

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA ECONÓMICA Y CIENCIAS SOCIALES
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA



**IDENTIFICACIÓN DE ESTRUCTURAS Y RELACIONES ESPACIALES EN
LOS COMPONENTES DEL SISTEMA DE LA CORRIENTE DE
HUMBOLDT: UNA COMPARACIÓN DE MÉTODOS ESTADÍSTICOS**

TESIS PARA OPTAR POR EL TÍTULO PROFESIONAL DE INGENIERÍA ESTADÍSTICA

DANIEL ISAIAS GRADOS PAREDES

LIMA – PERÚ

2010

*A mi madre y mis hermanos.
Y a todos aquellos que hacen
investigación para el desarrollo del Perú.*

AGRADECIMIENTOS

Gracias a IMARPE por haberme permitido utilizar sus instalaciones para poder desarrollar este trabajo y por haberme permitido utilizar la información necesaria para la investigación. Al IRD por financiar mi estadía en IMARPE.

Agradezco profundamente a Arnaud por la orientación en un mundo nuevo para mí que es la ecología. Por su paciencia, las enseñanzas y su amistad. Te agradezco por las críticas y los consejos pues me ayudan a mejorar. Gracias Sophie por tu amistad. Muchas gracias Arnaud y Sophie, pues nos enseñan a valorar y cuidar las pequeñas cosas que tenemos, gracias por aquellas conversaciones, no solamente de ecología, sino también por aquellas concernientes a política, cultura, etc.

A Michael, por aquellas primeras conversaciones de ecología marina, por tu apoyo y consejos en muchas etapas de mi tesis y también por aquellos chistes que solamente tú entiendes.

A mi profesora Magen Por su constante ayuda, por sus recomendaciones y por lo enseñado en mi etapa de estudiante en la UNI.

A Rocío por ser la mejor amiga que he tenido, por su constante apoyo y por su gran paciencia, por buscar un país con más igualdad.

A mis amigos de la universidad que me acompañan siempre. A todos aquellos que han estado, que están y que estarán en mi largo camino de la vida.

Resumen

Para lograr la sostenibilidad de los recursos marinos es necesario comprender las relaciones que existen entre los diferentes procesos que ocurren dentro de los ecosistemas (procesos físicos o procesos biológicos). La mayoría de los datos que provienen de la ecología tienen una fuerte autocorrelación, la cual hace imposible aplicar métodos estadísticos tradicionales que requieren el supuesto de independencia de las observaciones.

En este trabajo hemos aplicado y comparado 4 metodologías que nos permiten trabajar con datos autocorrelacionados; estas metodologías son: la correlación con ajuste del tamaño de muestra, los métodos geoestadísticos, los PCNM (análisis de coordenadas principales) y los métodos wavelets. Dichas metodologías fueron aplicadas para estudiar las estructuras presentes en la profundidad de la zona mínima de oxígeno (ZMO), el zooplancton y los peces del Sistema de la Corriente de Humboldt.

Como resultado se pudo observar las ventajas que ofrecen los métodos PCNM para estudiar las estructuras para cada variable y las interacciones entre la ZMO y el zooplancton, el zooplancton y peces y entre la ZMO y peces. También se observó que los cuatro métodos dieron como resultado la presencia de tres estructuras presentes en cada una de las tres variables. Siendo estas estructuras desde los 500 m hasta los 30 km, desde los 30 km hasta los 60 o 70km y finalmente desde los 70 km hasta los 130 km. Estas estructuras corresponden a dos escalas espaciales que son: la sub-meso escala (100s m a kms) y la mesoescala (10s km a 100s km). Finalmente se observó que existe una relación positiva entre la ZMO y el zooplancton, también se observó relaciones positivas entre el zooplancton y peces.

ÍNDICE

Resumen y abstract	1
Capítulo I: Introducción	
1. Antecedentes	5
2. Planteamiento del problema	9
3. Objetivos	9
4. Hipótesis	10
Capítulo II: Marco Teórico	
1 Correlación con ajuste del tamaño de muestra	11
2 Geoestadística	13
2.1 Teoría de las variables aleatorias regionalizadas	13
2.1.1 Variables aleatorias	13
2.1.2 Funciones aleatorias	14
2.2 Covarianza espacial	14
2.2.1 Estacionariedad	14
2.2.2 Variación Intrínseca	15
2.3 Equivalencia entre el variograma y la covarianza	16
2.4 Análisis Estructural	16
2.4.1 Cálculo del Variograma	16
2.4.2 Modelando el Variograma	18
2.4.3 Modelos de Variogramas	19
2.5 Anisotropía	24
2.6 Variograma robusto	26
2.7 Identificación de las escalas relevantes mediante variogramas anidados	26
2.8 Modelando la covarianza espacial entre la variable respuesta y las variables predictoras: Variogramas cruzados	27
3 Análisis de Coordenadas Principales de Matrices Vecinas (PCNM)	28
3.1 Cálculo de la matriz de distancias	31
3.2 Modificando la Matriz de Distancias	31
3.3 Análisis de Coordenadas Principales	32
3.4 Caracterización de las escalas espaciales mediante los PCNM variables	33
3.5 Relación entre los submodelos y otras variables	34
4 Análisis Wavelets	35
4.1 La transformada de wavelets	35

4.2 Espectro de la Potencia de Wavelets	36
4.3 Espectro Global de wavelets	37
4.4 Análisis wavelets cruzado	37
Capítulo III: Metodología	
1 Tipo de estudio	40
2 Definición de la población de estudio y pre-tratamiento de los datos	40
2.1 Área de estudio	40
2.2 Datos Acústicos	41
2.3 Pre-procesamiento	41
2.4 Discriminación de los grupos	42
3 Descripción de las variables	44
3.1 Profundidad de la zona mínima de oxígeno	44
3.2 Estimación de biomاسas	45
3.2.1 Abundancia de fluid like	45
3.2.2 Abundancia de fish	46
3.3 Base de datos	46
Capítulo IV: Aplicación	
1 Distribución del fluid like según periodo	47
2 Aplicación del test de correlación con ajuste del tamaño de muestra	47
2.1 Correlación con modificación del tamaño de muestra	48
3 Aplicación del método geoestadístico	51
3.1 Modelamiento geoestadístico	52
3.1.1 Análisis de estructuras espaciales de la variable física	52
3.1.2 Análisis de estructuras espaciales del log fluid like	53
3.1.3 Análisis de estructuras espaciales del log fish	57
3.2 Analizando la estructura de coregionalización	60
3.2.1 Coregionalización entre variable física y log fluid like	60
3.2.2 Coregionalización entre variable física y log fish	61
3.2.3 Coregionalización entre log fluid like y log fish	63
4 Aplicación de los PCNM	65
4.1 Descomposición del espacio	66

4.2 Análisis de estructuras	68
4.2.1 Estructuras espaciales para la variable física	68
4.2.2 Estructuras espaciales para el log fluid like	70
4.2.3 Estructuras espaciales para el log fish	72
4.3 Análisis de los submodelos (relaciones entre variables)	75
4.3.1 Relación entre el log fluid like y la variable física	75
4.3.2 Relación entre el log fish, el log fluid like y la variable física	76
5 Aplicación de los análisis wavelets	78
5.1 Análisis wavelets	78
5.2 Análisis del Wavelets cruzado	79
Capítulo V: Discusión	
1 De las metodologías	84
2 Identificación de las estructuras espaciales	88
3 Relación entre Zooplankton y Zona Mínima de Oxígeno (ZMO)	89
4 Relación entre fish con covariables el zooplankton y la ZMO	90
Capítulo VI: Conclusiones	94
Bibliografía	96
Anexo 1: Regresión robusta	100
Anexo 2: Suavización por spline	101
Anexo 3: Resultados wavelets	102
Anexo 4: Otros métodos para tratar datos con autocorrelación espacial	108
Anexo 5: Pruebas de permutación	112

CAPITULO I:
INTRODUCCIÓN

1. Antecedentes

En los últimos años la pesca marítima está sufriendo una sobreexplotación global¹. Según los informes emitidos por la FAO, el 47% de las especies están siendo explotadas, el 18% se encuentran en estado de sobreexplotación y el 10% se encuentra en recuperación del agotamiento (FAO, 2002). Como ejemplos de esta sobreexplotación podemos mencionar a los stocks de sardinas de California y Japón que colapsaron en los años 40 (Figura 1). Un ejemplo más reciente es lo ocurrido en Canadá, en la década de los 80, la sobreexplotación industrial de la pesca del Bacalao trajo como consecuencia la desaparición de gran parte de la población de esta especie. A pesar de que la pesquería de bacalao fue cerrada en 1994, a la fecha la población no se recupero.

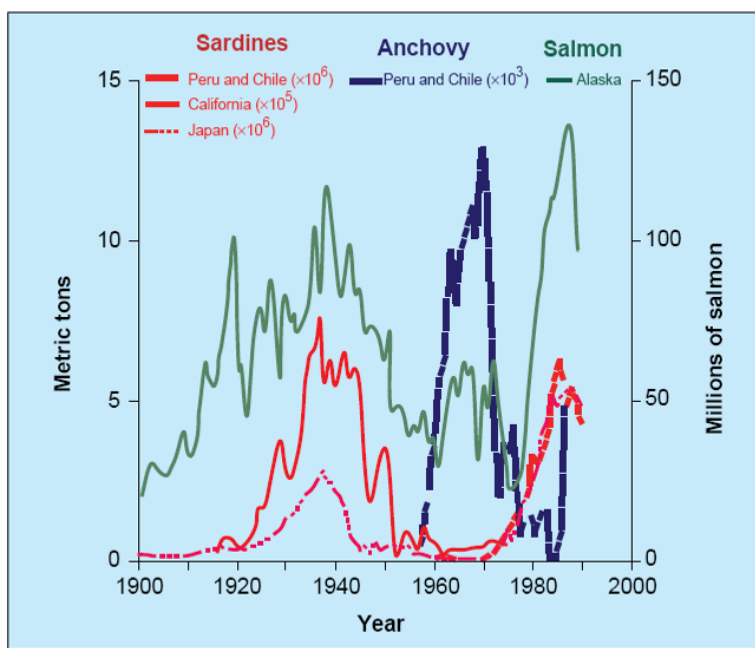


Figura 1. Serie temporal de captura para varias especies de peces del pacífico; fuente, Botsford et al. 1997

Para el Perú, la actividad pesquera es muy importante económicamente. Representa la segunda fuente generadora de divisas después de la minería. La pesquería industrial se inicio en los años 50, teniendo un crecimiento lento. En los años 60, la pesca basada sólo en anchoveta (*Engraulis ringens*) se cuadruplicó de 3.5 millones a 12 millones de toneladas capturadas. Sin embargo a

¹<http://web.worldbank.org/WBSITE/EXTERNAL/BANCOMUNDIAL/NEWSSPANISH/0,,print:Y~isCURL:Y~contentMDK:20626008~pagePK:64257043~piPK:437376~theSitePK:1074568,00.html>

principios de los años 70, la conjunción de (i) el exceso de las capturas, (ii) la ocurrencia de un fenómeno El Niño fuerte (1972 - 1973) y (iii) el inicio de un periodo decadal menos favorable a la anchoveta, llevaron al colapso de esta pesca (Alheit y Ñiquen, 2004; Bertrand et al., 2004b). Recién a comienzos de los 90 se evidenció una recuperación de la actividad². En el año 2001, estos ingresos representaron 1,124 millones de dólares en exportaciones³.

