

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Identificação de estruturas de processos multi-escala
em ecossistemas marinhos utilizando ondaletas**

Daniel Isaias Grados Paredes

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Estatística
e Experimentação Agronômica

**Piracicaba
2011**

Daniel Isaias Grados Paredes
Engenheiro Estatístico

**Identificação de estruturas de processos multi-escala
em ecossistemas marinhos utilizando ondaletas**

Orientador:
Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Estatística
e Experimentação Agronômica

**Piracicaba
2011**

Dedicatória

A minha mãe, Lidia Paredes,
A meus irmãos, José, Richard, Javier e Ronald,
que sempre me apoiaram e estiveram ao meu lado.
A Rocio por todos os momentos felizes.

AGRADECIMENTOS

Primeiramente a minha mãe Lidia e meus irmãos por estar sempre ao meu lado, mesmo a quilômetros de distancia.

A Arnaud e Sophie que acreditaram em mim, Obrigado Arnaud pela ajuda e os conselhos. Obrigado Sophie por a amizade.

A Ronan Fablet por teu constante apoio e constante ensino da programação é interpretação dos meus dados.

Ao Paulo Justiniano por teu ensino da área de estatística espacial. Por o continuo apoio do meu trabalho, por as sugestões em meu trabalho.

A minha namorada Rocio, você é a pessoa que esta quando eu mais preciso de alguém, é minha companheira do meu caminho da vida. Obrigado por acreditar nas coisas que eu acredito.

A minhas amigas Zayda e Rosmery que fizeram muito confortável minha estadia em IMARPE, obrigado Zayda pelas brincadeiras.

A meus amigos Juan, Nancy, Silvia, Veronika, Andres, Adriana, Javier e a todos os Peruanos e Colombianos pela companhia em Piracicaba.

A meus amigos do mestrado Ana Julia, Gabriel, João, Josy, Faby, Kuang, Luiz, por sempre estarem ao meu lado.

A Ana Julia, muito obrigado Ana pela continua revisão do português, se no for por você meu trabalho teria ainda mais portunhol, eu devo muitos piscos a você.

Às secretárias do programa de pós graduação: Luciane e Solange por toda a ajuda, em especial a Luciane por as muitas ajudas que eu solicitei.

Aos funcionários do programa de pós graduação: Eduardo e Jorge.

Aos professores do programa de pós graduação em Estatística e Experimentação Agronômica: Carlos Tadeu dos Santos Dias, Edwin Moisés Marcos Ortega, Silvio Sandoval Zocchi, Sônia Maria De Stefano Piedade e Taciana Villela Savian pelos conselhos e ajuda nos momentos de dúvidas.

Agradeço em especial à professora Clarice Garcia Borges Demétrio por sua dedicação nas aulas, você vai ser um exemplo para mim.

Ao IRD pela constante apoio oferecida a minha pessoa.

Ao IMARPE por fazer possível meu trabalho no campo de ecologia marinha.

Por fim a todos que sempre acreditaram em mim.

OBRIGADO!!!!!!

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 DESENVOLVIMENTO	23
2.1 Materiais e Métodos	23
2.1.1 Área de estudo	23
2.1.2 Variáveis	23
2.1.3 Extração multi-escala das estruturas características	24
2.1.4 Caracterização da distribuição das estruturas	30
2.1.4.1 Indicador de eficiência do método de extração de estruturas	33
2.1.5 Estudo do impacto das estruturas físicas sobre a distribuição do zooplâncton	33
2.1.5.1 Independência dos processos físicos e a distribuição do zooplâncton	33
3 RESULTADOS E DISCUSSÃO	37
3.1 Validação do método de extração de estruturas	37
3.1.1 Aplicação da metodologia aos dados do Ecossistema marinho	42
3.2 Tipologia das estruturas	43
3.2.1 Estudo do impacto das estruturas sob a distribuição do zooplâncton	49
3.2.1.1 Estudo da independência entre os dois processos	56
3.3 Discussão	57
3.3.1 Da metodologia	57
3.3.2 Tipologia das estruturas físicas	58
3.3.3 Do impacto sobre a biomassa de zooplâncton	59
4 CONCLUSÕES	61
REFERÊNCIAS	63

RESUMO

Identificação de estruturas de processos multi-escala em ecossistemas marinhos utilizando ondaletas

Ecossistemas marinhos de *upwelling* são muito heterogêneos e apresentam uma intensa atividade de mesoescala de dimensão de dezenas de quilômetros e submesoescala que variam de centenas de metros até quilômetros dos processos físicos. A importância das estruturas dos processos físicos está na estruturação que eles exercem sob a biomassa de zooplâncton. O presente trabalho está relacionado a um estudo levado a cabo no Norte do Sistema da Corrente de Humboldt (Peru). Utilizou-se duas variáveis, a profundidade do limite superior da zona de mínimo oxigênio (ZMO) e a biomassa de zooplâncton. É desenvolvida uma metodologia de análise baseada no uso de ondaletas para a identificação e extração das estruturas dos processos físicos em suas diferentes escalas. O método foi aplicado aos dados de ZMO. Estudos de simulação mostraram que o método tem a capacidade de identificar as estruturas de interesse, tendo erro de estimação nas bordas do espectro da potência de ondaleta. A tipologia das estruturas identificadas mostraram que existe três tipos de estruturas, estruturas grandes e muito energéticas, duas estruturas pequenas com profundidades diferentes. Outro resultado importante foi que dentro das estruturas pequenas e mais profundas existe maior biomassa de zooplâncton, principalmente nas estruturas de *downwelling*.

Palavras-chave: Processos físicos; Zooplâncton; Análise de ondaletas; Espectro de ondaleta

ABSTRACT

Identification of structures of multi-scale processes in marine ecosystems using wavelet analysis

Marine upwelling ecosystems are very heterogeneous and have intense physical processes of mesoscale, of tens of kilometers, and submesoscale, ranging from hundreds of meters to kilometers. The importance of the structures of the physical processes is the role they play in structuring the biomass of zooplankton. This work was carried out in northern Humboldt Current System (Peru). Two variables were used: the depth of the upper limit of the oxygen minimum zone (ZMO) and biomass of zooplankton. In this work we developed an analysis method based on wavelet analysis for identification and extraction of structures of physical processes at their different scales. The method was applied to the ZMO data. Simulation studies showed that the method has the ability to identify the structures of interest with estimation errors in the boundaries of the power spectrum of the wavelet. A typology of the structures give us three large and very energetic structures corresponding to mesoscale processes, and two other small ones of different depths corresponding to submesoscale processes. Another important result was that within the smaller and deeper structures there is more biomass of zooplankton, especially in downwelling structures.

Keywords: Physical processes; Zooplankton; Wavelet methods; Wavelet spectrum

LISTA DE FIGURAS

Figura 1 - Esquema do conceito de transferência bottom-up. Fonte: Bertrand, S. et al., (2008)	18
Figura 2 - Esquema das estruturas físicas e seus impactos sobre os organismos	19
Figura 3 - Representação de um ecograma acústico	24
Figura 4 - Elementos do sinal, a amplitude e o período	24
Figura 5 - Esquema da obtenção de estruturas de meso e submesoescala	29
Figura 6 - Sinal simulado com período de 47 e amplitude de 7	38
Figura 7 - Sinal simulado com período de 20 e amplitude de 15	38
Figura 8 - Sinal simulado utilizando 4 séries diferentes	41
Figura 9 - Resultado da aplicação do método ao um transecto	42
Figura 10 - Representação do espaço dos componentes principais	45
Figura 11 - Dendograma com os três <i>cluster</i>	46
Figura 12 - Representação dos 3 grupos de estruturas identificadas	49
Figura 13 - K-cruzada de Ripley	57

LISTA DE TABELAS

Tabela 1 - Percentagem de identificação de estruturas e EQM para sinais seno	38
Tabela 2 - Percentagem de identificação de estruturas e EQM para sinais com ruído branco	39
Tabela 3 - Percentagem de identificação de estruturas e EQM para sinais com ruído vermelho	40
Tabela 4 - Variáveis endógenas e exógenas	43
Tabela 5 - Matriz variância covariância	44
Tabela 6 - Porcentagem de variância explicada	45
Tabela 7 - Cargas das variáveis	46
Tabela 8 - Descrição das variáveis categóricas para os <i>cluster</i>	47
Tabela 9 - Variáveis contínuas nos grupos	48
Tabela 10 - Média da biomassa de zooplâncton dentro do grupo 1	50
Tabela 11 - Anova para o zooplâncton no grupo 1	51
Tabela 12 - Teste de Tukey para o zooplâncton no grupo 1	51
Tabela 13 - Anova no grupo 1 para a diferencial de biomassa	51
Tabela 14 - Teste de Tukey para o diferencial de biomassa	52
Tabela 15 - Média da biomassa segundo os fatores (dia, zona e fase) no grupo 2	52
Tabela 16 - Anova para a biomassa do zooplâncton dentro do grupo 2	53
Tabela 17 - Teste de Tukey para a biomassa de zooplâncton no grupo 2	53
Tabela 18 - Anova para o diferencial da biomassa no grupo 2	54
Tabela 19 - Teste de Tukey para o diferencial de biomassa no grupo 2	54
Tabela 20 - Média da biomassa segundo os fatores (dia, zona e fase) no grupo 3	55
Tabela 21 - Anova da biomassa de zooplâncton no grupo 3	55
Tabela 22 - Teste de Tukey para a biomassa de zooplâncton dia no grupo 3	55
Tabela 23 - Anova do diferencial da biomassa no grupo 3	56
Tabela 24 - Teste de Tukey para o diferencial da biomassa de zooplâncton no grupo 3 .	56

1 INTRODUÇÃO

O acontecimento das interações entre predadores e presas é dado pela sobreposição espacial deles (predadores e suas presas). Há evidências que em ecossistemas marinhos existe uma transferência 'bottom-up' (de baixo para cima) da estruturação espacial (RUSSELL et al., 1992; BERTRAND, S. et al., 2008): A transferência começa nos processos físicos, os processos físicos estruturam a paisagem da comunidade pelágica mediante as turbulências nas massas de água. As partículas inertes (nutrientes) e uma parte dos seres vivos (fitoplâncton e grande parte do zooplâncton) são passivamente organizadas no espaço por causa dos processos físicos. Logo, pela necessidade de se alimentar, os peixes condicionam sua distribuição segundo a distribuição de suas presas (FRONTIER et al., 2004) e pela sucessão de relações predador-presa, a estrutura espacial originada pelos processos físicos são transferidos a todos os níveis tróficos do ecossistema (Figura 1). A identificação e explicação da variação espacial da distribuição dos organismos é um tema muito importante em ecologia (DALE et al., 2002). Populações são espacialmente estruturadas por diversos fatores, em particular (i) os processos ecológicos são intrinsecamente espaciais, dado que as interações entre indivíduos vizinhos gera auto-correlação (TILMAN; KAREIVA, 1997), e (ii) as espécies respondem às variações meio-ambientais, como estes últimos estão estruturados, induzem dependência espacial da distribuição das espécies. A resposta das espécies às variações de seu entorno acontece a diferentes escalas características e estas respostas podem variar em função do alcance de escalas em consideração.

Dentro destes ecossistemas marinhos estão os sistemas de afloramento, os quais são particularmente heterogêneos a pequenas escalas de espaço e tempo (LAVANIEGOS et al., 2002; CARR; KEARNS, 2003; BERTRAND, A. et al., 2008; BERTRAND, S. et al., 2008; CHAIGNEAU; GIZOLME; GRADOS, 2008). A intensa atividade de mesoescala de dimensão de dezenas de quilômetros e submesoescala, que variam de centenas de metros até quilômetros geram frentes entre águas costeiras (com muito nutriente) e águas oceânicas de poucos nutrientes (CHAIGNEAU; PIZARRO, 2005), filamentos (THOMAS, 1999), redemoinhos (HORMAZABAL; SHAFFER; LETH, 2004; CHAIGNEAU; PIZARRO, 2005) e ondas internas (BERTRAND, A. et al., 2008, BERTRAND; BALLON; CHAIGNEAU, 2010). Nesta paisagem, as atividades de meso e submesoescala condicionam a distribuição espacial dos or-

ganismos (ROGACHEV; SALOMATIN; CARMACK, 1996; SEKI et al., 2002; FOSSHEIM et al., 2005; BERTRAND, A. et al., 2008) que usualmente apresentam distribuições agregadas (e.g. WIEBE, 1970; HAURY; MCGOWAN; WIEBE, 1978; MACKAS; BOYD, 1979; MACKAS; DENMAN; ABBOTT, 1985; RUSSEL et al., 1992; BERTRAND, A. et al., 2008; BERTRAND, S. et al., 2008).

