um espectro de energias, permite caracterizar um padrão de variabilidade complexa, isto é, cujas densidades espectrais são proporcionais a $1/\nu^{\beta}$, com β assumindo diferentes valores a partir do tipo de processo considerado. Sinais gerados por processos estocásticos do tipo $1/\nu^{\beta}$ são encontrados na física, meteorologia, biologia, engenharia, economia, etc. Por esta razão, quando se trata de propriedades que visem a caracterizar séries temporais provenientes de sistemas físicos, como é o caso das aplicações consideradas neste trabalho, é fundamental incorporar ao conjunto de medidas analíticas que compõem o modelo de representação baseado em grades numéricas generalizadas parâmetros capazes de identificar padrões de variabilidade típicos de determinados comportamentos estocásticos como os descritos anteriormente. Neste sentido, a técnica da análise não tendencial de flutuações[2] (DFA), proposta por (PENG et al., 1994), vem se destacando na caracterização de padrões temporais que parecem ocorrer devido a processos estocásticos com memória de longo alcance, e tem sido aplicada a dados biológicos, financeiros, sinais de física espacial, etc.

A técnica DFA mede o expoente de escala[3] α de padrões de variabilidade de uma série temporal não estacionária para determinar auto-afinidade em um processo não linear dinâmico. Devido a sua robustez, especialmente quando a série temporal analisada é composta por poucos pontos (VERONESE et al., 2011b), esta técnica foi escolhida como medida analítica padrão para séries do tipo $\pounds^2$ no modelo de representação desenvolvido neste trabalho. O algoritmo utilizado para o cálculo do expoente de escala α através da aplicação da técnica DFA é descrito no Apêndice C.

6.2.2 Grau variacional 4

A escolha de uma ferramenta padrão para o módulo que analisa as grades $\pounds^4$ não é trivial como no caso das grades $\pounds^2$. O primeiro problema corresponde à escolha de um critério para gerar o parâmetro analítico de segunda ordem a ser obtido a partir de uma série de valores provenientes da análise de primeira ordem (isto é, extraídos de cada imagem).

[2] A denominação original em Inglês da técnica *Detrended Fluctuation Analysis* vem sendo traduzida livremente como "Análise de Flutuação Destendenciada" ou ainda "Análise Destendenciada de Flutuações", embora a palavra "destendenciada" não conste nos dicionários oficiais da Língua Portuguesa. Neste trabalho, adotamos o termo "não tendencial" como tradução para *detrended*, em respeito ao vocabulário oficial e ao Acordo Ortográfico da Língua Portuguesa.

[3] A relação entre os valores de α e β é dada por $\beta \approx 2\alpha - 1$, conforme descrito no Apêndice C.

No caso de padrões relacionados à evolução de estruturas provenientes de processos não lineares, uma medida de segunda ordem que muitas vezes pode ser útil é um caracterizador de relaxação espaçotemporal, comparando valores médios da variável de primeira ordem com a finalidade de caracterizar a emergência de estabilidade do padrão ao longo de sua evolução.

Uma técnica capaz de quantificar este processo é a Análise de Padrões Gradientes (GPA) (ROSA et al., 2003). Por se tratar de uma técnica já consolidada na literatura e desenvolvida pelo grupo de computação científica do LAC, a GPA foi escolhida como a metodologia padrão para analisar grades com grau variacional $\kappa = 4$ incorporada ao modelo para representação analítica de GNG.

O método de Análise de Padrões Gradientes é capaz de caracterizar a formação e a evolução de padrões de assimetria bilateral em uma imagem através do cálculo do chamado Coeficiente de Assimetria Gradiente, G_A. O coeficiente é obtido a partir do campo gradiente de uma matriz. Os detalhes do cálculo deste coeficiente, bem como informações sobre o algoritmo utilizado em nosso modelo são fornecidos no Apêndice D.

6.3 Representação Analítica de um Conjunto de Dados Canônicos

6.3.1 Teste do sistema

Para verificar a relação entre a quantidade total de medidas e o tempo de execução do algoritmo para geração do modelo de representação baseado em GNG, foram utilizados três sistemas heterogêneos de dados a partir de séries temporais sintéticas $\pounds^2$ e $\pounds^4$ geradas pela aplicação dos modelos descritos na Seção 6.2. A Figura 6.3 mostra os resultados obtidos.

O primeiro SHD, composto por 10 séries $\pounds^{2,500}$ e 10 séries $\pounds^{4,10,64,64}$, foi analisado em um tempo total de 820 segundos. Para o segundo SHD, composto por 10 séries $\pounds^{2,1000}$ e 10 séries $\pounds^{4,20,64,64}$, o modelo de representação foi gerado em 1102 segundos, enquanto o último SHD, composto por 10 séries $\pounds^{2,10000}$ e 10 séries $\pounds^{4,40,64,64}$, consumiu um tempo de execução de 2240 segundos. É possível observar, portanto, que o desempenho do algoritmo cai à medida que aumentamos a quantidade de medidas tomadas sobre as grades, que interferem no tamanho da entrada do programa *DLattice++*, porém o tempo de execução mais alto observado não inviabiliza a aplicação do modelo, especialmente se considerarmos que, em geral, o sistema computacional

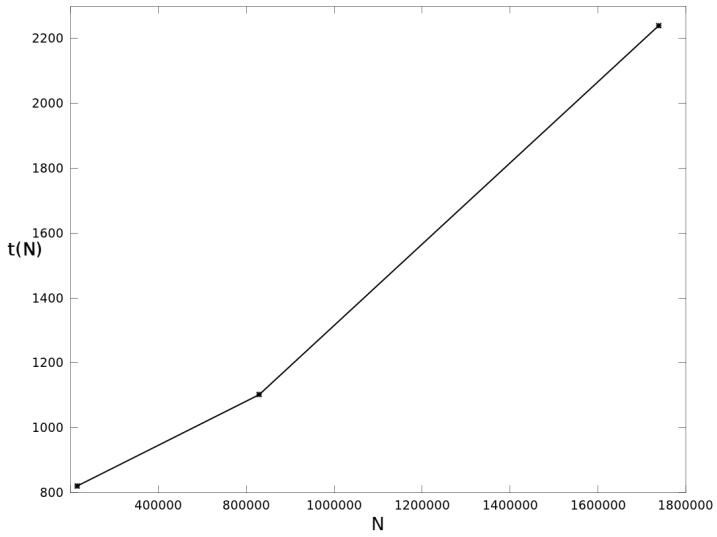


Figura 6.3 - Medida simplificada de desempenho do sistema.

a ser utilizado apresentará configuração superior. O teste descrito foi executado em um computador pessoal com a seguinte configuração: processador Intel®Core™2 Duo 1.66 GHz, HD de 120 GB e 1 GB de memória RAM.

6.3.2 Aplicação do modelo

O SHD formado por dados canônicos utilizado para gerar um exemplo de representação e visualização analítica a partir do modelo desenvolvido neste trabalho contém, como é possível observar na Figura 6.4, três séries temporais com grau variacional $\kappa = 2$, sendo a primeira composta por 10000 pontos, a segunda composta por 50000 pontos, e a terceira composta por 1000 pontos. As GNG correspondentes a estas séries estão representadas na primeira linha do modelo, que traz, na coluna mais à esquerda, o valor de κ para as séries distribuídas em cada linha. Conforme sugerido no Capítulo 5, o valor de κ foi omitido na representação individual de cada GNG.

Na segunda linha, correspondente a $\kappa = 4$, conforme indicado na coluna mais à esquerda, estão representadas as três séries $\pounds^4$ utilizadas, compostas, respectivamente, por sequências de 100, 16 e 20 mapas acoplados, sendo a primeira e a terceira compostas por mapas de 64×64 píxeis, e a segunda, composta por mapas de 128×128 píxeis. O uso do símbolo de grade $\pounds$ é opcional.

2	$L^{10000}_{(1.24)}$	$L^{50000}_{(1.48)}$	$L^{1000}_{(1.02)}$	3
4	$L^{100,64,64}_{(0.0)}$	$L^{16,128,128}_{(0.0)}$	$L^{20,64,64}_{(0.001)}$	3
				6

Figura 6.4 - Exemplo de saída da visualização analítica gerada através do modelo de representação baseado em grades numéricas generalizadas para conjunto de dados canônicos.

2	$L^{5600}_{(1.44)}$	$L^{8736}_{(1.15)}$	$L^{2988}_{(1.32)}$	$L^{5760}_{(1.28)}$	4
4	$L^{3,64,64}_{(0.014)}$	$L^{9,64,64}_{(0.021)}$			2
					6

Figura 6.5 - Saída do algoritmo *DLattice++*: Modelo de representação baseado em grades numéricas generalizadas para conjunto de dados de clima espacial para o evento SAJ6, descrito no Capítulo 2.

6.4 Representação Analítica de Estudos de Caso em Clima Espacial

O modelo de visualização analítica baseado em GNG foi aplicado ao sistema heterogêneo de dados descrito na seção 2.4, considerando apenas os dados de séries temporais dos tipos $\pounds^2$ e $\pounds^4$. O resultado é exibido na Figura 6.5. É possível observar, a partir desta visualização, que os padrões de variabilidade em todas as séries do tipo $\pounds^2$ analisadas são persistentes ($\alpha > 0.5$), enquanto para as séries do tipo $\pounds^4$ os valores de Δ_{G_A} indicam a possibilidade de ter ocorrido processo de relaxação espaçotemporal (ROSA et al., 1999).

Outro exemplo de aplicação para visualização analítica em clima espacial envolve a análise de dados relacionados a uma tempestade geomagnética (KIVELSON; RUSSEL, 1995). Após um período de atividade geomagnética, que pode se estender por mais de um mês, geofísicos espaciais procuram comparar todos os padrões de variabilidade para entender o fenômeno como um todo. Embora não haja ainda uma ferramenta de análise adotada de maneira consensual pela comunidade, a técnica DFA poderia ser usada para, numa primeira análise global, verificar os níveis de persistência ou anti-persistência para cada variável ao final do período de alta atividade geomagnética.

Este estudo poderia ser útil, por exemplo, ao permitir identificar associações entre padrões de variabilidade de algumas variáveis. Para analisar a aplicabilidade do modelo a esse tipo de SHD, selecionamos um conjunto de dados composto por nove séries temporais medidas durante abril e maio de 2000 com resoluções entre 1 e 30 minutos. Este conjunto de dados envolve séries temporais medidas a partir do satélite ACE e também a partir de instrumentos de superfície. Os dados podem ser obtidos a partir do arquivo do SPDF gerenciado pelo Goddard Space Flight Center – NASA (`http://spdf.gsfc.nasa.gov/data_orbits.html`).

Na Figura 6.6 são mostradas todas as séries temporais relacionadas ao período de atividade geomagnética considerado. Tempestades e sub-tempestades são observadas quando ocorrem flutuações significativas nas variáveis medidas.