Los casos anteriormente mencionados dan evidencia de que los métodos tradicionales de “manejo pesquero” son muy ineficientes. Estos se focalizaban en la optimización de la captura de una especie y frecuentemente ignoraban el hábitat, los predadores y presas de esta especie (Pikitch et al., 2004). Como claramente lo demuestran los casos anteriormente descritos, ya no es necesario brindar información a los pescadores para que extraigan más. Por el contrario se debe de realizar investigaciones que permitan brindar recomendaciones sobre la sostenibilidad del recurso y del ecosistema. Como respuesta a este fracaso a nivel mundial, la FAO propone el “Manejo Pesquero Basado en el ecosistema” que tiene como objetivo la sostenibilidad de los ecosistemas marinos, evitar su degradación, minimizar el riesgo del cambio irreversible de los procesos en los ecosistemas y especies, y generar suficiente conocimiento y entendimiento de las consecuencias de las acciones del hombre sobre el ecosistema (Pikitch et al., 2004). En el Perú, el Instituto del Mar del Perú (IMARPE) es el encargado de monitorear e investigar la actividad pesquera, biológica y dinámica de poblaciones que habitan en el mar peruano.

Como es sintetizado por Chavez et al. (2008), el océano tropical de la costa oeste de Sud América es notable por muchas razones. Primero, el afloramiento de las aguas frías trae los nutrientes hacia el fitoplancton a la superficie, incrementando dramáticamente la productividad biológica en esta región. Segundo, como resultado de la alta productividad, el norte del Sistema de la Corriente de Humboldt (SCH) frente al Perú produce más peces por unidad de área que cualquier otra región oceánica del mundo. Representando menos de 0.1% de la superficie mundial del océano, el SCH produce alrededor del 10% de la captura mundial de peces. Tercero, como

² http://www.fao.org/fishery/countrysector/FI-CP_PE/es

³ http://www.fao.org/fishery/countrysector/FI-CP_PE/es

resultado de la precipitación de materia orgánica derivados de la producción primaria superficial, así como de una pobre ventilación, existe una extensa y extremadamente superficial Zona Mínima de Oxígeno (ZMO) ubicada debajo de estas aguas oxigenadas superficiales. La ZMO es una barrera para muchos animales por los que muchos se concentran cerca de la superficie. En el otro extremo del espectro diferentes formas de vida marina se han adaptado a este duro medio el que utilizan como zona de refugio de predadores. Cuarto, debajo la ZMO, en los sedimentos de la plataforma continental la ausencia de oxígeno preserva organismos muertos, creando un récord del ecosistema pasado y de los cambios climáticos. Quinto, a través del acoplamiento atmósfera/océano el SCH está íntimamente vinculado a las dinámicas del Pacífico ecuatorial y está sujeto a grandes fluctuaciones interanuales a multidecadales en el clima, los ecosistemas y sus pesquerías.

Para comprender el SCH es necesario estudiar la interacción entre el componente físico y los componentes biológicos, teniendo cada variable un componente espacial. Para los ecólogos, un gran desafío es comprender el comportamiento o la dinámica de las poblaciones marinas (interacciones dentro y entre poblaciones). Para tal fin es necesario desarrollar métodos de observación de dicho ecosistema, Como los recientes avances de las técnicas y métodos de observación acústica submarina (Bertrand et al., 2003). Estos métodos consisten en la emisión de pulsos (ondas de sonido) desde transductores ubicados debajo del casco del barco en dirección al fondo y la recepción de los ecos devueltos por los obstáculos que las ondas encontraron en el camino. A partir de estos ecos de sonido (de diferentes intensidades y a diferentes frecuencias) se realizan estimaciones para así inferir la biomasa de diferentes especies (Simmonds y MacLennan, 2005).

Dichas interacciones ecológicas también tienen un componente espacial, el cual es considerado como factor determinante de estas interacciones (Rose y Legget, 1990; Levin, 1992; Bertrand et al., 2004a). Estas interacciones ocurren a diferentes escalas. Levin (1992) estima que para cada una de las escalas de análisis corresponde la descripción de un mecanismo dado. Sin embargo, no

se puede decir que un proceso o mecanismo esté restringido a una escala particular, sino que existen escalas en las cuales un proceso resulta relevante o prevalece sobre otros (Wiens, 1989). Por ese motivo el análisis de estructuras espaciales es de gran interés para los ecólogos porque muchas investigaciones de fenómenos ecológicos son estructuradas por fuerzas que tienen componentes espaciales (Legendre y Legendre, 1998).

En el presente estudio tenemos datos recolectados por el IMARPE mediante técnicas acústicas. Procesamientos adecuados (Ballón et al., in prep) permitieron extraer información de la profundidad de la zona mínima de oxígeno (componente físico), abundancia del zooplancton y abundancia de peces (ambos componentes biológicos). Para hacer un estudio válido de dichas estructuras es necesario plantear métodos estadísticos que nos permitan valorar los datos colectados *in situ*. Sin embargo el uso de herramientas estadísticas ha sido tradicionalmente muy limitada por los ecólogos (López de Luzuriaga et al., 2006). La indiferencia de los ecólogos por el problema de la escala en sus estudios podría haberse debido, en gran medida, a que la teoría ecológica fue desarrollada sin tenerla en cuenta (Cueto, 2006).

A menudo, la estadística clásica no es adecuada para el análisis de datos ecológicos pues la presencia de estructuras o autocorrelación espacial implica la violación del supuesto de independencia de las observaciones (Legendre y Legendre, 1998; Dormann et al., 2007), fundamental para muchas técnicas clásicas. Tratar de eliminar la autocorrelación de los datos que se posee no es adecuado pues la importancia de la heterogeneidad espacial radica en la importancia para las teorías ecológicas (Legendre y Fortín, 1989). Por tal motivo, se deben usar métodos que incorporen el componente espacial en el análisis y también permitan identificar dichas escalas características.

Comúnmente la asociación entre dos variables es cuantificada mediante la correlación lineal. Esta correlación presenta serias deficiencias cuando las variables se encuentran autocorrelacionadas. Clifford et al. (1989) presenta el cálculo de la correlación lineal con ajuste del

tamaño de muestra para evaluar el test de significancia estadística en presencia de autocorrelación en las variables. Un método que nos permite cuantificar la variabilidad espacial es la geoestadística. El cual muestra cómo la dependencia espacial de la varianza y la de la covarianza cambian entre observaciones. Para Cressie (1993), la estructura espacial es comúnmente modelada por el ajuste de una función de variograma a un variograma empírico. Borcard y Legendre (2002) desarrollaron recientemente un método estadístico llamado “Coordenadas Principales de Matrices Vecinas” (PCNM son sus siglas en inglés), que puede ser usado para describir la escala espacial dominante. El PCNM consiste en construir una matriz de distancias euclidianas a partir de las coordenadas geográficas de los puntos muestreados, luego se procede a extraer las coordenadas principales de dicha matriz. La ventaja de esta metodología es que nos permite trabajar a un continuum de escalas, y estudiar el efecto de otras variables sobre la variable respuesta. Por otro lado, los métodos wavelets no necesitan suponer estacionariedad en la serie (Dale y Mah, 1998). Su principal ventaja se basa en la descomposición del espectro de la serie en frecuencia y escala, lo que nos permite observar la serie a pequeña escala y a gran escala. Y finalmente mediante los métodos de wavelets cruzados podemos estudiar la interacción entre dos variables.

2 Planteamiento del problema

Lo anteriormente descrito nos lleva a plantearnos las siguientes preguntas:

¿Qué metodología estadística nos permite identificar estructuras y relaciones espaciales entre componentes del SCH?

¿Cuáles son las estructuras espaciales presentes en los componentes físicos y biológicos del SCH?

¿Cuáles son las escalas espaciales presentes en los componentes físicos y biológicos del SCH?

¿Cómo son las relaciones espaciales entre los componentes físicos y biológicos en el SCH?

3 Objetivos

General

Hacer una comparación de metodologías estadísticas que cuantifican estructuras y relaciones espaciales y determinar la más adecuada para la elección de estructuras espaciales.

Específicos

- Identificar las estructuras espaciales presentes en los componentes físicos y biológicos del SCH.
- Identificar las escalas espaciales que caracterizan los componentes físicos y biológicos en el SCH.
- Determinar la relación existente en las interacciones entre los componentes físicos y biológicos del SCH.

4 Hipótesis

La profundidad del límite superior de la Zona mínima de oxígeno condiciona la distribución espacial del Zooplancton, así como el zooplancton, también, condiciona la distribución espacial de los peces. Estas condicionantes ocurren a diferentes escalas. Ocurren diferentes relaciones a cada escala de medición.

CAPITULO II:
MARCO TEÓRICO

Diferentes técnicas estadísticas han sido implementadas para trabajar con fenómenos que se encuentran autocorrelacionados (la autocorrelación refleja el grado de asociación entre observaciones cercanas). Entre estos métodos desarrollaremos la correlación con ajuste del tamaño de muestra, métodos geoestadísticos, métodos de PCNM y wavelets. Estos métodos fueron elegidos porque nos permiten realizar el estudio desde distancias pequeñas a grandes distancia.