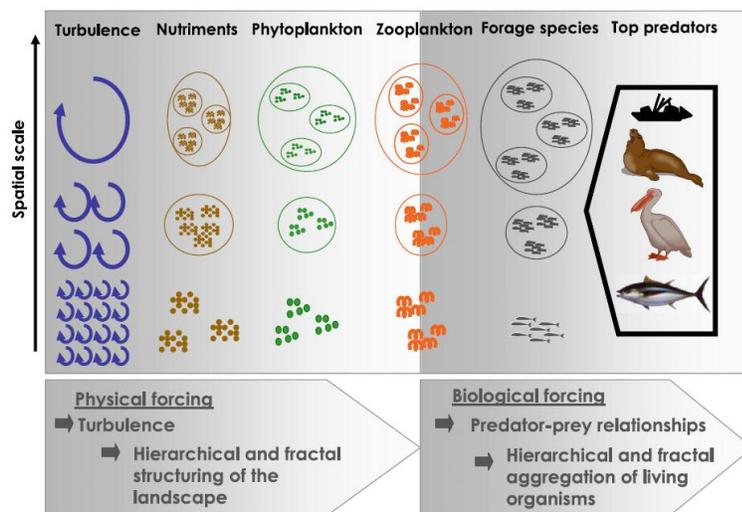


Figura 1 - Esquema do conceito de transferência bottom-up. Fonte: Bertrand, S. et al., (2008)

Este é o caso do ecossistema de afloramento mais produtivo do mundo em termos de pesca (CHAVEZ et al., 2008), o norte do Sistema da Corrente de Humboldt (NSCH) do Peru. O NSCH é uma região onde a oscilação sul de **El Niño** e a variabilidade climática são mais notáveis (CHAVEZ et al., 2008). O NSCH e seus recursos vivos são diretamente impactados por uma intensa e grande variabilidade climática em diferentes escalas. O NSCH, também é caracterizado pela presença de uma muito pronunciada e extensa zona de mínimo oxigênio (ZMO) (HELLY; LEVIN, 2004; CHAVEZ et al., 2008; KARSTENSEN; STRAMMA; VISBECK, 2008; PAULMIER; RUIZ-PINO, 2009). A ZMO é uma barreira para alguns animais (em particular a anchova) e concentra a maioria dos recursos vivos cerca da superfície (BERTRAND; BALLON; CHAIGNEAU, 2010).

Estruturas físicas de mesoescala (filamentos, redemoinhos) acrescentam a produtividade primária (FALKOWSKI et al., 1991; STRASS, 1992; DADOU et al., 1996; MCGILLICUDDY et al., 1998, 1999, 2007; OSCHLIES; GARÇON, 1998; MAHADEVAN;

ARCHER, 2000; WILLIAMS; FOLLOWS, 2003; BENITEZ-NELSON et al., 2007; BIBBY et al., 2008), e condicionam a distribuição do zooplâncton. Através destas estruturas acontece a sucessão de relações predador-presas, que se estendem aos níveis tróficos superiores (BERTRAND, S. et al., 2008; TEW-KAI et al., 2009). No NSCH os processos físicos de mesoescala são mais estáveis e menos intensos que em outros ecossistemas (CHAIGNEAU; ELDIN; DEWITTE, 2009). Isto pode ser importante para uma eficiente transferência da produtividade primária aos níveis tróficos superiores (CHAVEZ et al., 2008). As estruturas físicas de convergência de mesoescala acumulam o zooplâncton no seu centro (YEBRA; ALMEIDA; HERNANDEZ-LEON, 2005) e as estruturas de divergência dispersam o zooplâncton. Ao nível de submesoescala, existe a possibilidade de que processos físicos como as ondas internas ou pequenos redemoinhos poderiam favorecer a produtividade (KLEIN; LAPEYRE, 2009) e condicionam a estruturação vertical e horizontal do zooplâncton (BERTRAND, A. et al., 2008). Pois, estas estruturas físicas de submesoescala (de convergência ou divergência) podem ter um impacto sobre a distribuição do zooplâncton similar às estruturas de mesoescala, uma representação gráfica da estruturação é observada na Figura 2.

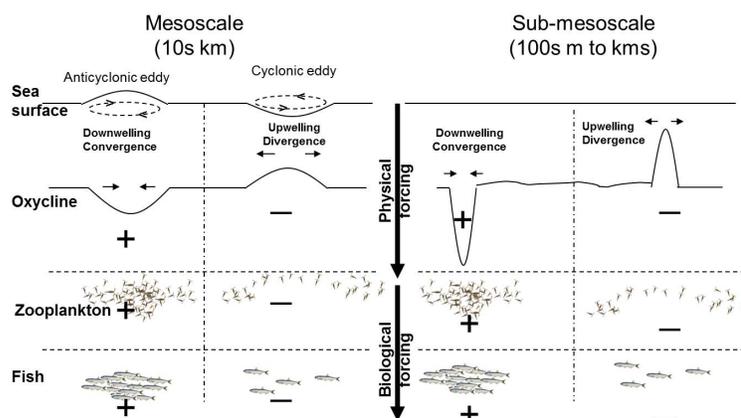


Figura 2 - Esquema das estruturas físicas e seus impactos sobre os organismos

Muitos estudos realizados no NSCH estão focalizados nas dinâmicas físicas e ecológicas em escalas temporais e grandes escalas espaciais. Isto foi graças à existência de informação nessas escalas. É assim que existe evidência que os processos físicos em mesoes-

cala jogam um rol fundamental na estruturação e funcionamento dos ecossistemas marinhos (BERTRAND, A. et al., 2008). Mas, existe uma grande necessidade de ampliar os estudos nas pequenas escalas espaciais, onde acontecem as relações predador-presas. Mas, os estudos de submesoescala foram difíceis devido à falta de dados de alta resolução (CAPET et al., 2008a; KLEIN; LAPEYRE, 2009).

Felizmente, novos métodos baseados na acústica e desenvolvidos em Peru proporcionam um novo caminho para estudar os ecossistemas a submesoescala. Utilizando dados acústicos de duas frequências, Bertrand, Ballon e Chaigneau (2010) e Ballon et al., (2011) desenvolvem um método para extrair dados simultâneos de grande resolução (um dado por segundo) de: (i) a profundidade do limite superior da ZMO e (ii) o zooplâncton, além de outras variáveis. Os efeitos dos processos físicos sobre os componentes do ecossistema podem ser estudados utilizando a profundidade do limite superior da ZMO (BERTRAND; BALLON; CHAIGNEAU, 2010). Os métodos acústicos propõem a visualização de uma ampla gama de estruturas de meso e submesoescalas tais como redemoinhos, frentes, filamentos e ondas internas (BERTRAND; BALLON; CHAIGNEAU, 2010; BALLON et al., 2011).

Pelo descrito anteriormente, nos ecossistemas marinhos existe uma estruturação hierárquica. Devido à estruturação hierárquica processos de submesoescala podem estar contidos em processos de mesoescala (por exemplo, uma onda interna pode estar dentro de um redemoinho de mesoescala), esta estruturação acrescenta a complexidade dos ecossistemas. Assim, para entender a dinâmica (estruturação e funcionamento) do ecossistema é preciso considerar múltiplos níveis de escalas no estudo. Para poder analisar as diferentes escalas temos que usar ferramentas estatísticas ou matemáticas que tenham a capacidade de ter em consideração as diferentes escalas. Como foi falado linhas acima, ao nível de mesoescala estruturas de convergência (*downwelling*) agregam mais zooplâncton no seu interior e estruturas de divergência dispersam o zooplâncton, então além de ter métodos multi-escala há a necessidade de identificar estas estruturas para um estudo mais profundo. O método tem que ter a capacidade de identificar e extrair estruturas que ficam dentro de outras estruturas.

Métodos como os variogramas aninhados (CRESSIE, 1993) são usadas para identificar estruturas espaciais (BELLIER; MONESTIEZ; DURBEC, 2007) ou PCNM (BORCARD; LEGENDRE, 2002; BORCARD et al., 2004) que ainda tem pouca aplicação no campo

de ecologia pela complexidade de seus análises. Estes dois métodos permitem trabalhar com diferentes escalas tendo como suposto básico a estacionariedade da variável (GRADOS, 2010). Mas muitas vezes acontece que as variáveis provenientes do campo da ecologia não são estacionários. Um método que não precisa que a variável seja estacionária é o método de ondaletas (TORRENCE; COMPO, 1998). O método de ondaleta foi desenvolvido na área da física e permite trabalhar a múltiplas escalas espaciais (ou temporais). Aplicações das ondaletas em campos de geofísica e climatologia são muito frequentes. Não obstante, os métodos de ondaletas estão sendo aplicados aos campos de ecologia (MENARD et al., 2007; BERTRAND, A. et al., 2008). Os três métodos dão resultados comparáveis mas são as ondaletas os que permitem identificar estruturas multi-escala no espaço (GRADOS, 2010).

Os métodos de ondaletas fazem uma decomposição em tempo-escala da série (ou sinal) permitindo identificar todas as escalas presentes. O método de ondaleta tem a capacidade de identificar pequenas escalas que estão dentro de grandes escalas. Pelas características deste método, neste trabalho vai-se usar as ondaletas para desenvolver um algoritmo que vai permitir identificar estruturas dos processos físicos em diferentes escalas.

O descrito acima deixa claro que há necessidade de compreender a dinâmica dos processos no NSCH. É assim que surgem as seguintes perguntas:

- Existe uma característica geral dos processos físicos de meso e submesoescala?
- Qual é o impacto que tem os processos físicos de meso e submesoescala sobre a distribuição espacial na abundância de zooplâncton?

Neste estudo se tem dois tipos de objetivos, o primeiro é um objetivo metodológico e os outros dois são objetivos ecológicos, eles são:

- Desenvolver uma metodologia para a identificação e extração de estruturas dos processos físicos no espaço.
- Fazer uma tipologia da estruturação espacial dos processos físicos de meso e submesoescala.

- Estudar o impacto que tem os processos físicos de meso e submesoescala sobre a distribuição espacial da abundância do zooplâncton.

2 DESENVOLVIMENTO

2.1 Materiais e Métodos

2.1.1 Área de estudo

O estudo foi realizado no norte do Sistema da Corrente de Humboldt (NSCH), frente às costas do Peru. A coleta dos dados foi realizado pelo **Instituto do Mar do Peru (IMARPE)**, que mediante suas embarcações de pesquisa monitora as condições bióticas do ecossistema marinho. As embarcações de pesquisa fazem trajetórias paralelas à costa do Peru, as trajetórias são chamadas de transectos. A pesquisa tem transectos de até 180 km, e inter-transectos de até 25 km de distância. Os dados usados nesta pesquisa provem de um cruzeiro acústico que foi feito entre fevereiro e abril do ano 2005. Neste trabalho utiliza-se informação simultânea do proxi (aproximação) da profundidade do limite superior da Zona Mínima de Oxigênio (ZMO) (BERTRAND; BALLON; CHAIGNEAU, 2010) e a abundância de zooplâncton (BALLON et al., 2011). Estes dados proporcionados por **IMARPE** foram obtidos por métodos acústicos e a coleta foi cada 14 metros ao longo do trajeto do barco.

2.1.2 Variáveis

No trabalho se tem principalmente 2 variáveis, as quais são:

O proxi da profundidade do limite superior da Zona de Mínima Oxigênio (ZMO):

No NSCH, a extensão vertical da comunidade epipelágica é limitada pela presença de uma Zona de mínimo oxigênio (CRIALES-HERNÁNDEZ, 2008). A estimacão da profundidade vertical da comunidade epipelágica é mediante a metodologia proposta por Bertrand, Ballon e Chaigneau (2010). A ZMO é uma estimacão dos processos físicos, pois eles estruturam ao limite superior da zona de mínimo oxigênio. A variável ZMO é medida em metros.

Abundância do Zooplâncton: O zooplâncton é uma espécie biológica, a abundância é o volume de zooplâncton que está presente no SCH, geralmente são as presas dos peixes. Sua estimacão é feita mediante a proposta de Ballon et al. (2011), que utiliza os dados acústicos para estimar esta variável. As unidades da abundância de zooplâncton são $\text{gramas.metros}^{-2}$

($\text{g}\cdot\text{m}^{-2}$).

Na Figura 3 observa-se a representação das duas variáveis, as linhas pretas representam a ZMO que é um estimador do limite superior da zona de mínimo oxigênio. As cores amarelas representam a abundância do zooplâncton.

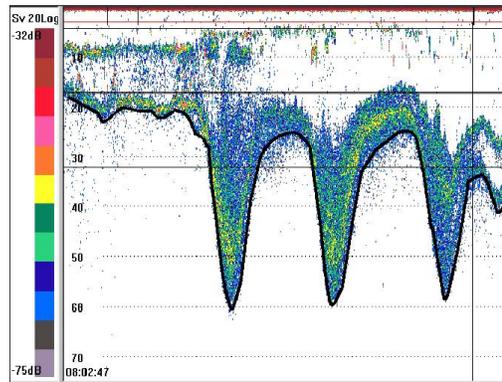


Figura 3 - Representação de um ecograma acústico

2.1.3 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.

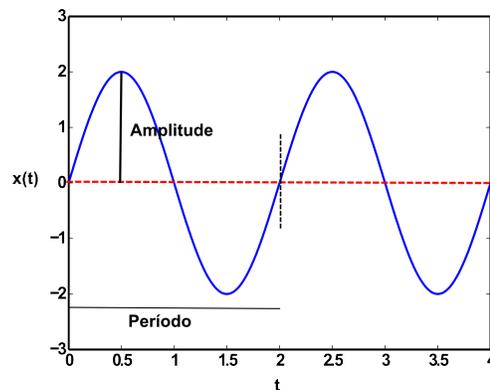


Figura 4 - Elementos do sinal, a amplitude e o período

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 arbitraria (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.