Na Figura 6.7, é exibido o resultado da aplicação do modelo baseado em GNG sobre o conjunto de dados apresentado na Figura 6.6, representando todo o sistema de dados composto por GNG, neste caso, do tipo $\pounds^2$. É possível observar que os padrões de variabilidade em todas as séries são persistentes ($\alpha > 0.5$), sendo que alguns apresentam o mesmo expoente de escala.

6.5 Generalização das Aplicações

6.5.1 Dados ambientais

As mudanças climáticas e a previsão de eventos extremos relacionados ao meio ambiente têm se apresentado como temas de relevância mundial nas últimas duas décadas. Neste contexto, o Painel Intergovernamental sobre Mudança Climática (IPCC), criado em 1988 pelo Programa das Nações Unidas para o Meio Ambiente e pela Organização Meteorológica Mundial, tem produzido relatórios sobre mudanças climáticas visando a orientar os poderes públicos e privados na questão da interferência humana sobre os processos relacionados às questões ambientais, onde se destacam a produção e a emissão de gases do Efeito Estufa. Um projeto que se destaca no contexto brasileiro, com participação do INPE, é o chamado Projeto Furnas, cujo principal objetivo é investigar o balanço de carbono nos reservatórios de FURNAS Centrais Elétricas S/A.

Pesquisas recentes sobre produção e emissão de gases do Efeito Estufa em reservatórios têm demonstrado que estes sistemas apresentam emissões consideráveis, particularmente de metano (CH_4), gás carbônico (CO_2) e óxido nitroso (N_2O). Este

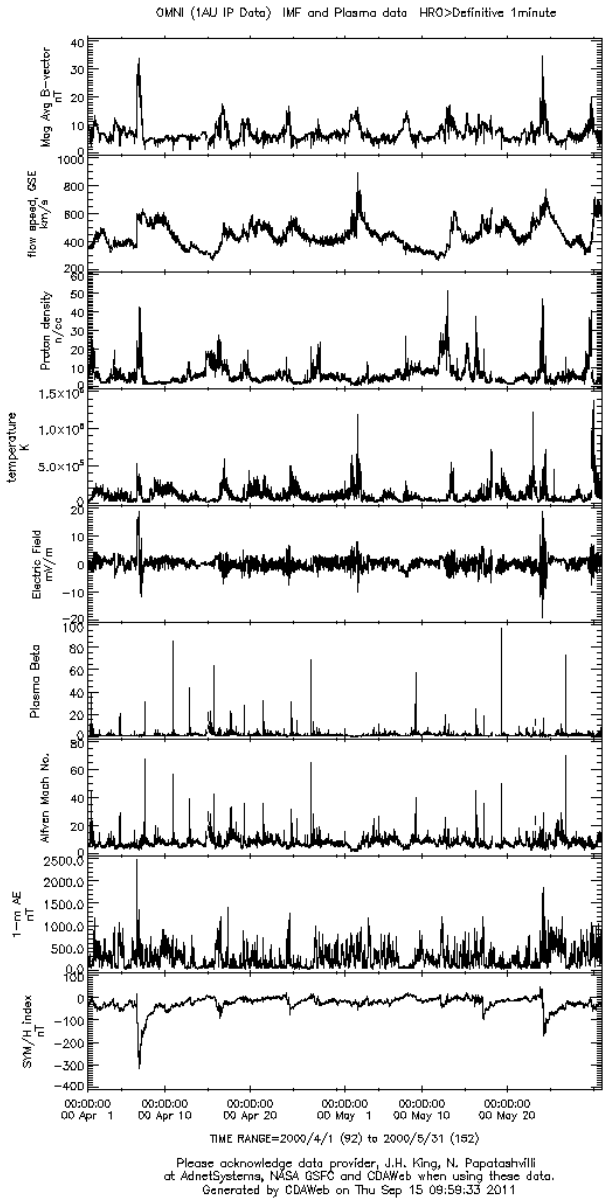


Figura 6.6 - Conjunto de séries temporais $\pounds^2$ para o período de atividade geomagnética de abril a maio de 2000.

2	$L^{10308}_{(1.11)}$	$L^{10093}_{(0.92)}$	$L^{10120}_{(1.13)}$	$L^{2101}_{(0.7)}$	$L^{1818}_{(0.92)}$	$L^{3426}_{(1.3)}$	$L^{1119}_{(1.02)}$	$L^{10120}_{(0.92)}$	$L^{2938}_{(1.29)}$	9
										9

Figura 6.7 - Saída do algoritmo *DLattice++*: Modelo de representação baseado em grades numéricas generalizadas para conjunto de dados geomagnéticos.

projeto está dimensionado para ser desenvolvido em seis anos, período em que serão realizadas medições de cerca de 55 variáveis em todos os 10 reservatórios da empresa (http://www.dsr.inpe.br/projetofurnas/). Trata-se de um SHD cuja modelagem utilizando a representação de dados introduzida nesta tese poderá ser útil para vários usuários do banco de dados do Projeto. Como exemplo, destacamos os vinte e cinco tipos de séries temporais do tipo $\pounds^2$ coletadas pelo SIMA (Sistema de Monitoramento Ambiental) sob responsabilidade do INPE. Este sistema de instrumentos integrados em uma única plataforma (Figura 6.8) realiza medidas no tempo – usualmente amostradas como médias diárias, semanais e mensais – das seguintes variáveis:

(1) Voltagem da bateria da plataforma (V);

(2) Clorofila (mg/l);

(3) Concentração de DO (mg/l);

(4) Porcentagem de DO (%);

(5) Concentração de NH_4 (mg/l);

(6) Concentração de NO_3 (mg/l);

(7) Condutividade (mS/cm);

(8) Direção do vento (^oNV);

(9) Intensidade do vento (m/s);

(10) pH;

(11) Pressão atmosférica (hPa);

(12) Radiação incidente (W/m^2);

(13) Radiação refletida (W/m^2);

(14) Temperatura da água (^{o}C) para $2m$, $5m$, $20m$ e $40m$ de profundidade;

(15) Temperatura da sonda (^{o}C);

(16) Temperatura do ar (^{o}C);

(17) Turbidez (NTU);

(18) Umidade relativa do ar (%);

(19) Velocidade meridional da corrente (cm/s);

(20) Velocidade zonal da corrente (cm/s);

(21) Velocidade meridional do vento (m/s); e

(22) Velocidade zonal do vento (m/s).

Indentifica-se neste exemplo de estudo de caso, portanto, um típico SHD para o qual a aplicação dos conceitos e modelos introduzidos nesta tese poderão ser aplicados sem restrições, uma vez realizado preliminarmente um teste de modelo adequado, conforme sugerido no início deste capítulo.

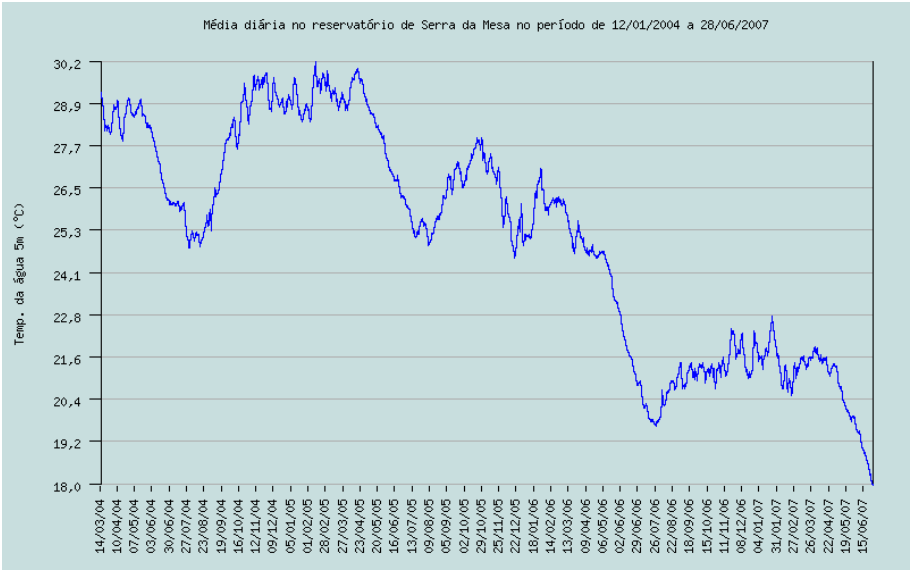
6.5.2 Aplicações em outras abordagens e outros domínios

Uma aplicação que surge como um desdobramento natural do modelo de representação e visualização analítica de GNG é a sua utilização para casos de monitoramento em tempo real de processos complexos, como é o caso do estudo de caso em clima espacial apresentado anteriormente. Entretanto, numa abordagem dinâmica, a representação sistêmica deve ser realizada, por exemplo, através da visualização panorâmica, como discutido na Seção 5.2.3. Neste caso, para um dado período de tempo, a visualização do sistema de GNG é atualizada para que uma análise integrada possa ser realizada tendo como objetivo uma tomada de decisão. Note que, em tal abordagem dinâmica, os planos de representação podem ser visualizados em uma projeção 3D com manipulação digital feita diretamente na tela (Figura 6.9).

Outro fator importante a ser considerado é a aplicação do modelo desenvolvido neste trabalho em outros domínios que não o temporal. Um exemplo que se destaca são os espectros de objetos astronômicos. Em geral, um espectro estelar ou galáctico pode ser interpretado como uma GNG do tipo $\pounds^2$, onde a linha de emissão ou absorção de um determinado elemento químico é identificada no domínio da frequência ou do comprimento de onda (Figura 6.10).

Nesse caso, uma ferramenta analítica a ser adaptada ao programa poderia ser simplesmente a identificação e a extração automática do comprimento de onda para o

(a)



(b)

Figura 6.8 - (a) Sistema de Monitoramento Ambiental (SIMA) instalado em um dos reservatórios de FURNAS. (b) Exemplo de uma $\pounds^2$ coletada pelo SIMA: temperatura da água medida na profundidade de 5 m.
Fonte: http://www.dsr.inpe.br/projetofurnas/

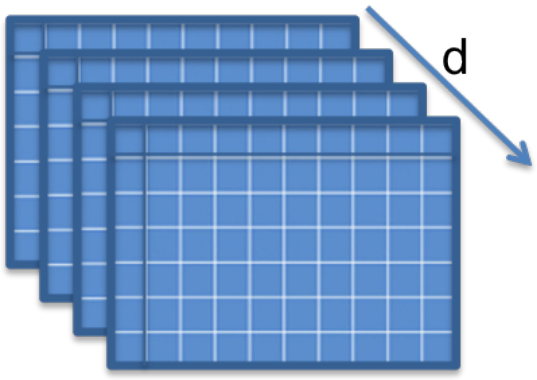


Figura 6.9 - Uma abordagem de projeção 3D permitiria apresentar um conjunto de representações de GNG ao longo de um domínio d, que, para o caso de monitoramento de um processo, pode ser um período de amostragem, dando um caráter dinâmico ao uso do modelo de representação de GNG. No caso do exemplo em astronomia, o domínio d pode ser o *redshift* ou uma classificação de galáxias. Em um exemplo de aplicação médica, cada representação poderia ser um dado paciente classificado por idade, por exemplo.