Abordaremos los conceptos teóricos de los 4 métodos, haciendo hincapié en el uso de estas.

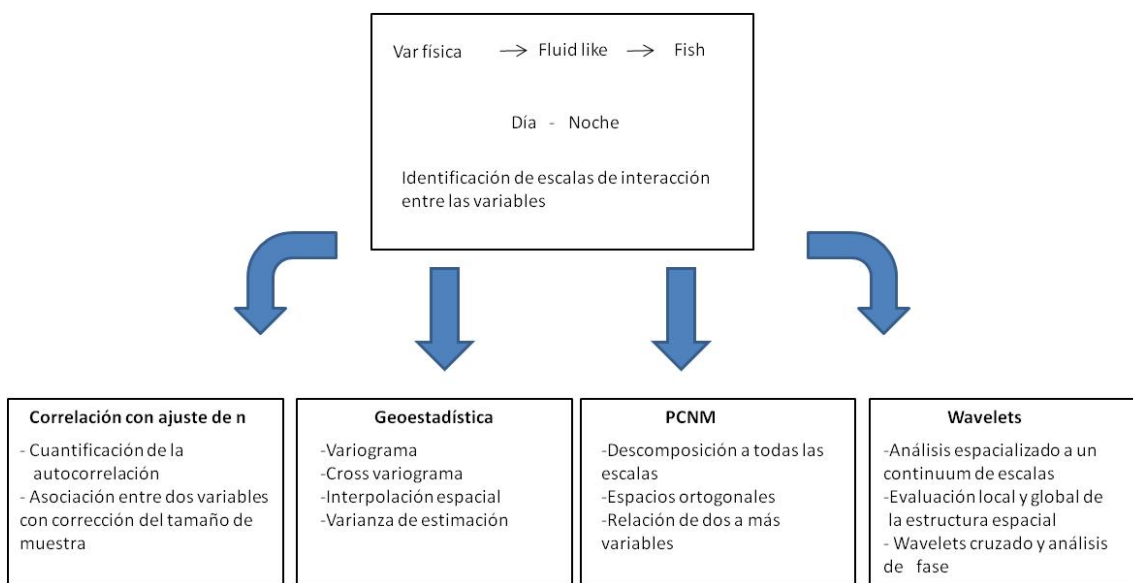


Figura 2. Esquema gráfico de los métodos a utilizar para analizar los datos. Métodos estadísticos con sus respectivas utilidades para una serie de datos.

1 Correlación con ajuste del tamaño de muestra

La correlación lineal es utilizada para evaluar la relación lineal entre dos variables. Teniendo como supuesto la independencia de las observaciones dentro de cada variable (no autocorrelacionadas). Pero la autocorrelación (espacial o temporal) está presente en las variables ecológicas (Legendre y Legendre 1998). Ante el problema del cálculo de la correlación lineal entre dos series autocorrelacionadas, Clifford et al. (1989) proponen calcular esta correlación y para evaluar la significancia de esta correlación propone ajustar el tamaño de muestra. Este tamaño de muestra ajustado toma en consideración la autocorrelación presente en cada serie de datos.

Sea S_{xx} y S_{yy} las varianzas muestrales de dos variables x y y , y S_{xy} la covarianza de estas dos variables. El coeficiente de correlación es usualmente calculado como:

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \dots (1)$$

$$\text{donde } S_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N}, S_{xx} = \frac{\sum (X - \bar{X})^2}{N} \text{ y } S_{yy} = \frac{\sum (Y - \bar{Y})^2}{N}$$

$$\text{Y su test de significancia estadística está dado por } t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Donde la distribución t tiene $n-2$ grados de libertad (esto es para series sin autocorrelación).

Cuando las variables tienen autocorrelación espacial el tamaño de la muestra (n) puede ser reemplazado por un tamaño de muestra efectivo que toma en cuenta esta autocorrelación. Una propuesta del tamaño de muestra efectivo fue dado por Clifford et al. (1989) el cual viene dado por: $n'_r = 1 + S_r^{-2}$ (S_r^2 es la varianza de la autocorrelación). Un estimador de la varianza de la correlación está basado en la autocorrelación que existe en las variables x y y . Y haciendo la semejanza de las variables x y y con una serie temporal (igualmente espaciados) se calcula la autocorrelación presente a cada distancia k , entonces $r_x(k)$ y $r_y(k)$ son las autocorrelaciones para las variables x y y , con $n(k)$ el número de pares para una distancia de clase k .

La autocorrelación se calcula como:

$$r_y(k) = \frac{1}{n-k-1} \sum_{i=1}^{n-k} (y_{i+k} - \bar{y}') (y_i - \bar{y}'') / \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \dots (2)$$

$$\text{con } \bar{y}' = \frac{1}{n-k} \sum_{i=k+1}^n y_i \text{ y } \bar{y}'' = \frac{1}{n-k} \sum_{i=1}^{n-k} y_i$$

El cálculo del tamaño de muestra efectivo usando las autocorrelaciones presentes en cada variable es como sigue:

$$n'_r = 1 + \frac{n^2}{n + 2 \sum_{k=1}^{n-1} (n-k)r_x(k)r_y(k)} \dots (3)$$

Finalmente, para calcular la significancia de la autocorrelación se reemplaza el tamaño de muestra efectivo n'_r en el test de significancia.

2 Geoestadística

La geoestadística se define como el estudio de fenómenos regionalizados, es decir, fenómenos que se extienden en el espacio. A la variable que describe un fenómeno regionalizado se le llama variable regionalizada. La geoestadística asocia la aleatoriedad a la variable regionalizada, para lo cual se plantea un modelo estocástico según el cual se toma a la variable regionalizada como una posible realización de una función aleatoria. El conocimiento de que la variabilidad espacial es una fuente de incertidumbre, implica la necesidad de cuantificar y modelar esta variabilidad espacial. La geoestadística cuantifica y modela la variabilidad espacial por medio de un análisis estructural del proceso. Su herramienta principal (en el análisis estructural) es el variograma, que permite cuantificar la autocorrelación espacial de los datos.

2.1 Teoría de las variables aleatorias regionalizadas

2.1.1 Variables aleatorias

Una variable aleatoria es una variable que puede tomar ciertos valores de acuerdo a una distribución de probabilidad.

Formalmente se define de la siguiente manera: dado un espacio de probabilidad (E, A, P) , una variable aleatoria x es la aplicación del espacio muestral E en \mathfrak{R} que es medible respecto de la σ – álgebra de Borel en \mathfrak{R} y A .

2.1.2 Funciones aleatorias

Una función aleatoria es un proceso estocástico definido en un espacio, o también, una función cuyos valores son realizaciones de variables aleatorias en cualquier punto del espacio. Una variable aleatoria regionalizada es una realización de una función aleatoria.

2.2 Covarianza espacial

Cuando se tiene 2 variables $(z_1$ y $z_2)$, el cálculo de la covarianza es dado por:

$$Cov(z_1, z_2) = E[(z_1 - \bar{z}_1)(z_2 - \bar{z}_2)] \dots (4)$$

donde \bar{z}_1 y \bar{z}_2 son las medias de z_1 y z_2 respectivamente. La covarianza es una función simétrica.

Haciendo la extensión a un proceso espacial, sea x_1 y x_2 los valores de $Y(x_1)$ y $Y(x_2)$, la covarianza es:

$$Cov(x_1, x_2) = E[(Y(x_1) - \mu(x_1))(Y(x_2) - \mu(x_2))] \dots (5)$$

donde $\mu(x_1)$ y $\mu(x_2)$ son las medias de $Y(x_1)$ y $Y(x_2)$ respectivamente. Sin embargo la solución de la ecuación 6 no es simple porque solamente se tiene una realización del proceso Y en cada punto (Webster, 2007). Lo anterior puede solucionarse tomando el supuesto de estacionariedad. Es en el supuesto de estacionariedad en que se basa la teoría geoestadística.

2.2.1 Estacionariedad

Un proceso estacionario es un fenómeno homogéneo que se desarrolla en el espacio. Se dice que un proceso es **estrictamente estacionario** cuando una distribución de dimensiones finitas

permanece invariante ante cualquier translación arbitraria de los puntos por medio de h (distancia de separación entre observaciones). Esto se expresa por medio de la probabilidad de:

$$P(Y(x_i) < y_i, \dots, Y(x_k) < y_k) = P(Y(x_i + h) < y_i, \dots, Y(x_k + h) < y_k) \dots (6)$$

Una condición menos restrictiva está dada por la estacionariedad débil o también llamado **segundo orden de estacionariedad**. Esta condición es más frecuente en la práctica. Hablamos de un proceso espacial $Y(x)$, de estacionariedad débil si:

- i) $E(Y(x)) = \mu$ (media constante que no depende de la ubicación en el espacio).
- ii) $Cov(Y(x), Y(x+h)) = C(h)$ para todo $h \in \mathfrak{R}^2$.

La estacionariedad débil implica que la covarianza entre dos observaciones puede ser resumida por la función de covarianza $C(h)$. Y esta función $C(h)$ depende únicamente de la separación h entre las dos observaciones.

2.2.2 Variación Intrínseca

Representamos un proceso estacionario por el modelo:

$$Y(x) = \mu + \varepsilon(x) \dots (7)$$

donde μ es la media del proceso y $\varepsilon(x)$ es el componente aleatorio de una distribución con media cero. Cuya función de covarianza se puede estimar de la forma siguiente:

$$C(h) = E[\varepsilon(x)\varepsilon(x+h)] \dots (8)$$

Matheron (1965) plantea que para valores pequeños de h se cumple:

$$E[Y(x+h) - Y(x)] = 0 \dots (9)$$

Reemplazando las covarianzas por las varianzas de la diferencia, tenemos:

$$E[Y(x+h) - Y(x)]^2 = Var(Y(x+h) - Y(x)) = 2\gamma(h) \dots (10)$$

Las dos ecuaciones anteriores (9 y 10) constituyen las hipótesis intrínsecas de Matheron (Webster, 2007). En la ecuación 10 podemos notar que solamente el lado derecho depende de h .

La función $2\gamma(h)$ es llamada variograma y $\gamma(h)$ es el semi variograma (en la práctica también es llamado variograma).