A partir da função mãe é que podemos chegar à transformada de ondaleta a qual 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.

Um ponto muito importante nas funções mãe é a escolha destas, mas existem muitos critérios que tem que se definir na hora da escolha da ondaleta, como por exemplo, ondaleta real vs ondaleta complexa, ondaleta continua vs ondaleta discreta, ortogonal vs não ortogonal. Em resumo, as ondaletas contínuas freqüentemente 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 continua 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. Estudos de comparação das duas funções chegaram a conclusões que as duas funções são similares. 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 é a frequência angular central ($\omega_0 = 2\pi f_0$) 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, e 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.

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

A partir da potência da transformada de ondaleta podem-se obter zonas de maior energia, para isto tem-se que comparar 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 usando 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 se 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.

Extração das estruturas de convergencia e divergencia

Aplicou-se o método ondaleta com a função Morlet às séries de dados (ZMO) com o objetivo de decompor a série em múltiplas escalas. As análises de ondaletas foram aplicadas em todos os transectos e dependendo das características da série foi utilizado um valor diferente do

parâmetro ω_0 . Os valores 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 5b 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). Também, a zona ao redor do máximo local de energia foi usada para estimar a amplitude do sinal. Para isto foi reconstruída o sinal para aquela zona. Para a estimação da amplitude foi usada a propriedade de reconstrução do sinal a partir da decomposição redundante do espectro de ondaleta continua. A reconstrução foi feita como segue:

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^J \frac{\Re(W_x(a, \tau_j))}{\tau_j^{1/2}} \quad (8)$$

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

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

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

Agora para conhecer a característica da estrutura (fase positiva, negativa ou *downwelling*, *upwelling*) utilizou-se a informação proporcionada pela fase da série $x(t)$. A fase tem uma variação de $-\pi$ ate π e a forma de calcular a fase é com o razão da parte imaginaria \Im entre a parte real \Re dos coeficientes de ondaletas $W_x(a, \tau)$ e esta dado como:

$$\phi_x(a, \tau) = \tan^{-1} \frac{\Im(W_x(a, \tau))}{\Re(W_x(a, \tau))} \quad (10)$$

A informação proporcionada nesta etapa é importante e vai servir para fazer a caracterização das estruturas obtidas.

Um exemplo do processo de obtenção de objetos esta na Figura 5, onde o sinal da Figura 5a foi gerada mediante a soma de três séries, duas delas foram geradas da função

$\text{seno}(\text{período})$, para o qual utilizou-se dois valores de período (12 e 30) e a terceira série foi um ruído branco com média zero e variância um. A Figura 5b é a representação do espectro de ondaleta, onde pode-se observar as caixas de cor vermelho e dentro delas as zonas de maior energia do espectro. A energia de cada caixa é utilizada para a reconstrução do sinal. As Figuras 5c e d representam a reconstrução do sinal original, cada um delas representa um período diferente. Sendo que a figura foi reconstruída utilizando um período de 12 e do lado direito um período de 32. Nestas duas figuras os círculos de cor vermelho estão representando a fase do sinal.

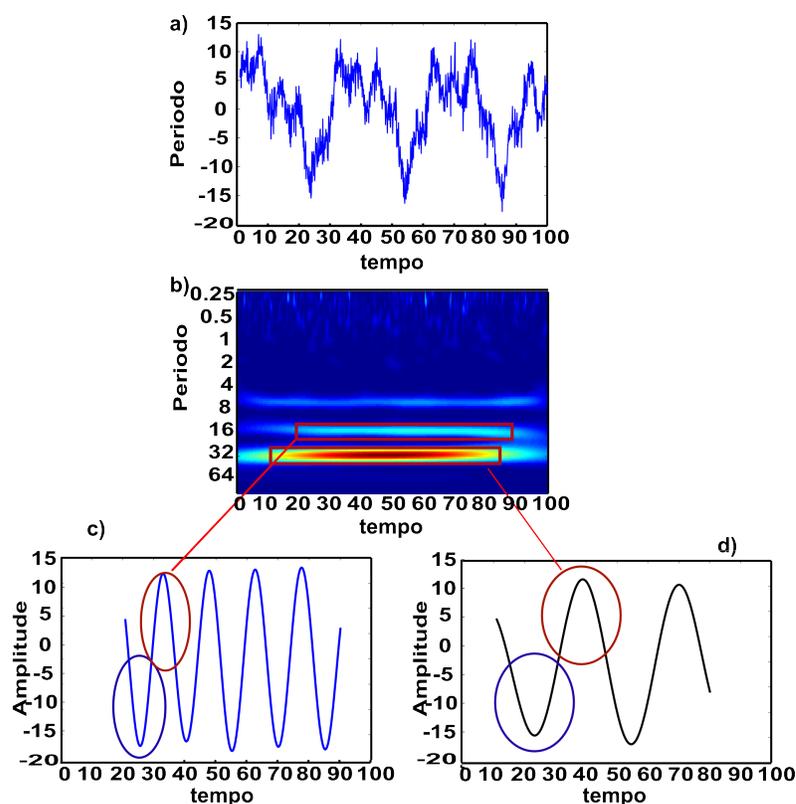


Figura 5 - Esquema da obtenção de estruturas de meso e submesoescala

Na série aplicou-se a análises de ondaletas e sua potência do espectro pode ser observada na Figura 5b, onde também é observada duas zonas (em caixas vermelhas), as quais são as zonas de máxima energia. As energias dentro das zonas foram utilizadas para reconstruir a série nas escalas que compreendem aqueles intervalos das caixas, assim como resultado temos as Figuras 5c e d, onde são reconstruções da série original, para as reconstruções utilizou-se um período de 12 (5c) e outro período de 32 5d. Finalmente também estimou-se a fase, os

círculos de cor vermelho representam às fases positivas e as de cor azul representam às fases negativas.

Para avaliar a capacidade de identificar estruturas nos sinais vão-se simular diferentes séries com diferentes parâmetros, o método vai-se aplicar às séries simuladas e estudar a capacidade de reconhecer os parâmetros.

2.1.4 Caracterização da distribuição das estruturas

A caracterização das estruturas obtidas pelo método desenvolvido vai ser realizada utilizando métodos exploratório de dados (métodos multivariados). Estes métodos exploratório de dados estudam as igualdades e diferenças de indivíduos desde um ponto de vista multidimensional. Três métodos são usados, os quais são: componentes principais (ACP), análises *cluster* e partição dos *cluster* (HUSSON; JOSSE; PAGÈS, 2010). Geralmente considera-se que os indivíduos estão distribuídos em uma grande nuvem de pontos num espaço euclidiano, o ACP leva aqueles indivíduos num novo espaço, fazendo uma redução do número de variáveis. Outro método que também permite a redução do espaço de estudo é a análises *cluster*, que mediante um critério de distância e algoritmos de aglomeração reduz o espaço dos indivíduos. Uma distância muito usada em análises *cluster* é a distância euclidiana, e critérios de aglomeração como o centróide ou ward são usados em *cluster*. A hierarquia que é apresentada no *cluster* é chamada dendograma, um terceiro método usado é a partição do *cluster*, geralmente o método usado para a partição é o k-medias.

Neste estudo utilizaremos os três métodos para caracterizar nossas estruturas.

Análises de Componentes Principais (ACP) A ideia principal do ACP é fazer descrição de um conjunto de dados X com I indivíduos e K variáveis utilizando um número pequeno ($R < K$) de variáveis não correlacionadas ficando com a maior informação possível. A análise *cluster* pode ser realizada sobre os resultados do ACP. Se a análise *cluster* é feita sobre todos os componentes do ACP os resultados serão similares ao realizado com *cluster* sobre os dados originais (indivíduos), um resultado mais interessante é usar o *cluster* sobre os R componentes principais. O ACP pode ser olhado como um método que vai separar a informação do ruído (HUSSON; JOSSE; PAGÈS, 2010), sendo que o primeiro componente fica com a informação

relevante para o conjunto de dados e os outros componentes estão restringidos ao ruído.

A obtenção dos componentes principais é mediante a decomposição (ou diagonalização) da matriz variância-covariância de nosso conjunto de variáveis originais.

Seja Z a nova matriz obtida, estas novas variáveis estão associadas à matriz original mediante:

$$Z = XA \quad (11)$$

Sendo que $A^t A = I$

Então, calcular os componentes principais é igual a aplicação de uma transformação ortogonal A às variáveis X , a transformação vai dar a nova matriz de variáveis no correlacionadas Z .

Porcentagem de variância associada a cada componente

A porcentagem de variância associada a cada componente é dada pela divisão entre seu valor próprio e a soma dos valores próprios. $\lambda_i / \sum_{i=1}^K \lambda_i$. Sendo que o primeiro componente tem o valor próprio maior, é assim que este componente tem um % de explicação maior.

Variáveis suplementares ou complementares

Além das variáveis usadas para a construção dos componentes (as quais chamaremos de variáveis ativas), podem-se usar outras variáveis as quais ajudarão a interpretar os dados. A estas variáveis chamaremos de variáveis suplementares (complementarias). As variáveis suplementares não são usadas para o cálculo dos componentes, mas são projetadas sobre o espaço gerado pelos componentes, esta projeção mostra com que componente estão correlacionado. Existem dois tipos de variáveis suplementares, as variáveis suplementares contínuas e as variáveis categóricas.

Cluster e partição dos cluster.

Outro método para estudar base de dados é o método de agrupamento ou análises *cluster*. Que por critério de similaridade vai formando grupos ou *cluster*.

Basicamente a análises *cluster* esta baseada num critério de distância euclidiana. Um critério de aglomeração utilizado é o método Ward.

O método Ward é baseado no teorema de Huygens que faz uma decomposição da variância total em variância entre grupos e dentro grupos. A variância total é decomposta como:

$$\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^{I_q} (x_{iqk} - \bar{x}_k)^2 = \sum_{k=1}^K \sum_{q=1}^Q I_q (\bar{x}_{qk} - \bar{x}_k)^2 + \sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^{I_q} (x_{iqk} - \bar{x}_{qk})^2 \quad (12)$$

Por enquanto, x_{iqk} é o valor da variável k para o individuo i dentro do grupo q , \bar{x}_{qk} é a média da variável k para o *cluster* q , \bar{x}_k á média total da variável k e I_q é o numero de indivíduos para o *cluster* q .

O método Ward esta baseado na agregação dos grupos de um jeito que o incremento da variância dentro de grupo é mínimo para cada passo do algoritmo, a variância entre grupos vai aumentando em cada passo.

Eleição do número de *cluster*

Escolher o número de *cluster* é uma questão central e várias abordagens têm sido propostas. Alguns deles são baseados na árvore hierárquica. De fato, uma árvore hierárquica pode ser considerada como uma sequência de partições aninhadas daquele em que cada indivíduo é um *cluster* para aquele em que todos os indivíduos pertencem ao mesmo *cluster*. O número de *cluster* pode, então, ser escolhido olhando para o aspecto global da árvore.

O critério da inércia sugere uma divisão em Q *cluster* quando o incremento da inércia entre $Q-1$ e Q *cluster* é muito maior do que Q e $Q+1$ *cluster*.

Um critério empírico que formaliza a idéia é dado por: $\frac{\Delta Q}{\Delta(Q+1)}$ onde ΔQ representa o incremento da inércia de $Q-1$ a Q *cluster*.

Partição dos *cluster*

Uma forma simples é ficar com os *cluster* dado pelo critério da inércia. Uma segunda estratégia é baseada no algoritmo de k-means. Neste algoritmo é fixado o numero de *cluster* Q . Outra estratégia combina os dois métodos descritos anteriormente. A partição do critério da inércia é introduzida como partição inicial ao método de k-means.

2.1.4.1 Indicador de eficiência do método de extração de estruturas

Depois de ter desenvolvido a metodologia de extração de estruturas temos que estudar a sua eficiência para identificar e extrair as estruturas dos processos físicos. Um indicador muito utilizado é o erro quadrado médio (EQM), Então nós também utilizaremos este indicador para nosso método. O indicador é dado a seguir como:

$$EQM = \frac{1}{N} \sum_{i=1}^N (\theta - \hat{\theta})^2 \quad (13)$$

Sendo que θ é o parâmetro teórico e $\hat{\theta}$ é o parâmetro estimado, i representa a estrutura e N o numero de estruturas identificadas. Este indicador foi usado para os dois parâmetros (θ vai representar separadamente o período e amplitude), o valor de EQM pode ser utilizado como método de comparação ante outros métodos. Mas ate agora não existe método algum que permita identificar e extrair estruturas.

2.1.5 Estudo do impacto das estruturas físicas sobre a distribuição do zooplâncton

Tendo identificado as características dos processos físicos vai-se estudar o impacto destas estruturas sobre a distribuição do zooplâncton. Assim o estudo foi realizado para cada *cluster*. Para estudar isto se utilizou análises de variância onde além de estudar o impacto das estruturas sobre a distribuição do zooplâncton estudou-se o impacto do dia (dia, noite) e a zona (costa, fora da costa) sobre o zooplâncton. Para as analises de varinância foi utilizada a variável biomassa de zooplâncton sendo esta transformada com a raiz cúbica.

