

MC3 – ELAC 2015

© F.M.Ramos & R.R.Rosa-INPE-MCTI

Proibida a reprodução sem a autorização do autor

Visualização e Análise Computacional de Séries Temporais em Ciências Ambientais e Espaciais

Reinaldo R. Rosa

CONTEUDO REFERENTE À PARTE A (Aulas 1 e 2)

Resumo

Este Minicurso aborda aspectos relacionados ao processo de medida e a representação de um processo físico na forma de série temporal, bem como apresenta uma síntese sobre metodologias de análise, básicas e avançadas, tomando como exemplo um conjunto representativo de diferentes padrões de variabilidade medidos no domínio do tempo. Discute-se ainda, de forma geral, metodologias mais avançadas para análise de dois casos mais específicos: (i) séries temporais curtas e (ii) medidas provenientes de observações e experimentos realizados no domínio espaço-temporal, gerando uma seqüência de imagens digitais. Como exemplos de aplicação apresentam-se dados, relacionados às explosões solares, medidos através de telescópios em terra e abordo de satélites, nas faixas de radiofrequência e raios-X. Descrevem-se também as principais abordagens computacionais para caracterização de complexidade e processos caóticos dissipativos através da análise de séries temporais. Como exemplo de aplicação apresenta-se caracterização de complexidade em séries temporais de variáveis medidas em ciência espacial e física ambiental.

1. Processos Não-Lineares e Padrões “Complexos” de Variabilidade Temporal

Uma variedade de padrões periódicos, quase-periódicos e com variabilidades mais complexas podem ser gerados por uma variedade de processos não lineares onde se destacam os chamados processos *estocásticos* e *caóticos*. Os primeiros são aqueles associados ao comportamento de variáveis aleatórias cuja dinâmica é regida por uma grande quantidade de graus de liberdade (por isso, são também chamados de processos de *alta dimensão*). Os caóticos são, grosso modo, aqueles regidos por leis determinísticas que apresentam alta sensibilidade às variações de parâmetros do sistema dinâmico subjacente e geralmente são processos envolvendo um conjunto pequeno de variáveis (por isso, são também chamados de *baixa dimensão*). Além desses, existem também os chamados processos *reativos-difusivos* que envolvem processos de inibição e ativação de reagentes mediante a ação de um agente catalisador. Em geral, os três

tipos de processos podem ser representados por sistemas de equações diferenciais e também admitem as suas diferentes combinações em processos híbridos ainda mais complicados. Uma relação de processos não-lineares, relacionados à dinâmica de diferentes sistemas, é apresentada no Apêndice A. A partir da Seção 8.2 apresentamos um conjunto de técnicas que podem ser aplicadas na caracterização de processos não lineares, tanto de alta como de baixa dimensão, representados através dos seus respectivos padrões de variabilidade.

1.1 Processos Oscilatórios Periódicos e Quase Periódicos

Iniciaremos o estudo de padrões de variabilidade no tempo mediante um exemplo matemático relativamente simples, que descreve a dinâmica de uma membrana oscilatória circular. Este processo é descrito pela solução, em coordenadas polares, da equação de uma onda com amplitude variável $A(t)$ e velocidade de fase $c: A_{tt}=(1/c^2)\nabla^2 A$. A descrição matemática detalhada do problema pode ser estudada diretamente no *site* referenciado em [1] e um código em Matlab para simulações da membrana elástica está disponível na página da ELAC-MC8 [2]. Medindo os valores da variação da **amplitude** A da membrana ao longo do tempo teremos como mostrado na Figura 1, uma série temporal com padrão de variabilidade periódico caracterizado por uma função senoidal regular. Do ponto de vista funcional trata-se de um processo não-linear (pois a função seno é não-linear) que, entretanto, apresenta um padrão de variabilidade relativamente simples. Portanto, podemos notar que existem processos não-lineares que geram séries temporais com padrões simples de variabilidade. Em nossa abordagem, os termos, “simples” e “complexo”, estão principalmente relacionados à presença ou não de periodicidades e assimetrias na evolução temporal de $A(t)$, respectivamente. Nesse caso, o padrão temporal regular, mostrado na Figura 1, pode ser caracterizado por apenas dois parâmetros **medidos a partir da série**: (i) a sua **faixa dinâmica** definida em torno da **média**, igual a $2|A|$ e o seu período de oscilação T definido como o intervalo entre dois máximos consecutivos (ou dois mínimos) da amplitude.

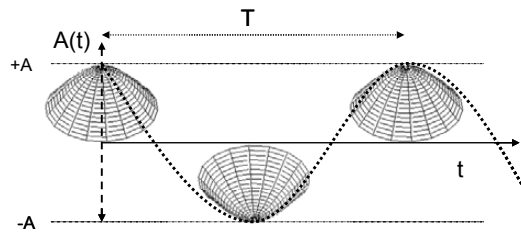


Figura 1 - Padrão mais simples de variabilidade temporal de uma membrana elástica.

Porém, a série da Figura 1 é o resultado mais simples da solução da equação de onda em coordenadas polares. Modos mais complicados podem ser obtidos. Fica como um exercício, você refletir sobre as séries $A(t)$ que seriam geradas a partir de padrões de oscilações mais complicados como aqueles mostrados na Figura 2.

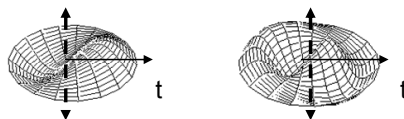


Figura 2 - Outros modos de oscilação da membrana elástica circular.

Outro sistema, relativamente simples, que gera uma série temporal equivalente à da membrana elástica é o *Pêndulo Simples*. Como exercício, tente imaginar a série temporal gerada pela medida da velocidade (ou do *momento*), ou da *amplitude projetada*, do pêndulo simples em movimento. Aqui cabem duas observações importantes:

- (i) Repare que tanto no caso da membrana como no caso do pêndulo, não consideramos efeitos de amortecimento, por exemplo, devido ao atrito, resultando em *dissipação de energia*. Como fica o padrão de variabilidade nesses casos?
- (ii) Tomando a dinâmica pendular como exemplo *canônico*, é oportuno introduzir a noção de *espaço de fases*: uma forma mais sofisticada e abstrata de caracterizar o padrão de variabilidade de um sistema dinâmico. Trata-se de um espaço composto pelas *variáveis de estado* de um sistema e seu estudo rigoroso só pode ser realizado mediante uma abordagem computacional. Mas isso será feito mais adiante, na Seção 2.3.

Análogo ao caso da membrana tente imaginar o padrão de variabilidade gerado por um pêndulo mais complicado conhecido como Pêndulo Composto regido por um processo Caótico.

Até aqui, demos dois exemplos concretos de processos oscilatórios não-lineares que geram séries temporais periódicas ou quase-periódicas. O caráter sinótico desses sistemas descreve, em primeira aproximação, um conjunto de vários fenômenos encontrados na natureza: vibrações eletromagnéticas; vibrações moleculares; comportamento dos sistemas massa-mola estudados em laboratórios de física e engenharia; vibrações de membranas físicas e biológicas em dispositivos mecânicos e tecidos orgânicos; geração de ondas e movimentos convectivos em meios contínuos; variações dos níveis médios dos oceanos, rios e lagos; variações climáticas moduladas pelo ciclo das estações; variação do número de manchas solares modulado pelo ciclo da atividade solar; etc.