2.3 Equivalencia entre el variograma y la covarianza

La relación entre el variograma y la función de covarianza viene dada por:

$$\begin{aligned}2\gamma(h) &= \text{Var}(Y(x+h) - Y(x)) \\ &= \text{Var}(Y(x+h) - Y(x)) - 2\text{Cov}(Y(x+h), Y(x)) \\ \gamma(h) &= C(0) - C(h) \dots (11)\end{aligned}$$

De la ecuación 11 observamos que el variograma es simplemente es una imagen reflejada sobre el eje de abscisas de la función de covarianza (Figura 3a).

El variograma también se relaciona con la función de autocorrelación de la siguiente manera:

$$\gamma(h) = \sigma^2 (1 - \rho(h)) \dots (12)$$

Mediante las ecuaciones 11 y 12, notamos que el variograma es siempre mayor a 0, aún si la correlación es 0. Por ello el variograma es la herramienta fundamental de la teoría geoestadística.

2.4 Análisis Estructural

El análisis estructural está compuesto por:

- 1) el cálculo del variograma.
- 2) El ajuste de un modelo teórico al variograma.

El variograma permite cuantificar el comportamiento espacial de un variable sobre un área.

2.4.1 Cálculo del Variograma

El variograma es la varianza de los incrementos de la variable regionalizada en las ubicaciones separadas una distancia h .

El variograma es la piedra angular de la geoestadística por lo que es necesario estimarlo.

De la ecuación 10 nosotros notamos que el cálculo del variograma es:

$$\gamma(h) = \frac{1}{2} E[Y(x) - Y(x+h)]^2 \dots (13)$$

El estimador del variograma viene representado por:

$$\bar{\gamma}(h) = \frac{1}{2} \text{mean}[Y(x) - Y(x+h)]^2 \dots (14)$$

donde $Y(x)$ y $Y(x+h)$ representan los valores de Y separados por una distancia h .

Para un conjunto de datos $Y(x_i)$, $i=1, \dots, n$ calculamos $\bar{\gamma}(h)$ como:

$$\bar{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{Y(x_i) - Y(x_i+h)\}^2 \dots (15)$$

donde h es el incremento y $m(h)$ es el número de pares de datos separados por una distancia h .

El variograma puede ser calculado para diferentes direcciones (se llaman variogramas direccionales). Cuando el variograma es calculado en todas las direcciones se llama variograma omnidireccional, el cual representa de manera global la estructura del proceso.

Características del variograma

Simetría. Nosotros podemos representar la función de covarianza como:

$$\begin{aligned} C(h) &= E[\{Y(x) - \mu\}\{Y(x+h) - \mu\}] \\ &= E[\{Y(x-h)Y(x) - \mu^2\}] \\ &= E[\{Y(x)Y(x-h) - \mu^2\}] \\ &= -C(h) \end{aligned} \dots (16)$$

En otras palabras la autocovarianza es simétrica en el espacio (Webster et al., 2007). Esto también implica que $\gamma(h) = -\gamma(h)$ para todo h .

Semi-definida Positiva. La matriz de covarianza es semi-definida positiva. Esto es necesario porque una varianza no puede ser negativa.

Se presenta cuando el límite de $C(h)$ para $h \rightarrow 0$ es diferente de cero (Figura 3c y d)

Incremento Monótono. Los variogramas son funciones crecientes, es decir, la varianza incrementa con los incrementos de la distancia (Figura 3a).

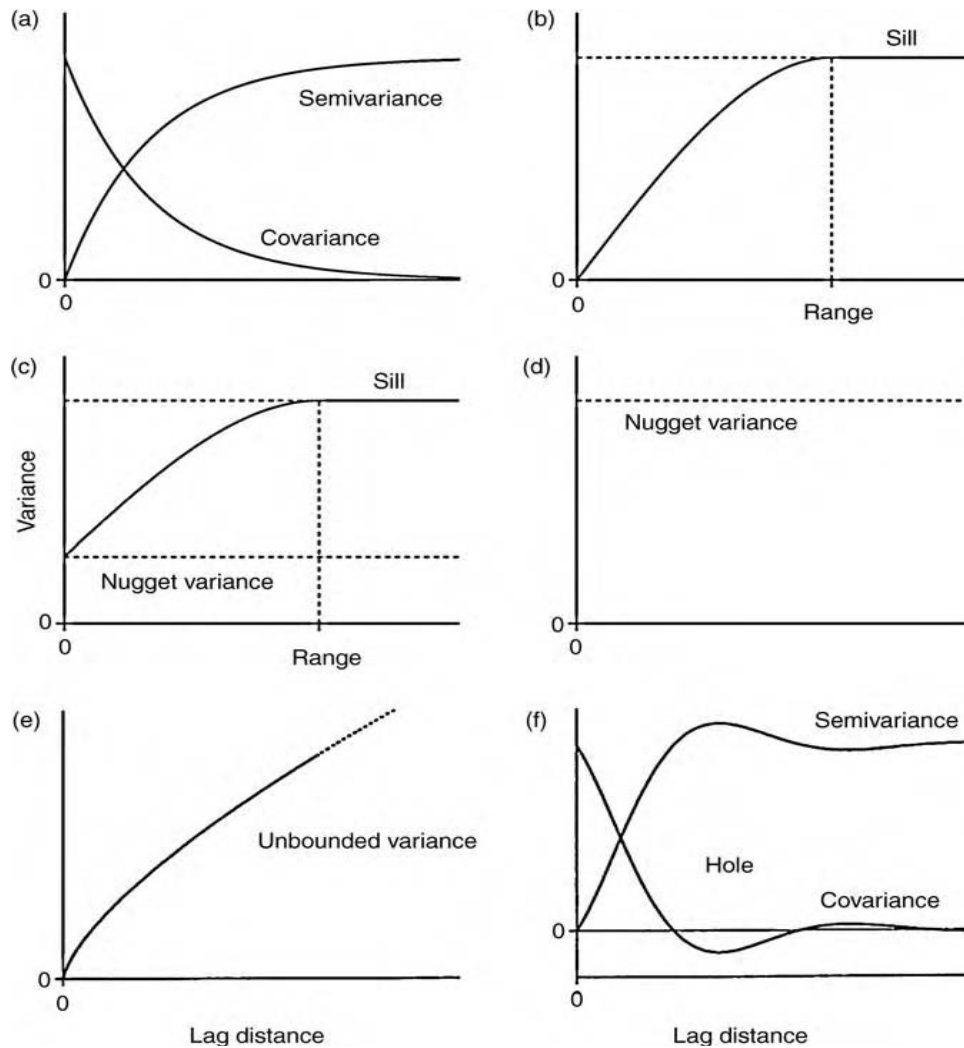


Figura 3. Características de los variogramas. a) Incrementos monótonos de la varianza, b) variograma sin efecto pepita (o varianza nugget), c) variograma con efecto pepita, d) efecto pepita puro, e) variograma no acotado, f) variograma con efecto hueco.

Variograma no acotado (Unbounded). Ocurre cuando el variograma incrementa indefinidamente.

En estos casos no existe la función de covarianza (Figura 3e).

Efecto hueco (Hole). En algunos casos los variogramas decrecen de un máximo a un mínimo local (Figura 3f).

2.4.2 Modelando el Variograma

Una vez construido el variograma el siguiente paso es ajustar a este un modelo teórico, con el objetivo de determinar los parámetros descriptivos del variograma.

Los parámetros son:

- **Efecto pepita (nugget).** El variograma teóricamente debe de ser nulo en el origen, pero en la práctica las funciones obtenidas pueden presentar discontinuidad en el origen. A esta discontinuidad se la conoce como el efecto pepita. El efecto pepita es la discontinuidad del variograma para distancias menores a h (Figura 3c).
- **Meseta (Sill) y Rango (Range).** Los variogramas de procesos estacionarios de segundo orden son limitados superiormente. El máximo valor superior es conocido como meseta de la varianza. El rango es el límite de la dependencia espacial (Figura 3b y c).

Denotaremos h por t y definiremos los modelos más importantes.

2.4.3 Modelos de Variogramas

Los modelos de variogramas que frecuentemente se usan son la función esférica, la exponencial y la gaussiana. Sin embargo cuando se modelan datos provenientes de comunidades ecológicas, el modelamiento de la varianza a lo largo de la distancia (variograma) frecuentemente presenta estructuras cíclicas. Este comportamiento puede ser modelado con efectos sinusoidales usando la función Bessel.

Lineal

No posee un valor de meseta (es infinito), el variograma continúa aumentando mientras la distancia t sigue aumentando.

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{si } t > 0, \tau^2 > 0, \sigma^2 > 0 \\ 0 & \text{otro caso} \end{cases} \dots (17)$$

Esférica

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{si } t \geq 1/\phi, \\ \tau^2 + \sigma^2 \left\{ \frac{3}{2}(\phi t) - \frac{1}{2}(\phi t)^3 \right\} & \text{si } 0 < t \leq 1/\phi, \\ 0 & \text{otro caso} \end{cases} \dots (18)$$

La función de covarianza de la forma esférica puede ser fácilmente calculada. Por definición $\gamma(0^+) = \lim_{t \rightarrow 0^+} \gamma(t) = \tau^2$; esta cuantificación es el efecto pepita: $\lim_{t \rightarrow \infty} \gamma(t) = \tau^2 + \sigma^2$, Este valor es llamado meseta y finalmente $t = 1/\phi$ evaluado en $\gamma(t)$ es el rango.

Este modelo constituye el modelo más utilizado. Muestra un crecimiento casi lineal hasta una cierta distancia (rango), donde se estabiliza. La Figura 4a es una representación de un modelo esférico.

Exponencial

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi t)) & \text{si } t \geq 0, \\ 0 & \text{otro caso} \end{cases} \dots (19)$$

La ventaja de la familia exponencial sobre la familia esférica radica en la simplicidad de la función.

En la Figura 4d podemos observar una imagen de este modelo exponencial.

La función de covarianza de la familia exponencial viene dado de la siguiente forma:

$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{si } t = 0, \\ (\sigma^2 \exp(-\phi t)) & \text{si } t \geq 0 \end{cases} \dots (20)$$

Gausiano

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi^2 t^2)) & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases} \dots (21)$$

El variograma gausiano es una función analítica y produce procesos espaciales muy suaves.

Este modelo es muy empleado para modelar fenómenos continuos (Figura 4b). Muestra un comportamiento parabólico cerca del origen.

