

IDENTIFICAÇÃO DE PADRÕES ESPACIAIS UTILIZANDO ONDALETAS.

Daniel GRADOS¹
Ronan FABLET²
Paulo Justiniano RIBEIRO JR ³
Arnaud BERTRAND ⁴

- RESUMO: blablaba
- PALAVRAS-CHAVE: Estruturas, Transformada de ondaleta, Potência da ondaleta.

1 Introdução

Em estudos ecológicos, a consideração de estrutura espacial é muito importante pois permite estudar a dinâmica dos processos ecológicos. Populações são espacialmente estruturadas em especial porque processos ecológicos interagem com seus vizinhos criando autocorrelação (TILMAN; KAREIVA, 1997). Processos ecológicos operando numa escala tem contribuição na formação de estruturas ao longo de diferentes escalas (WIENS, 1989; LEVIN, 1992; DUNGAN et al., 2002), sendo que os ecossistemas de upwelling são muito heterogêneos a processos de submesoescala.

Assim existe a necessidade de identificar estruturas dos processos do ecossistema para compreender o funcionamento deles. Neste sentido existem métodos usados para identificar as escalas características em processos, dentro destes métodos temos os variogramas. Os variogramas representam a estrutura

¹Programa de Pós-Graduação em Estatística e Experimentação Agrônoma, Universidade de São Paulo, CEP:81531-990, Piracicaba, São Paulo, Brasil , E-mail: *danny.grados@gmail.com*

²Departamento de Sinais e Comunicações, Telecom Bretagne, Brest, Brest, France , E-mail: *ronan.fablet@telecom-bretagne.eu*

³Laboratório de estatística e geoinformação, Universidade Federal do Paraná, CEP:81531-990, Curitiba, Paraná, Brasil , E-mail: *paulojus@leg.ufpr.br*

⁴Institut de Recherche pour le Développement, EME UMR212, Teruel N 357, Miraflores, Lima, Peru, E-mail: *arnaud.bertrand@ird.fr*

espacial num parâmetro chamado alcance, outro método desenvolvido por Bellier et al., (2002) permite fazer uma decomposição multiescala da variável estudada. Mas quando as variáveis não são estacionários estes métodos tem dificuldade metodológica na estimação dos parâmetros de interesse. Um método que permite trabalhar com variáveis não estacionárias são as ondaletas. As ondaletas tem a capacidade de decompor um sinal em múltiplas escalas

2 Desenvolvimento do método

Nesta seção serão descritos o procedimentos do método desenvolvido.

2.1 Extração multi-escala das estruturas características

Pela necessidade de identificar escalas características de um processo, onde pequenas estruturas podem estar dentro de grandes estruturas, no trabalho vai-se desenvolver uma metodologia baseada no método de ondaletas o qual vai permitir identificar múltiplas escalas dos processos físicos.

Na Figura 1 observa-se os parâmetros de interesse, as quais são a amplitude e período. O método desenvolvido vai estimar aqueles dois parâmetros.

2.1.1 Transformada de ondaleta

A transformada de ondaleta envolve uma operação linear que pode ser utilizada na análise de sinais não estacionários (DAUBECHIES, 1990; TORRENCE; COMPO, 1998). A transformada permite extrair informações das variações em frequência dos sinais e para detectar estruturas localizadas no tempo ou espaço.

Neste sentido é usado uma função ondaleta mãe, representado por $\psi(t)$. A escolha da função ondaleta ψ_t não é arbitrária (CAZELLES et al., 2007). Esta função é normalizada $\int \|\psi(t)\|^2 dt = 1$ e verifica-se que $\int \psi(t) dt = 0$. A decomposição da ondaleta é, portanto, uma representação linear onde a variância é preservada (DAUBECHIES, 1990). Isto quer dizer que a série (ou sinal) original pode ser reconstruída mediante uma transformada inversa de ondaleta.