1.2 Processos Caóticos Canônicos

Vamos agora discutir o padrão de variabilidade gerado por uma simples equação matemática conhecida como *mapa logístico* (ou *mapa quadrático*), muito estudado pelos matemáticos e físicos. O mapa logístico é uma equação de iteração discreta que pode ser escrita da seguinte forma:

$$A_{n+1} = \rho A_n (1 - A_n) \quad | \quad A_n \geq 0 \quad (1)$$

Curiosamente, o padrão de variabilidade da amplitude A como função discreta de n (que representa aqui, um valor discreto do tempo) torna-se “complexo” (não periódico) para valores de $\rho \geq 3,5$. Além disso, o seu comportamento apresenta alta sensibilidade a diferentes escolhas do valor da condição inicial A_0 . Na Figura 3 mostramos os padrões de variabilidade representados pelos primeiros 100 pontos gerados pela Equação 1 com diferentes valores de k . Na condição $\rho \geq 3,8$ o processo encontra-se no regime “caótico” [3,4]. Como discutido anteriormente, o termo “caótico”

representa, grosso modo, que um *processo plenamente determinístico* é capaz de apresentar uma dinâmica equivalente aos processos chamados *estocásticos*: aqueles onde $A(t)$ comporta-se de forma imprevisível. Isto é, a não-linearidade do processo, observada na série como “quebra de periodicidade” ou “duplicação de períodos”, impõe certo grau de incerteza na previsão dos valores futuros de $A(t)$. A Figura 3 mostra os padrões complexos de variabilidade gerados pela equação 1. Note que, a propriedade de complexidade aqui é exclusivamente a variação do período acompanhada de pequenas variações da amplitude ao longo do tempo, implicando também uma alteração na *autocorrelação* da série quando comparada com ela própria em diferentes escalas (intervalos) de tempo. Diferentes graus de autocorrelação irão quantificar diferentes níveis de “aleatoriedade” da série. Séries como aquelas dos osciladores são puramente determinísticas, por isso apresentam máxima autocorrelação em escalas proporcionais ao período de oscilação. No outro extremo estão variáveis geradas por processos estocásticos completamente sem memória, para os quais a *autocorrelação* é muito baixa ou nula em qualquer escala. Pense na série temporal gerada pelo lançamento sucessivo de um dado não viciado por um longo intervalo de tempo. Não há modelo que possa prever qual o próximo valor a ser obtido após o último lançamento. O que existe, do ponto de vista da caracterização matemática do processo, é apenas uma medida da probabilidade de um dos valores pertencentes ao domínio de valores possíveis ser obtido após o lançamento. A *autocorrelação* e outros parâmetros baseados nessa medida podem ser calculados através de operações computacionais sobre a série. Este fato será abordado na próxima seção.

Outro processo usual de mapeamento não linear é conhecido como “Mapa de Hénon”, estudado por Michel Hénon em 1976, composto por duas equações de recorrência acoplando duas amplitudes $A(n)$ e $B(n)$:
 (i) $A(n+1)=1-k_1A^2(n) + B(n)$
 com (ii) $B(n+1)=k_2A(n)$. O padrão de variabilidade mostrado na Figura 6 foi gerado através deste processo em regime caótico: $k_1=1.4$ e $k_2=0.3$, $A(0)=B(0)=0$.

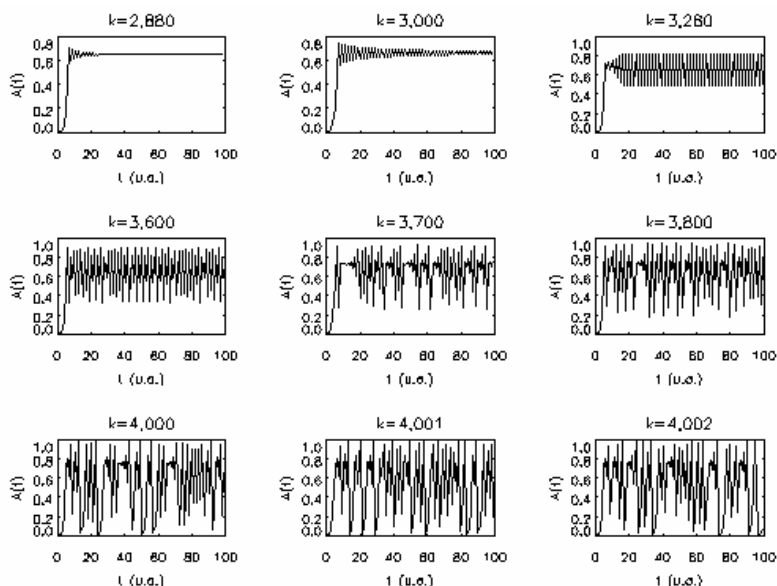


Figura 3 – Diferentes séries temporais geradas pela Equação 1 considerando diferentes valores da constante k. A condição inicial é a mesma em todos os casos $A_0=0.001$.

Uma propriedade também muito importante das séries temporais geradas pelos mapeamentos logísticos e de Hénon em regime caótico é o padrão de variabilidade *antipersistente*. Isto é, para cada escala de iteração $L_X = |(n+1) - n|$ o valor da amplitude tende a aumentar se estiver baixo ou a diminuir se estiver alto. Esse comportamento *antipersistente* resulta um padrão de variabilidade, apesar de complexo, restrito a uma faixa dinâmica quase constante, representada por uma única escala média $\langle L_Y \rangle \sim 2|A|$. Nesse caso, podemos dizer que a faixa dinâmica é aproximadamente *estacionária*. Nos padrões estacionários os valores dos *momentos estatísticos* do conjunto de dados são aproximadamente constantes para *subamostras* da amostra principal. Em linguagem da física estatística chamamos o conjunto de valores, que compõe uma série temporal, de *ensembles*. No formato de *ensemble* a série pode ser embaralhada e a ordem temporal das amplitudes desfeita, por isso, um *ensemble* representa, grosso modo, uma coleção de valores identificada apenas pelos seus momentos estatísticos. Esse conceito é muito importante na construção da teoria de sistemas termodinâmicos (no equilíbrio e fora do equilíbrio), cuja discussão está fora do escopo deste minicurso, mas pode ser estudado em outros textos [5,6].

Note que, a propriedade de *antipersistência*, no caso do mapa logístico e de outros processos caóticos descritos no Apêndice A, implica um *ensemble* muito pobre em relação à diversidade de valores da variável A_n , fora da escala da faixa dinâmica. O mesmo acontece para os casos dos osciladores discutidos anteriormente. Por maior que seja a diversidade de valores, no domínio real com alta precisão em ponto flutuante, os valores calculados ficam restritos ao domínio da escala L_Y compreendido entre $+A$ e $-A$, flutuando em torno de uma média característica $\langle A \rangle$. Quando essa característica é violada e, além disso, para cada passo no processo de geração de um novo valor de amplitude pode ocorrer *persistência*, a faixa dinâmica torna-se instável e a *variância* do ensemble começa a crescer. A persistência, nesse caso, é caracterizada quando um valor de amplitude aumenta e continua aumentando até um intervalo longo de tempo, valendo o mesmo comportamento para a diminuição da amplitude. O valor da variância, nesse caso, expressa uma medida da *energia* intrínseca do sistema responsável pela evolução do processo dinâmico de geração de amplitudes. Essa característica implica em uma nova propriedade que pode aparecer no padrão de variabilidade da série temporal: o comportamento *multiescalar*. Portanto, um padrão de variabilidade *multiescalar* apresenta persistência em determinadas escalas de intervalo temporal $\Delta t_1, \Delta t_2, \dots, \Delta t_i$, resultando em uma variedade de faixas dinâmicas caracterizadas por diferentes escalas de amplitude $\Delta A_1, \Delta A_2, \dots, \Delta A_j$. Aqui chegamos a um ponto extremamente importante do nosso estudo: discutir o significado de “espectro de energias” de uma série temporal. Note que, as diferentes escalas temporais irão definir, quando recorrentes na série, diferentes “períodos” de variação da energia (representada pela faixa dinâmica ou variância da série em cada intervalo temporal). Como o inverso do período temporal nos fornece uma medida de frequência ($\omega_i=1/\Delta t_i$) teremos uma nova representação do processo não-linear subjacente, que pode ser extraída diretamente da série temporal, que se chama Espectro de Energias (gráficos de $\Delta A_i \times \omega_i$). Esta representação permite analisar os processos no domínio das frequências. A análise espectral de processos não-lineares aperiódicos baseia-se na aplicação do algoritmo conhecido como *Transformada*

Rápida de Fourier (do termo em inglês *Fast Fourier Transform*, normalmente abreviado como FFT) discutido na Seção 2.2.