Outro método que foi utilizado para estudar o impacto dos processos físicos sobre o zooplâncton foi o duplo K-Ripley.

2.1.5.1 Independência dos processos físicos e a distribuição do zooplâncton

Para este caso, além de aplicar a metodologia à série da estimação do limite superior da zona de mínimo oxigênio foi aplicado também à biomassa de zooplâncton, assim temos dois processos no espaço. Estes dois processos podem ser considerados como ao acaso. Para testar a independência deles vamos usar a função dupla K-Ripley.

A função K-Ripley esta baseada na teoria dos processos pontuais, um processo

pontual segundo sua distribuição pode ser um espaço completamente aleatório, um processo regular ou um processo agrupado. Dentro dos processos pontuais temos os processos homogêneos de poisson. Um processo homogêneo de poisson é um processo espacial completamente aleatório (CRESSIE, 1993) o qual tem duas características importantes que são estacionários e isotrópicos.

Função K-Ripley

A função K-Ripley estuda a dependência espacial ao longo de uma variedade de escalas de um processo pontual. A função K-Ripley é definida como:

$$K_N = \lambda^{-1}(h)E(N_0(h)) \quad (14)$$

Sendo que $N_0(h)$ representa o número esperado de eventos de um processo aleatório N a uma distância h . Teoricamente a função K-Ripley é representada como $K_N = \pi h^2$.

Para estimar a função K-Ripley (suposto um processo de poisson homogêneo) é considerado uma intensidade constante, então a função K-Ripley estimada é:

$$\widehat{K}_N(h) = (\lambda n)^{-1} \sum_{i=1}^n \sum_{j \neq i} I(u_{ij} \leq h) \quad (15)$$

Considerando a função K-Ripley teórica e estimada se testa a hipótese nula que a distribuição espacial de um processo é completamente aleatória, isto quer dizer que a distribuição dos eventos observados é um processo de Poisson homogêneo e portanto $\widehat{K}_N(h) = \pi h^2$. A significância de $\widehat{K}_N(h)$ é calculada mediante métodos de simulação.

Para o estudo de dois processos no espaço é usado a função K-cruzada de Ripley, a qual tem como definição:

$$K_{ij}(h) = \lambda_j^{-1}(h)E(N_{0ij}(h)) \quad (16)$$

Sendo que $N_0(h)$ é o número de eventos de tipo j dentro de uma distância h de um evento aleatório de tipo i , e λ_j é a intensidade do tipo de evento j . Igualmente que no caso da função K-Ripley existe uma função teórica para o caso bivariado, sendo que a função K-cruzada de Ripley teórica esta definido como $K^{12}(h) = \pi h^2$. A estimação com o suposto de um processo homogêneo de Poisson, novamente, é considerado a intensidade constante, então a função

K-cruzada de Ripley é:

$$\widehat{K}_{ij}(h) = (\lambda_i \lambda_j A)^{-1} \sum_k \sum_l I(u_{ik,jl} \leq h), i = 1, j = 2 \quad (17)$$

Por enquanto $u_{ik,jl} = \| x_{ik} - x_{jl} \|$ é a distância entre a posição k do evento i e a posição l do evento j , A é a área da região estudada. Para o estudo, o evento i e evento j representam as posições no espaço das estruturas da ZMO e do zooplâncton respectivamente.

Software utilizado

A implementação do método de identificação e extração de estruturas foi realizado no *software MATLAB*, a caracterização das estruturas dos processos físicos foram feitas no *software R* utilizando o pacote *FactoMineR*. Os análises de variância foram feitos no *software R* e o para teste de K-cruzada de Ripley usou-se o pacote *spatstat*.

3 RESULTADOS E DISCUSSÃO

Para a implementação do método desenvolvido, como foi mencionado anteriormente, foi utilizado o *software MATLAB* versão 2009b. As análises estatísticas para a caracterização das estruturas foram realizadas no *software R* versão 2.13.0 onde foi usado o pacote *FactoMineR* para fazer a caracterização das estruturas extraídas, se utilizou principalmente as funções ACP e HCPC do pacote já mencionado. Os resultados são mostrados nas seguintes seções do trabalho.

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 gero 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. A estas séries foi aplicado o método de identificação de estruturas. As séries y tem a seguinte estrutura: $y = Amplitude * seno(periodo)$. 1000 amostras foram geradas com períodos que vão desde 2 ate 80 e amplitudes que tem variação entre 5 e 40, a escolha deles foi ao acaso. Os resultados indicam que a identificação foi boa (Tabela 1). Pode-se ver exemplos de dois sinais nas Figuras 6 e 7, nas quais se observo duas séries simuladas (séries senos) com período de 47 e 20 respectivamente, e amplitude de 7 e 15 respectivamente. Nas duas figuras, as linhas de cor vermelha representam fases positivas (ou *upwelling*) e as linhas de cor azul representam a fase negativa (ou *downwelling*).

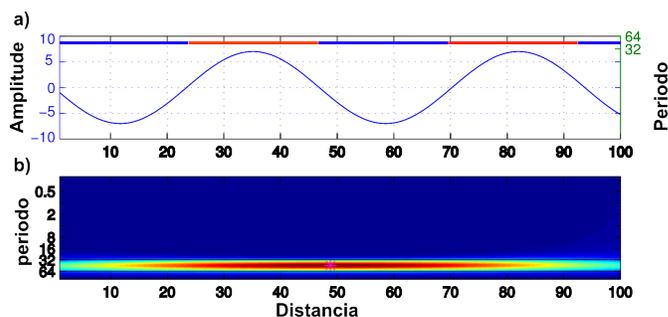


Figura 6 - Sinal simulado com período de 47 e amplitude de 7

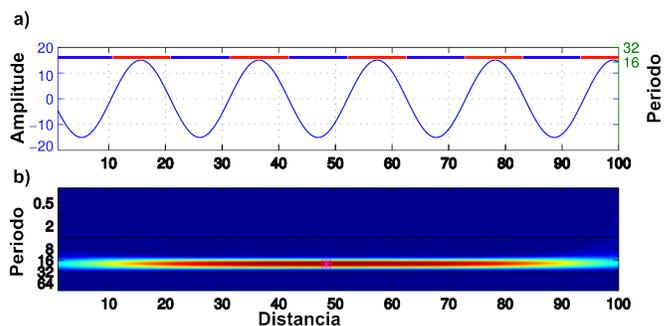


Figura 7 - Sinal simulado com período de 20 e amplitude de 15

Os resultados das 1000 simulações são mostrados na Tabela 1, onde observa-se que existe um Erro quadrático médio (EQM) pequeno para a o período e a amplitude. Este resultado indica que o método esta tendo uma boa estimaco dos parâmetros de amplitude e período para séries sim ruído. Existindo erro de estimaco nas bordas do espectro da potência de ondaleta.

Tabela 1 - Percentagem de identificaco de estruturas e EQM para sinais seno

Parâmetro	Identificaco	EQM
Período	99%	1.7
Amplitude	99%	0.4

Outro resultado interessante foi o numero de estruturas que foram mal identificadas, se simularam 1000 séries (com 1000 valores de amplitude e período conhecidos) sendo que das 1000 só o 1% das estruturas foram mal identificado (os valores de amplitude e período estimados são distintos aos valores usados na simulação).

Identificação de estruturas em funções seno com ruído branco

Para fazer mais complexa as séries de dados, se adiciono um ruído branco à séries de dados senos. Assim temos: $y = Amplitude * seno(perodo) + \varepsilon$ sendo que $\varepsilon \sim N(0, 1)$. Diferentes séries de seno foram geradas utilizando valores de amplitude e período distintos em cada simulação.

Os valores de amplitude e período foram gerados de uma distribuição uniforme. Sendo que o intervalo de variação do período é de 2 ate 80 e o intervalo de variação da amplitude foi de 5 ate 40. Os resultados indicam que existe uma boa identificação das estruturas (Tabela 2). Simulou-se 1000 séries com valores de período e amplitude conhecidos, das 1000 simulações que representam o 100% só o 4% foram mal identificadas, novamente observou-se que as estruturas não estão sendo bem identificadas próximo as bordas do espectro da potência de ondaleta. Também, os resultados indicam que o EQM da amplitude acrescentou-se com a complexidade do sinal.

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

Parâmetro	Identificação	EQM
Período	96%	1.7
Amplitude	96%	4.6

Identificação de estruturas em funções seno com ruído vermelho

Tendo já o método contrastado frente a sinais sim ruído e com ruído branco, então agora se vai testar o método frente a sinais com ruído vermelho. Assim é que se tem a seguinte estrutura da série $y = Amplitude * seno(perodo) + ar(1)$, sendo que $ar(1)$ é um processo autoregressivo de ordem 1. Para isto se gerou 1000 séries com valores de amplitude e período conhecidos com intervalos iguais aos casos anteriores.

Na Tabela 3 observa-se que dos 1000 sinais simulados o 91% de estruturas foram

identificada corretamente, também observa-se que o EQM para o período é similar aos encontrados nos casos anteriores (< 2) mas no caso da amplitude o EQM foi muito maior do que os dois casos anteriores (sem ruído e com ruído branco), isto foi por causa da complexidade das sinais. Existiram mais erros de identificação perto das bordas do espectro da potência de ondaleta, como nos casos anteriores está-se tendo uma ruim identificação de estruturas perto do borde do espectro da potência de ondaleta.

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

Parâmetro	Identificação	EQM
Período	91%	1.9
Amplitude	91%	14.9

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 8 do lado esquerdo se observa 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.

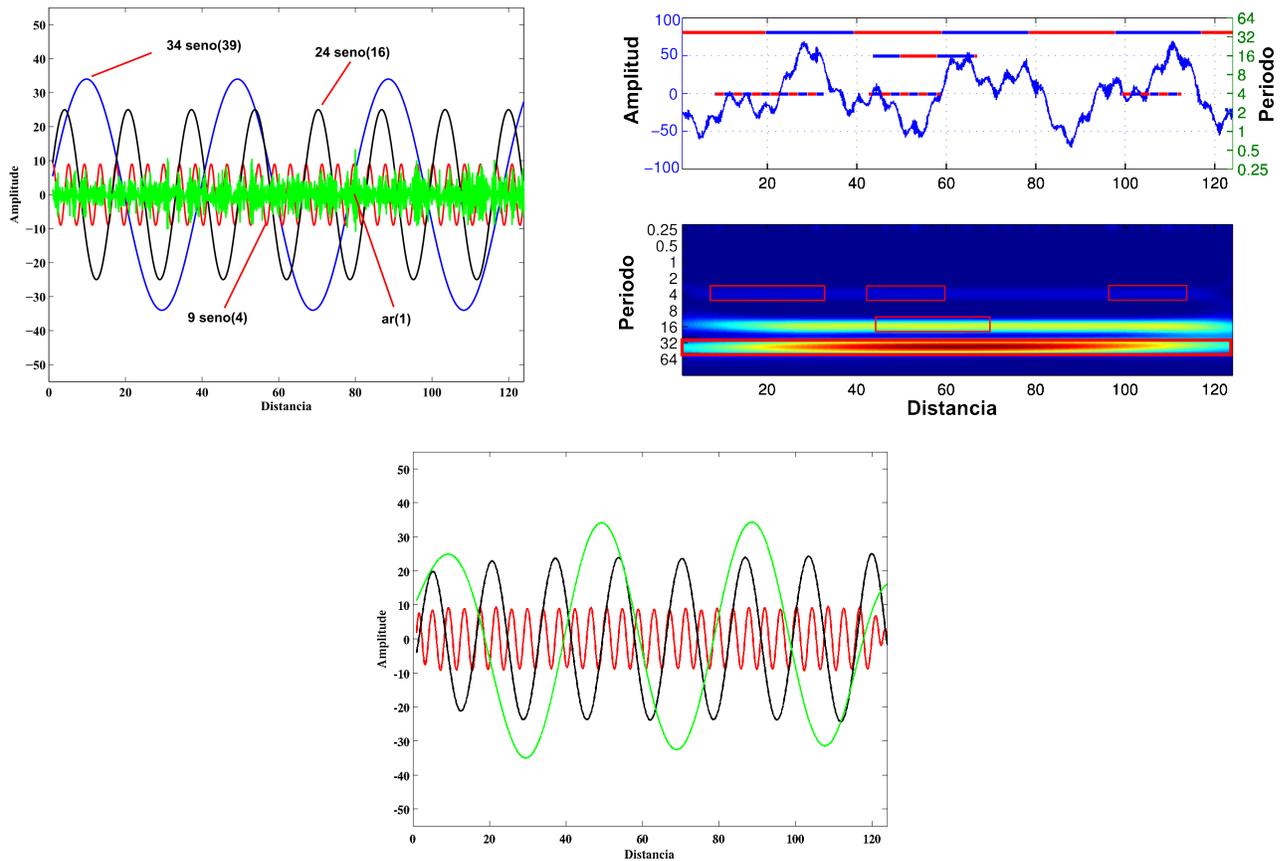


Figura 8 - Sinal simulado utilizando 4 séries diferentes

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.