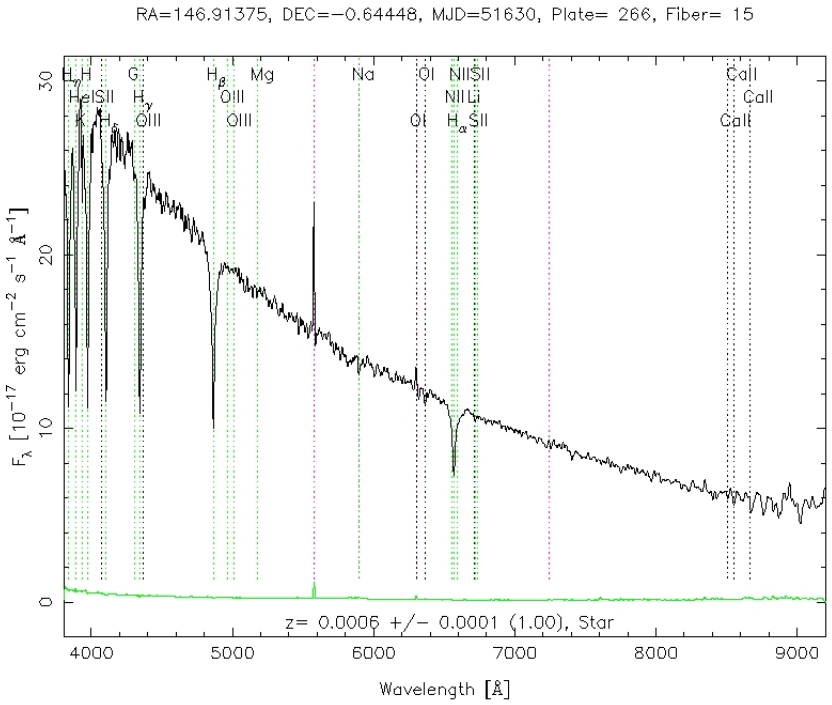


Figura 6.10 - Espectro estelar típico.
Fonte: http://cas.sdss.org/dr4/en/proj/advanced/spectraltypes/

qual um determinado elemento químico tenha sido identificado. Uma projeção 3D de diferentes sistemas de espectros poderia, ainda, ser realizada em função de diferentes galáxias ou distâncias cosmológicas, denominadas "desvios para o vermelho" (*redshift*). Esta aplicação poderia ser testada, por exemplo, dentro do novo paradigma de Observatórios Virtuais (CARVALHO et al., 2010).

7 CONSIDERAÇÕES FINAIS

"What we need is not more information but the ability to present the right information to the right people at the right time in the most effective and efficient form."

(Robert E. Horn)

A pesquisa desenvolvida neste projeto de doutorado em Computação Aplicada foi motivada por dois fatores principais: (i) a necessidade de um novo formalismo capaz de generalizar dados de séries temporais, muito comuns no estudo experimental e observacional de processos não lineares encontrados na natureza; e (ii) a possibilidade de aplicar o paradigma da Analítica Visual sobre um modelo que represente de forma integrada um conjunto de séries temporais já analisadas. Assim, no estudo de um protótipo teórico e computacional, as rotinas para cálculo de medidas analíticas foram incorporadas ao módulo de análise do modelo, denominado *DLattice++*. Resultados preliminares desta nova abordagem foram discutidos no contexto de duas aplicações que estão entre as principais áreas de pesquisa do INPE: a física espacial e a física ambiental. Entre os principais resultados da pesquisa, podemos destacar, além da formalização de uma nova ideia sobre o tratamento de conjuntos de séries temporais, introduzindo o novo conceito de *grade numérica generalizada*, a implementação do método DFA no módulo de análise para $\pounds^2$, respeitando um estudo comparativo com outras técnicas. Além disso, um novo tipo de simulador de sequências de imagens (grades do tipo $\pounds^4$) foi proposto e testado, integrando pela primeira vez um mapa acoplado a sinais do tipo ruído *Browniano*. Este resultado secundário poderá ser explorado e aprimorado dentro de um novo projeto de doutorado dentro do grupo de pesquisa em computação científica do LAC.

Uma tarefa futura, com enfoque puramente matemático e por isso não associada ao escopo desta tese, diz respeito à necessidade de aprimorar, dentro do rigor da linguagem matemática, a definição formal para o conceito de grade numérica generalizada. Nesse sentido, parece mais apropriado tratar o tema dentro da Teoria das Categorias (HERRLICH; STRECKER, 2007), de forma que um matemático possa construir uma teoria rigorosa para a generalização proposta.

Do ponto de vista da inovação, o projeto resultou na disponibilização do código

fonte desenvolvido para geração do modelo de representação baseado em GNGs. Esta implementação poderá ser aprimorada e incorporada ao Programa de Estudo e Monitoramento do Clima Espacial do INPE, cuja fase operacional será estabelecida a partir do próximo ano. Uma vez escolhidas as ferramentas de análise mais apropriadas, o modelo em funcionamento permitirá representar um conjunto de dados para o monitoramento da atividade solar observada em diferentes frequências e resoluções. No módulo de análise, o modelo já traz, para a representação de $\pounds^2$, a técnica DFA, validada para as aplicações em Clima Espacial conforme publicado por (VERONESE et al., 2011b).

Como trabalhos futuros que podem ser desenvolvidos a partir desta pesquisa de doutorado, destacam-se os seguintes:

a) Testar e sugerir ferramentas para séries $\pounds^4$ e $\pounds^5$ em diferentes contextos de aplicações. Embora não haja ainda consenso sobre ferramentas e metodologias para grades do tipo $\pounds^5$, várias técnicas para $\pounds^4$, além da GPA, incorporada no presente modelo, já podem ser testadas. Neste contexto, podemos citar, por exemplo, correlações matriciais do tipo Moran/Geary (GEARY, 1954), expoentes de Hurst (BARABáSI; STANLEY, 1995), funcionais de Minkowisky (MECKE; STOYAN, 2000), etc. Entretanto, para todas as técnicas, o resultado corresponde, em geral, a uma função ou valor analítico associado a cada imagem. Portanto, é essencial definir, para cada técnica, qual a medida de segunda ordem (equivalente ao fator de relaxação apresentado para aplicação da análise de padrões gradientes) a ser calculada sobre todas aquelas extraídas de cada $\pounds^4$. No caso do Clima Espacial, espera-se que esta seja uma tarefa realizada dentro do Programa EMBRACE.

b) Pesquisar ferramentas e recursos computacionais para o desenvolvimento de um módulo para projeção 3D, em domínios de interesse, que permita a visualização dinâmica e panorâmica de vários sistemas de GNGs.

c) Otimizar o processo computacional de análise dos dados, bem como a integração entre os módulos de análise e visualização, incorporando o programa *DLattice++* maior conveniência para o usuário e flexibilidade na incorporação de novas ferramentas.

d) Explorar através de outras técnicas de análise o modelo introduzido neste

trabalho para geração de séries espaçotemporais do tipo *mapa acoplado Browniano*, verificando seu potencial para validação e testes sobre outras aplicações.

Espera-se que os desafios propostos acima, bem como a aplicação do sistema a outros programas de monitoramento do INPE, como, por exemplo, o Projeto FURNAS, sejam associados a bolsas de desenvolvimento.

No contexto da comunidade internacional, as novidades apresentadas nesta tese e publicadas nos artigos Veronese et al. (2009) e Veronese et al. (2011b) têm despertado interesse nos programas mais atuais para pesquisa do clima espacial, em grande parte devido ao volume de dados gerados pelas missões espaciais e ao novo desafio de observação e análise integrada da atividade solar para fins de monitoramento.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDREWS, D. F.; HERZBERG, A. M. **Data: a collection of problems from many fields for the student and research worker**. Nova Iorque: Springer-Verlag, 1985. 7

ASSIREU, A. T.; ROSA, R. R.; VIJAYKUMAR, N. L.; LORENZETTI, J. A.; REMPEL, E.; RAMOS, F. M.; Sá, L. D. A.; BOLZAN, M. J. A.; ZANANDREA, A. Gradient pattern analysis of short nonstationary time series: an application. **Physica D**, v. 168, n. 1, p. 397–403, 2002. 89

BAR-YAN, Y. **Dynamics of complex systems**: studies in nonlinearity. Reading, Massachusetts: Perseus Books, 1997. 2

BARABáSI, A.-L.; STANLEY, H. E. **Fractal concepts in surface growth**. Grã Bretanha: Cambridge University Press, 1995. 62

BASHAN, A.; BARTSCH, R.; KANTELHARDT, J. W.; HAVLIN, S. Comparison of detrending methods for fluctuation analysis. **Physica A**, v. 387, p. 5080–5090, 2008. 85

BLOECHLE, J. luc; RIGAMONTI, M.; HADJAR, K.; LALANNE, D.; INGOLD, R. Xcdf: a canonical and structured document format. In: INTERNATIONAL ASSOCIATION FOR PATTERN RECOGNITION (IAPR) WORKSHOP ON DOCUMENTO ANALYSIS SYSTEMS (DAS2006), 2006, Nelson, New Zealand. **Proceedings...** Nelson: Springer, 2006. p. 141–152. Springer Lecture Notes in Computer Science. 45

BLYTHE, J.; PATWARDHAN, M.; OATES, T.; DESJARDINS, M.; RHEINGANS, P. Visualization support for fusing relational, spatio-temporal data: Building career histories. In: INTERNATIONAL CONFERENCE ON INFORMATION FUSION, 2006, Florença, Itália. **Proceedings...** Florença: IEEE, 2006. p. 1–7. 77

BOLZAN, M. J. A.; ROSA, R. R.; SAHAI, Y. Multifractal analysis of low-latitude geomagnetic fluctuations. **Annales Geophysicae**, Copernicus GmbH, Paris, v. 27, p. 569–576, 2009. 22

BROCKWELL, P. J.; DAVIS, R. A. **Time series: theory and methods**. 2. ed. Nova Iorque: Springer-Verlag, 1991. 6

BURNHAM, K. P.; R., A. D. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. Nova Iorque: Springer, 2002. 22

CAI, Y.; STUMPF, R.; WYNNE, T.; TOMLINSON, M.; CHUNG, S. H.; BOUTONNIER, X.; IHMIG, M.; FRANCO, R.; BAUERNFEIND, N. Visual transformation for interactive spatio-temporal data mining. **Journal of Knowledge and Information Systems**, v. 11, n. 5, 2007. 78

CAO, Y.; AMES, D. P.; YANG, P. Towards virtual watersheds: integrated data mining, management, mapping and modeling. In: AWRA 2010 SPRING SPECIALTY CONFERENCE, 2010, Orlando, USA. **Proceedings...** Orlando, 2010. p. 1–5. 32

CARVALHO, R. R. de; GAL, R. R.; Campos Velho, H. F. de; CAPELATO, H. V.; BARBERA, F. L.; VASCONCELLOS, E. C.; RUIZ, R. S. R.; KOHL-MOREIRA, J. L.; LOPES, P. A.; SOARES-SANTOS, M. The brazilian virtual observatory - a new paradigm for astronomy. **Journal of Computational Interdisciplinary Sciences**, v. 1, n. 3, p. 187–206, 2010. 59