1.3 Processo Estocástico do Tipo $1/\omega^\alpha$

De forma equivalente à equação logística, existem processos puramente matemáticos capazes de gerar padrões complexos de variabilidade estocástica. Porém, com variabilidade do tipo *persistente*. A Equação 2 é um modelo espectral **multifractal** inspirado por um processo estocástico conhecido como “movimento Browniano” [6], muito investigado por físicos e matemáticos, incluindo A. Einstein.

Vimos que o espectro de um processo estocástico pode ser representado por uma série discreta de frequências e a sua função aleatória correspondente pode ser escrita como uma série discreta para tempos $t_i = i\Delta t$ com $i=1, \dots, M$. Dessa forma, a série temporal pode ser escrita como uma simples superposição de oscilações harmônicas dado por:

$$A(t_i) = \sum_{k=1}^{M/2} [P(\omega_k)\Delta\omega]^{1/2} \cos(\omega_k t_i + \phi_k), \quad i=1, \dots, M \quad (2)$$

onde $\omega_k = k \Delta\omega$ ($k=1, \dots, M/2$), com $\Delta\omega=2\pi/M\Delta t$, e os ϕ_k `s são as fases escolhidas aleatoriamente. O **espectro de potências** $P(\omega_k) \sim \omega_k^{-\alpha}$ é escolhido *a priori* e o valor $M\Delta t$ determina o tamanho da série necessário para que o espectro tenha um corte de alta frequência no valor de *Nyquist* ($\pi/\Delta t$). Esse processo tem uma *estrutura auto-afim* em seus incrementos e representa uma classe de flutuações do tipo Browniano Fracionário com índice α . A partir de **princípios de simetria** codificados em **teoria de grupos**, prova-se que processos auto-afins apresentam auto-similaridades apenas sob algumas transformações escalares específicas [4]. Nesse sentido, flutuações fracionárias Brownianas representam uma generalização do movimento Browniano ordinário, que corresponde ao caso $\alpha = 1/2$ [4]. Trata-se, portanto, de um exemplo canônico de processo multiescalar introduzido anteriormente.

Nesse modelo *multiescalar* o processo estocástico pode apresentar memória (*autocorrelação*) em diferentes escalas, apesar de conter, devido à sua natureza *multiescalar*, um alto grau de imprevisibilidade em relação à evolução do processo em determinadas escalas. Aqui o termo **fractal** aparece pela primeira vez e, grosso modo, significa a propriedade de **autosimilaridade** que um **ensemble** de medidas organizadas em uma dada representação geométrica, apresenta em diferentes escalas [4]. Na Figura 4 temos dois exemplos de séries temporais geradas a partir da equação 2, para $\alpha = 1,5$ e $\alpha = 2,0$.

O exemplo mais importante de processo não-linear com padrão de variabilidade **persistente** é o movimento turbulento de fluidos: gases e líquidos sujeitos as diferentes condições de instabilidades. O fenômeno da turbulência é ainda assunto de profundas pesquisas teóricas e experimentais e está presente em diversos processos físicos relacionados ao meio ambiente quando observado em diversas escalas. As características da turbulência aparecem desde processos observados na escala quântica até processos observados em escala cosmológica. No Apêndice B mostramos uma série temporal proveniente de valores de temperatura medidos próximo à copa das arvores na

floresta amazônica (veja referências na 2ª Parte do Minicurso), em alta frequência de amostragem: $f = 60$ medidas por segundo = 60Hz. Nessa série identificamos as diferentes escalas de energia distribuídas em diferentes escalas de tempo. Na seção 2.2 introduzimos o conceito de Lei de Potência, associado a um espectro de energias, que permite caracterizar um padrão de variabilidade do tipo turbulento, isto é do tipo $1/\omega^\alpha$ com α assumindo diferentes valores a partir do tipo de turbulência considerado. Note que a riqueza de escalas impõe um caráter estocástico, como discutido anteriormente, às séries temporais geradas por processos turbulentos. Isto é, a amplitude medida a partir de um processo turbulento comporta-se como uma variável aleatória dificultando a tarefa de fazer previsões sobre o seu comportamento. Um exemplo típico, fora da física, que apresenta comportamento análogo à turbulência é a variação dos preços no mercado de ações [7].

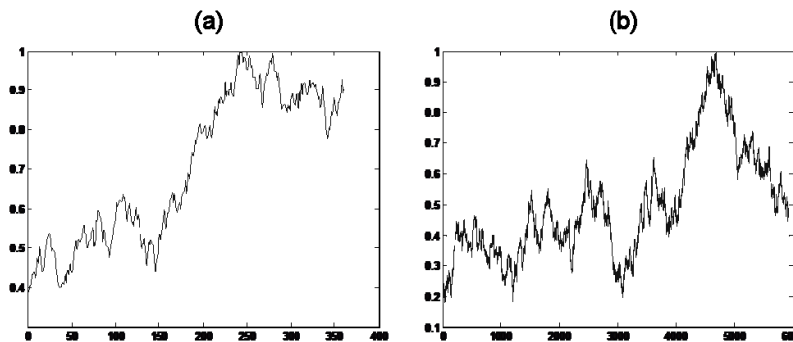


Figura 4 - Exemplo de série temporais com padrão complexo de variabilidade incluindo a propriedade da persistência devido à sua riqueza de escalas. Ambas as séries são do tipo $1/f^\alpha$ sendo geradas a partir da equação 2: (a) para $\alpha = 1,5$ e $\alpha = 2,0$.

Portanto, um processo $1/\omega^\alpha$, em especial o turbulento, está associado a sistemas ricos em dinâmicas multiescalares que apresentam similaridades em algumas escalas. Tomemos como exemplo o movimento multiescalar de uma árvore em resposta à dinâmica multiescalar do meio-ambiente representado pelos diferentes valores do **gradiente** de pressão, que determinam diferentes valores para a intensidade dos ventos. Em comparação considere o comportamento de um edifício em resposta à mesma dinâmica multiescalar. Nesse caso, o edifício não responde ao processo multiescalar. Não há, portanto transferência de energia, conseqüentemente a faixa dinâmica de movimento do edifício é praticamente nula. Ao contrário, a árvore responde ao movimento multiescalar do vento praticamente em todas as escalas. Cabe destacar que talvez exista uma *superescala* para a qual o edifício responda, configurando assim um evento extremo, pois a faixa dinâmica nesse caso passa por uma variação abrupta levando a energia (variância) para o infinito. Entretanto, na maioria dos casos, eventos extremos são eventos raros, pois a valor da superescala associada tem baixa probabilidade de ocorrer. Como exercício, reflita sobre a dinâmica de terremotos. Você irá encontrar uma lei muito importante para o nosso estudo conhecida como Lei de **Gutenberg-Richter**, muito fácil de entender se você está apto para obter e interpretar um **histograma**, nossa ferramenta de partida na próxima seção.

1.4 Sistemas Complexos

A partir da discussão que fizemos até aqui, sobre processos não-lineares (representados pelas dinâmicas oscilatórias, caóticas e estocásticas “ $1/\omega^{\alpha}$ ”, a última contendo como exemplo canônico a dinâmica turbulenta), chegamos ao ponto de discutir a formalização atual sobre “Sistemas Complexos”.