3.1.1 Aplicação da metodologia aos dados do Ecossistema marinho

O método desenvolvido foi aplicado aos dados contínuos da ZMO. Depois de aplicar o método aos transectos foram obtidas uma série de variáveis as quais classificamos em dois tipos. O primeiro tipo é chamado de variáveis endógenas, nela temos as variáveis próprias da estrutura como são a tamanho da estrutura, a altura da estrutura, a profundidade, a razão que foi da divisão entre a altura e o tamanho, a energia que é o produto da altura e o tamanho da estrutura e a última variável endógena é a razãoinf que foi a divisão entre o tamanho da estrutura e estruturas mais pequenas. O segundo tipo de variáveis obtidas foi chamado de variáveis endógenas e são as variáveis obtidas do espaço como a posição no espaço (latitude e longitude), a distância à costa, a distância ao talude, a biomassa de zooplâncton dentro da estrutura identificada e o diferencial da biomassa que é obtida como a diferença da biomassa da estrutura e a biomassa média das estruturas vizinhas.

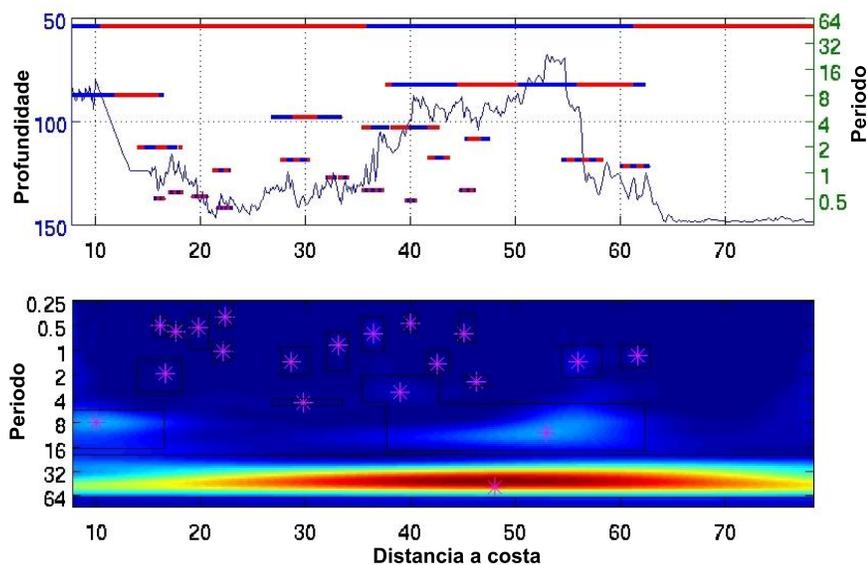


Figura 9 - Resultado da aplicação do método ao um transecto

Na Figura 9 observou-se o resultado da aplicação do método a um transecto. No

eixo y representa a profundidade do sinal da ZMO que vai de 50 até os 150 m e no eixo x representa a distância à costa que vai desde os 10 até os 80 km. Também está identificada o valor do período. As linhas de cores (azul e vermelho) representam o tamanho da estrutura e a fase, linha vermelha indica estrutura de *downwelling* e as linhas de cor azul indicam estruturas de *upwelling*. Também se observa que dentro de estruturas grandes estão estruturas pequenas. Na figura de embaixo é observado o espectro da potência de ondaleta. As estrelas indicam a posição do máximo local.

3.2 Tipologia das estruturas

Para fazer a tipologia das estruturas dos processos físicos se utilizo as variáveis endógenas, estas variáveis são chamadas de variáveis ativas. As variáveis exógenas foram usadas como variáveis contínuas suplementares as quais foram projetadas sobre o espaço dos componentes encontrados com as variáveis contínuas.

As variáveis q obtivemos foram classificadas em dois tipos, as quais chamamos, como já se menciona, de variáveis endógenas e variáveis exógenas (Tabela 4):

Tabela 4 - Variáveis endógenas e exógenas

Variáveis endógenas	Variáveis exógenas
Tamanho	Latitude
Altura	Longitude
Profundidade (prof)	Distância à costa (DC)
Razão	Distância ao talude (Dtalude)
Razãoinf	Biomassa
Energia	dif. Biomassa
	fase (<i>upwelling</i> , <i>downwelling</i>)
	zona (costa, fora da costa)
	dia (dia, noite)

Análises de Componentes Principais

Calculou-se a matriz de correlação entre as variáveis endógenas, observando-se uma forte cor-

relação entre a variável energia e o tamanho da estrutura. A variável energia ao ter uma correlação forte com a variável tamanho ia gerar um componente artificial, então foi o motivo por a qual a variável energia foi classificada como variável suplementaria. A matriz de correlação das variáveis ativas está na Tabela 5. Aquela matriz foi usada para obter os componentes principais.

Tabela 5 - Matriz variância covariância

	Tamanho	Altura	Prof	razão	razãoinf
Tamanho	1.00				
Altura	0.38	1.00			
Prof	-0.00	0.06	1.00		
razão	-0.35	0.31	0.04	1.00	
razãoinf	-0.15	0.09	0.12	0.20	1.00

Para as análises de ACP foram usadas as variáveis tamanho, altura, profundidade, razão e razãoinf como variáveis contínuas, como variáveis suplementares contínuas foram usadas as variáveis exógenas e como variáveis suplementares categóricas foram usadas o dia (dia, noite), a zona (costa, fora da costa) e a fase (*upwelling*, *downwelling*).

Assim, depois de fazer a decomposição da matriz variância covariância se pode olhar que temos três valores próprios maiores a um (Tabela 6). O critério dos valores próprios maiores a um indica que temos que ficar só com 3 componentes. Com os três componentes se podem explicar o 77.5 % da variação total dos dados. Observa-se que o primeiro componente explica um 29.9% da variância, o segundo componente explica o 27.5% da variância e o terceiro componente explica o 20%.

Depois de ter elegido os três componentes agora se podem observar as cargas das variáveis contínuas (Tabela 7) dos componentes, onde se observa que o primeiro componente tem mais peso as variáveis de razão, no segundo componente ficam com mais peso as variáveis de altura e ancho da estrutura, finalmente o terceiro componente tem mais peso a profundidade da estrutura.

Tabela 6 - Porcentagem de variância explicada

	valor próprio	% de variância	% de variância acumulada
comp 1	1.49	29.89	29.89
comp 2	1.38	27.52	57.41
comp 3	1.01	20.17	77.58
comp 4	0.81	16.29	93.87
comp 5	0.31	6.13	100.00

De igual forma a informação apresentada na Tabela 7 é amostrado na Figura 10, onde é amostrado cada combinação das coordenadas e seu respectivo % de variância, também estão as variáveis contínuas.

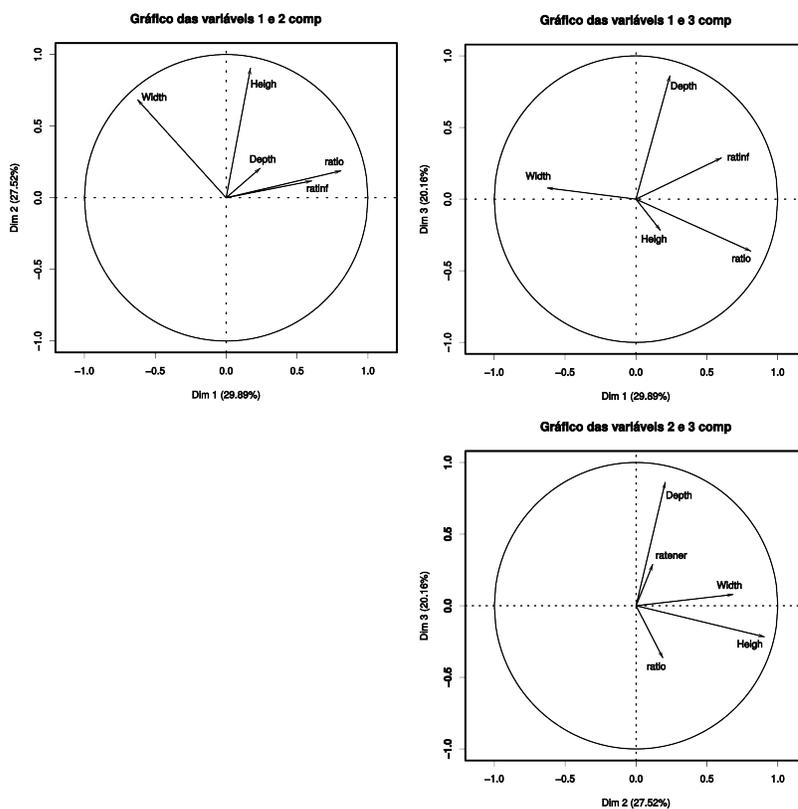


Figura 10 - Representação do espaço dos componentes principais

Tabela 7 - Cargas das variáveis

	Dim.1	Dim.2	Dim.3
Tamanho	-0.63	0.68	0.08
Altura	0.17	0.90	-0.22
Prof	0.24	0.21	0.86
razão	0.81	0.19	-0.36
razãoinf	0.60	0.12	0.29

Análises de *Cluster*

Para a análise *cluster* foi utilizada os resultados dos três componentes. Obteve-se 3 *cluster* os quais são mostrados na Figura 11.

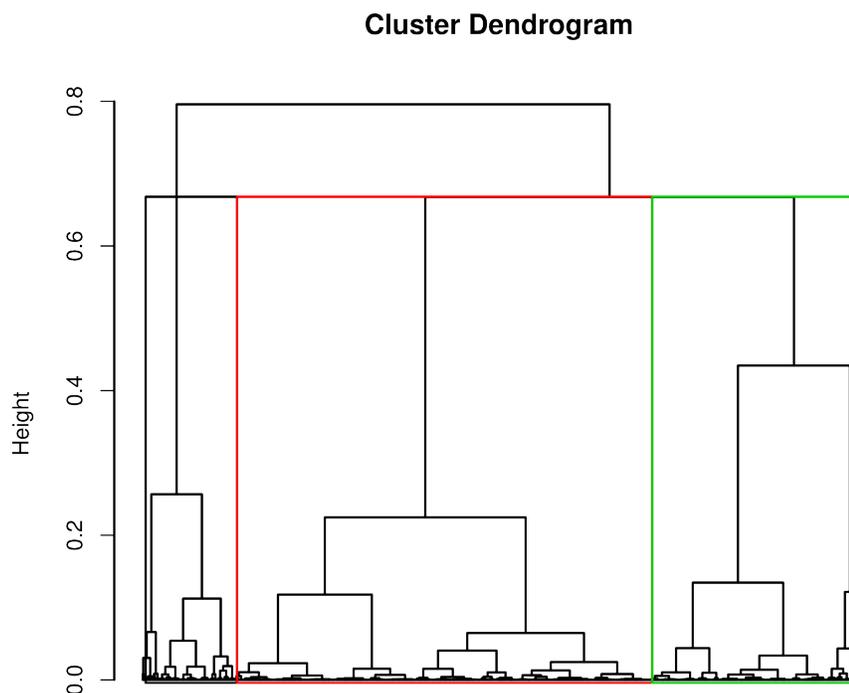


Figura 11 - Dendrograma com os três *cluster*

Fazendo as análises das variáveis suplementares categóricas para os três *cluster* se observa que para o *cluster* 1 são significativos a zona (costa, fora da costa), para o *cluster* 2

são significativos o dia (dia, noite) e a fase (*upwelling* ou *downwelling*). Finalmente no *cluster* 3 resultaram significativas as três variáveis categóricas (zona, dia e fase) (Tabela 8).

Tabela 8 - Descrição das variáveis categóricas para os *cluster*

Grupo		Cla/Mod	Mod/Cla	Global	p.value	v.test
I	Costa	11.1	39.8	25.8	0.00	3.06
	fora da costa	5.8	60.2	74.2	0.00	-3.06
II	<i>Upwelling</i>	58.1	56.7	50.9	0.00	4.33
	dia	55.5	72.1	67.9	0.00	3.40
	noite	45.4	27.9	32.1	0.00	-3.40
	<i>Downwelling</i>	46.2	43.3	49.0	0.00	-4.33
III	<i>Downwelling</i>	45.7	55.2	49.0	0.00	3.75
	fora da costa	43.1	78.9	74.2	0.00	3.22
	noite	45.4	35.9	32.1	0.01	2.44
	dia	38.3	64.1	67.9	0.01	-2.44
	Costa	33.2	21.1	25.8	0.00	-3.22
	<i>Upwelling</i>	35.6	44.8	50.9	0.00	-3.75

Do mesmo jeito apresentamos as variáveis contínuas com estatísticas descritivas para cada *cluster* (Tabela 9). Se pode observar que no 1 esta conformado por as estruturas mais grandes, que tem um tamanho médio de 23 km e uma altura média de 18 m, estas estruturas são as mais energéticas (energia média 358) que foram encontradas. O *cluster* 2 esta composta por estruturas pequenas (tamanho médio 3.3 km e altura média 5m), e pouco profundas (profundidade média 36m). Estas estruturas tem uma energia média de 20 e representa o segundo tipo de estruturas mais energéticas. No caso das variáveis suplementares contínuas observa-se uma biomassa de zooplâncton de 20 gr/m^2 .