CASTLE, C. J. E.; CROOKS, A. T. Principles and concepts of agent-based modelling for developing geospatial simulations. **UCL Working Paper Series**, 2006. 8

CHEN, Y.; YANG, J.; RIBARSKY, W. Toward effective insight management in visual analytics systems. In: PACIFIC VISUALIZATION SYMPOSIUM (PACIFICVIS), 2009, Beijing, China. **Proceedings...** Beijing: IEEE, 2009. p. 49–56. 33

CHINCHOR, N. A.; THOMAS, J. J.; WONG, P. C.; CHRISTEL, M. G.; RIBARSKY, W. Multimedia analysis + visual analytics = multimedia analytics. **IEEE Computer Society**, p. 52–60, 2010. 31, 33

CLADIS, P. E.; PALFFY-MUHORAY, P. **Spatio-Temporal Patterns in Nonequilibrium Systems**. Reading, Massachusetts: Addison-Wesley, 1995. 2

COHEN, J.; HOLLIS, M. **NCCS Support of Space Exploration: Improving Space Weather Modeling Required for Interplanetary Travel**. 2011. Disponível em: <http://science.gsfc.nasa.gov/606/docs/newsletter/cisto_news.winter-spring08.pdf>. Acesso em: 27 de setembro de 2011. 15, 16

de Carvalho, A. C. P. de L. F.; BRAYNER, A.; LOUREIRO, A.; FURTADO, A. L.; von Staa, A.; de Lucena, C. J. P. et al. **Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016**. São Paulo: SBC, 2006. 1

de Pinho, R. D. **Espaço incremental para a mineração visual de conjuntos dinâmicos de documentos**. Tese (Doutorado) — ICMC - USP, São Carlos, Abril 2009. 31

DENIZET, C.; LEDRU, S. **JLaTeXMath – A Java API to render LaTeX**. novembro 2011. Disponível em: <http://forge.scilab.org/index.php/p/jlatexmath/>. 41

DUNN, P. F. **Measurement and Data Analysis for Engineering and Science**. Nova Iorque: McGraw-Hill, 2005. 22

EARNSHAW, R. A.; WISEMAN, N. **An introductory guide to scientific visualization**. Berlin: Springer-Verlag, 1992. 29

EISENKOLB, A.; SCHILL, K.; RÖHRBEIN, F.; BAIER, V.; MUSTO, A.; BRAUER, W. Visual processing and representation of spatio-temporal patterns. In: et al., C. F. (Ed.). **Proceedings...** Berlin: Springer-Verlag, 2000. p. 145–156. 76

FAYYAD, U.; HAUSSLER, D.; STOLORZ, P. Mining scientific data. **Communications of the ACM**, v. 39, n. 11, p. 51–57, 1996. 29

FRY, B. J. **Computational Information Design**. Tese (Doutorado) — Massachusetts Institute of Technology, 2004. 78

GABER, M. M. (Ed.). **Scientific data mining and knowledge discovery**. Berlin: Springer-Verlag, 2010. 29

GEARY, R. C. The contiguity ratio and statistical mapping. **The Incorporated Statistician**, v. 5, n. 3, p. 115–145, 1954. 62

GERSHENFELD, N. **The nature of mathematical modeling**. Cambridge: Cambridge University Press, 2000. 82

GOLDFARB, L.; GAY, D.; GOLUBITSKY, O.; KORKIN, D.; SCRIMGER, I. **What is a structural representation?** Fredericton, Canada: [s.n.], June 2008. 77

GRAVES, S.; RAMACHANDRAN, R.; LYNNES, C.; MASKEY, M.; KEISER, K.; PHAM, L. Mining scientific data using the internet as the computer. In: IEEE INTERNATIONAL GEOSCIENCE & REMOTE SENSING SYMPOSIUM, 2008, Boston, USA. **Proceedings...** Boston: IEEE, 2008. 1, 8

GREEN, T. M.; RIBARSKY, W.; FISHER, B. Visual analytics for complex concepts using a human cognition model. In: IEEE SYMPOSIUM ON VISUAL ANALYTICS SCIENCE AND TECHNOLOGY, 2008, Columbus, USA. **Proceedings...** Columbus: IEEE, 2008. p. 91–98. 33

GRÉGIO, A. R. A.; FILHO, B. P. de C.; MONTES, A.; SANTOS, R. In: _____. **Minicursos SBSEG 2009**. Brasília: Sociedade Brasileira de Computação, 2009. cap. Técnicas de Visualização de Dados aplicadas à Segurança da Informação, p. 195–236. 31

GUERRA, J. de M. **Caractrizaćão fina dos padrões de variabilidade do ECG para validaćão de modelos e aplicaćões em microgravidade**. 143 p. Dissertaćão de Mestrado — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2009. 7

HANISCH, R. J.; FARRIS, A.; GREISEN, E. W.; PENCE, W. D.; SCHLESINGER, B. M.; TEUBEN, P. J.; THOMPSON, R. W.; Warnock III, A. **Definition of the Flexible Image Transport System (FITS)**. San Francisco, 2001. 76

HANSLMEIER, A. **The Sun and Space Weather**. 2. ed. Dordrecht, The Netherlands: Springer, 2007. 32

HEIJMANS, J. **An Introduction to Distributed Visualization**. [S.l.: s.n.], february 2002. 31

HERRLICH, H.; STRECKER, G. **Category Theory: an introduction**. Berlim: Heldermann Verlag, 2007. 3rd ed. 61

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: Domain-analysis. **Journal of the American Society for Information Science**, v. 46, n. 6, p. 400–425, 1995. 30

HOUAISS, A.; VILLAR, M. de S. **Dicionário Houaiss da língua portuguesa**. 1. ed. Rio de Janeiro: Objetiva, 2001. 8, 32

ISTP. **The NASA International Solar-Terrestrial Physics**. 2011. Disponível em: <http://istp.gsfc.nasa.gov/istp/events/2000jun6/>. Acesso em: 27 de setembro de 2011. 12

JACOBSON, R. **Information Design**. Cambridge: MIT Press, 1999. 3, 78

JÄHNE, B. **Spatio-Temporal Image Processing: Theory and Scientific Applications**. Berlin: Springer, 1993. (Lecture Notes in Computer Science, v. 751). ISBN 3-540-57418-2. 12

JIRICKA, K.; KARLICKY, M.; KEPKA, O.; TLAMICHA, A. Fast drift burst observations with the new ondrejov radiospectrograph. **Solar Physics**, v. 147, p. 203–208, 1993. 13

IV'06. **Reviewing Data Visualization: an Analytical Taxonomical Study**. 31

KANTZ, H.; SCHREIBER, T. **Nonlinear time series analysis**. 2. ed. Nova Iorque: Cambridge University Press, 2003. 296 p. 8, 22

KAY, S. M.; MARPLE, S. L. Spectrum analysis - a modern perspective. **Proceedings of the IEEE**, v. 69, p. 1380–1419, 1981. 86, 87

KEIM, D. A. **Databases and Visualization**. Montreal, 1996. Disponível em: <http://www.cip.ifi.lmu.de/~keim/>. 29

KEIM, D. A.; MANSMANN, F.; SCHNEIDEWIND, J.; ZIEGLER, H. Challenges in visual data analysis. In: INFORMATION VISUALIZATION (IV'06), 2006, Londres, Inglaterra. **Proceedings...** Londres: IEEE, 2006. 3, 29, 31

KEIM, D. A.; SIPS, M.; ANKERST, M. Visual data mining techniques. In: _____. Oxford: Elsevier, 2005. p. 831–843. ISBN 0-12-387582-X. 29

KENDALL, M. G.; ORD, J. K. **Time Series**. 3. ed. Nova Iorque: Oxford University Press, 1990. 6

KESHNER, M. S. 1/f noise. **Proceedings of the IEEE**, v. 70, n. 3, p. 212–218, 1982. 22

KIVELSON, M. G.; RUSSEL, C. T. **Introduction to Space Physics**. Nova Iorque: Cambridge University Press, 1995. 52

LAROSE, D. T. **Data mining methods and models**. Nova Jersey: John Wiley & Sons, 2006. 30

LAROUSSE DO BRASIL. **Dicionário enciclopédico ilustrado Larousse**. São Paulo, 2007. 19

LAST, M.; KANDEL, A.; BUNKE, H. (Ed.). **Data mining in time series databases**. Singapore: World Scientific, 2004. 30

MECKE, K. R.; STOYAN, D. (Ed.). **Statistical Physics and Spatial Statistics**: The art of analyzing and modeling spatial structures and pattern formation. Berlim: Springer-Verlag, 2000. 8, 23, 62

MORETTIN, P. A.; TOLOI, C. M. C. **Séries temporais**. 2. ed. São Paulo: Atual, 1987. 6

NOAA. **Oil Spill Case Histories 1967 through 1991**. Seattle: NOAA, 1992. Disponível em: <http://response.restoration.noaa.gov/book_shelf/26_spilldb.pdf>. 11

DOYLE, A.; REED, C. (Ed.). **Introduction to OGC Web Services**. Wayland, maio 2001. 31

OSBORNE, A.; PROVENZALE, A. Finite correlation dimension for stochastic systems with power-law spectra. **Physica D**, n. 35, p. 357–381, 1989. 46

PAULOVICH, F. V. **Mapeamento de dados multi-dimensionais - integrando mineração e visualização**. Tese (Doutorado) — ICMC - USP, São Carlos, 2008. 31

PEITGEN, H.-O.; JÜRGENS, H.; SAUPE, D. **Chaos and Fractals: New Frontiers of Science**. Nova Iorque: Springer, 2004. 22

PENG, C.-K.; BULDYREV, S. V.; HAVLIN, S.; SIMONS, M.; STANLEY, H. E.; GOLDBERGER, A. L. Mosaic organization of dna nucleotides. **Physical Review E**, v. 49, n. 2, p. 1685–1689, 1994. 24, 49, 85

PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P. **Numerical Recipes: The Art of Scientific Computing**. Nova Iorque: Cambridge University Press, 2007. 41

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**, v. 283, p. 171–174, 2000. 47

ROBERTSON, P. K. A methodology for choosing data representations. **IEEE Computer Graphics and Applications**, 1991. 29, 31

RODDICK, J. F.; SPILIOPOULOU, M. A bibliography of temporal, spatial and spatio-temporal data mining research. **SIGKDD Explorations**, v. 1, n. 1, p. 34–38, 1999. 5, 8

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. 23

ROSA, R. R.; CAMPOS, M. R.; RAMOS, F. M.; FUJIWARA, S.; SATO, T. Gradient pattern analysis of structural dynamics: application to molecular system relaxation. **Brazilian Journal of Physics**, v. 33, p. 605–609, 2003. 10, 23, 24, 50

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 solar radio bursts. **Advances in Space Research**, v. 42, p. 844–851, 2008. 11

ROSA, R. R.; RAMOS, F. M. Análise de padrões gradientes e a física estatística da formaćão de estruturas espaćo-temporais. In: _____. **Tendências da física estatística no Brasil**. São Paulo: Livraria da Física, 2003. 9