Exponencial Potencia

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-|\phi t|^2)) & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases} \dots (22)$$

Podemos notar que la familia exponencial y gausiana son casos especiales de esta familia.

Cuadrático racional

$$\gamma(t) = \begin{cases} \tau^2 + \frac{\sigma^2 t^2}{(1+\phi t)} & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases} \dots (23)$$

Onda

Representa una función periódica

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t}\right) & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases} \dots (24)$$

La función de covarianza es: $C(t) = \sigma^2 \sin(\phi t) / \phi t$. Una representación de un modelo Onda podemos ver en la Figura 4c.

Función Materna La función Materna está dado por:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu t \phi})^\nu}{2^{(\nu-1)} \Gamma(\nu)} K_\nu(2\sqrt{\nu t \phi}) \right] & \text{si } t > 0, \\ \tau^2 & \text{otro caso} \end{cases} \dots (25)$$

Aquí ν ($\nu > 0$) es un parámetro que controla la suavización, mientras que ϕ es el parámetro de escala espacial. La función $\Gamma(\bullet)$ es la función gamma usual, mientras que K_ν es la función Bessel modificada de orden ν . Se puede decir que esta familia da origen a la función exponencial ($\nu = 1/2$) y para la función Gaussiana ($\nu \rightarrow \infty$). Cuando $\nu = 3/2$ obtenemos la forma cerrada de la función Materna, la cual tiene la forma:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 [1 - (1 + \phi t) \exp(-\phi t)] & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases} \dots (26)$$

Hacemos un resumen de estas familias e incluimos algunas otras. En dicho resumen se encuentra la función del variograma y la función de covarianza.

Tabla 1. Resumen de las funciones de covarianza de modelos isotrópicos

Modelo	Función de Covarianza $C(t)$
Lineal	$C(t)$ no existe
Esférico	$C(t) = \begin{cases} 0 & \text{si } t \geq 1/\phi, \\ \sigma^2 \left(1 - \left\{\frac{3}{2}(\phi t) + \frac{1}{2}(\phi t)^3\right\}\right) & \text{si } 0 < t \leq 1/\phi, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$
Exponencial	$C(t) = \begin{cases} \sigma^2 (\exp(-\phi t)) & \text{si } t \geq 0, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$
Gausiano	$C(t) = \begin{cases} \sigma^2 (\exp(-\phi^2 t^2)) & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$
Exponencial Potencia	$C(t) = \begin{cases} \sigma^2 (\exp(- \phi t ^2)) & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$
Cuadrático Racional	$C(t) = \begin{cases} \sigma^2 \left(1 - \frac{t^2}{(1+\phi t^2)}\right) & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$
Onda	$C(t) = \begin{cases} \sigma^2 \left(\frac{\sin(\phi t)}{\phi t}\right) & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$
Materna	$C(t) = \begin{cases} \frac{\sigma^2}{2^{(\nu-1)}\Gamma(\nu)} (2\sqrt{\nu t\phi})^\nu K_\nu(2\sqrt{\nu t\phi}) & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{otro caso} \end{cases}$

Tabla 2. Resumen de las funciones de covarianza de modelos isotrópicos

Modelo	Variograma $\gamma(t)$
Lineal	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{si } t > 0, \tau^2 > 0, \sigma^2 > 0 \\ 0 & \text{otro caso} \end{cases}$
Esférico	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{si } t \geq 1/\phi, \\ \tau^2 + \sigma^2 \left\{ \frac{3}{2}(\phi t) - \frac{1}{2}(\phi t)^3 \right\} & \text{si } 0 < t \leq 1/\phi, \\ 0 & \text{otro caso} \end{cases}$
Exponencial	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi t)) & \text{si } t \geq 0, \\ 0 & \text{otro caso} \end{cases}$
Gausiano	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi^2 t^2)) & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases}$
Exponencial Potencia	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(- \phi t ^2)) & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases}$
Cuadrático Racional	$\gamma(t) = \begin{cases} \tau^2 + \frac{\sigma^2 t^2}{(1+\phi t)} & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases}$
Onda	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t}\right) & \text{si } t > 0, \\ 0 & \text{otro caso} \end{cases}$
Materna	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu t \phi})^\nu}{2^{(\nu-1)} \Gamma(\nu)} K_\nu(2\sqrt{\nu t \phi}) \right] & \text{si } t > 0, \\ \tau^2 & \text{otro caso} \end{cases}$

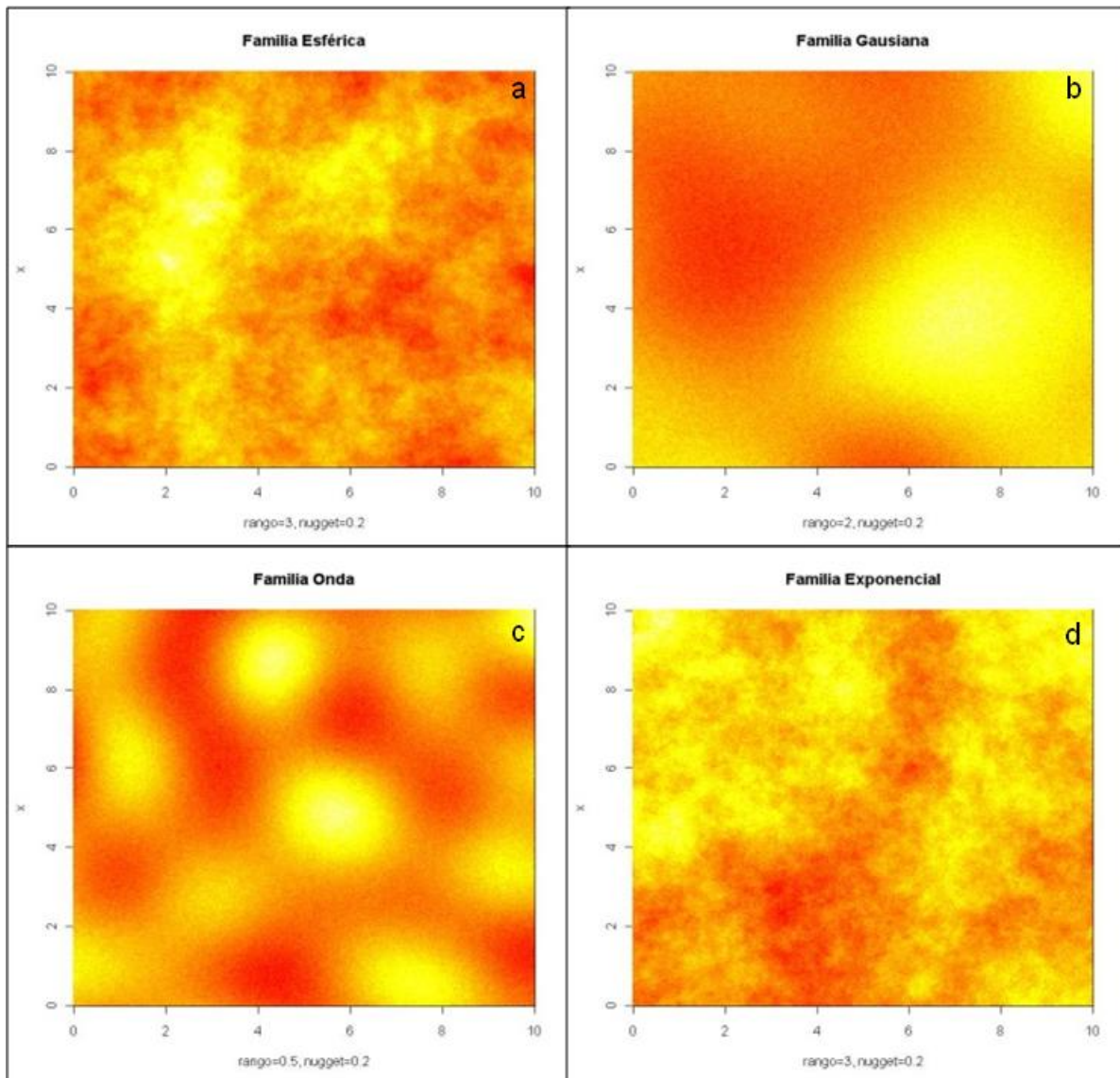


Figura 4. Representación grafica de simulación de los procesos espaciales según diferentes familias de modelos: a) familia esférica b), familia gausiana, c) familia onda, d) familia exponencial. Las simulaciones de estos procesos fueron realizados en R mediante la librería RandomFields.

2.5 Anisotropía

Cuando la autocorrelación de un proceso varía según la dirección, se dice que es un proceso anisotrópico. Hablamos de anisotropía geométrica cuando los variogramas calculados en diferentes direcciones tienen rangos diferentes pero similar meseta. Este tipo de anisotropía se

corrige mediante una rotación de las coordenadas. Para visualizar esto utilizaremos la Figura 5 y como ejemplo un variograma esférico.

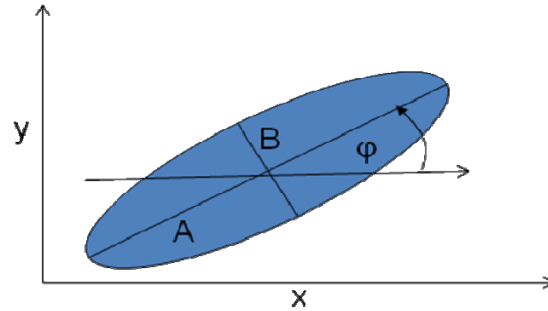


Figura 5. Representación geométrica de la anisotropía

Donde A es el máximo diámetro de la elipse (rango mayor) y B es el mínimo diámetro (rango menor) A y B son perpendiculares; φ es el ángulo en dirección al rango mayor.

La Ecuación usada para la transformación es:

$$\Omega(\mathcal{G}) = \{A^2 \cos^2(\mathcal{G} - \varphi) + B^2 \text{sen}^2(\mathcal{G} - \varphi)\}^{1/2} \dots (27)$$

En esta ecuación Ω define la anisotropía, y \mathcal{G} es la dirección de las distancias.