Em geral, padrões complexos de variabilidade no tempo são gerados por processos não lineares relacionados ao comportamento de sistemas dito “complexos”. Mas o que são sistemas complexos? Como caracterizá-los? Podemos tentar começando pela caracterização de suas propriedades. Com base no que foi discutido anteriormente, podemos afirmar que as propriedades principais relacionadas diretamente à variabilidade complexa são as seguintes: (a) Não-linearidade na dinâmica dos processos (ativação, reação, inibição, sincronismo, dissipação) e (b) diversidade **escalar** dos processos dinâmicos (ex. pêndulo não-linear, mesmo contendo poucos elementos apresenta comportamento complexo). Outras duas propriedades que estão mais inerentemente próximas do sistema propriamente dito são: (c) A Quantidade e Diversidade de elementos envolvidos na dinâmica (ex. osciladores acoplados; partículas interagentes no nível quântico). Essas propriedades geralmente levam à emergência de uma nova fenomenologia; e (d) Quantidade e Diversidade de processos – aqueles onde a mistura de uma infinidade de diferentes processos gera queda da **entropia** e faz emergir uma nova estrutura dinâmica coerente e estável em relação ao meio ambiente – Ex. Universo e sua hierarquia de escalas e processos presentes na formação de *Grupos de Galáxias, Galáxias, Estrelas, Sistemas Estelares, Planetas, Atmosferas* com sua evolução química e biológica, *moléculas complexas: DNA e enzimas, células e organismos*. Nesse processo de evolução hierárquica nota-se o papel fundamental das quatro propriedades: (a) não linearidade, (b) multiescalonamento, (c) quantidade e diversidade de elementos interagentes e (d) quantidade e diversidade de processos. Note quanta informação pode estar por trás de uma simples medida na forma de série temporal. Em nossa discussão preliminar, sem ainda utilizarmos formalmente a notação matemática, partimos da análise de uma membrana simples e chegamos aos organismos vivos. Dentre os mais simples está o vírus e dentre os mais complexos estão aqueles que possuem neurônios. Estes dois sistemas são dinâmicos, ricos em elementos, subestruturas e processos físicos, químicos e biológicos. Sem falar dos organismos que possuem outras propriedades mais sutis como a consciência e o livre arbítrio combinados. Finalmente, chegamos a um ponto “complexo”: estamos falando de nós mesmos.

Antes de passarmos para a próxima seção onde abordaremos formalmente algumas técnicas de análise, vamos discutir três últimos pontos, todos ainda de caráter epistemológico:

- (i) Apesar de não haver dúvidas que um neurônio biológico é mais complexo que um modelo de membrana, para que serviria quantificarmos a complexidade de sistemas? E como fazê-lo de forma sistemática e universal? Fica como mais um exercício de reflexão aceitar o desafio de responder qual sistema é mais complexo: uma galáxia ou uma célula?
- (ii) É possível, em primeira aproximação, reduzir um sistema complexo a fim de entendê-lo melhor pelo menos parcialmente? Como exercício, tente construir

uma série temporal a partir de uma **representação binária** da vida de uma pessoa? Como podemos fazer isso e quais são os riscos e vantagens da abordagem reducionista de um processo tão complexo como esse?

- (iii) Na evolução da nossa discussão aparece no texto o termo “entropia”. Se você não tem idéia sobre o seu significado ou tem idéia, mas não o compreende, ficará surpreso quando entendê-lo. Isso poderá mudar a sua visão de mundo! Como?

2. A Caracterização de Processos em Séries Temporais

Um tratamento convencional e completo sobre séries temporais engloba uma centena de conceitos, métodos e técnicas estatísticas, matemáticas e computacionais. Exemplos de textos atualizados com riqueza de detalhes teóricos são referenciados em [7]. Devido à quantidade reduzida de páginas disponíveis para a abordagem do assunto neste livro, iremos discutir um conjunto compacto de técnicas voltadas para a caracterização do padrão de variabilidade de séries temporais discretas e não-estacionárias que representam processos não-lineares equivalentes àqueles gerados pelos processos relacionados no Apêndice A. Em [2], listamos um conjunto de técnicas não abordadas em nosso texto principal, a serem estudadas como complemento ao assunto aqui tratado.

2.1 Conceito de Série Temporal

Na abordagem proposta aqui, uma série temporal é um conjunto discreto de valores numéricos ordenados no tempo de forma regular, onde cada valor representa uma amplitude $A(n)$ gerada através de um processo não linear no tempo. Portanto, uma dada série temporal de amplitudes:

$$\{A(n)\}_{n \in Z} \text{ ,}$$

satisfaz as seguintes condições:

$$(i) \sum_{n \in Z} |A(n)|^2 < \infty ; (ii) \exists A(n) \mid A(n) > A(n-1) \text{ and } A(n) > A(n+1)$$

$$\text{OU } \exists A(n) \mid A(n) < A(n-1) \text{ and } A(n) < A(n+1)$$

A condição (i) garante a “energia finita” e, portanto a existência de uma faixa dinâmica da ST, enquanto a condição (ii) garante a quebra da linearidade do processo gerador da série. A faixa dinâmica admite a existência de uma média $\langle A \rangle = [\sum A(n)]/N$ em torno da qual os valores de $A(n)$ flutuam: $\Delta A_n = |A(n) - \langle A \rangle|$. A quebra de linearidade admite a existência de flutuações locais periódicas e/ou quase-periódicas também relacionadas ao processo subjacente.

A presença de flutuações permite realizar a contagem de valores semelhantes de amplitude ao longo da série e representar a sua frequência relativa através de um **histograma**. Destaca-se aqui que ST geradas por processos determinísticos admitem tratar as suas respectivas amplitudes $A(n)$ como **variáveis aleatórias** no tempo. Dessa forma, mesmo um padrão de variabilidade discreto gerado por um processo

determinístico admite uma modelagem estatística através de uma **Distribuição de Probabilidades**. Uma distribuição de probabilidades é modelada por uma Função de Probabilidades (FP), que pode ser discreta ou contínua. Quando a FP (discreta ou contínua) é usada para modelar um processo aleatório discreto ela é chamada, na literatura internacional, de *Probability Mass Function* (PMF). No caso do processo ser contínuo, ele é chamado de *Probability Density Function* (PDF). Note que, apesar da ST representar o processo não-linear de forma discreta através de $\{A(n)\}$ isso não implica que o processo em si seja discreto. Processos caóticos, por exemplo, discutidos na Seção anterior, podem ser gerados tanto por processos discretos (Ex.: Eq. 1) como por processos contínuos (ver os exemplos da Equação de Kuramoto-Sivashinsky no Apêndice A e o da Equação de Lorenz, tratada na Parte B deste minicurso). Um texto complementar incluindo uma relação das principais Funções de Probabilidade está disponível na *webpage* deste minicurso [2].

Na Figura 5 são mostrados três padrões de variabilidade distinguíveis pela inspeção visual, cada um representando um processo diferente: (a) oscilações periódicas; (b) superposição uniforme de oscilações periódicas; (c) processo de geração de números aleatórios via *computação de Monte Carlo*. São padrões distinguíveis, pois representam dois processos extremos: (a) e (c), e um intermediário (b). Entretanto, entre os mesmos existe uma riqueza de outros processos cujos padrões de variabilidade são semelhantes apesar da grande diferença entre os tipos de processos subjacentes. Como exemplo, dois padrões semelhantes são mostrados na Figura 6.

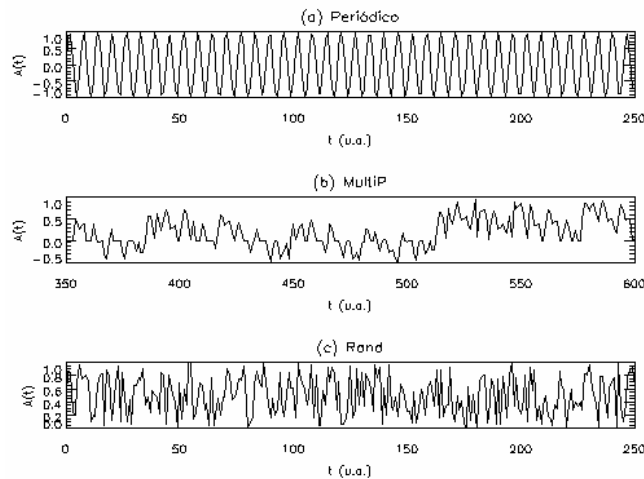


Figura 5 – Padrões de variabilidade representando dois processos extremos: (a) Uma série temporal periódica para processos oscilatórios periódicos e (c) Uma série temporal gerada por um processo pseudo-aleatório tipo Monte Carlo. O padrão intermediário mostrado em (b) representa um processo multi-periódico gerado pela simples superposição de vários sinais periódicos respeitando uma hierarquia uniforme de períodos e amplitudes com pequenos deslocamentos de fase.