Tabela 9 - Variáveis contínuas nos grupos

Grupo		v.test	média grupo	média geral	sd grupo	sd geral	p.value
I	Tamanho	27.5	23.2	4.2	14.5	7.1	0.00
	energia	27.1	357.7	41.5	293.6	119.9	0.00
	Altura	22.1	17.6	6.7	9.9	5.1	0.00
	DC	-2.3	63.6	73.7	40.6	44.7	0.02
	Dtalude	-2.6	19.7	32.2	44.7	49.4	0.01
	razão	-5.6	1.7	3.7	3.9	3.5	0.00
II	DC	-4.6	68.3	73.7	42.8	44.7	0.00
	Tamanho	-4.6	3.3	4.2	3.2	7.1	0.00
	energia	-7.2	19.2	41.5	23.8	119.9	0.00
	Dtalude	-7.5	22.6	32.2	42.1	49.4	0.00
	Biom	-9.6	20.3	30.8	29.8	41.5	0.00
	razão	-11.2	2.7	3.7	1.9	3.5	0.00
	Altura	-11.8	5.15	6.7	2.4	5.1	0.00
III	Prof	-19.1	35.8	50.1	18.1	28.8	0.00
	razãoinf	19.9	0.8	0.6	0.2	0.2	0.00
	Prof	18.8	67.7	50.1	30.3	28.8	0.00
	razão	14.4	5.4	3.7	4.3	3.5	0.00
	Biom	9.9	44.2	30.8	50.4	41.5	0.00
	Dtalude	9.0	46.8	32.2	55.0	49.4	0.00
	DC	5.9	82.5	73.7	46.4	44.8	0.00
	energia	-6.9	14.3	41.5	19.3	119.9	0.00
	Tamanho	-9.7	1.9	4.2	1.9	7.1	0.00

Finalmente no *cluster* 3 estão conformado por pequenas (tamanho médio 2km) e razão médio de 5 o qual indica que a altura média é de 10m, estas estruturas são as mais

profundas (profundidade média 68m), outra característica destas estruturas é que representam às menos energéticas (energia média 14). A média da biomassa de zooplâncton foi de $44 \text{ gr}/\text{m}^2$. Observou-se que as estruturas do *cluster* 3 tem a capacidade de ter mais biomassa de zooplâncton, sendo estas menos energéticas do que as outras categorias. Na Figura 12 são mostradas em forma gráfica os 3 tipos de estruturas.

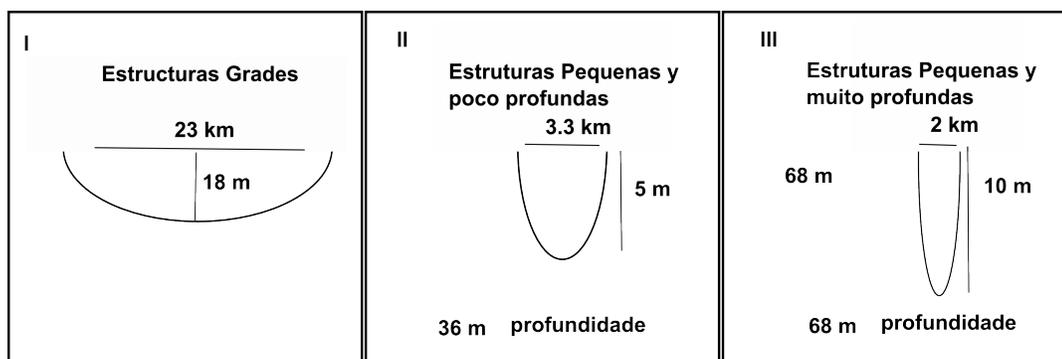


Figura 12 - Representação dos 3 grupos de estruturas identificadas

3.2.1 Estudo do impacto das estruturas sob a distribuição do zooplâncton

Já definidos os três tipos de estruturas, se estudo seu impacto sobre a biomassa de zooplâncton. Para isso se uso os resultados dos três *cluster* e usando anovas estudo-se se existe diferença entre características (dia, zona e fase).

Para o *cluster* 1, na Tabela 10 observa-se que pela noite tem mais biomassa que no período de dia (média da noite e $53.2 \text{ gr}/\text{m}^2$ e do dia e $15.0 \text{ gr}/\text{m}^2$). Igualmente observa-se que há maior biomassa na zona fora da costa que na costa, sendo a diferença mais forte nos dados de noite. Também, se observa que os resultados em fase não tem muita diferença, mas observa-se que em noite existe mais biomassa de zooplâncton. Para contrastar se, efetivamente, existem diferenças significativas na biomassa de zooplâncton ou não utilizamos anovas com desenhos aninhados. Considerou-se o dia como fator principal, dentro do dia (dia e noite) foi aninhado à zona (costa e fora da costa), finalmente dentro de cada zona foi aninhada a fase (*upwelling* ou *downwelling*). Os resultados (Tabela 11) indicam que existe diferença significativa em dia (dia e noite), dia:zona e dia:zona:fase. A análise post-hoc (Tabela 12)

mostra que existe maior abundância em noite. Do mesmo jeito a análise post-hoc amostra que existe maior biomassa de zooplâncton fora da costa (seja em dia ou noite). Outro resultado interessante é que as estruturas de *upwelling* fora da costa tem maior biomassa de zooplâncton do que estruturas de *downwelling* na costa (em noite)

Tabela 10 - Média da biomassa de zooplâncton dentro do grupo 1

Dia	Fase	Zona		Média fase	Média dia
		Costa	Fora costa		
Dia	<i>Upwelling</i>	6.5	12.1	10.3	15.0
	<i>Downwelling</i>	13.8	22.1	18.7	
	Média zona	11.3	17.2		
Noche	<i>Upwelling</i>	12.2	71.9	35.4	53.2
	<i>Downwelling</i>	16.4	91.8	67.8	
	Média zona	13.8	85.5		
	Média total zona	12.3	42.7		

Para um melhor estudo na fase (*downwelling* ou *upwelling*), foi feito o diferencial da biomassa entre estruturas. O diferencial foi feito da seguinte forma. Se pego uma estrutura e a biomassa do zooplâncton dessa estrutura foi retirada com a biomassa média das estruturas vizinhas. Logo com um anova aninhado foi analisado os resultados do diferencial de biomassa do zooplâncton (Tabela 13). Os resultados indicam que existe diferença significativa em dia:zona:fase. Depois com o teste post-hoc (onde foram utilizadas só os resultados significativos) (Tabela 14) se observa que as ondas de *downwelling* tem maior impacto no diferencial de biomassa do zooplâncton. Isto quer dizer que dentro das estruturas de *downwelling* pode-se encontrar mais biomassa de zooplâncton que nas estruturas de *upwelling*, isto foi observado em dia como em noite.

Tabela 11 - Anova para o zooplâncton no grupo 1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dia	1	28.9	28.9	63.5	0.0000
dia:zona	2	45.9	22.9	50.4	0.0000
dia:zona:fase	4	6.5	1.6	3.5	0.0092
Residuais	90	40.9	0.5		

Tabela 12 - Teste de Tukey para o zooplâncton no grupo 1

		diff	lwr	upr	p adj
Dia	Noite-Dia	1.1	0.8	1.4	0.00
Dia:Zona	Noite:fora costa-Noite:costa	2.1	1.6	2.7	0.00
	Noite:fora costa:Upwe-Noite:costa:Upwe	2.1	1.1	3.0	0.00
Dia:Zona:Fase	Noite:fora costa:Down-Noite:costa:Upwe	2.4	1.5	3.2	0.00
	Noite:fora costa:Upwe-Noite:costa:Down	1.7	0.6	2.8	0.00
	Noite:fora costa:Down-Noite:costa:Down	2.0	1.1	2.9	0.00

Tabela 13 - Anova no grupo 1 para a diferencial de biomassa

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dia	1	132.8	132.8	1.8	0.1781
dia:zona	2	333.9	166.9	2.3	0.1046
dia:zona:fase	4	1207.3	301.8	4.1	0.0037
Residuais	90	6492.1	72.1		

Tabela 14 - Teste de Tukey para o diferencial de biomassa

	diff	lwr	upr	p adj
Dia:costa:Down-Dia:fora costa:Upwe	12.05	2.45	21.64	0.00

Para o *cluster 2* também foi feita a análises descritiva da biomassa de zooplâncton (Tabela 15). Se observo que dentro do *cluster 2* há mais biomassa de zooplâncton em noite que em dia (31.7 e 15.9 respectivamente). Sendo que também há mais biomassa fora da costa que dentro da costa, a diferença foi maior na noite. Na análise de fase observou-se resultados similares em estruturas de *downwelling* e *upwelling* isto foi em dados de dia e noite.

Para contrastar se efetivamente existem diferenças significativas foi feito um teste de médias, o resultado pode ser observado na Tabela 15.

Tabela 15 - Média da biomassa segundo os fatores (dia, zona e fase) no grupo 2

Dia	Fase	Zona		Média fase	Média dia
		Costa	Fora costa		
Dia	<i>Upwelling</i>	6.1	18.1	15.2	15.9
	<i>Downwelling</i>	5.6	21.0	17.1	
	Média zona	5.9	19.3		
Noche	<i>Upwelling</i>	8.1	41.7	30.4	31.7
	<i>Downwelling</i>	14.3	42.4	33.4	
	Média zona	10.9	42.1		
	Média total zona	7.6	25.2		

Os resultados das análises de variância mostraram que existe diferenças significativas na biomassa de zooplâncton em dia e também em dia:zona (Tabela 16). Os testes post-hoc (Tabela 17) amostraram diferenças significativa entre noite e dia, sendo que se tem

mais biomassa em noite que em dia, além disso, é observado que fora da costa existe maior abundância. Estes resultados são similares aos observados no *cluster* 1, onde há diferenças significativas segundo o dia e, também, segundo a zona

Tabela 16 - Anova para a biomassa do zooplâncton dentro do grupo 2

	Df	SQ	QM	F value	Pr(>F)
dia	1	51.41	51.41	63.38	0.0000
dia:zona	2	101.56	50.78	62.61	0.0000
dia:zona:fase	4	3.64	0.91	1.12	0.3450
Residuais	705	571.78	0.81		

Tabela 17 - Teste de Tukey para a biomassa de zooplâncton no grupo 2

Factor		diff	lwr	upr	p adj
Dia	Noite-Dia	0.6	0.5	0.7	0.00
Dia:Zona	Dia:fora costa-Dia:costa	0.7	0.5	0.9	0.00
	Noite:fora costa-Noite:costa	1.1	0.8	1.5	0.00

Para estudar a relação com a fase, se trabalho com o diferencial da biomassa, assim a análise de variância (Tabela 18) indicaram que existe diferença significativa em dia:zona:fase. O teste post-hoc mostraram que as estruturas de *downwelling* tem maior impacto na agregação de zooplâncton, seja na costa ou fora da costa (Tabela 19). Observou-se que estruturas de *downwelling* tem mais biomassa que estruturas de *upwelling* (em dia e fora da costa).

Tabela 18 - Anova para o diferencial da biomassa no grupo 2

	Df	SQ	QM	F value	Pr(>F)
dia	1	9.38	9.38	0.20	0.6567
dia:zona	2	134.37	67.18	1.42	0.2434
dia:zona:fase	4	1444.05	361.01	7.61	0.0000
Residuais	705	33454.41	47.45		

Tabela 19 - Teste de Tukey para o diferencial de biomassa no grupo 2

	diff	lwr	upr	p adj
Dia:fora costa:Down-Dia:fora costa:Upwe	3.09	0.92	5.26	0.00

Finalmente no *cluster 3*, as análises descritivas da biomassa no *cluster 3* (Tabela 20) mostra que existe mais biomassa em noite que em dia, mais biomassa fora da costa que na costa, sendo estes valores maiores aos obtidos nos *cluster 1* e *2*. Também é observado que as estruturas de *downwelling* e *upwelling* tem mais biomassa que as estruturas do mesmo tipo nos *cluster 1* e *2*.

Para testar se efetivamente existem diferenças significativas nos fatores (dia, zona e fase) aplicamos uma análise de variância (Tabela 21). Os resultados do anova indicaram que existe diferenças significativas em dia e dia:zona. O teste post-hoc (Tabela 22) indicou que existe mais biomassa em noite que em dia, sendo também que existe mais biomassa de zooplâncton fora da costa que dentro da costa.