Utilizando a função mãe pode-se chegar à transformada de ondaleta, a transformada de ondaleta decompõe uma sinal utilizando a função mãe. A transformada tem a seguinte forma:

$$W_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t - \tau}{a}\right) dt = \int_{-\infty}^{\infty} x(t) \psi_{a, \tau}^*(t) dt \quad (1)$$

Por enquanto * denota a conjugada complexa, os parâmetros a e τ representam a dilatação (o fator de escala) e translação, respectivamente. Os coeficientes de ondaletas, $W_x(a, \tau)$, representam a contribuição das escalas a a diferentes posições do tempo τ . Na equação anterior o fator $1/\sqrt{a}$ esta normalizando a função ondaleta assim tem-se variância igual a 1 e, portanto, são comparáveis a todas as escalas a . A idéia central da transformada de ondaleta é de decompor

o sinal analisado em diferentes níveis de frequência; tal processo é conhecido como multiresolução. Esta forma de representação do sinal fornece uma estrutura hierárquica simples.

Como foi mencionado, a transformada de ondaleta utiliza a função mãe para decompor a sinal, assim um ponto muito importante é a escolha da função mãe. Existem muitos critérios que tem que se definir na escolha da função mãe, como por exemplo, ondaleta real vs ondaleta complexa, ondaleta contínua vs ondaleta discreta, ortogonal vs não ortogonal. Em resumo, as ondaletas contínuas frequentemente fazem uma decomposição redundante, isto quer dizer que a informação extraída em um determinado intervalo de escalas pega informação das escalas dos vizinhos (TORRENCE; COMPO, 1998), mas são mais robustos ao ruído em comparação com outras ondaletas. As ondaletas discretas tem a vantagem de uma rápida implementação, mas pelo geral o número de escalas e a característica invariante no tempo dependem fortemente do número de dados. Todas as famílias das ondaletas tem uma característica geral: oscilações lentas têm uma boa frequência e uma ruim resolução temporal, por enquanto que, as oscilações rápidas têm uma boa resolução temporal, mas uma ruim resolução em frequência.

Assim, para a ondaleta contínua as duas funções mais conhecidas são: ondaleta Hat e ondaleta Morlet. A escolha da função ondaleta pode influir na resolução do tempo e escala da decomposição. Portanto, como a função Morlet tem uma boa localização em escalas (e frequência), então se espera uma resolução de alta frequência. Em contraparte, a função ondaleta Hat tem uma boa localização no tempo, mas uma ruim localização na frequência. No presente trabalho utilizou-se a função Morlet, que tem a seguinte forma:

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2} \quad (2)$$

Por enquanto ω_0 representa a frequência angular central ($\omega_0 = 2\pi f_0$) e é o valor que determina a forma do sinal modular. A ondaleta é o produto de uma função senoidal complexa $e^{i\omega_0\eta}$ e um envelope gaussiano $e^{-\eta^2/2}$. Assim quando a frequência angular central (ω_0) é perto de 2π , a escala da função ondaleta é inversamente proporcional à frequência central da ondaleta, $f \approx 1/a$. Esta aproximação é de muita ajuda na hora de interpretação das análises de ondaletas. Uma vantagem de usar a função ondaleta Morlet é que permite o estudo de sinais a cada escala, suas fases e a amplitude do sinal (CAZELLES et al., 2008). Muitas vezes a forma de observar a informação da transformada de ondaleta é mediante o espectro da potência de ondaleta. A potência de ondaleta é definido como:

$$S_x(a, \tau) = |W_x(a, \tau)|^2 \quad (3)$$

a $S_x(a, \tau)$ estima a distribuição da variância entre a e diferentes τ . Uma representação usada para o espectro da potencia de ondaleta é fazer um gráfico 2D do $S_x(a, \tau)$, onde no eixo x representa o tempo (ou espaço) e no eixo y representa as escalas (ou frequências). O espectro de Fourier pode ser comparado com o espectro global da potência de ondaleta, que é definido como a média de energia (variância média) contido em todos os coeficientes de ondaletas de igual escala a .

$$\bar{S}_x(a) = \frac{\sigma_x^2}{T} \int_0^T \|W_x(a, \tau)\|^2 d\tau \quad (4)$$

Com σ_x^2 variância da série x e T a duração do tempo na série. A equação anterior apresenta a variância para cada período (ou escala). Mas, também se pode representar a variância para cada tempo, ficando a equação do seguinte forma.

$$\bar{S}_x(\tau) = \frac{\sigma_x^2 \pi^{1/4} \tau^{1/2}}{C_g} \int_0^T a^{1/2} \|W_x(a, \tau)\|^2 da \quad (5)$$

Para identificar zonas significativas, de maior energia, do espectro da potência da transformada de ondaleta comparou-se o espectro da potência de ondaleta com outro espectro apropriado. Para muitos processos (físicos, geográficos, etc) um espectro de comparação muito utilizado é o ruído vermelho (TORRENCE; COMPO, 1998), assim o espectro do sinal em estudo é comparado com o espectro de ruído vermelho. A série do ruído vermelho pode ser modelada utilizando métodos de Monte Carlo. Um modelo usado é um processo auto regressivo $AR(1)$ ($x_n = \alpha x_{n-1} + z_n$) por enquanto α é o $lag - 1$.