ROSA, R. R.; REMPEL, E. L.; STRIEDER, C.; FREITAS, R. M. Pattern analysis of 3d turbulence. 2011. Em preparação. 10

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. 23, 52

SANTOS, R. Introdução à mineração de dados com aplicações em ciências ambientais e espaciais. In: _____. **Computação e Matemática Aplicadas às Ciências e Tecnologias Espaciais**. São José dos Campos: MCT/INPE, 2008. cap. 1, p. 15–38. ISBN 978-85-17-00037-9. 1

SAWANT, H. S.; GOPALSWAMY, N.; ROSA, R. R.; SYCH, R. A.; ANFINOGENTOV, S. A.; FERNANDES, F. C. R.; CECATTO, J. R.; COSTA, J. E. R. The brazilian decimetric array and space weather. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 73, n. 11-12, p. 1300–1310, 2011. 26

SILVA, A. F. da; ROSA, R. R.; ROMAN, L. S.; VEJE, E.; PEPE, I. Characterization of asymmetric fragmentation patterns in sfm images of porous silicon. **Solid State Communication**, v. 113, n. 12, p. 703–708, 2000. 10, 11

(SPDF), S. P. D. F. **CDF Frequently Asked Questions**. junho 2011. Disponível em: <http://cdf.gsfc.nasa.gov/html/FAQ.html>. 76

SPIDR. **Space Physics Interactive Data Resource**. 2011. Disponível em: <http://spidr.ngdc.noaa.gov/spidr/home.do>. Acesso em: 27 de setembro de 2011. 12

STROUD, J. R.; MÜLLER, P.; SANSÓ, B. Dynamic models for spatiotemporal data. **J. R. Statist. Soc. B**, v. 63, n. 4, p. 673–689, 2001. 12

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009. ISBN 978-85-7393-761-9. 7, 29, 30, 32

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 2. ed. San Diego, CA: Elsevier, 2003. 30

THOMAS, J. J.; COOK, K. A. A visual analytics agenda. **IEEE Computer Society**, p. 10–13, 2006. 31

UMBANHOWAR, P. B.; MELO, F.; SWINNEY, H. L. Localized excitations in a vertically vibrated granular layer. **Nature**, v. 382, p. 793–796, 1996. 9

VERHEIN, F. Mining complex spatio-temporal sequence patterns. **SIAM**, p. 605–616, 2009. 8, 78

VERONESE, T. B.; BOLZAN, M. J. A.; ROSA, R. R. 2011. Em preparação. 46

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. 25, 49, 62, 63, 87

VERONESE, T. B.; ROSA, R. R.; VIJAYKUMAR, N. L.; BOLZAN, M. J. A. Generalized numerical lattices for time series representation in complex data systems. **Journal of Computational Interdisciplinary Sciences**, v. 1, n. 2, 2009. ISSN 1983-8409. 19, 34, 35, 63

WEHREND, S.; LEWIS, C. A problem-oriented classification of visualization techniques. 1990. 8

WELLS, D. C.; GREISEN, E. W.; HARTEN, R. H. Fits: A flexible image transport system. **Astronomy and Astrophysics Supplement Series**, v. 44, p. 363–370, 1981. 76

WHITE, J. V. **Edição e Design: para designers, diretores de arte e editores**. São Paulo: JSN Editora, 2006. ISBN 85-85985-17-8. 19

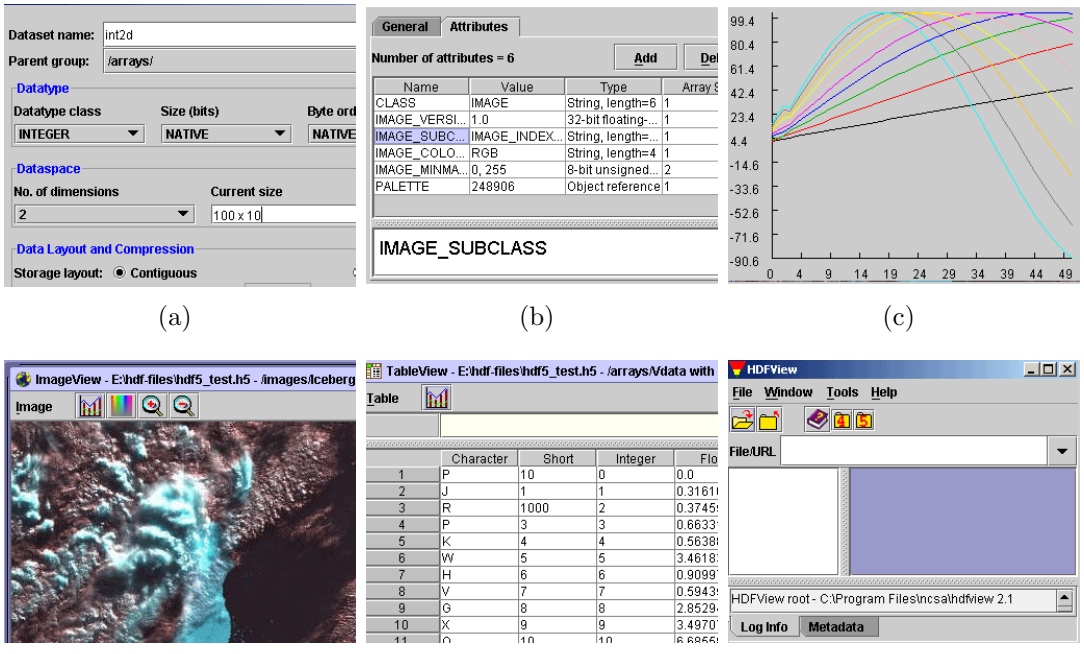
WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2. ed. San Francisco, CA: Morgan Kaufmann Publishers, 2005. ISBN 0-12-088407-0. 6, 30

WONG, P. C.; THOMAS, J. Visual analytics. **IEEE Computer Society**, p. 20–21, 2004. 33

ZHIZHIN, M.; KIHN, E.; REDMON, R.; MEDVEDEV, D.; MISHIN, D. Space physics interactive data resource—spidr. **Earth Sci Inform**, p. 79–91, 2008. 32

APÊNDICE A - VISUALIZAÇÃO DE DADOS CIENTÍFICOS

HDF (*Hierarchical Data Format*) denomina o conjunto de formatos de dados e bibliotecas projetados para armazenar e organizar grandes quantidades de dados numéricos. Originalmente desenvolvido pelo Centro Nacional para Aplicações de Supercomputação (NCSA), da Universidade de Illinois, e atualmente mantido pelo Grupo HDF (*The HDF Group*), o HDF é reconhecido por muitas plataformas de software comerciais e não comerciais, incluindo Java, Matlab, IDL e Python. A versão mais recente, HDF5, organiza os arquivos em uma estrutura hierárquica, com duas estruturas primárias: grupo (*groups*), contendo instâncias de zero ou mais grupos ou conjuntos de dados e suporte a metadados; e conjunto de dados (*datasets*), uma matriz multidimensional de elementos de dados com suporte a metadados. Grupos e membros de grupos funcionam em grande parte de maneira similar a diretórios e arquivos em UNIX. A Figura A.1 ilustra um exemplo de arquivo em formato HDF e suas visualizações geradas pela ferramenta HDFView. Informações detalhadas sobre o formato HDF podem ser encontradas na página do Grupo HDF (http://www.hdfgroup.org).



(a) (b) (c)

(d) (e) (f)

Figura A.1 - Exemplo de visualização de dados no formato HDF.
Fonte: http://www.hdfgroup.org/hdf-java-html/hdfview/

CDF (*Common Data Format*) é uma abstração conceitual de dados para armazenamento, manipulação e acesso a conjuntos de dados multidimensionais desenvolvida pela NASA. O componente básico do CDF é uma interface de programação que fornece ao usuário controle completo sobre como os valores de dados são armazenados no CDF ((SPDF), 2011). A Figura A.2 ilustra um exemplo de visualização obtida pelo uso da ferramenta CWIT (*CDF Windows Imaging Tool*), disponibilizada pelo projeto SPDF (*Space Physics Data Facility*) da NASA.

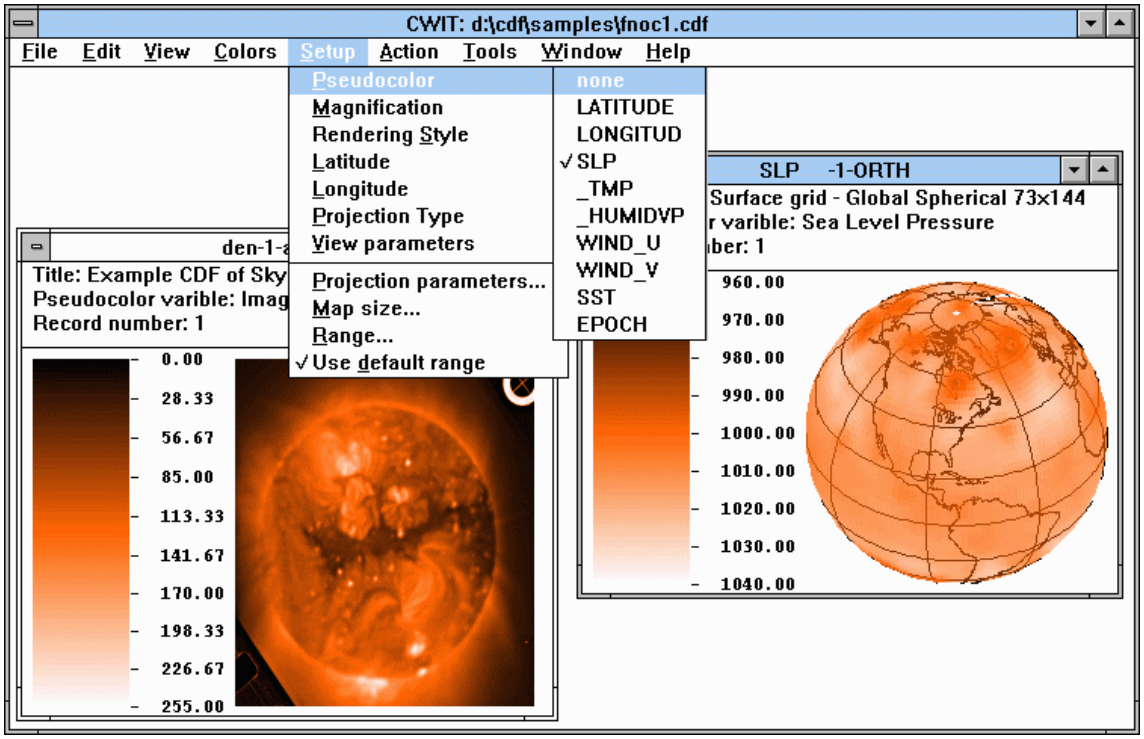


Figura A.2 - Exemplo de representação de dados no formato CDF.
Fonte: `http://cdf.gsfc.nasa.gov/`

O formato de arquivo FITS (Flexible Image Transport System) é amplamente utilizado em astronomia para o armazenamento e troca de imagens astronômicas, oferecendo um número ilimitado de parâmetros que podem ser associados a essas imagens (HANISCH et al., 2001; WELLS et al., 1981).