Insertando la ecuación 30 en el modelo esférico tenemos:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{si } t \geq 1/\Omega(\mathcal{G}), \\ \tau^2 + \sigma^2 \left\{ \frac{3}{2}(\Omega(\mathcal{G}) * t) - \frac{1}{2}(\Omega(\mathcal{G}) * t)^3 \right\} & \text{si } 0 < t < 1/\Omega(\mathcal{G}), \\ 0 & \text{Otro caso} \end{cases} \dots (28)$$

Sea $R = A/B$ el ratio de anisotropía, reescribimos la ecuación 28 de la siguiente manera.

$$\Omega(\mathcal{G}) = \{A^2 \cos^2(\mathcal{G} - \varphi) + \left(\frac{A}{R}\right)^2 \text{sen}^2(\mathcal{G} - \varphi)\}^{1/2} \dots (29)$$

Esta transformación, en efecto, nos permite representar toda la variación en una forma isotrópica, transformando la elipse en una circunferencia.

2.6 Variograma robusto

La estimación del variograma es una etapa crucial de la predicción espacial, ya que determina los pesos a utilizarse en la interpolación espacial por el método de Kriging. Es por eso que es importante contar con un estimador de variograma que no esté afectado por la presencia de valores extremos en la serie de datos. De lo contrario puede producir mapas interpolados erróneos.

El variograma propuesto por Matheron, basado en el método de los momentos (ver ecuación 13) es inadecuado ante la presencia de observaciones extremas (Genton, 1998). Cressie y Hawkins (1980) proponen un estimador robusto del variograma.

$$2 \hat{\gamma}_{CH}^{\wedge}(h) = \frac{\left\{ \frac{1}{m(h)} \sum_{i=1}^{m(h)} |Y(x_i) - Y(x_j)|^{1/2} \right\}^4}{0.475 + \frac{0.494}{m(h)} + \frac{0.045}{m^2(h)}} \dots (30)$$

Este variograma robusto reduce el efecto de los datos extremos. El denominador de la ecuación 30 es una corrección basada en el supuesto de que los procesos que se estiman tienen diferencias normalmente distribuidas sobre todas las clases de distancias.

Para reducir el efecto de valores extremos en la data, nosotros hemos utilizado este variograma robusto.

2.7 Identificación de las escalas relevantes mediante variogramas anidados

Se calcula un variograma. Para cada estructura presente en el variograma es ajustado un modelo teórico. Cada rango del modelo teórico ajustado al variograma representa una estructura espacial (por ejemplo el tamaño de un 'cluster' de peces - un cluster es un grupo de cardúmenes de peces).

Un proceso aleatorio, $Y(x)$, puede ser tratado como una combinación de múltiples procesos independientes, uno anidado con otro y representando diferentes características espaciales. De

igual modo el variograma del proceso $Y(x)$ es también una combinación anidada de dos o más variogramas.

$$\gamma(h) = \gamma^1(h) + \gamma^2(h) + \gamma^3(h) + \dots + \gamma^s(h) \dots (31)$$

Donde los números se refieren a variogramas y no a una potencia.

Cada variograma anidado tiene diferentes parámetros (meseta, rango, etc.).

Bajo el supuesto de que los procesos son independientes, podemos definir la ecuación 31 como:

$$\gamma(h) = \sum_{k=1}^s b^k g^k(h) \dots (32)$$

Donde $g^k(h)$ es la función básica del variograma y b^k es un coeficiente que cuantifica la contribución relativa de la varianza de $g^k(h)$ a la suma en (32).

2.8 Modelando la covarianza espacial entre la variable respuesta y las variables predictoras: Variogramas cruzados

Los variogramas cruzados permiten estudiar el comportamiento espacial entre dos variables ($Y_u(x)$ y $Y_v(x)$).

Para formalizar la idea, definimos dos variables regionalizadas, $Y_u(x)$ y $Y_v(x)$. Si extendemos la función del variograma al estudio de dos variables $Y_u(x)$ y $Y_v(x)$ tenemos como resultante el variograma cruzado. La forma de la ecuación es:

$$\gamma_{uv}(h) = \frac{1}{2} E[\{Y_u(x) - Y_u(x+h)\}\{Y_v(x) - Y_v(x+h)\}] \dots (33)$$

Si ambas variables son estacionarias de segundo orden con medias μ_u y μ_v , entonces la función de covarianza de ambas es:

$$C_{uv}(h) = E[\{Y_u(x) - \mu_u\}\{Y_v(x+h) - \mu_v\}] \dots (34)$$

La función $C_{uv}(h)$ es asimétrica, lo cual quiere decir que la covarianza cruzada en una dirección a aquella evaluada en una dirección opuesta: $C_{uv}(h) \neq C_{vu}(h)$

Finalmente la función de correlación cruzada es dado por:

$$\rho_{uv}(h) = \frac{C_{uv}(h)}{\sqrt{C_{uu}(0)C_{vv}(0)}} \dots (35)$$

Otra forma de expresar la relación espacial entre dos variables es mediante el coeficiente de codispersión. Para una distancia h tenemos:

$$v_{uv}(h) = \frac{\gamma_{uv}(h)}{\sqrt{\gamma_{uu}(h)\gamma_{vv}(h)}} \dots (36)$$

3 Análisis de Coordenadas Principales de Matrices Vecinas (PCNM)

Recientemente, un nuevo enfoque llamado de Análisis de Coordenadas Principales de Matrices Vecinas (PCNM) ha sido propuesto como una alternativa a los análisis de las tendencias de superficie (Borcard y Legendre, 2002). PCNM permite descomponer la estructura espacial (a partir de sus coordenadas) de las variables ecológicas a distintas escalas espaciales. Estos modelos permiten obtener un conjunto de variables espaciales, las cuales pueden ser usadas en análisis de regresión múltiple como variables respuesta. Este método ya ha sido utilizado con éxito en varias aplicaciones ecológicas (Borcard et al., 2004; Brind'Amour et al., 2005; Legendre et al., 2005).

Las variables espaciales son obtenidas mediante escalamiento multidimensional clásico, también llamado Análisis de Coordenadas Principales (PCOA; Gower, 1966), de una matriz truncada de pares de distancia geográfica entre los sitios muestreados. Teniendo como resultados un conjunto de valores y vectores propios. Los vectores propios asociados a los

valores propios positivos son utilizados como variables respuesta en el análisis de regresión lineal multivariada. Los vectores propios asociados a los valores propios positivos corresponden a la representación euclidiana de la matriz de distancia truncada D . Los vectores propios son ortogonales entre sí, motivo por el cual no están correlacionados.

Los pasos a seguir para aplicar el método de PCNM son:

- 1) Calcular la matriz de distancias euclidianas D a partir de las coordenadas geográficas de los datos.
- 2) Truncar la matriz D la que se llamará D^*
- 3) Construir una matriz $Q = -1/2 P D^* P$ y mediante escalamiento multidimensional (o análisis de coordenadas principales) obtener los valores propios (λ_i) de Q y también los vectores propios v_i generados por los valores propios.
- 4) Realizar regresiones lineales múltiples entre la variable de interés (Y) y los vectores propios generados por los valores propios positivos. Los vectores propios que resultan ser significativos en este paso son llamados las PANM variables.

Podemos observar los pasos del método PCNM en la Figura 6.

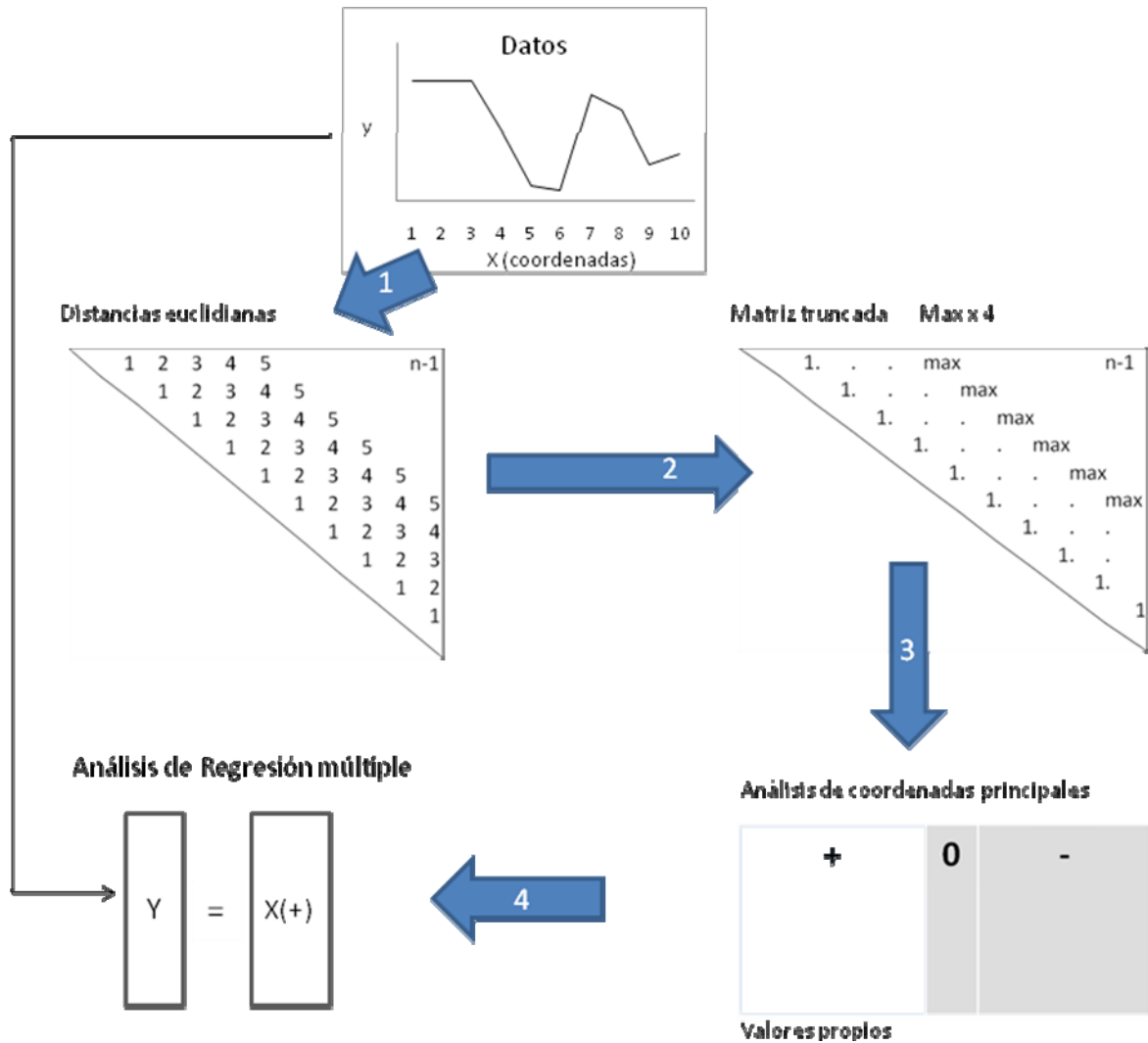


Figura 6. Esquema gráfico de los PCNM: (1) Cálculo de la matriz de distancias euclidianas a partir de las coordenadas geográficas (latitud y longitud); (2) Truncamiento de la matriz de distancias, (3) Análisis de las coordenadas principales a la matriz truncada, y (4) Análisis de regresión lineal múltiple entre la variable respuesta (Y) y los vectores propios (asociados a valores propios significativos). Los vectores propios que resultan ser significativos en la regresión son llamadas las PCNM variables.