Segundo Gilman, Fuglister e Mitchell (1963), a transformada da potência discreta de Fourier é dado como:

$$P_k = \frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos(2\pi k/N)} \quad (6)$$

A equação anterior mostra a forma de calcular o espectro de energia de fundo vermelho em função da autocorrelação da série analisada. Na qual k pode tomar valores de 0, 1, ... $N/2$. O nível de 95% no espectro local da ondaleta implica em dizer que a energia na potência de ondaleta é maior do que a energia no ruído vermelho. O nível de significância é obtido baseado na seguinte aproximação:

$$\frac{|W_x(a, \tau)|^2}{\sigma^2} \approx \frac{1}{2} P_k \chi_2^2 \quad (7)$$

χ_2^2 representa a distribuição chi-quadrada com 2 graus de liberdade.

2.1.2 Estimação do período e amplitude

As zonas do espectro da potência de ondaletas maiores que 95% de significância foram usados para identificar os máximos locais de energia, assim ao redor de cada máximo local de energia definiu-se um área que contem a energia. Na Figura 2b observa-se duas caixas, elas representam a informação de dois máximos locais, os quais foram gerados com diferentes períodos. A informação proporcionada pelo máximo local de energia foi usada para estimar o valor do período (com a ideia de que com este valor chega-se a gerar um valor de energia similar ao do máximo local). Matematicamente a estimação do período tem a seguinte forma:

$$\hat{\theta} = \theta_{max(S_x(a,\tau))} \quad (8)$$

sendo que θ representa ao período e os valores de a são os limites de período das caixas identificadas anteriormente.

A estimação da amplitude do sinal foi realizada utilizando a informação contéudo na zona ao redor do máximo local de energia. Para isto foi reconstruída o sinal para aquela zona. A propriedade de reconstrução do sinal foi utilizada para estimara amplitude. A reconstrução da sinal 'e feita como segue:

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^J \frac{\Re(W_x(a, \tau_j))}{\tau_j^{1/2}} \quad (9)$$

Por enquanto, o fator $\psi_0(0)$ remove a energia da escala, assim também, o fator C_δ é calculado como:

$$C_\delta = \frac{\delta j \delta t^{1/2}}{\psi_0(0)} \sum_{j=0}^J \frac{\Re(W_x(a, \tau_j))}{\tau_j^{1/2}} \quad (10)$$

para cada estrutura foi estimado a amplitude utilizando o sinal reconstruído.

Finalmente para estimar a fase da série $x(t)$ calculou-se a fase, o qual tem uma variação de $-\pi$ ate π e a forma de calcular a fase é com o razão da parte imaginária \Im entre a parte real \Re dos coeficientes de ondaletas $W_x(a, \tau)$ e esta dado como:

$$\phi_x(a, \tau) = \tan^{-1} \frac{\Im(W_x(a, \tau))}{\Re(W_x(a, \tau))} \quad (11)$$

a informação proporcionada nesta etapa é importante e vai servir para fazer a caracterização das estruturas obtidas. Sendo estas de côncavas ou convexas.

Um exemplo do processo de obtenção de objetos esta na Figura 2, onde o sinal da Figura 2a foi gerada mediante a soma de três séries, duas delas foram geradas da função $seno(periodo)$, para o qual utilizou-se dois valores de $periodo$ (12 e 30) e a terceira série foi um ruído branco com média zero e variância um. À série da Figura 2a aplicou-se a transformada de ondaleta, na Figura 2b é a representação do espectro de ondaleta onde com linhas pretas estão representadas as zonas significativas. Estas zonas foram onde a energia do espectro de ondaleta foi maior do que o espectro do ruído vermelho utilizado como nível de significancia. O espectro de ondaleta apresentada na Figura 2b foi utilizado para identificar as zonas de máximos locais de energia. Sendo que na Figura 2c estão representadas com caixas de cor vermelho. Assim, utilizando a infomação contida em cada caixa estimou-se o período (para cada caixa). Finalmente, para cada caixa recontriuse os sinais a cada escala (ou período). Na Figura 2b pode-se observar dois sinais reconstruídas as quais corresponden às duas séries utilizadas para gerar o sinal original, também estimou-se a fase de cada série, sendo a fase representada por os círculos pretos.