Em Eisenkolb et al. (2000), são conduzidos experimentos para investigar a representação e o processamento de informações espaçotemporais através de um enfoque não convencional, enfatizando, ao invés de algoritmos visuais, as habilidades cognitivas

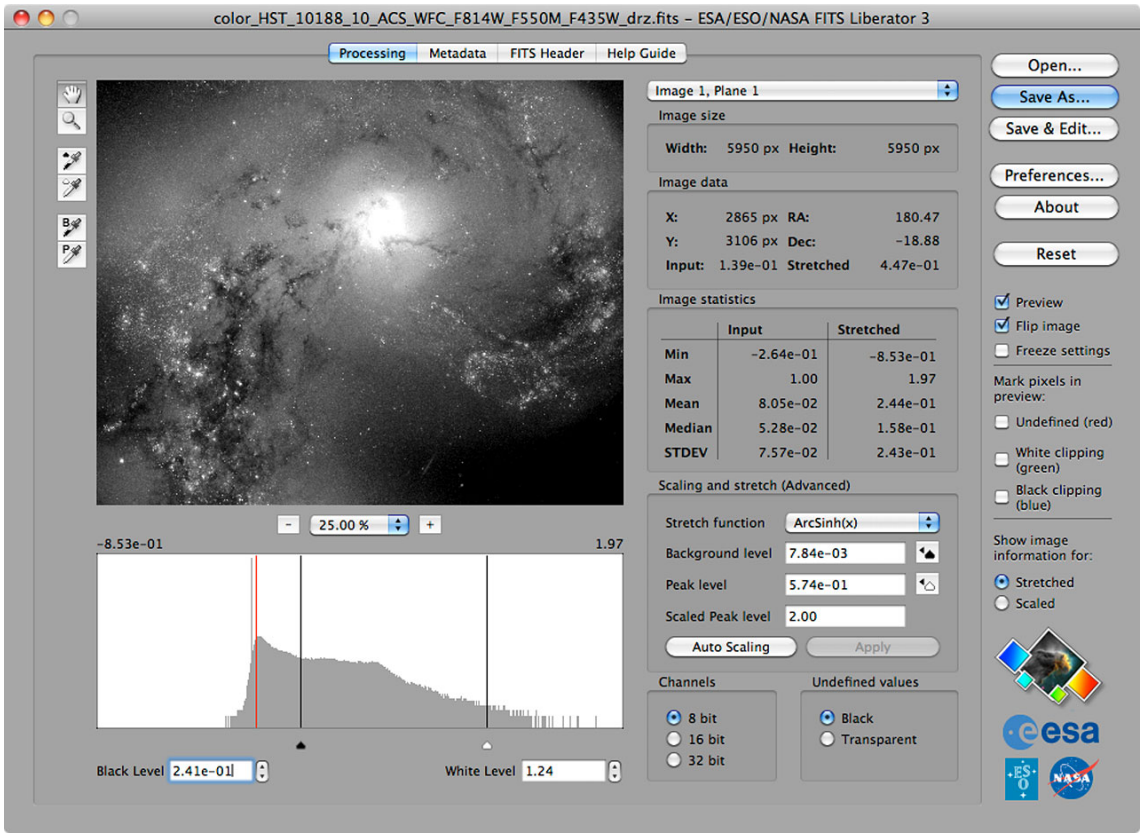


Figura A.3 - Exemplo de representação de dados no formato FITS.
Fonte: http://www.spacetelescope.org/images/ann1013a/

baseadas em memória, como o reconhecimento de gestos. Os autores utilizam os parâmetros de saída destes experimentos para treinar um modelo de rede neural artificial e, alternativamente, para determinar faixas de descrições simbólicas capazes de gerar uma interface de usuário similar às condições da visão humana.

Blythe et al. (2006) apresentam dois tipos de visualização baseadas em metáforas físicas que facilitam a fusão, a análise e o entendimento profundo de dados relacionais espaçotemporais. A primeira visualização baseia-se na metáfora do fluxo de um fluido através de tubos elásticos, e a segunda baseia-se na propagação de onda. As visualizações são discutidas no contexto de fundir informações sobre as atividades desempenhadas por cientistas ao longo do tempo com o objetivo de construir seus históricos profissionais. Um formalismo baseado em eventos temporais, denominado ETS (*Evolving Transformation System*, é descrito em Goldfarb et al. (2008). O objetivo deste formalismo é proporcionar uma representação estrutural (simbólica),

destacando sua relevância em áreas como inteligência artificial e reconhecimento de padrões, bem como nas ciências naturais, especialmente a Biologia. Cai et al. (2007) realizam um estudo de caso em oceanografia para validar um mapeamento de dados de um espaço abstrato para um espaço intuitivo, obtendo maior robustez em rastreamento de objetos e na precisão da predição em detecção de objetos, além da diminuição do tempo gasto pelo usuário para processar imagens em relação à análise manual. Segundo os autores, estes resultados sugerem que interações humanas mínimas associadas a transformações computacionais adequadas podem aumentar significativamente a produtividade geral.

Verhein (2009) utiliza o conceito de Regras de Associação Espaçotemporal (STARs – *Spatio-Temporal Association Rules*) para descrever o deslocamento de objetos entre regiões ao longo do tempo. Baseada em grafos não direcionados, uma representação denominada Grade k-STAR (*k-STAR Lattice*) descreve movimentos sequenciais de grupos de objetos. O método é validado por dois experimentos: o primeiro, realizado sobre um conjunto de dados sintéticos para análise de desempenho e escalabilidade e verificação do comportamento em ambiente com ruído; e o segundo, um conjunto de dados reais de rastreamento de animais usado para validar a motivação e demonstrar a utilidade do padrão proposto.

O trabalho de Fry (2004), criador do ambiente *Processing*[1], trouxe para o contexto das ciências computacionais a arquitetura de informação, disciplina definida por Jacobson (1999) como a arte e a ciência de preparar a informação para ser usada mais eficiente e efetivamente por seres humanos. A Figura A.4 mostra um exemplo de visualização, utilizando *Processing*, de dados de ligações IP e telefônicas entre Nova Iorque e cidades de todo o mundo.

[1] O ambiente *Processing* (http://processing.org) foi projetado especificamente para simplificar a construção de softwares voltados à visualização de informações provenientes de sistemas nos quais a grande quantidade de dados torna difícil obter um quadro geral para a compreensão do seu significado.

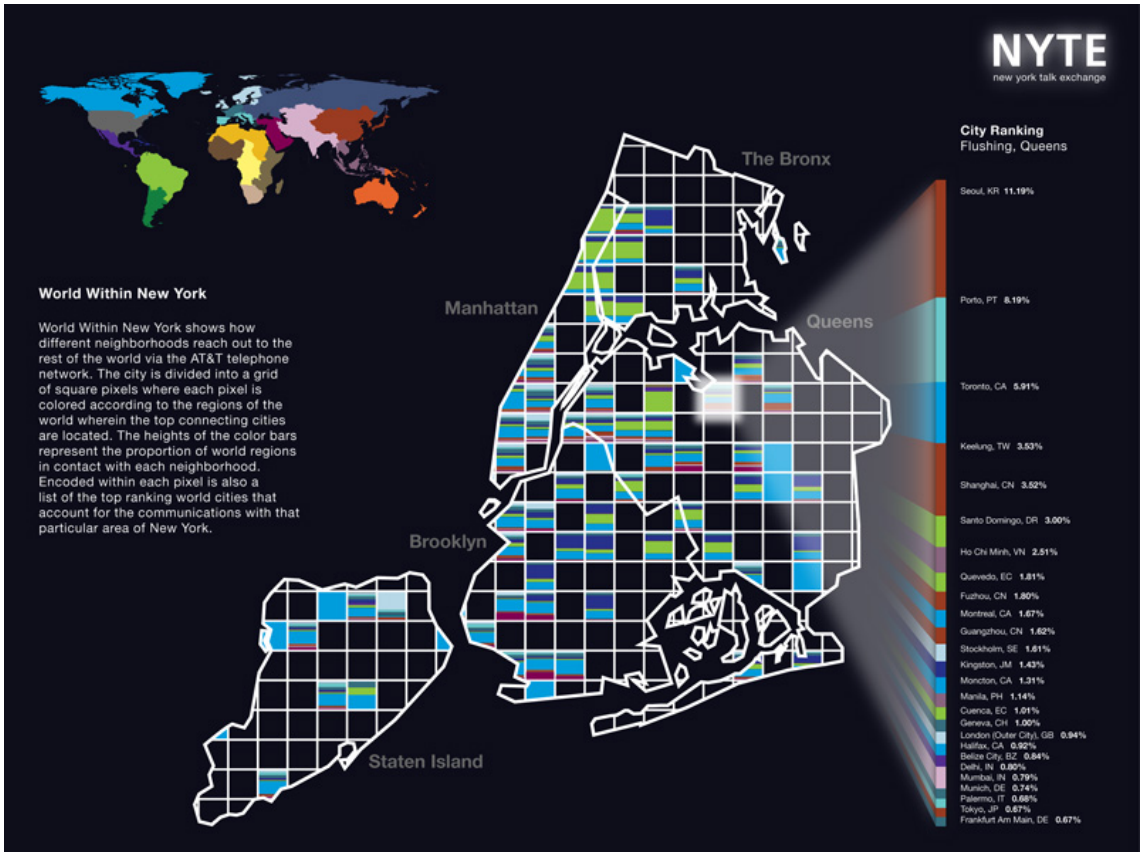


Figura A.4 - Exemplo de visualização utilizando o ambiente *Processing*.
Fonte: NYTE (http://senseable.mit.edu/nyte/)

APÊNDICE B - O RUÍDO *BROWNIANO*

O termo "Ruido Browniano" é proveniente do processo estocástico conhecido como Movimento Browniano, que descreve a dinâmica de um grão de pólen em movimento numa superfície líquida, observada em 1827 pelo físico e biólogo escocês Robert Brown. Albert Einstein, em 1905, propôs corretamente que a flutuação observada se deve à interação entre a partícula (o pólen) e as moléculas do líquido onde o mesmo se desloca. Além de sua relação direta com o fenômeno de difusão de partículas na física, em biologia o *Brownianoe*stá diretamente relacionado ao processo de difusão responsável pela formação de proteínas, à síntese de ATP e ao transporte intracelular de moléculas. Num contexto mais geral, o ruído *Brownianoe*tem interesse prático pois sua caracterização pode explicar a componente aleatória que impõe limitações à exatidão das medidas instrumentais.