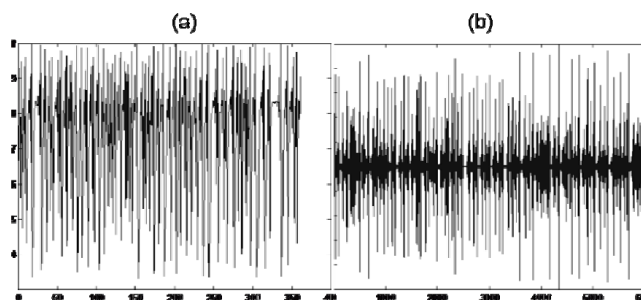


Figura 6 – Exemplos de padrões complexos de variabilidade, com amplitudes normalizadas e provenientes de dois diferentes processos caóticos. São semelhantes à inspeção puramente visual. O padrão em (a) foi gerado pelo Mapeamento de Henon, enquanto o em (b) pela Equação diferencial de Kuramoto-Sivashinsky.

Tomando como motivação as severas limitações da simples inspeção visual das séries temporais, é importante discutir os problemas relacionados à compatibilidade entre a duração e resolução do processo e a duração e resolução da medida. Nesse sentido, no Apêndice B, o conceito de frequência de amostragem é apresentado considerando as diferenças entre processos reais e sintéticos.

2.2 Técnicas Matemáticas e Computacionais

A seguir relacionamos um conjunto de técnicas para análise de séries temporais que permitem caracterizar diretamente, em primeira aproximação, o tipo de processo subjacente levando em conta apenas as características contidas nas séries. A explicação detalhada de cada uma incluindo exemplos de algoritmos em C e Matlab estão disponíveis em [2].

- Histogramas e Modelagem Estatística através de PDFs

A Figura 7 mostra o procedimento de montagem do histograma de um ruído branco (processo de flutuação aleatória em torno da média zero) que é descrito estatisticamente pelo modelo “Gaussiano” (por isso também conhecido como “flutuação Gaussiana”). Diferentemente de uma distribuição uniforme (processo gerado pela seqüência de lançamento de um dado não viciado) a distribuição Gaussiana informa que o valor mais provável no processo é dado pela média das amplitudes. Mas, como se sabe, existem vários processos chamados “não-Gaussianos” que ditam outras regras para o valor mais provável. Um dos processos não-Gaussianos mais interessantes é chamado de “Lei de Potência”: Considere, de forma geral, um modelo de lei de potência $Y(X)=kX^{\alpha}$. Logo, sua representação linear fica $\log(Y)=\alpha\log X+\log k$. Graficamente, a saída desse modelo pode ser representada nos eixos $\log Y$ x $\log X$, respeitando as variações de Y e X , já incorporando o coeficiente linear da melhor reta, k . Vamos fazer um exercício? O material está na web [2]. Nesse caso, os valores mais raros são os menos frequentes. Como exemplo, considere os movimentos das placas tectônicas que levam ao terremoto: esta é a Lei de Gutenberg-Richter, já mencionada anteriormente, mostrada na Figura 8.

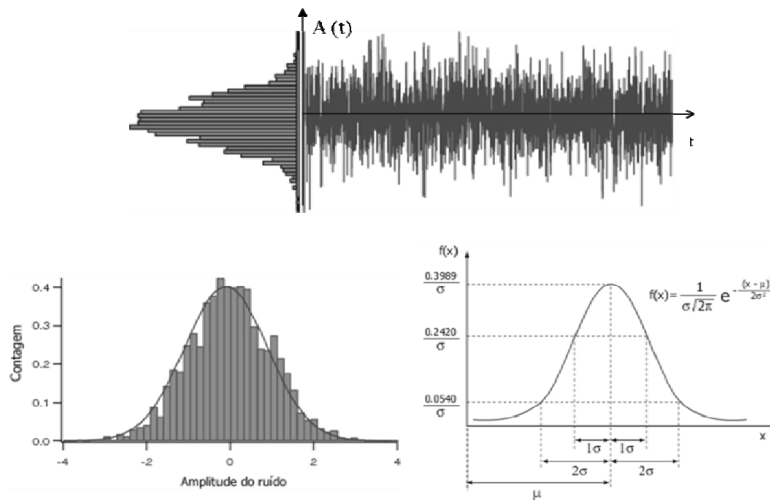


Figura 7- Histograma e modelagem Gaussiana de um ruído branco.

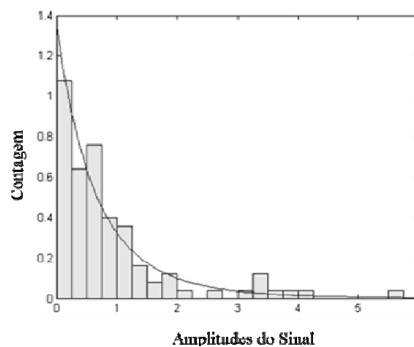


Figura 8 – Histograma e modelagem por Lei de Potência de um sinal estocástico não-Gaussiano gerado a partir de medidas geofísicas (contagem dos movimentos tectônicos na escala Richter).

- *Espectro de Potências e Função de Auto-Correlação*

O espectro de potências pode ser obtido através da aplicação da FFT sobre a coleção de valores A_n , que consiste no cálculo da série $A_k = 1/N^{1/2} \sum A_n \exp [i(2\pi nk)/N]$ para $k=1,2,\dots,N$. Obtém-se então o chamado “espectro de potências” $P(\omega) = |A_k|^2$ (veja os resultados para as séries discutidas ao longo deste texto, na próxima seção).

A Função de Auto-Correlação é definida a partir da soma das flutuações em torno da média normalizadas pela variância de tal modo que seu valor fique no intervalo de -1 a +1, significando, respectivamente o extremo de anti-correlações e correlações, passando pelo valor nulo que expressa a falta de correlações entre as escalas de tempo (“lags”) envolvidas na composição da série.

Ambas as técnicas permitem extrair parâmetros que devem ser avaliados mediante o chamado “teste de nulidade”, que consiste na repetição do mesmo cálculo sobre a série embaralhada (“surrogate” data). O embaralhamento deve ser aleatório.

Técnicas associadas: “box counting” para obter dimensões fractais e expoentes de Hurst, DFA (“Detrended Fluctuation Analysis), entre outras (veja lista de referências em [2]).

- *Reconstruções de Primeira Ordem (Portraits)*

Esta técnica permite obter o grau de espalhamento “*ergódico*” da série quando é comparada com ela mesma deslocada de uma escala previamente selecionada. O algoritmo utilizado para obter os “*portraits*” das séries apresentadas ao longo do texto baseia-se na diferenciação da variável $A(n)$ gerando a sua $A(n)$ diferenciada. Existem outras técnicas associadas: reconstrução de espaço de fases (detalhado na 2ª parte deste MC). Uma delas é a *Plotagem de Recorrência* que distribui os valores de $A(n)$ no espaço simétrico composto pelos valores da série selecionados em todas as escalas de tempo: da resolução à duração total da amostra. Na Figura 9 são mostrados os: (a) *portrait* e (b) plotagem de recorrência para a série multi-periódica da Figura 5b. Os padrões são tipicamente intermediários entre os casos extremos periódico e aleatório, mostrados na próxima Seção.

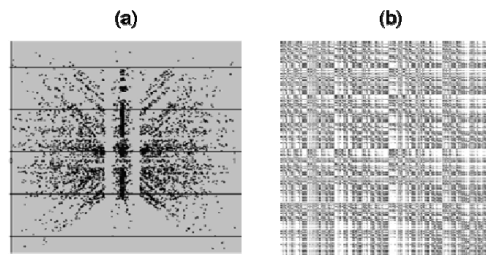


Figura 9 – Exemplos de saídas após as operações de reconstrução de primeira ordem (“portrait”) e plotagem de recorrência do processo multiperódico.

- *Análise de Padrões-Gradientes*

A análise de padrões-gradientes (“Gradient Pattern Analysis”- GPA) permite obter, por intermédio de operadores computacionais sobre o gradiente de matrizes, quatro momentos gradientes que descrevem a complexidade estrutural dos padrões espaço-temporais através das suas assimetrias, diversidade de módulo e fase, e entropia do gradiente [8].