2.1.3 Indicador de eficiência

Já desenvolvido o método de identificação de estruturas vai-se propor um índice de eficiência. Assim, um indicador de eficiência muito utilizado é o erro quadrado médio (EQM). O indicador está representado como:

$$EQM = \frac{1}{N} \sum_{i=1}^N (\theta - \hat{\theta})^2 \quad (12)$$

sendo que θ representa ao parâmetro teórico e $\hat{\theta}$ representa ao parâmetro estimado, i é o número de estrutura e N o número de estruturas identificadas. Este indicador vai-se usar para quantificar o erro de estimação para o período e amplitude.

2.1.4 Validação do método

Para estudar a capacidade de identificar os parâmetros período e amplitude vai-se gerar diferentes séries com diferentes características dos sinais. A primeira família de séries tem a seguinte forma $y = Amplitude \times seno(periodo)$, onde o valor de amplitude foi elegido ao acaso de uma distribuição uniforme ($Amplitude \sim U(a, b)$), da mesma forma o parâmetro período foi gerado de uma distribuição uniforme ($periodo \sim U(m, n)$). Estas séries são sem ruído. A segunda família gerada foram as séries com ruído branco, as séries tem a seguinte estrutura $y = Amplitude \times seno(periodo) + \varepsilon$, onde ε representa ao ruído branco. Sendo que $\varepsilon \sim N(0, 1)$. Finalmente a terceira família utilizada foram séries com ruído vermelho e tem a seguinte forma $y = Amplitude \times seno(periodo) + ruido$, sendo que $ruido$ é um modelo autoregressivo de ordem um ($ruido \sim Ar(1)$). O método foi aplicado as séries geradas, assim estimou-se os parâmetros período e amplitude que foram usadas para gerar as séries.

3 Resultados

3.1 Validação do método de extração de estruturas

Com o objetivo de testar se a metodologia proposta tem a capacidade de identificar as estruturas de um sinal, foi gerado sinais com diferentes valores de amplitude e período e foi aplicado o método desenvolvido. Especificamente foi geradas séries de senos onde o valor de amplitude e período gerou-se aleatoriamente utilizando uma distribuição uniforme, depois foram simuladas séries de senos mais um ruído branco, em terceiro lugar simularam-se séries de senos mais um ruído vermelho, assim finalmente se gerou séries mais complexas, o qual consistiu na soma de senos com diferentes valores de amplitude e período, ao qual foi adicionado séries de ruído vermelho.

Identificação de estruturas em funções seno

Como primeiro modelo, simularam-se funções seno com parâmetros de amplitude e período conhecida. A amplitude e período foram gerados a partir de uma distribuição uniforme. O método desenvolvido foi aplicado às séries simuladas. As séries simuladas y tem a seguinte estrutura: $y = Amplitude * seno(período)$. 1000 amostras foram geradas com períodos que vão desde 2 até 60 ($período \sim U(2, 60)$) e amplitudes que tem variação entre 5 e 40 ($amplitude \sim U(5, 40)$), a escolha deles foi ao acaso. Os resultados das séries simuladas estão apresentadas na Tabela 1. Observa-se há um que 98.5% de identificação dos parâmetros utilizados, também existe um erro quadrático médio (EQM) igual que 4.8 para a estimação do período e de 7.5 para a estimação da amplitude pequeno para a o período e a amplitude. Um valor de correlação de 0.99 entre os parâmetros teóricos e estimados.

Identificação de estruturas em funções seno com ruído branco

Às séries já utilizadas acrescentaram-se ruído branco (ε) e aplicou-se o método de identificação de estruturas. O ruído branco foi gerado utilizando uma normal com média zero e variância um ($\varepsilon \sim N(0, 1)$). Os resultados podem ser observados na Tabela 2. Das 1000 séries que representam o 100% só o 3% foram mal identificadas, observou-se que existe erro de estimação nas bordas do espectro da potência de ondaleta.