O ruído *Brownianoe*apresenta espectro de frequências (ou potências[1].) onde a densidade espectral das escalas (ou potências) é inversamente proporcional à frequência ν do sinal $A(t_i)$ de acordo com a seguinte lei de potência:

$$S(\nu) \propto 1/\nu^2. \tag{B.1}$$

Na teoria da análise de sinais estocásticos, o termo generalizado do ruído $1/f^\beta$ descreve um conjunto de sinais cujo espectro de potência é dado por $S(f) \propto 1/f^\beta$. O caso fundamental, quando $\beta = 0$, é denominado "ruído branco", para o qual o espectro de potência é constante:

$$S(\nu) = S_0. \tag{B.2}$$

O sinal no domínio espectral apresenta informação adicional difícil de ser obtida apenas no domínio temporal. Por exemplo, ao analisar um sinal senoidal levemente distorcido em função do tempo, dificilmente se percebe essa imperfeição. Na análise no domínio da frequência, pequenas distorções ou modulações não lineares (que implicam em componentes de frequência diferentes) são facilmente identificadas,

[1]Em física, potência é a grandeza que determina a quantidade de energia concedida por uma fonte a cada unidade de tempo. Outras notações muito utilizadas para o espectro de potência são $S(f) = 1/f^\beta$ e $P(\omega) = 1/\omega^\beta$.

pois cada componente de frequência é visualizada separadamente. Usualmente, as escalas vertical (itensidade) e horizontal (frequência) de um espectro de frequência são em geral logarítmicas, o que facilita a leitura de sinais de baixa intensidade. Assim, a intensidade pode ser diretamente lida em dB (unidade mais usual em sistemas de comunicação), e um amplo espectro de frequência pode ser visualizado simultaneamente na escala horizontal.

Nas figuras B.1(a) e B.1(b) são mostrados exemplos de um sinal do tipo ruído *Brownianoe* outro do tipo ruído branco, respectivamente. Os respectivos espectros de potências são mostrados nas figuras B.1(c) e B.1(d). Considerando que, no espectro de potências, $S(\nu)$ equivale a uma medida de energia, notamos que uma característica fundamental do ruído *Brownianoé* o fato de apresentar maiores energias nas baixas frequências (ou escalas). O processo estocástico que gera um ruído *Brownianoa* partir de um ruído branco é conhecido como Processo de Wiener (GERSHENFELD, 2000). Nessa abordagem, o espectro de potência de um ruído *Brownianoé* expresso como a integral do espectro do ruído branco, logo o ruído *Brownianotem* espectro de potência igual a S_0^2/ν^2. Admitindo que $S_0 = 1$, temos então que $S(\nu) \propto 1/\nu^2$.

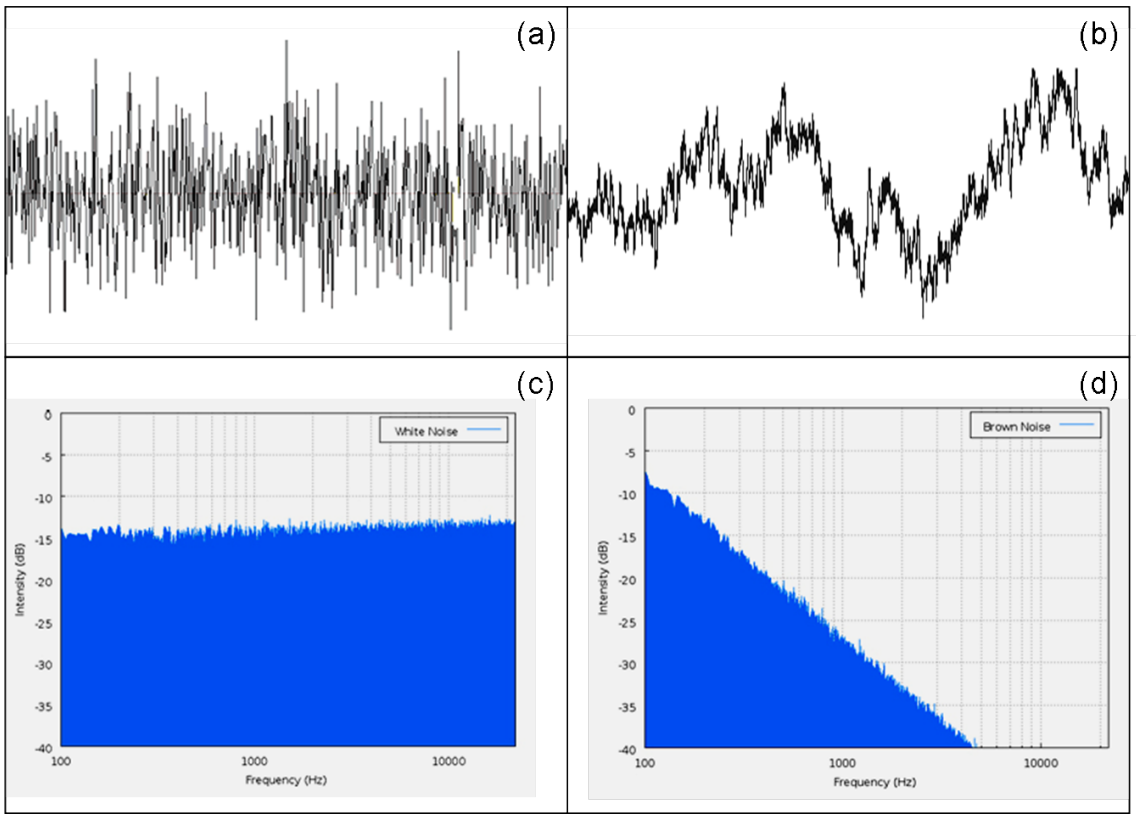


Figura B.1 - Exemplos de sinais genéricos (intensidade × tempo, por exemplo) com os seus respectivos espectros de potências. Ruído Branco: (a) e (c). Ruído Browniano (b) e (d).

APÊNDICE C - ANÁLISE NÃO TENDENCIAL DE FLUTUAÇÕES

Métodos de destendenciamento para análise de flutuações têm sido propostos e aplicados para detecção de correlações persistentes em análise de séries temporais não estacionárias (BASHAN et al., 2008). O algoritmo considerado nesta abordagem, introduzido por Peng et al. (1994), é composto de seis operações computacionais a partir de uma série temporal discreta de amplitudes $\{A_i\}$, descritas a sequir.

- *Integração discreta*: Calcule a representação cumulativa de $\{A_i\}$ como:

$$C(k) = \sum_{i=1}^{k} \left(A_i - \langle A \rangle\right), \qquad (k = 1, 2, \ldots, N) \tag{C.1}$$

  onde $\langle A \rangle = \sum_{i=1}^{N} A$ corresponde à média $\{A_i\}$.

- *Janelamento*: Usando uma janela local arbitrária de tamanho n, divida $C(k)$ em $N_n = int(N/n)$ sub-intervalos não sobrepostos c_j $(j = 1, 2, ..., N_n)$. Note que cada sub-intervalo c_j tem tamanho n e N pode não ser o inteiro múltiplo de n. Então, a série $C(k)$ é dividida mais uma vez do lado oposto para garantir que todos os pontos sejam processados, obtendo ao final desta operação $2N_n$ sub-intervalos.

- *Ajuste*: Obtenha, para cada sub-intervalo, o ajuste de mínimos quadrados, como segue:

$$p_j^m(k) = b_{j_0} + b_{j_1} k + \cdots + b_{j_{m-1}} k^{m-1} + b_{j_m} k^m, \qquad m = 1, 2, \ldots \tag{C.2}$$

  onde m é interpretado como a *ordem da tendência destendenciada*, denotada aqui como DFA^m.

- *Variância*: Calcule a série do desvio acumulado em cada sub-intervalo, do qual a tendência foi subtraída: $C_j(k) = C(k) - p_j^m(k)$. Então, calcule a variância dos $2N_n$ sub-intervalos:

$$F^2(j, n) = \langle C_j^2(i) \rangle = \frac{1}{n} \sum_{i=1}^{n} \left[C((j-1)n + i) - p_j^m(i)\right]^2 \tag{C.3}$$

para $j = 1, 2, \ldots, N_n$, e

$$F^2(j,n) = \langle C_j^2(i) \rangle = \frac{1}{n} \sum_{i=1}^{n} \left[C(N - (j - N_n)n + i) - p_j^m(i) \right]^2 \qquad \text{(C.4)}$$

para $j = N_n + 1, N_n + 2, \ldots, 2N_n$.

- *Flutuação*: Calcule a média de todas as variâncias e a raiz quadrada para obter a função de flutuação $F(n)$:

$$F(n) = \left[\frac{1}{2N_n} \sum_{j=1}^{2N_n} F^2(j,n) \right]^{1/2}. \qquad \text{(C.5)}$$

- *Expoente de escala*: Execute novamente, recursivamente, o cálculo do janelamento até o cálculo do $F(n)$ correspondente com diferentes tamanhos de janelas $n([N/4] > n \geq 2m + 2)$. Em geral, na presença de flutuações na forma da lei de potências $F(n) = Kn^\alpha$, $F(n)$ aumenta linearmente com o aumento de n. Então, usando a regressão linear por mínimos quadrados sobre $\log F(n) = \log K + \alpha \log n$, é possível obter a inclinação α, que é o expoente de escala do método DFA.

C.1 PSD e DFA

Se $A(t_k)$ é o k-ésimo valor de uma série temporal composta por N amostras discretas com resolução temporal τ, sua energia é dada por $E(k) = \sum_0^{N-1} |A(t_k)|^2 \tau$ (para processos estocásticos estacionários de duração infinita, a energia é geralmente infinita) e a potência do sinal é definida como $P(k) = E(k)/N\tau$. Observe que as unidades de P correspondem ao quadrado das unidades da série temporal, e, para séries temporais com média nula, a potência P é igual à variância de $\{A(t_k)\}$ com $i = 1, ..., N$. Então, a distribuição de P (ou variância) de uma série temporal com frequência $1/T$ (em Hertz, com $T = \tau + \Delta t$) é a chamada densidade do espectro de potências (PSD) (KAY; MARPLE, 1981). Em termos práticos, admitindo todas as frequências possíveis de todas as escalas possíveis Δt, PSD é o módulo quadrado da transformada de Fourier da série redimensionada e pode ser estimado com base no cálculo direto via FFT.

A função de correlação $C(\Delta t)$ decai com um expoente $C(\Delta t) \approx (\Delta t)^{-\gamma}$ e a PSD decai com $P(f) \approx (f)^{-\beta}$, where $f = 1/\Delta t$. Então, a inclinação do espectro de potências β é geralmente usada para caracterizar processos estocásticos diferentes que são responsáveis pela faixa de autocorrelação em uma determinada série temporal. Com base no teorema de Wiener-Khinchin (KAY; MARPLE, 1981), é possível mostrar que os dois expoentes β (PSD) e α (DFA) são relacionados por $\beta = 2\alpha - 1$. Para o movimento *Browniano*, temos $1 \leq \beta \leq 3$, portanto $1 \leq \alpha \leq 2$. O potencial diagnóstico dos métodos PSD e DFA é testado sobre um subconjunto de sinais do tipo ruído *Browniano* em (VERONESE et al., 2011b), mostrando que, para séries temporais curtas, a técnica DFA é capaz de detectar o tamanho da correlação com maior precisão do que o expoente de escala PSD.