Pode ser aplicada em perfis espaciais e séries temporais para classificar padrões de variabilidade não-periódicos de acordo com as quebras de simetrias dos perfis periódicos (ver Figura 10). A extração de momentos gradientes baseia-se na contagem dos vetores assimétricos contidos na grade-gradiente. Quando aplicado nas decomposições do tipo multiresolução por *Ondeletas*, gera o chamado *espectro-gradiente* do sinal. Trata-se de uma metodologia alternativa para a análise de **Series Curtas**: aquelas compostas por poucos pontos ($N < 10^3$ pontos), desqualificando, portanto as outras técnicas apresentadas. Uma lista completa com cerca de 30 artigos científicos sobre a técnica está disponível em [2].

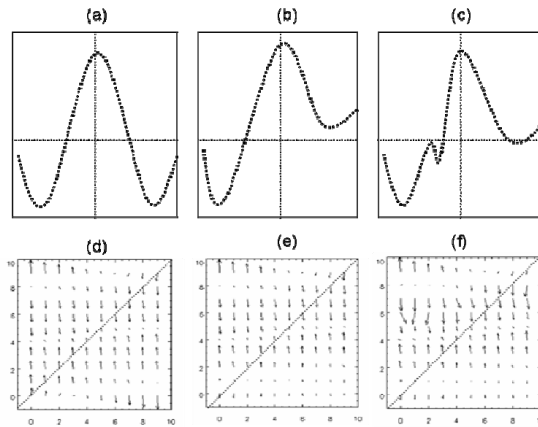


Figura 10 – Exemplos de padrões-gradientes (d,e,f) de perfil (temporal ou espacial) totalmente simétrico (a) e assimétricos (b e c).

- Outras técnicas são discutidas no texto disponível na webpage da ELAC [2].

2.3 Caracterização dos Processos Apresentados no Apêndice A

Nas Figuras 11, 12 e 13 são mostrados, ordenadamente, os padrões na forma de espectros de potências, “portraits” e plotagem de recorrência para os padrões de variabilidade (a) periódico, (b) pseudo-aleatório ($\alpha \sim 0$), (c) Hénon em regime caótico (anti-persistente com $\alpha > 0$) e (d) multifractal Browniano ($\alpha \sim 2$).

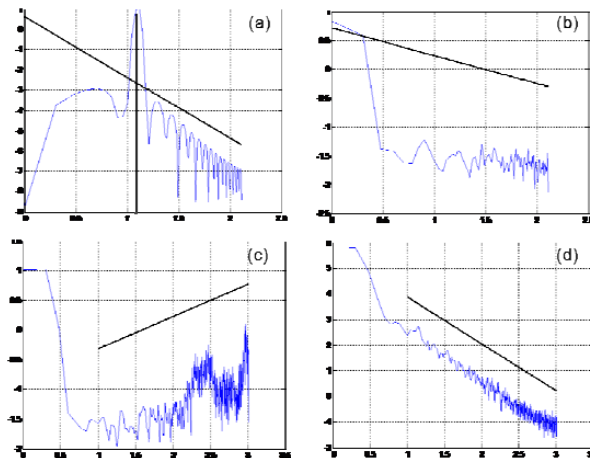


Figura 11 - Espectros de Potência para os processos: (a) oscilatório periódico; (b) pseudo-aleatório; (c) Hénon caótico ($k_1=1.4$ e $k_2=0.3$); e (d) multifractal Br $1/\omega^2$.

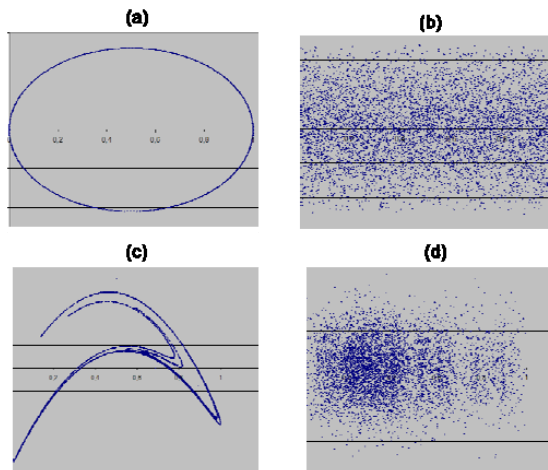


Figura 12 - Resultados da reconstrução de 1ª ordem (Portraits) para os processos: (a) oscilatório periódico; (b) pseudo-aleatório; (c) Hénon caótico ($k_1=1.4$ e $k_2=0.3$); e (d) multifractal $Br 1/\omega^2$.

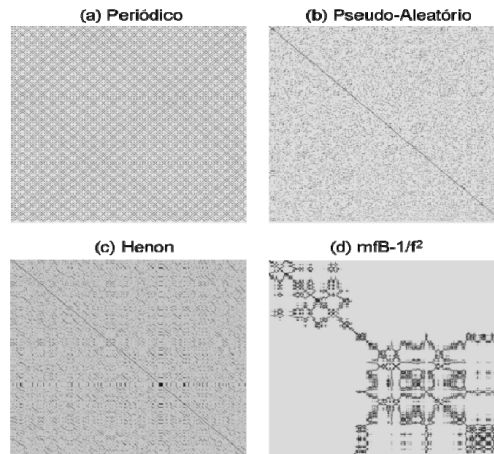


Figura 13- Resultados da plotagem de recorrência para os processos: (a) oscilatório periódico; (b) pseudo-aleatório; (c) Hénon caótico ($k_1=1.4$ e $k_2=0.3$); e (d) multifractal $Br 1/\omega^2$.

3. Caracterização de Processos Espaço-Temporais

3.1 Tipos de Processos

Os principais tipos de processos no domínio espaço-temporal são discutidos em detalhe em [2]. São importantes, no contexto da física espacial e ambiental, os seguintes: Difusivo-Convectivo, Ondas Planas Não-Lineares com Relaxação Espaço-Temporal, “Oscilons” resultantes de forças de contato entre partículas de matéria condensada, turbulência localizada dos tipos Euleriano e Langragiano, Reação-Difusão, entre outros.

Na Figura 14 são mostrados exemplos de imagens (“snapshots”) provenientes de diferentes processos espaço-temporais. (a) camadas granulares gerando “oscilons” por forças de contato (Cortesia do Prof. H.L. Swinney, Univ. Texas, Austin); (b) superfície rugosa gerada por processo difusivo estocástico (tipo KPZ) [9]; (c) Nano superfície de silício poroso observada por SFM e visualizada pelo sistema flyby desenvolvido pelo LAC [10]; explosão solar observada através do telescópio de raios-X moles da missão Yohkoh [11].

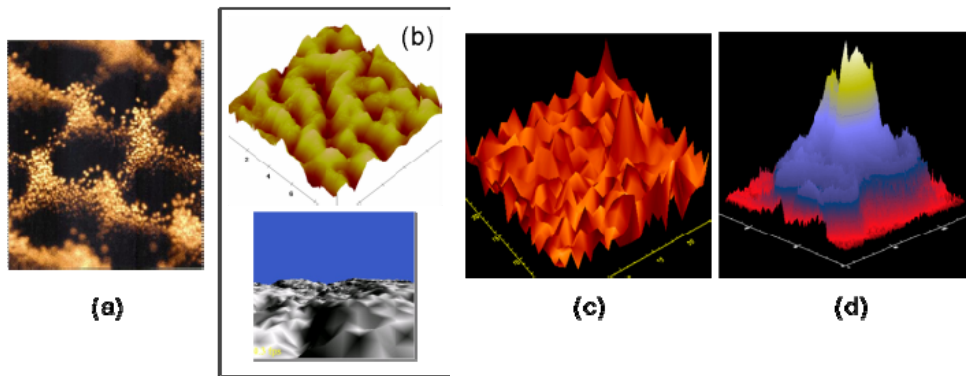


Figura 14 – “Snapshots” provenientes de diferentes processos espaço-temporais: (a) “oscilons” em camadas granulares; (b) crescimento difusivo estocástico 2D, tipo KPZ; (c) porosidade em amostras de silício poroso; (d) explosão solar.