Identificação de estruturas em funções seno com ruído vermelho

Na Tabela 3 observa-se que dos 1000 sinais simulados com ruído vermelho o 96% de estruturas foram identificadas corretamente, também observa-se que o EQM para o período é amplitude forma similares aos encontrados nos casos anteriores.

Identificação de estruturas em funções complexas

Séries complexas foram geradas, as quais são combinação de sub séries. Para os fins da simulação usamos 3 séries mais um ruído vermelho, o qual foi composto como: $y = Ampl_1 * seno(per_1) + Ampl_2 * seno(per_2) + Ampl_3 * seno(per_3) + ar(1)$ sendo que a escolha dos valores de $Ampl_1$, $Ampl_2$, $Ampl_3$, per_1 , per_2 e per_3 são aleatórios, para o ruído vermelho $ar(1)$ foi usada um valor de ρ de 0.85. Assim, se tem uma série complexa, o método de extração foi aplicado à séries de dados simulados. Na Figura 3 do lado esquerdo observa-se um exemplo de uma série simulada o qual tem três diferentes valores de amplitude e período como esta mostrada na figura, ao lado direito se observa o sinal resultante e seu respectivo espectro de ondaleta. As caixas vermelhas (dentro do espectro da potência de ondaleta) representam as zonas de máxima energia, onde são reconstruídas as séries originais. Na figura abaixo são mostradas as três séries reconstruídas. Os valores de período foram estimados corretamente, para a amplitude observou-se uma subestimação perto das bordas do espectro.

Em geral existe uma boa identificação dos parâmetros usados (período e amplitude), com efeito das bordas na estimação da amplitude.

Escolha do valor da frequência central ω_0

A escolha do valor da frequência central é muito importante na decomposição dos sinais, o valor vai depender das características do sinal estudado. Avaliaram-se os sinais simulados com um primeiro valor de $\omega_0=6$, e assim contrastamos outros valores do parâmetro. Como já foi falada a escolha do parâmetro depende das características observadas nos sinais estudados, para séries com alta frequência (alta variabilidade a pequena escala) foi usado valores de ω_0 entre 6, 7 e alguns casos 8. Estes valores permitem o estudo de estruturas de alcance pequeno. Para séries de baixa frequência (variabilidade a grandes escalas) foram usados valores entre 3, 4 e 5. os quais permitem identificar estruturas de grandes escalas.

Observou-se também que a estimação do parâmetro de amplitude foi sensível para mudanças do valor de ω mais a estimação do período mostrou-se robusto para mudanças do valor de ω_0 , isto foi porque a reconstrução do sinal leva como variável o valor de ω_0 e a partir do sinal reconstruída é estimado a amplitude.

Discussão e recomendações

No presente trabalho foi desenvolvido um método para identificar estruturas presentes em diferentes sinais. O método tem a capacidade de estimar dois parâmetros de interesse (período e amplitude). O método de ondaletas, além de estimar o tamanho da estrutura (período), permite localizar sua a posição no espaço ou tempo (TORRENCE; COMPO, 1998). Grados et al. (submetido) Mostrou que as ondaletas tem a capacidade de identificar diferentes escalas nas sinais físicas.

Os estudos de simulação permitiram observar que o método desenvolvido tem a capacidade de identificar corretamente os parâmetros de interesse, tendo resultados muito bons na identificação do período e a amplitude. Estes dois parâmetros são o tamanho e a altura da estrutura identificada.

A escolha do parâmetro de frequência central é um tema muito importante na decomposição do sinal, este parâmetro permite o ajuste a pequenas ou grandes escalas. Os resultados indicam que o valor da frequência tem influencia na estimação da amplitude. Outro aspecto importante foi a família usada, no trabalho utilizou-se à família Morlet, pois permite obter dois tipos de informação (a fase e amplitude) (CAZELLES at al, 2007; TORRENCE; COMPO, 1998).

Uma vantagem do método é que pode ser utilizada em outras áreas de estudo, como para estudar comportamentos dos peixes no ecossistema. Também se pode aplicar a dados de ecologia terrestre.

Agradecimentos

GRADOS, D. ; FABLET, R. ; RIBEIRO Jr., P.J.; BERTRAND, A., Identificação de padrões espaciais utilizando ondaletas. *Rev. Bras. de Biometria* São Paulo, v.xx, n.x, p.xx-xx, xxx. *Rev. Mat. Estat.* (São Paulo), v. x, n.x, p. xx-xx, x.