Desarrollaremos más detalladamente cada paso del método PCNM.

3.1 Cálculo de la matriz de distancias

Recordemos algunas definiciones del álgebra de matrices. Una matriz D (matriz cuadrada) es llamada matriz de distancias si y solo si es simétrica y $d_{ii} = 0$ y $d_{ij} \geq 0, i \neq j$. Si para algún p , existen puntos $y_1, y_2, \dots, y_n \in \mathfrak{R}^p$ y que:

$$d_{ij}^2 = (y_i - y_j)'(y_i - y_j) \dots (37)$$

donde d_{ij} es la distancia euclidiana.

Para nuestro caso, nosotros tenemos un conjunto de datos que se encuentran distribuidos espacialmente. De este conjunto de datos calculamos las distancias euclidianas entre sus coordenadas correspondientes (longitud y latitud), llegando a tener una matriz de distancias. Es así que esta matriz tiene como observaciones u elementos a:

$$D = \begin{cases} 0 & \text{si } i = j \\ d_{ij} & \text{si } i \neq j \end{cases} \dots (38)$$

donde d_{ij} es la distancia euclidiana desde el punto i al punto j .

3.2 Modificando la Matriz de Distancias

Definimos un umbral para las distancias con el fin de quedarnos sólo con las distancias entre sitios (puntos) vecinos más cercanos. Solo las distancias menores al umbral son retenidas en una matriz (matriz modificada D^*). Aquellas distancias que resultan mayores al umbral serán consideradas como muy grandes (Borcard y Legendre, 2002) y serán reemplazadas por un valor arbitrario. Borcard y Legendre (2002) plantean que este valor arbitrario es 4 veces el umbral.

Los reemplazos se realizan de acuerdo a la siguiente regla.

Sea D la matriz de distancias euclidianas, la cual tiene como elementos a d_{ij} y t el umbral. La matriz modificada D^* es:

$$D^* = \begin{cases} d_{ij} & \text{si } d_{ij} \leq t \\ 4 * t & \text{si } d_{ij} > t \end{cases} \dots (39)$$

Entonces se procede al cálculo de las coordenadas principales de la matriz modificada.

3.3 Análisis de Coordenadas Principales

Recordemos previamente que: “Sea una matriz cuadrada A , se dice que esta es diagonalizable si existe un P tal que $A = PMP^{-1}$ donde P , llamada matriz de paso, es invertible y M es una matriz diagonal”.

Entonces a partir de la matriz de distancias modificadas D^* , definimos una nueva matriz llamada D_2^* , donde:

$$\begin{aligned} D_2^* &= [(d_{ij}^*)^2] = \sum_{a=1}^m (y_{ia} - y_{ja})^2 = \sum_{a=1}^m (y_{ia}^2 + y_{ja}^2 - 2y_{ia}y_{ja}) \\ &= C1' + 1C' - 2 \sum_{a=1}^m y_a y_a' = C1' + 1C' - 2YY' \dots (40) \end{aligned}$$

Aquí C es un vector que contiene los elementos $\sum_{a=1}^m y_{ia}^2$ y la matriz $B = YY'$ es llamada la matriz producto escalar.

Multiplicando en la ecuación (40) por $J = I - \frac{11'}{n}$ (con el objetivo de conseguir una matriz

centrada) en ambos lados de la misma y por el factor $-1/2$ nos queda:

$$\frac{-1}{2} JD_2^* J = \frac{-1}{2} J(C1' + 1C' - 2YY') J$$

$$\frac{-1}{2} JD_2^* J = YY' = B \dots (41)$$

Haciendo una analogía de la ecuación (40) con la definición de una matriz diagonalizable, podemos decir que B es una matriz diagonalizable.

Pasemos ahora al problema de obtener las coordenadas principales. Dada la matriz B (Ecuación 41) la vamos a expresar de la siguiente manera:

$$B = YY' = V\Lambda V' = (V\Lambda^{1/2})(V\Lambda^{1/2})' \dots (42)$$

Entonces de (42) podemos notar que $Y = V\Lambda^{1/2}$, donde $\Lambda^{1/2}$ es la matriz diagonal con elementos $\lambda_i^{1/2}$ (donde los λ_i son los valores propios de la matriz B). Las filas de Y contienen las coordenadas principales de las observaciones y las columnas de Y contienen los vectores propios v_i correspondientes a cada valor propio λ_i .

Los v_i asociados a los mayores valores propios representan la variación a gran escala; mientras que los vectores propios generados por los menores valores propios representan la variación a menor escala.

La matriz de distancias euclidianas ha sido truncada lo que puede generar valores propios negativos. Sin embargo, los valores propios negativos no se pueden representar en un espacio euclidiano (Bellier et al., 2007), por tal motivo solamente los valores propios positivos son utilizados como descriptores espaciales.

3.4 Caracterización de las escalas espaciales mediante los PCNM variables

Una prueba, para la significancia de los v_i viene explicado por Dray et al. (2006) quienes demuestran la asociación de los PCNM con los índices de Moran. Dray et al. (2006) proponen estimar la significancia de la autocorrelación espacial para cada v_i (asociado a valores propios positivos) mediante test de permutación (ver nexos 5). Aquellos v_i que resultan ser significativos

son utilizados como variables explicativas en regresiones múltiples. La regresión múltiple planteada entre los v_i significativos y la variable respuesta (Y) nos indicará que vectores son relevantes para el modelo de regresión planteado. Estos vectores relevantes son llamados PCNM variables.

Una deficiencia de los PCNM ocurre al momento de definir el tamaño de las estructuras espaciales, pues por el momento no existen métodos desarrollados que nos permitan agrupar a las PCNM variables en submodelos. Realizándose la elección de los submodelos de manera arbitraria. Para la elección de los submodelos, Bellier et al. (2007) propone calcular un variograma para cada PCNM variable. Luego ajustar un modelo teórico e identificar los rangos en cada variograma. Entonces para poder decidir el número de submodelos hay que apoyarse visualmente en la figura generada entre los rangos encontrados (para cada PCNM variable) vs los valores propios (que generaron los PCNM variables). Cada submodelo elegido estará constituido por un conjunto de las PCNM variables (de acuerdo a su rango). Estos submodelos nos están indicando las estructuras presentes en la variable Y.

3.5 Relación entre los submodelos y otras variables

Con el objetivo de conocer la importancia y la relación de covariables (pueden ser variables medioambientales u otras) con la variable en estudio (Y), nosotros podemos aplicar regresiones lineales múltiples para cada submodelo. Para tal motivo se aplica una regresión lineal teniendo como variable respuesta a Y y como variables explicativas a las covariables conjuntamente con las PCNM variables pertenecientes a cada submodelo. Los resultados de las regresiones nos indicarán si las covariables son relevantes para explicar la variabilidad de la variable Y. También, el coeficiente obtenido en la regresión para cada covariable nos indicará la relación de estas covariables con la variable en estudio Y.

4 Análisis Wavelets

Frecuentemente las relaciones entre abundancia de poblaciones y procesos medioambientales se basan en la correlación estadística o el análisis espectral (Ménard et al., 2007). Ambos métodos asumen estacionariedad en las variables. Al contrario, los métodos wavelets no requieren el supuesto de estacionariedad de las variables (Torrence y Campo, 1998; Dale y Mah, 1998) lo que es útil para muchos métodos estadísticos pues es un supuesto de estos métodos. Los métodos wavelets tienen una gran aplicación en diversos campos de las ciencias tales como detección de señales sísmicas, procesamiento de imágenes, mecánica cuántica, investigación médica, etc. Callezes et al. (2005) utilizaron los wavelets para estudiar la asociación entre el fenómeno del Niño y la propagación del dengue en Tailandia. Ménard et al. (2007) aplicó los wavelets para cuantificar la asociación entre el clima y las poblaciones de atunes en el océano Índico. Estos dos últimos estudios son algunas de las aplicaciones de los métodos wavelets en ecología.

Los wavelets permiten descomponer una señal en sus diferentes componentes de frecuencia y proveen una evaluación local y global de la estructura espacial (Fortín y Dale, 2005). Desarrollaremos una breve descripción del análisis.

4.1 La transformada de wavelets

El análisis wavelets consiste en un estudio o un análisis de la señal en tiempo y frecuencia de un parámetro, proporcionando información de las diferentes frecuencias de la señal de la serie. La función wavelets $\psi(t)$ es expresada mediante dos parámetros τ y a que vienen a ser la posición y la escala, respectivamente.

La Wavelet $\psi_{a,\tau}(t)$ es definida como:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \dots (43)$$

Se puede obtener una extracción de la amplitud y la fase de las series utilizando la familia Morlet¹ (Menard et al., 2007). Esta familia genera información continua en el conjunto de los números imaginarios:

$$\psi_{a,\tau}(t) = \pi^{-1/4} e^{(-i2\pi f_0 t)} e^{(-t^2/2)} \dots (44)$$

donde f_0 es la frecuencia central de la función wavelets

Una transformación wavelets de una serie $x_{(t)}$ es definida como una convergencia de funciones wavelets $\psi_{a,\tau}(t)$ y viene expresada por:

$$W_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-\tau}{a} \right) dt = \int_{-\infty}^{\infty} x(t) \psi_{a,\tau}^*(t) dt \dots (45)$$

donde el * denota la conjugada compleja. Acá $W_x(a, \tau)$ denota la contribución de las escalas de la señal a diferentes posiciones de tiempo.