3.2 Técnicas de Análise

- Espectro-Gradiente (EG): permite classificar a complexidade estrutural de um perfil temporal ou espacial, de uma estrutura 2D ou 3D, calculando o coeficiente de assimetria em cada componente espectral do sinal ou imagem. A decomposição é realizada através da análise de multiresolução com uma ondeleta discreta (por exemplo, DB8), e o cálculo do coeficiente de assimetria é obtido através da Análise de Padrão-Gradiente. Através do espectro de assimetria é possível classificar a complexidade estrutural de um processo que tende a se estabilizar em um padrão médio [12].
- Característica de Euler Generalizada (CEG): através do terceiro funcional de Minkowski, caracteriza-se a evolução da coalescência de subestruturas em uma estrutura maior [8,13].
- Redes Neurais Híbridas (RNH): trata-se de um classificador neural de texturas híbrido para reconhecimento e identificação de classes de texturas e granularidade [14]. No processo de classificação, um mapa auto-organizável de Kohonen – integrado ao algoritmo LVQ (Linear Vector Quantization) – recebe como entrada os atributos de textura gerados sobre cada pixel do conjunto de dados e, através de regras de aprendizagem competitiva, gera o mapa temático da imagem.

- Morphing-Cross-Dissolve (MCD): esta ferramenta possibilita a extração de atributos geométricos (escalares e angulares) ao longo de uma estrutura espacial na forma de imagem 2D e 3D e permite, escolhendo um ponto do atributo, simular alterações na estrutura (rotações e deformações) para teste da evolução morfométrica [15].

3.3 Estudo de Caso: Explosões Solares

No INPE, vários projetos voltados para o estudo observacional e teórico das explosões solares estão em andamento, com destaque para Brazilian Decimetric Array (<http://www.das.inpe.br/fimi/BDA/>) (Figura 15). O projeto BDA fornecerá dados em radiofrequência de explosões solares observadas também em raios-X e UV, além de sinais de variabilidade das tempestades magnéticas próximas à atmosfera terrestre observadas pela NASA. A Figura 16 mostra um exemplo de explosão solar geofetiva, observada em diferentes frequências, cuja caracterização sistêmica do fenômeno e previsões de suas diferentes fases depende de uma análise morfométrica integrada de todos os dados envolvidos. Os resultados obtidos, de forma integrada, evidenciam que o fenômeno da explosão solar é rico em escalas, comportando-se como um processo turbulento localizado regido pela interação dos arcos coronais em diversas escalas (Figuras 17 e 18).



Figura 15 – O arranjo inicial de antenas do BDA-INPE para observação de explosões solares em radiofrequência com alta resolução temporal e espacial.

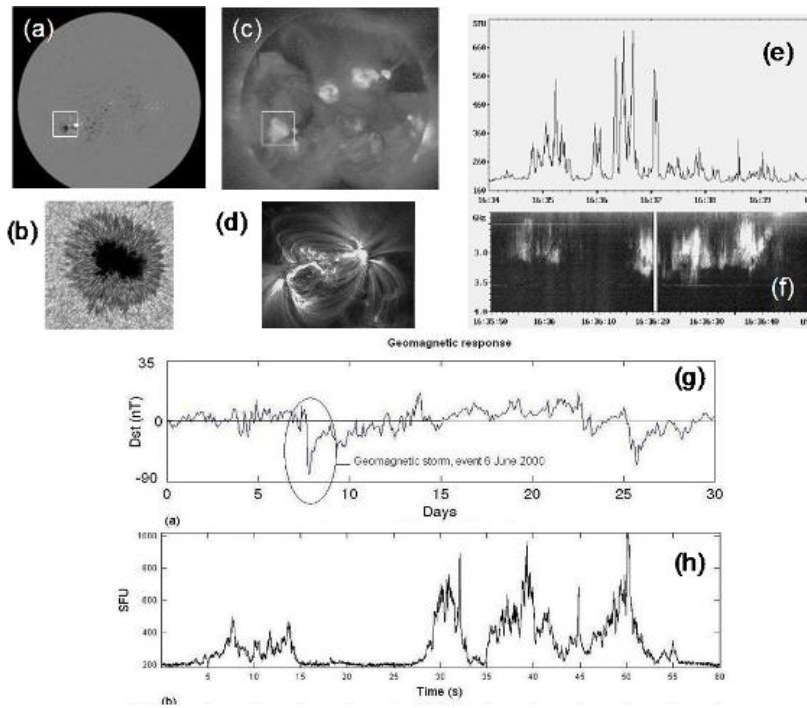


Figura 16 – Explosão solar geofetiva, observada em diferentes padrões, relacionada ao aparecimento de uma mancha solar, observada em (a) e com alta resolução na janela óptica ($H\alpha$) (b). Os respectivos padrões em Raios-X moles ($< 10\text{KeV}$) e UV são mostrados nas figuras (c) e (d). Os sinais captados em rádio são mostrados na frequência de 1.6 GHz – (e) e (f) – e na frequência de 3 GHz – (h). A resposta ao evento é detectada na magnetosfera terrestre em nT (nano Tesla) através do índice Dst (g).

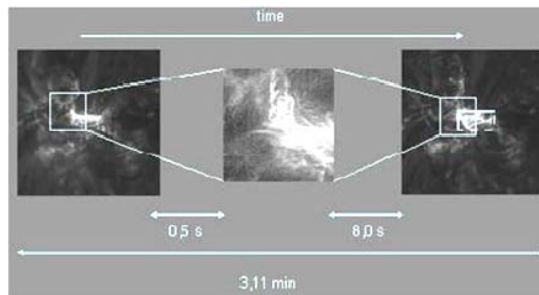


Figura 17 – Evolução espaço-temporal da região ativa solar observada em 3 GHz. Fonte: [17].

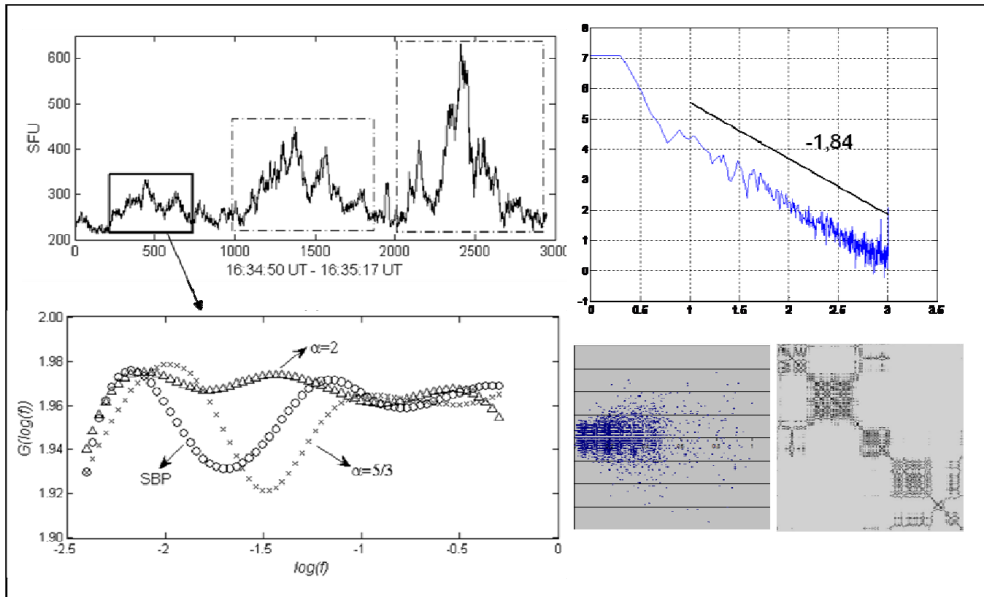


Figura 18 - Análise de uma pré-explosão evidenciando que um tipo de processo turbulento ocorre em todas as escalas de acordo com processo não-Gaussiano com espectro $1/f^{1.84}$. Veja o estudo completo sobre este evento em [13] e nos artigos indicados em [2].

Apêndice A – Exemplos de Processos Não-Lineares

Entre os processos discutidos no texto (oscilatórios periódicos e não-periódicos, caóticos, reativo-difusivos e multifractais tipo Browniano) destacamos, com especial interesse em ciências espaciais e atmosféricas: o *mapeamento de Henon*, a *equação KdV*, o *sistema de Lorenz*, o *sistema de Rössler*, as chamadas Equações de Amplitude (*Kuramoto-Sivashinsky*, *Ginzburg-Landau*, *Swift-Hohenberg*), além de outros. Uma descrição detalhada de cada um, contendo comentários sobre técnicas numéricas de solução, encontra-se disponível em [2].