■ *ABSTRACT: blablaba*

■ *KEYWORDS:*

Recebido em XX.XX.XXXX. Aprovado após revisão em XX.XX.XXXX.

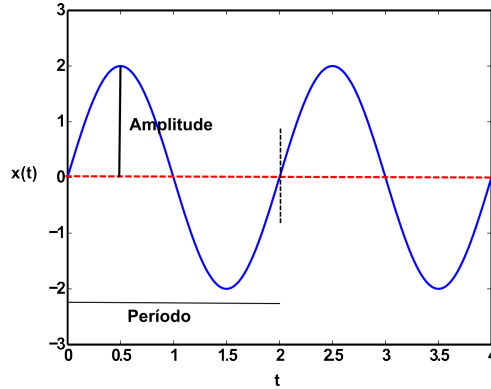


Figura - 1: Elementos do sinal, a amplitude e o período

Tabela - 1: Percentagem de identificação de estruturas e EQM para sinais seno

| Parâmetro | Identificação | EQM | ρ |
|-----------|---------------|-----|--------|
| Período | 98.5% | 4.8 | 0.99 |
| Amplitude | 98.5% | 7.5 | 0.99 |

Tabela - 2: Percentagem de identificação de estruturas e EQM para sinais com ruído branco

| Parâmetro | Identificação | EQM | ρ |
|-----------|---------------|-----|--------|
| Período | 97% | 3.2 | 0.9 |
| Amplitude | 97% | 7.9 | 0.9 |

Tabela - 3: Percentagem de identificação de estruturas e EQM para sinais com ruído vermelho

| Parâmetro | Identificação | EQM | ρ |
|-----------|---------------|-----|--------|
| Período | 96% | 3.2 | 0.9 |
| Amplitude | 96% | 7.9 | 0.9 |