Tabela 20 - Média da biomassa segundo os fatores (dia, zona e fase) no grupo 3

Dia	Fase	Zona		Média fase	Média dia
		Costa	Fora costa		
Dia	<i>Upwelling</i>	17.2	26.3	24.5	24.9
	<i>Downwelling</i>	21.5	26.0	25.3	
	Média zona	19.4	26.1		
Noche	<i>Upwelling</i>	24.6	100.2	84.2	78.4
	<i>Downwelling</i>	28.1	96.1	73.7	
	Média zona	26.8	98.2		
	Média total zona	22.9	49.9		

Tabela 21 - Anova da biomassa de zooplâncton no grupo 3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dia	1	220.9	220.9	267.8	0.0000
dia:zona	2	94.3	47.2	57.2	0.0000
dia:zona:fase	4	1.8	0.4	0.5	0.7061
Residuais	546	450.5	0.8		

Tabela 22 - Teste de Tukey para a biomassa de zooplâncton dia no grupo 3

		diff	lwr	upr	p adj
Dia	Noite-Dia	1.3	1.2	1.5	0.00
Dia:zona	Noite:fora costa-Noite:costa	1.51	1.14	1.88	0.00

Igualmente para estudar se existe diferença significativa induzido pela fase (*downwelling* ou *upwelling*) se uso o diferencial da biomassa de zooplâncton. A Tabela 23

indica que existe diferença significativa em fase para a variável diferencial de biomassa, sendo que em estruturas de fase *downwelling* é maior do que estruturas *upwelling* (Tabela 24).

Tabela 23 - Anova do diferencial da biomassa no grupo 3

	Df	SQ	QM	F value	Pr(>F)
dia	1	12.8	12.8	0.22	0.6385
dia:zona	2	97.2	48.6	0.84	0.4330
dia:zona:fase	4	2021.2	505.3	8.71	0.0000
Residuais	546	31665.5	58.0		

Tabela 24 - Teste de Tukey para o diferencial da biomassa de zooplâncton no grupo 3

	diff	lwr	upr	p adj
Noite:costa:Down-Noite:costa:Upwe	9.45	2.88	16.02	0.00
Noite:costa:Down-Noite:fora costa:Down	5.05	0.33	9.77	0.03

3.2.1.1 Estudo da independência entre os dois processos

Finalmente para testar se existe relação entre os processos físicos e os processos biológicos foram feitas análises de K-cruzado de Ripley, os resultados se podem olhar na Figura 13. Para evitar o problema da aleatoriedade do processo se juntaram todos os transectos, assim se obteve um grande transecto.

O resultado indica que os processos físicos condicionam a distribuição dos processos biológicos, ou em outras palavras, os processos físicos estão agregando aos processos biológicos.

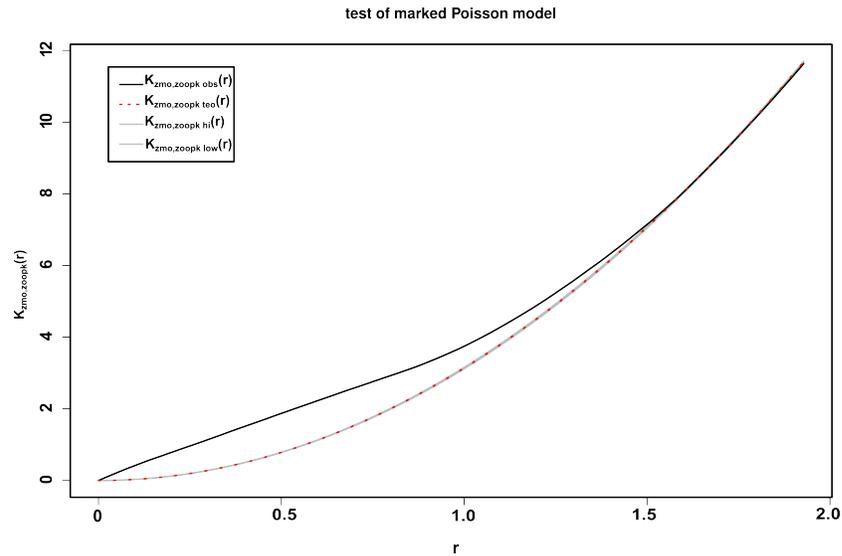


Figura 13 - K-cruzada de Ripley

3.3 Discussão

3.3.1 Da metodologia

Como foi mencionado num início, para compreender a dinâmica dos ecossistemas se tem que entender o comportamento de seus componentes. Assim com o objetivo de estudar a estruturação espacial dos processos físicos se tem a necessidade de procurar métodos que permitam trabalhar a múltiplas escalas. Métodos de geoestatística como os variogramas aninhados (CRESSIE, 1993) permitem fazer estudos a múltiplas escalas. O variograma aninhado identifica as estruturas características que estão presentes num processo. Igualmente que os métodos dos variogramas aninhados, se têm os métodos PCNM (BORCARD; LEGENDRE, 2002) os quais também identificam diferentes estruturas espaciais dos processos. Outro método que identificam diferentes estruturas espaciais são as ondaletas. O método de ondaletas, além de estimar o tamanho da estrutura, permite localizar sua a posição no espaço ou tempo (TORRENCE; COMPO, 1998). Estudos de comparação entre o PCNM e os variogramas aninhados foram desenvolvidos por Bellier et al. (2007) onde mostra resultados comparáveis entre eles, igualmente Grados (2010) mostrou que os três métodos estimam as escalas do processo, só que a vantagem do método de ondaletas sobre os PCNM e os variogramas aninhados é a

capacidade de identificar a escala e sua posição no espaço. Outra vantagem dos métodos de ondaletas é que pode identificar estruturas de pequenos alcances até grandes alcances, também pode identificar estruturas de pequenas escalas que estão dentro de estruturas de grandes escalas. Estas vantagens dos métodos de ondaletas sobre os outros dois métodos foram o motivo pela qual neste trabalho se utilizou as ondaletas para desenvolver o método de identificação e extração de estruturas.

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 influência na estimação da amplitude. Outro aspecto importante foi a família usada, no trabalho utilizou-se a família Morlet, pois permite obter dois tipos de informação (a fase e amplitude) (CAZELLES et 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.

3.3.2 Tipologia das estruturas físicas

Nos ecossistemas marinhos os processos físicos tem um rol fundamental na estruturação espacial (BERTRAND, A. et al., 2008; BERTRAND, S. et al., 2008). Os processos físicos de meso e submesoescala condicionam a distribuição espacial do limite superior da zona de mínimo oxigênio, então estudar o comportamento deste limite é estudar o comportamento dos processos físicos. Bertrand, Ballon e Chaigneau (2010) desenvolvem um método para estimar o limite superior da zona de mínimo oxigênio.

A partir deste estimador se começou a extrair as estruturas físicas, os resultados indicam a presença de três tipos de estruturas, grandes estruturas de tamanho médio de 23 km as quais correspondem aos processos de mesoescala (BERTRAND, A. et al., 2008; CHAIGNEAU;

GIZOLME; GRADOS, 2008). Segundo Chaigneau, Gizolme e Grados, (2008) estruturas de mesoescala são as mais energéticas do ecossistema de Humboldt, o qual foi observado com os resultados do trabalho.

O segundo tipo de estruturas tem um tamanho menor a $<10\text{km}$, dentro deste grupo há uma classificação do que tem variação segundo a profundidade. Estruturas pouco e muito profundas foram encontradas. Estas estruturas correspondem aos processos de submesoescala (BERTRAND, A. et al., 2008; BALLON et al., 2011). Estas estruturas correspondem a ondas internas ou pequenos redemoinhos (BERTRAND, A. et al., 2008) e são as mais abundantes no ecossistema, sendo que as estruturas pouco profundas são mais energéticas do que as estruturas mais profundas.

3.3.3 Do impacto sobre a biomassa de zooplâncton

Observou-se grande diferença entre a biomassa de zooplâncton em período de dia e noite, esta diferença é devido à migração vertical do zooplâncton, pois durante o dia aproximadamente o 77% da biomassa de zooplâncton migra para baixo da ZMO (BALLON et al., 2011), também observa-se diferença segundo a zona, sendo que tem mais abundância fora da costa que na costa.

Os três tipos de estruturas tem um impacto diferente sobre a biomassa de zooplâncton assim, as estruturas grandes que são as mais energéticas (CHAIGNEAU; GIZOLME; GRADOS, 2008) agregam o zooplâncton. Os resultados indicam que estruturas de *downwelling* tem mais biomassa de zooplâncton que as estruturas de *upwelling*, sendo que as estruturas de *downwelling* representam os processos físicos de convergência e as de *upwelling* representam aos processos físicos de divergência. Estes resultados corroboram o dito por Bakun (2006).

As estruturas pequenas têm impactos diferentes sobre a biomassa de zooplâncton, sendo que o comportamento da biomassa tem variação segundo o período de dia (mais biomassa em noite que em dia) e zona (mais biomassa fora da costa que na costa). A diferença das grandes estruturas em estas existe uma diferença pequena na agregação do zooplâncton. O terceiro grupo de estruturas são as mais profundas e pouco energéticas, são de pouco tamanho e muito altas. Estas estruturas tem mais abundância de zooplâncton, tendo um impacto maior que os outros dois tipos de estruturas. Até agora as estruturas de submesoescala são

pouco estudadas, alguns pesquisadores tem realizado estudos a este nível (submesoescala) (BERTRAND, A. et al., 2008), mas o estudo realizado por ele só foi uma zona pequena do ecossistema de Humboldt. O presente trabalho foi realizado a tudo o ecossistema da parte do Peru.

A análise de K-cruzada de Ripley indicou que existe uma agregação que é exercida pelos processos físicos aos processos biológicos, isto da evidencia da estruturação de baixo para cima que existe no ecossistema (RUSSEL et al., 1992). Sendo que os componentes físicos estruturam a paisagem do ecossistema.

4 CONCLUSÕES

No presente trabalho foi desenvolvida uma metodologia para identificar e extrair estruturas de processos físicos, este método pode ser utilizado em outros campos de aplicação. Utilizaram-se análises de ondaletas para o desenvolvimento do método. Como conclusão do trabalho observou-se que:

- O método desenvolvido mostrou a capacidade de identificar e extrair estruturas multi-escalas no espaço. Observou-se também que o parâmetro de localização é importante para a estimação da altura da estrutura.
- Depois de ter feito a caracterização das estruturas obtidas, observou-se que no ecossistema da corrente de Humboldt existem três tipos de estruturas. O primeiro tipo de estruturas são as de mesoescala as quais são as mais energéticas do ecossistema, o segundo e terceiro tipo são as estruturas de submesoescala, as quais têm diferença na profundidade, umas são localizadas aos 20 ou 30 metros de profundidade no mar e as outras são localizados aos 60 ou 70 metros de profundidade no mar. As estruturas de submesoescalas são as mais abundantes e menos energéticas no ecossistema da corrente de Humboldt.
- Observou-se que as estruturas mais profundas tem a capacidade de agregar mais abundância de zooplâncton no seu centro. Observou-se que estruturas de *downwelling* tem a capacidade de agregam mais abundância de zooplâncton que as estruturas de *upwelling*, isto foi observado mais nas estruturas mais profundas de submesoescala.

O método foi aplicado aos dados de um ano, para conhecer se as tipologias das estruturas são estacionarias no tempo deveria de fazer análises para diferentes anos e estudar o comportamento das estruturas.

REFERÊNCIAS

- BADDELEY, A.; TURNER, R. **Spatstat**: an R package for analyzing spatial point patterns. Disponível em: <www.jstatsoft.org>. Acesso em: 10 maio 2011.
- BAKUN, A. Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptive response. **Scientia Marina**, Barcelona, v.70, p.105-122, 2006.
- BALLON, M. **Acoustic study of macrozooplankton off Peru**: biomass estimation, spatial patterns, impact of physical forcing and effect on forage fish distribution. 2010. 155p. Tese de doutorado (Biologia de populações e ecologia) - University Montpellier II, Montpellier, 2010.
- BALLON, M.; BERTRAND, A.; LÉBOURGÈS-DHAUSSY, A.; GUTIÉRREZ, M.; AYÓN, P.; GRADOS, D.; GERLOTTO, F. Is there enough zooplankton to feed forage fish populations off Peru? An acoustic (positive) answer. **Progress in Oceanography**, Oxford, [s.n.], 2011.
- BENITEZ-NELSON, C.; BIDIGARE, R.; DICKEY, T.; LANDRY, M.; LEONARD, C.; BROWN, S.; NENCIOLI, F.; RII, Y.; MAITI, K.; BECKER, J.; BIBBY, T.; BLACK, W.; CAI, W.; CARLSON, C.; CHEN, F.; KUWAHARA, V.; MAHAFFEY, C.; McANDREW, P.; QUAY, P.; RAPPÉ, M.; SELPH, K.; SIMMONS, M.; YANG, E. Mesoscale eddies drive increased silica export in the subtropical pacific ocean. **Science**, Washington, v.316, p.1017-1021, 2007.
- BELLIER, E.; MONESTIEZ, P.; DURBEC, J.-P. Identifying spatial relationships at multiple scales: principal coordinates of neighbour matrices (PCNM) and geostatistical approaches, **Ecography**, Malden, v.30, p.385-399, 2007.
- BERTRAND, A.; BALLON, M.; CHAIGNEAU, A. Acoustic Observation of Living Organisms Reveals the Upper Limit of the Oxygen Minimum Zone. **PLoS ONE**, San Francisco, v.5, p.e10330, 2010. Disponível em: <<http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0010330>>. Acesso em: 20 mar. 2011.
- BERTRAND, A.; GERLOTTO, F.; BERTRAND, S.; GUTIERREZ, M.; ALZA, L.; CHIPOLLINI, A.; DIAZ, E.; ESPINOZA, P.; LEDESMA, J.; QUESQUEN, R.; PERALTILLA, S.; CHAVEZ, F. Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales, **Progress in Oceanography**, Oxford, v.79, p.264-277, 2008.
- BERTRAND, S.; DEWITTE, B.; TAM, J.; DIAZ, E.; BERTRAND, A. Impacts of kelvin wave forcing in the Peru Humboldt Current System: scenarios of spatial reorganizations from physics to fishers, **Progress in Oceanography**, Oxford, v.79, p.278-289, 2008.
- BIBBY, T.; GORBUNOV, M.; WYMAN, K.; FALKOWSKI, P. Photosynthetic community responses to upwelling in mesoscale eddies in the subtropical North Atlantic and Pacific Oceans, **Deep Sea Research Part II: Topical Studies in Oceanography**, Oxford, v.55, p.1310-1320, 2008.
- BORCARD, D.; LEGENDRE, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices, **Ecological Modelling**, Amsterdam, v.153, p.51-68, 2002.
- BORCARD, D.; LEGENDRE, P.; AVOIS-JACQUET, C.; TUOMISTO, H. Dissecting the spatial structure of ecological data at multiple scales, **Ecology**, Ithaca, v.85, p.1826-1832, 2004.