APÊNDICE D - ANÁLISE DE PADRÕES GRADIENTES (GPA)

O coeficiente de assimetria gradiente G_A é calculado sobre uma matriz quadrada de amplitudes a partir de 5 operações:

(1) Gera-se a campo gradiente da matriz a partir da regra mostrada na Figura D.1;

(2) Removem-se todos os pares de vetores com simetria bilateral em relação aos eixos principais, tanto da imagem como em torno dos máximos locais;

(3) Substitui-se cada vetor do campo gradiente por um ponto escalar localizado em seu centro ou na sua ponta (esta é uma variante do método, que não implica grandes alterações nos valores calculados). Esses pontos são chamados de "pontos de gradiente assimétricos" e a sua quantidade é denotada por N_V. Compare os exemplos (a) e (b) com os exemplos (c) e (d) na Figura D.2.

(4) Realiza-se a triangulação de Delaunay entre todos os N_V pontos de gradiente assimétricos, que serão as arestas de um campo de triangulação contendo N_C linhas de conexão. Veja os exemplos (e) e (f) na Figura D.2.

(5) Calcula-se o coeficiente G_A a partir da seguinte fórmula:

$$G_A = N_C - N_V/N_V \tag{D.1}$$

Padrões elementares assimétricos são mostrados nos exemplos (c) e (d) da Figura D.2, onde seus respectivos campos de triangulação e valores de GA são mostrados nos quadros (e) e (f), respectivamente.

Para os casos simétricos o valor de G_A é nulo, como também pode ser visto no exemplo da Figura D.3. A partir de uma adaptação do algoritmo GPA++, incorporamos o cálculo do coeficiente de assimetria gradiente no programa para geração do modelo de representação baseado em grades numéricas generalizadas, descrito no Capítulo 5.

Nas Figuras D.4 e D.5 são mostrados outros exemplos que ilustram o comportamento do coeficiente de assimetria gradiente. A técnica GPA também foi adaptada para a análise de grades do tipo $\pounds^2$ (ASSIREU et al., 2002) e poderá, portanto, ser incorporada ao modelo de análise do programa *DLattice++*.

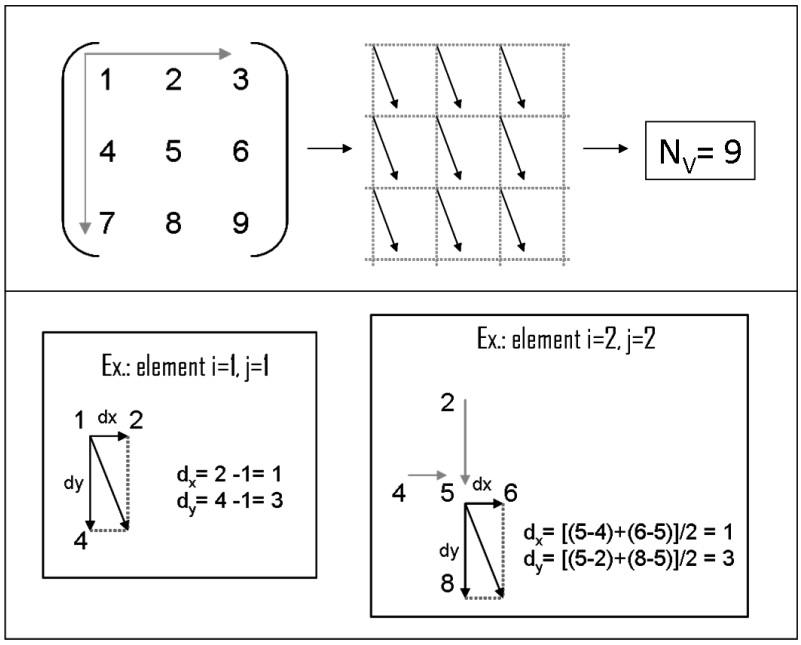


Figura D.1 - Exemplo de cálculo do campo gradiente para uma matriz elementar 3×3. Neste exemplo são mostrados o cálculo do gradiente em relação ao píxel localizado em $(1, 1)$ e para o píxel localizado em $(2, 2)$. No caso dos elementos das bordas as diferenças são simples, enquanto para os outros elementos as diferenças são ponderadas. A ordem dos valores nas diferenças respeita o critério de descida a partir do elemento $(1, 1)$. Como nenhum par de vetores é removido por assimetria bilateral, a quantidade de vetores assimétricos é dada por $N_V = 9$.

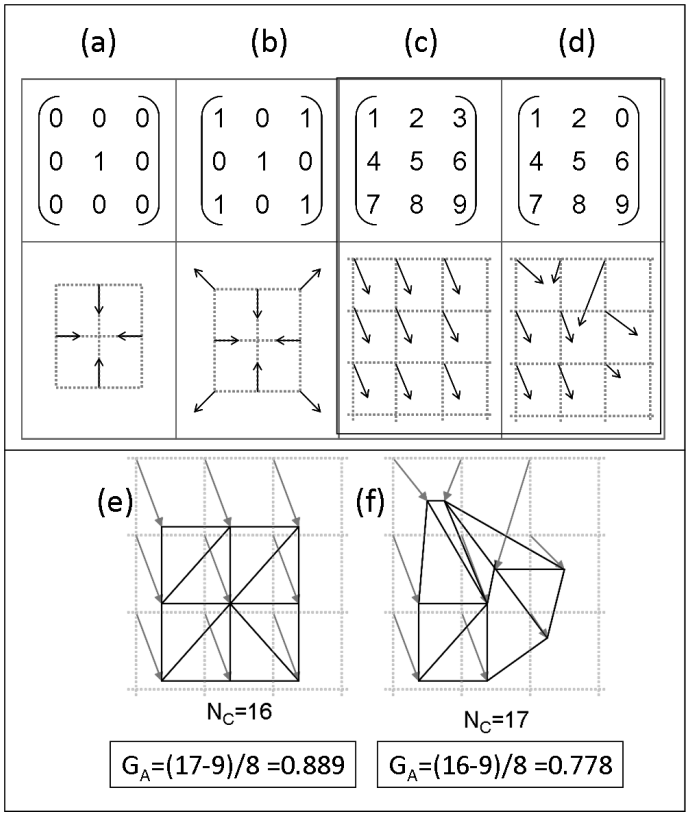


Figura D.2 - Exemplos de matrizes com padrões simétricos (a e b) e assimétricos (c e d). Em (e) e (f) são mostrados os respectivos campos de triangulação para os exemplos (c) e (d).

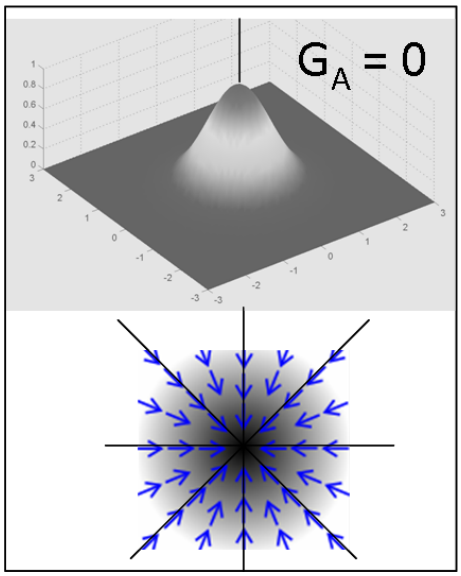


Figura D.3 - Para um padrão com simetria bilateral total, isto é, em relação aos eixos horizontal, vertical e diagonais na matriz, todos os vetores são removidos não restando nenhum par de vetores assimétricos resultando, por definição, um coeficiente de assimetria nulo.

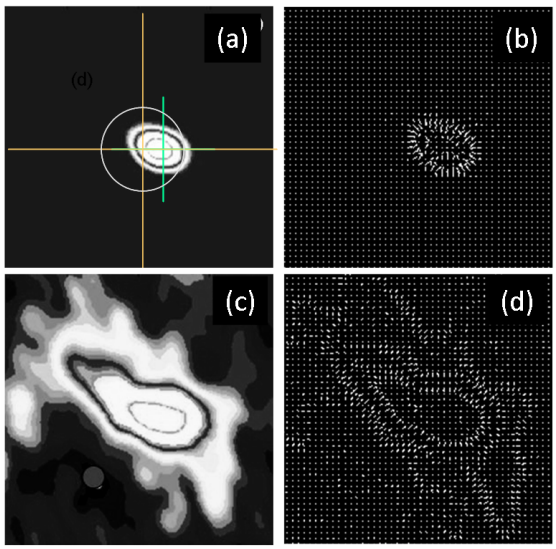


Figura D.4 - (a) Padrão com certo grau de assimetria em relação aos eixos de simetria bilateral global (passando pelo centro da imagem) e bilateral local (passando pelo centro de uma sub-região onde reside o píxel de maior intensidade. (b) O respectivo campo gradiente da imagem mostrada em (a). (c) Padrão mais assimétrico que o anterior possuindo alto grau de fragmentação assimétrica no respectivo campo gradiente assimétrico mostrado em (d).

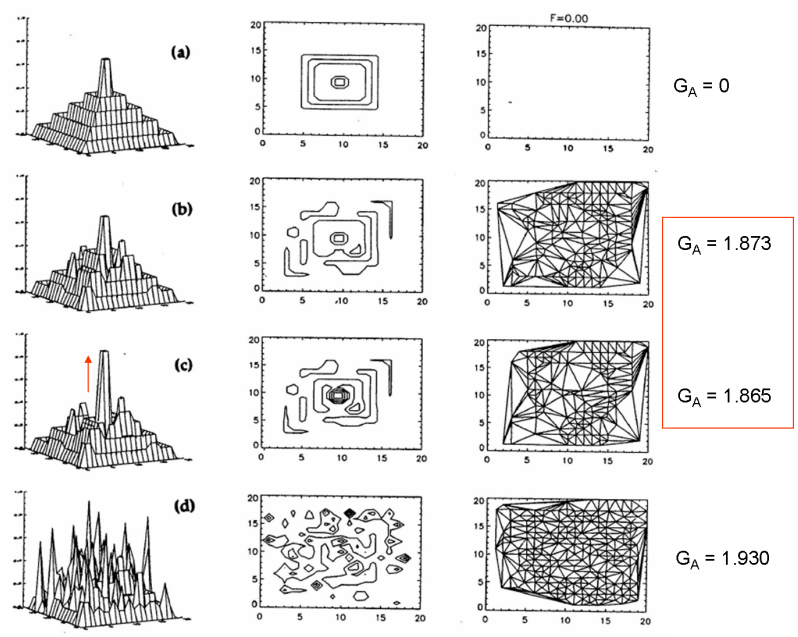


Figura D.5 - Sequência de quatro padrões de fragmentação provenientes de matrizes de tamanho 20×20. (a) Padrão com simetria bilateral total. (b) Padrão com alguma assimetria. (c) Padrão equivalente ao anterior mas com a estrutura central com maior amplitude, o que implica em um aumento da simetria bilateral reduzindo o valor de G_A. (d) caso extremo mais assimétrico e fragmentado.