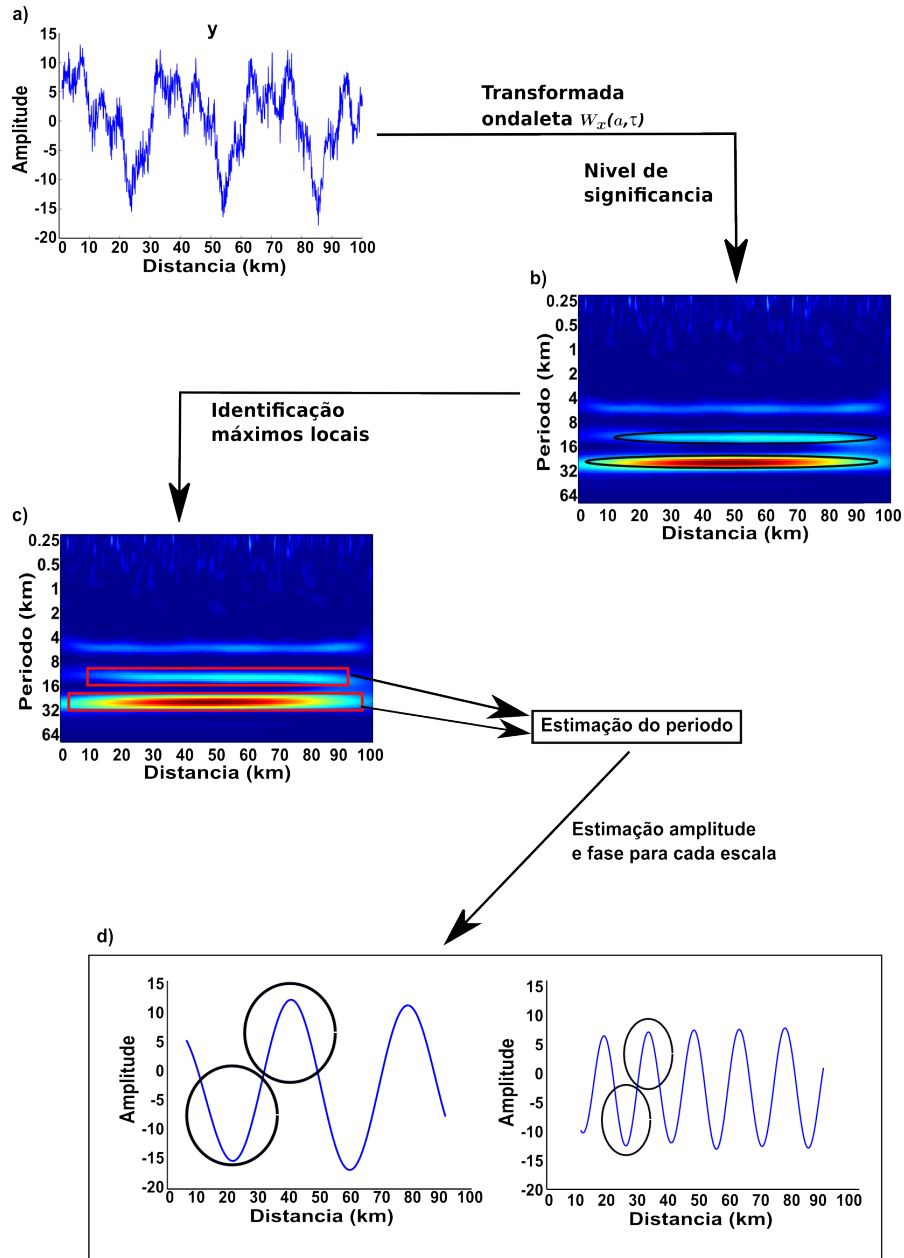


Figura - 2: Esquema da obtenção de estruturas de meso e submesoescala

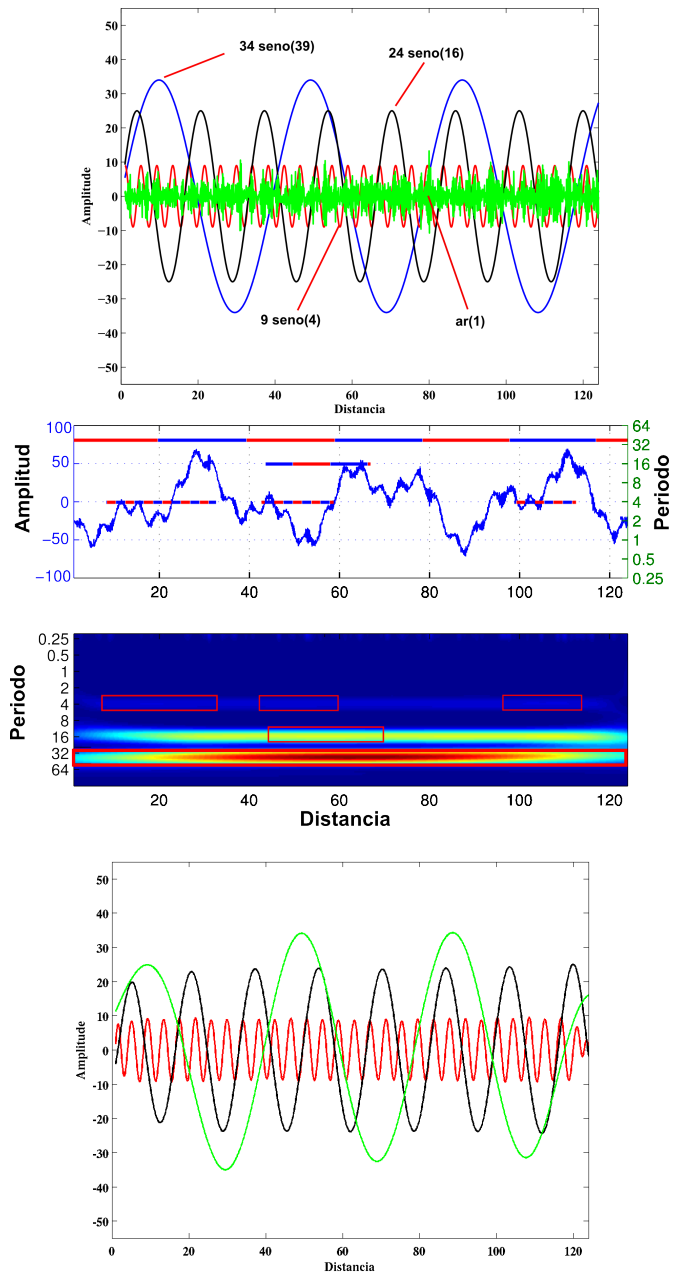


Figura - 3: Sinal simulado utilizando 4 séries diferentes