Um exemplo de processo de grande interesse para a pesquisa de fenômenos não-lineares em física espacial é estudado através da equação de Kuramoto-Sivashinsky.

$$\partial_t A(x,t) = -\partial_x^2 A(x,t) - \nu \partial_x^4 A(x,t) - \partial_x u^2, \quad (3)$$

Onde ν é o parâmetro de amortecimento devido à viscosidade. Usualmente, a equação de KS pode ser resolvida através do método espectral [16] e um exemplo do padrão de variabilidade da componente puramente temporal é o da Figura 6b (Seção 2.1).

Apêndice B - Resolução, Sensibilidade e Freqüência de Amostragem

B.1. Séries Temporais Provenientes de Experimentos Reais

- Resolução (Tempo)

Considere uma série de N medidas no tempo, com resolução $\Delta t = \tau$ e, portanto, com duração dada por $\tau \times (N-1)$. Essa resolução temporal define a “frequência de amostragem” (f.a.) da série, que quantifica o número de medidas, por unidade de tempo, tomados regularmente para compor a série. Por convenção de unidade, uma série com f.a. igual a 1Hz significa dizer que o instrumento coletou 1 medida por segundo, logo temos $\tau = 1$. Portanto, se o instrumento trabalhou por 1 hora (60 s), a série será composta por $N = 60 \times 1 + 1 = 61$ pontos (quantos pontos foram coletados, considerando que o instrumento trabalhou apenas por 10 s?). Considere agora uma f.a. de 60 Hz, isto é, um processo capaz de executar 60 medidas em um segundo. Portanto, nesse caso, temos τ igual a $1/60 \cong 0.0167$ e podemos calcular N através de $(D(s)/\tau(s)) + 1$. Alternativamente: $N = [D(s) \times \omega \text{ (Hz)}] + 1$, onde D é a duração da série em segundos e ω é a f.a. em Hz. Considerando um processo de coleta de medidas durando 1h (60x60s) teremos, para o caso de $\omega = 60\text{Hz}$, $N = 3600 \times 60 + 1 = 216.001$ pontos. Note que, a quantidade 1 adicionada na fórmula acima tem baixo peso no resultado final quando N é grande ($N > 103$ pontos). Nesse caso, a fórmula pode ser apresentada como $N = D(s) \times \omega \text{ (Hz)}$. Isso implica uma periodicidade (ou regularidade) no processo de medida, mas não implica que haja variabilidade periódica da amplitude medida.

- Sensibilidade (Amplitude)

A medida da variação da amplitude depende da sensibilidade do instrumento e pode variar para uma mesma f.a. Entretanto muitas vezes é a f.a. que permite observar ou não pequenas variabilidades da amplitude. Nos exemplos abaixo, dois aspectos são importantes: (i) dada a faixa dinâmica da amplitude, não adianta aumentar a sensibilidade do velocímetro; (ii) Repare que a f.a. é de 1 medida por minuto, isto é, igual a ≈ 0.017 Hz. Se a frequência fosse de 1 medida a cada 5 minutos (0.0033 Hz) não observaríamos a variabilidade que a resolução do velocímetro admite. Estaríamos, por isso, perdendo informação sobre as escalas mais finas (flutuações da velocidade). Dessa forma, f.a. e sensibilidade instrumental definem uma relação íntima entre Δt e Δv . Veja um exemplo prático no texto disponível na *webpage* da ELAC. A partir desse exemplo, cada conjunto de medidas pode ser caracterizado por meio da sua média estatística e sua variância. Portanto, o aumento da f.a. melhora a representação estatística da amostra, uma vez que permite aumentar o N . Desse modo, podemos formular a seguinte pergunta: conhecendo um determinado fenômeno de variabilidade, qual é f.a. de um instrumento a ser construído com o objetivo de observar o fenômeno e caracteriza-lo através da sua variabilidade média? Ou ainda, para um instrumento com f.a. conhecida quantas medidas são necessárias para obter um valor da média que seja robusto e portando caracterize estatisticamente o processo em questão? Note que último exemplo, a baixa f.a. induz ao erro na caracterização do instante onde ocorre uma transição do regime de velocidades, indicando, no caso em questão, o fim do engarrafamento naquele trecho. Nesse caso talvez, a caracterização da transição não careça de alta precisão (e se for uma prestação de socorro?). Mas em vários outros casos, a precisão pode ser importante ou até vital (pense em exemplos equivalentes envolvendo missões espaciais). Na Figura B1, temos dois exemplos de amostras de séries temporais observadas no meio ambiente Terrestre com as suas respectivas frequências de amostragem.

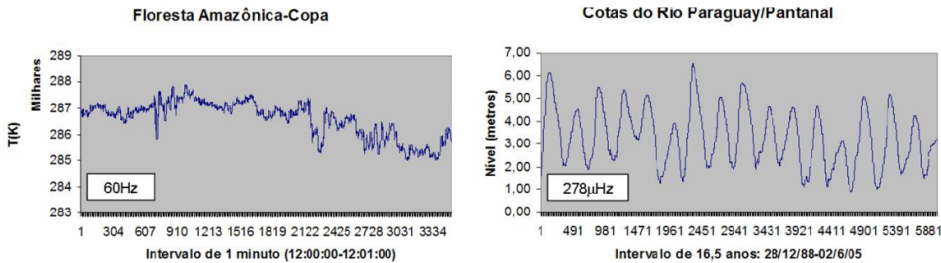


Figura B1 - Exemplos de séries provenientes de medidas em campo com instrumentos de alta (a) e baixa (b) frequências de amostragem.

B.2. Séries Temporais Provenientes de Experimentos Numéricos

No caos de séries temporais sintéticas, isto é, resultantes de uma simulação numérica, é muito importante estipular a sua frequência de amostragem tomando como referência, quando for o caso, a f.a. do dado experimental a ser modelado. Note que, para análise simples de séries temporais sintéticas não há limites para o valor de N . Portanto, tendo um par de valores de frequência e duração como referencia podemos gerar uma série temporal com N pontos. Por exemplo, gere uma série de mapa logístico que seja compatível com um sinal com frequência de 60 Hz observada durante 10 minutos. Isso nos dá $N=60 \times 10 \times 60 = 36.000$ pontos. Portanto, nessa série sintética, cada conjunto de 60 pontos equivale a um intervalo de 1s.

Agradecimentos – aos meus alunos Gustavo Zaniboni (Doutorando em Física pelo ITA), Mariana Baroni (Doutoranda CAP), Murilo Dantas (Mestrando CAP) e Rogério de Carvalho Brito (Mestre pela CAP) pelo apoio na produção de algoritmos e figuras. Aos professores H.S.Sawant, pelas figuras relacionadas ao projeto BDA, e H.L.Swinney, pelos dados das camadas granulares. Aos professores A.Assireu, A. Ferreira da Silva, A.Zanandrea, F.Ramos, J.Pontes, M.Bolzan e M.Domingues pela estimulante cooperação científica relacionada a alguns dos temas que inspiraram este minicurso.

Referências

- [1] <http://math.arizona.edu/~goriely/M322/M322-gui-cm.html>
- [2] <http://www.lac.inpe.br/ELAC/cursos.jsp>
- [3] Monteiro, L.H.A., Sistemas Dinâmicos, Editora Livraria da Física, 2002.
- [4] Peitgen, H-O; Jurgens, H.; Saupe, D., Chaos and Fractals, Springer-Verlag, 1992.
- [5] Sethena, J.P., Entropy, Order Parameters and Complexity, Oxford, 2006.
- [6] Salinas, S.R.A. Introdução à Física Estatística, Edusp, 2001.
- [7] Hamilton, J.D., Times Series Analysis, Princeton, 2003.
- [8] Rosa, R.R.; Baroni, M.P.M.A, Zaniboni, G.T., et al., Structural complexity of disordered surface, Physica A 386(2):666-673, 2007.