CAPET, X.; McWILLIAMS, J.; MOLEMAKER, M.; SHCHEPETKIN, A. Mesoscale to submesoscale transition in the California current system. part I: flow structure, eddy flux, and observational tests, **Journal of Physical Oceanography**, Boston, v.38, p.29-43, 2008.

CARR, M.; KEARNS, E. Production regimes in four eastern boundary current systems, **Deep Sea Research Part II: Topical Studies in Oceanography**, Oxford, v.50, p.3199-3221, 2003.

CAZELLES, B.; CHAVEZ, M.; BERTEAUX, D.; MÉNARD, F.; OLAV-VIK, J.; JENOUVRIER, S.; STENSETH, N. Wavelet analysis of ecological time series, **Oecologia**, Heidelberg, v.156, p.287-304, 2008.

CAZELLES, B.; CHAVEZ, M.; MAGNY, G.; GUÉGAN, J.-F.; HALES, S. Time-dependent spectral analysis of epidemiological time-series with wavelets, **Journal of the Royal Society Interface**, London, v.4, p.625-636, 2007.

CAZELLES, B.; STONE, L. Detection of imperfect population synchrony in an uncertain world, **Journal of Animal Ecology**, Oxford, v.72, p.953-968, 2003.

CHAIGNEAU, A.; GIZOLME, A.; GRADOS, C. Mesoscale eddies off Peru in altimeter records: Identification algorithms and eddy spatio-temporal patterns, **Progress in Oceanography**, Oxford, v.79 p.106-119, 2008.

CHAIGNEAU, A.; ELDIN, G.; DEWITTE, B. Eddy activity in the four major upwelling systems from satellite altimetry (1992-2007), **Progress in Oceanography**, Oxford, v.83, p.117-123, 2009.

CHAIGNEAU, A.; PIZARRO, O. Surface circulation and fronts of the South Pacific ocean, east of 120°W, **Geophysical Research Letters**, Washington, v.32, p.L08605.1-4, 2005.

CHAVEZ, F.; BERTRAND, A.; GUEVARRA-CARRASCO, R.; SOLER, P.; CSIRKE, J. The northern Humboldt Current system: brief history, present status and a view towards the future, **Progress in Oceanography**, Oxford, v.79, p.95-105, 2008.

CRESSIE, N. **Statistics for Spatial Data**, Iowa: Willey series, 1993. p.887.

CRIALES-HERNÁNDEZ, M.; SCHAWAMBORN, R.; GRACO, M.; AYÓN, P.; HIRCHE, H.; WOLFF, M. Zooplankton vertical distribution and migration off Central Peru in relation to the oxygen minimum layer, **Helgoland Marine Research**, Heidelberg, v.62, p.85-100, 2008.

DADOU, I.; GARÇON, V.; ANDERSEN, V.; FLIERL, G.; DAVIS, C. Impact of the North Equatorial Current meandering on a pelagic ecosystem: A modeling approach, **Journal of Marine Research**, New Haven, v.54, p.311-342, 1996.

DALE, M.; DIXON, P.; FORTIN, M.; LEGENDRE, P.; MYERS, D.; ROSENBERG, M. Conceptual and mathematical relationships among methods for spatial analysis, **Ecography**, Malden, v.25, p.558-577, 2002.

DAUBECHIES, I. The wavelet transform time-frequency localization and signal analysis, **IEEE Transactions on Information Theory**, Piscataway, v.36, p.961-1004, 1990.

FALKOWSKI, P.; ZIEMANN, D.; KOLBER, Z.; BIENFANG, P. Role of eddy pumping in enhancing primary production in the ocean, **Nature**, London, v.352, p.55-58, 1991.

FOSSEIM, M.; ZHOU, M.; TANDE, K.; PEDERSEN, O.; ZHU, Y.; EDVARDBSEN, A. Interactions between biological and environmental structures along the coast of northern Norway, **Marine Ecological Progress Series**, Oldendorf, v.300. p.147-158, 2005.

FRONTIER, S.; PICHOD-VIALE, D.; LEPRETRE, A.; DAVOULT, D.; LUCZAK, C. **Ecosystèmes**. Paris: Dunod, 2004. p.550.

GILMAN, D.; FUGLISTER, J., MITCHELL, J. On the power spectrum of red noise, **Journal of the Atmospheric Sciences**, Boston, v.20, p.182-184, 1963.

GRADOS, D. **Identificación de estructuras y relaciones espaciales en los componentes del sistema de la corriente de Humboldt**: una comparación de métodos estadísticos. 2010. 128p. Tese de graduação (Engenharia Estatística) - Universidad Nacional de Ingeniería, Lima, 2010.

HAURY, L.; MCGOWAN, J.; WIEBE, P. Patterns and processes in the time-space scales of plankton distributions. In: STEELE, J. H. (Ed.). **Spatial Pattern in Plankton Communities**, New York: Plenum Press, 1978. p.277-327.

HELLY, J.; LEVIN, L. Global distribution of naturally occurring marine hypoxia on continental margins, **Deep Sea Research Part I: Oceanographic Research Papers**, Oxford, v.51, p.1159-1168, 2004.

HORMAZABAL, S.; SHAFFER, G.; LETH, O. Coastal transition zone off Chile, **Journal of Geophysical Research**, Washington, v.109, p.1-13, 2004.

HUSSON, F.; JOSSE, J.; LE, S.; MAZET, J. **FactoMineR**: multivariate exploratory data analysis and data mining with R. R package version 1.16. Disponível em: <<http://CRAN.R-project.org/package=FactoMineR>>. Acesso em: 20 abr. 2011.

HUSSON, F.; JOSSE, J.; PAGÈS, J. **Principal component methods - hierarchical clustering - partitional clustering**: why would we need to choose for visualizing data?. Rennes: Agrocampus Ouest, 2010. 17p. Relatório técnico.

KARSTENSEN, J.; STRAMMA, L.; VISBECK, M. Oxygen minimum zones in the eastern tropical Atlantic and Pacific oceans, **Progress in Oceanography**, Oxford, v.77, p.331-350, 2008.

KLEIN, P.; LAPEYRE, G. The oceanic vertical pump induced by mesoscale and submesoscale turbulence, **Marine Science**, New York, v.1, p.351-375, 2009.

MACKAS, D.; BOYD, C. Spectral analysis of zooplankton spatial heterogeneity, **Science**, Washington, v.204, p.62-64, 1979.

MACKAS, D.; DENMAN, K.; ABBOTT, M. Plankton patchiness: biology in the physical vernacular, **Bulletin of Marine Science**, Florida, v.37, p.652-674, 1985.

MAHADEVAN, A.; ARCHER, D. Modeling the impact of fronts and mesoscale circulation on the nutrient supply and biogeochemistry of the upper ocean, **Journal of geophysical research**, Washington, v.105, p.1209-1225, 2000.

McGILLICUDDY, D.; ANDERSON, L.; BATES, N.; BIBBY, T.; BUESSELER, K.; CARLSON, C.; DAVIS, C.; EWART, C.; FALKOWSKI, P.; GOLDTHWAIT, S.; HANSELL, D.; JENKINS, W.; JOHNSON, R.; KOSNYREV, V.; LEDWELL, J.; LI, Q.; SIEGEL, D.; STEINBERG, D. Eddy/Wind interactions stimulate extraordinary mid-ocean plankton blooms, **Science**, Washington, v.316, p.1021-1026, 2007.

McGILLICUDDY, D.; JOHNSON, R.; SIEGEL, D.; MICHAELS, A.; BATES, N.; KNAP, A. Mesoscale variations of biogeochemical properties in the Sargasso Sea, **Journal of Geophysical Research**, Washington, v.104, p.13381-13394, 1999.

McGILLICUDDY, D.; ROBINSON, A.; SIEGEL, D.; JANNASCH, H.; JOHNSON, R.; DICKEY, T.; McNEIL, J.; MICHAELS, A.; KNAP, A. Influence of mesoscale eddies on new production in the Sargasso Sea, **Nature**, London, v.394, p.263-266, 1998.

MENARD, F.; MARSAC, F.; BELLIER, E.; CAZELLES, B. Climatic oscillations and tuna catch rates in the Indian ocean: a wavelets approach to time series analysis, **Fisheries Oceanography**, Oxford, v.16, p.95-104, 2007.

LAVANIEGOS, B.; JIMENEZ-PEREZ, L.; GAXIOLA-CASTRO, G. Plankton response to El Niño 1997-1998 and La Niña 1999 in the southern region of the California Current, **Progress in Oceanography**, Oxford, v.54, p.33-58, 2002.

OSCHLIES, A.; GARÇON, V. Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean, **Nature**, London, v.394, p.266-269, 1998.

PAULMIER, A.; RUIZ-PINTO, D. Oxygen minimum zones (OMZs) in the modern ocean, **Progress in Oceanography**, Oxford, v.80, p.113-128, 2009.

R. R **2.13.0**. New Jersey, 2011. Disponível em: <<http://www.r-project.org/>>. Acesso em: 20 maio 2011.

ROGACHEV, K.; SALOMATIN, A.; CARMACK, C. Concentration of pelagic organisms at mesoscale fronts in the western subarctic Pacific: small fish on long waves, **Fisheries Oceanography**, Oxford, v.5, p.153-162, 1996.

RUSSEL, R.; HUNT, G.; COYLE, K.; CONNEY, R. Foraging in a fractal environment: Spatial patterns in a marine predator-prey system, **Landscape Ecology**, Dordrecht, v.7, p.195-209, 1992.

SEKI, M.; POLOVINA, J.; KOBAYASHI, D.; BIDIGARE, R.; MITCHUM, G. 2002. An oceanographic characterization of swordfish (*Xiphias gladius*) longline fishing grounds in the springtime subtropical North Pacific, **Fisheries Oceanography**, Oxford, v.11, p.251-266, 2002.

STRASS, V. Chlorophyll patchiness caused by mesoscale upwelling at fronts, **Deep Sea Research Part I: Oceanographic Research Papers**, Oxford, v.39, p.75-96, 1992.

TEW-KAI, E.; ROSSI, V.; SUDRE, J.; WEIMERSKIRCH, H.; LOPEZ, C.; HERNANDEZ-GARCIA, E.; MARSAC, F.; GARÇON, V. Top marine predators track lagrangian coherent structures, **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v.106, p.8245-50, 2009.

THE MATHWORKS. **MATLAB R2009b**. Natick, 2009. Disponível em: <<http://www.mathworks.com/>>. Acesso em: 20 mar. 2011.

THOMAS, A. Seasonal distributions of satellite-measured phytoplankton pigment concentration along the Chilean coast, **Journal of Geophysical Research Oceans**, Washington, v.104, p.877-890, 1999.

TILMAN, D.; KAREIVA, P. **Spatial ecology**: the role of space in population dynamics and interspecific interactions. New Jersey: Princeton University Press, 1997. p.365.

TORRENCE, C.; COMPO, G. A practical guide to wavelet analysis, **Bulletin of the American Meteorological Society**, Boston, v.79, p.61-78, 1998.

WILLIAMS, R.; FOLLOWS, M. Physical transport of nutrients and the maintenance of biological production. In: FASHAM, M. J. R. (Ed.). **Ocean Biogeochemistry**: The Role of the Ocean Carbon Cycle in Global Change. New York: Springer, 2007. chap.2, p.19-51.

WIEBE, P. Small-scale spatial distribution in oceanic zooplankton, **Limnology and Oceanography**, Texas, v.15, p.205-217, 1970.

YEBRA, L.; ALMEIDA, C.; HERNANDEZ-LEON, S. Vertical distribution of zooplankton and active flux across an anticyclonic eddy in the Canary Island waters, **Deep Sea Research Part I: Oceanographic Research Papers**, Oxford, v.52, p.69-83, 2005.