4.2 Espectro de la Potencia de Wavelets

En la ecuación 44, cuando $f_0 = 6$ se puede aproximar la frecuencia mediante la ecuación

$$f = 1/a.$$

Reemplazando f por a calculamos el espectro de wavelets como:

$$WPS_x(f, \tau) = |W_x(f, \tau)|^2 \dots (46)$$

Mediante el WPS obtenemos una evaluación local de la serie de datos (Figura 7b)

En la Figura 7b, la línea negra que se observa es el cono de influencia (COI). El COI es la zona del gráfico por sobre la cual los picos de mayor intensidad no son afectados por los efectos del borde,

¹ La wavelet de Morlet es una función continua equivalente a una onda plana modulada por una gaussiana

mientras que por debajo de esta línea los resultados son fuertemente influenciados por los valores extremos de la muestra (Torrence y Compo, 1997).

También observamos en la Figura 7b la presencia de zonas encerradas por líneas negras. Estas zonas son estadísticamente significativas a un nivel del 95% respecto a un espectro de ruido blanco (Torrence y Compo, 1997).

4.3 Espectro Global de wavelets

Mediante el espectro de wavelet global se puede estudiar el comportamiento global de la serie de datos. Este espectro global es calculado como el tiempo promedio de los WPS para cada componente de la frecuencia (Figura 7c):

$$\overline{WPS}_x(f) = \frac{\sigma^2}{T} \int_0^T |W_x(f, \tau)|^2 dt \dots (47)$$

donde σ^2 es la varianza y T es la duración de la serie de tiempo $x_{(t)}$

4.4 Análisis wavelets cruzado (cross wavelets)

La transformada de wavelets puede ser extendido para cuantificar la relación estadística de dos series $x_{(t)}$ y $y_{(t)}$, esto se realiza mediante el cross wavelets (WCS) el cual viene dado por:

$$WPS_{x,y}(f, \tau) = WPS_x(f, \tau) \cdot WPS_y^*(f, \tau) \dots (48)$$

Acá WCS provee información local de la covarianza de ambas series de tiempo a frecuencias particulares.

Finalmente para comprender si estas dos series se encuentran en fase o en desfase se plantea la siguiente ecuación:

$$\phi_{x,y}(f, \tau) = \tan^{-1} \frac{\Im(WCS_{x,y}(f, \tau))}{\Re(WCS_{x,y}(f, \tau))} \dots (49)$$

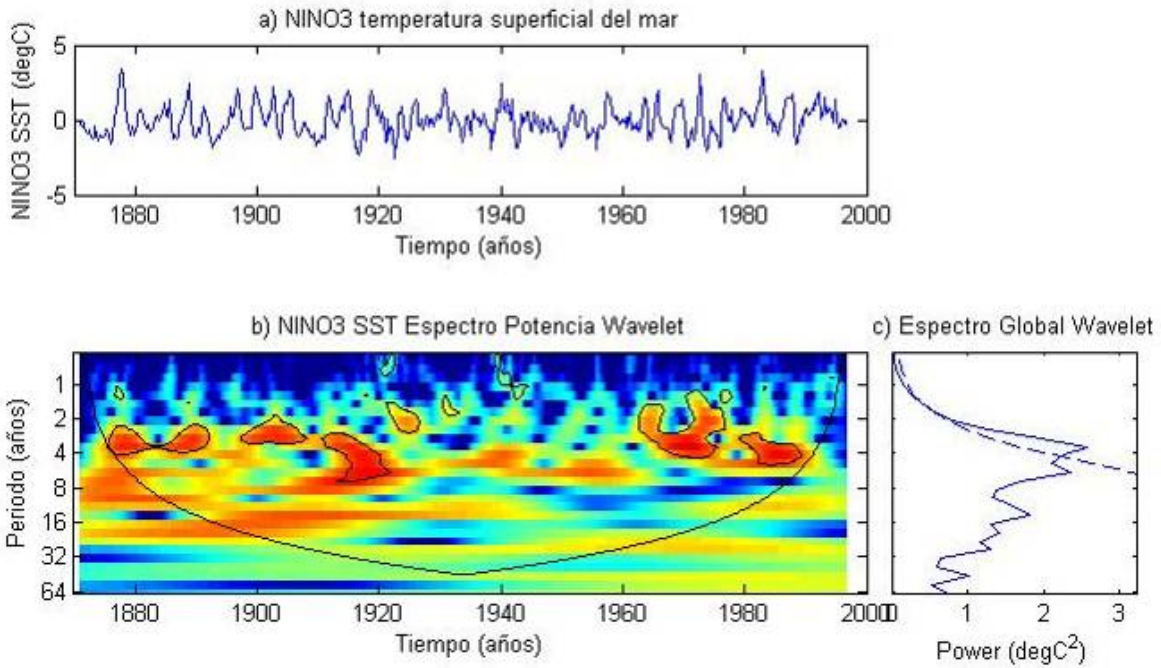


Figura 7. a) serie de tiempo de la temperatura superficial del mar usado para el análisis wavelets. b) el espectro de la potencia de wavelets usando el wavelets Morlet. Este espectro proporciona una evaluación local de la serie. c) Espectro global de wavelets de la serie, proporciona una evaluación global de la serie. Fuente: <http://paos.colorado.edu/research/wavelets/>

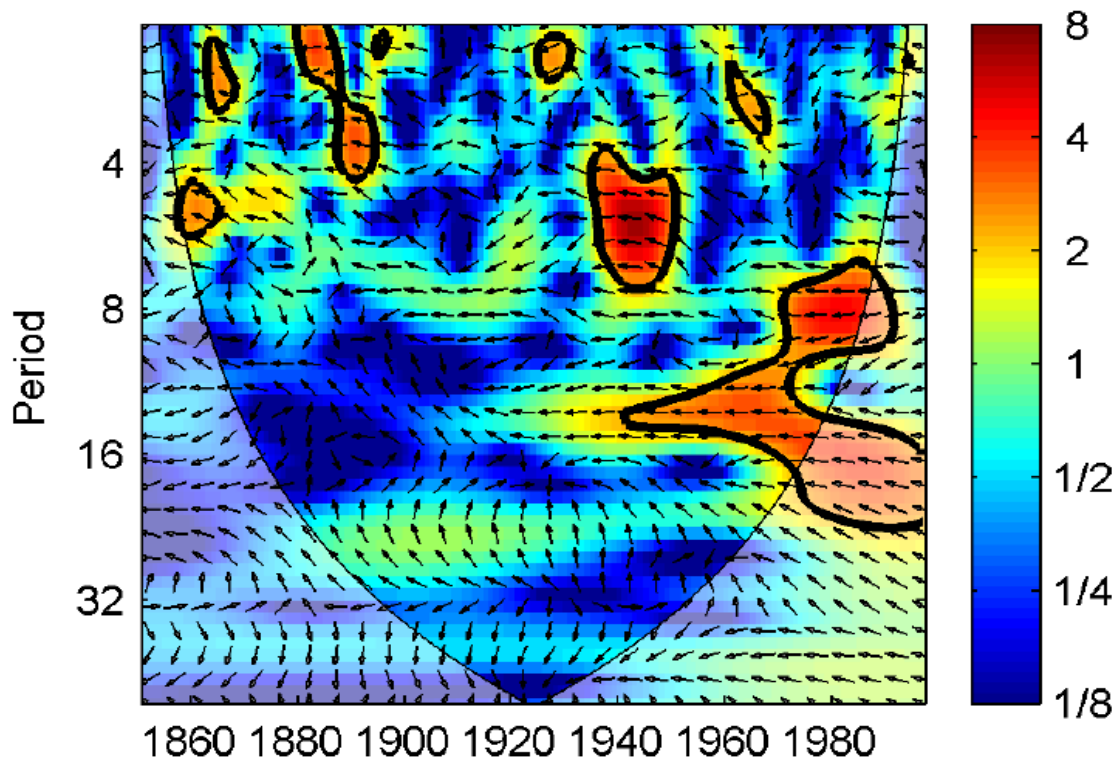


Figura 8. Análisis de wavelets cruzado de dos series temporales. El análisis de fase se presenta mediante flechas, flechas en dirección a la derecha indican series en fase y flechas en dirección a la izquierda indican series en desfase. Fuente: Grinsted et al. (2004).

En la Figura 8 podemos ver una aplicación del wavelet cruzado a dos series. La dirección de las flechas indica si estas dos series se encuentran en fase (flechas hacia la derecha) o en desfase (flechas hacia la izquierda). También podemos notar el COI y las regiones de confianza.

CAPITULO III:
METODOLOGIA

1 Tipo de estudio

La actual investigación responde a un análisis exploratorio para lo cual no se tienen hipótesis previas. El estudio es de tipo observacional y de corte transversal.

2 Definición de la población de estudio y pre-tratamiento de los datos

2.1 Área de estudio

El área de estudio se encuentra dentro del Sistema de la Corriente de Humboldt (SCH) el cual abarca las costas del Perú y de Chile. La recolección de los datos fue realizado por el Instituto del Mar del Perú (IMARPE), que mediante sus embarcaciones de investigación monitorea las condiciones bióticas y abióticas del ecosistema. IMARPE cuenta con tres barcos de investigación científica (BIC): Olaya (41 m de largo), SNP2 (21 m de largo) y Humboldt (76 m de largo).

Los datos utilizados en este estudio provienen de un crucero acústico que se llevó a cabo entre los días 20 de febrero y 04 de abril del 2005 a bordo del BIC Olaya. El área estudiada cubrió todo el litoral Peruano (entre los 3°29' S hasta los 18°3' S) hasta 100 millas náuticas (mn) de las costas (Figura 9).

El BIC Olaya rastreó 92 transectos, realizando 128 lances de pesca y recorriendo un total de 4 916 mn de rastreo acústico (Tabla 3).

Tabla 3. Datos de crucero acústico 2005

<i>Crucero</i>	Pelágicos 2005
<i>Fecha inicial</i>	20-02-2005
<i>Fecha final</i>	04-04-2005
<i>Área cubierta</i>	3°29' S -18° 03' S
<i>Horas muestreadas</i>	600
<i>Sistema de la ecosonda</i>	Simrad EK500 split-beam
<i>Frecuencias</i>	38 and 120 kHz

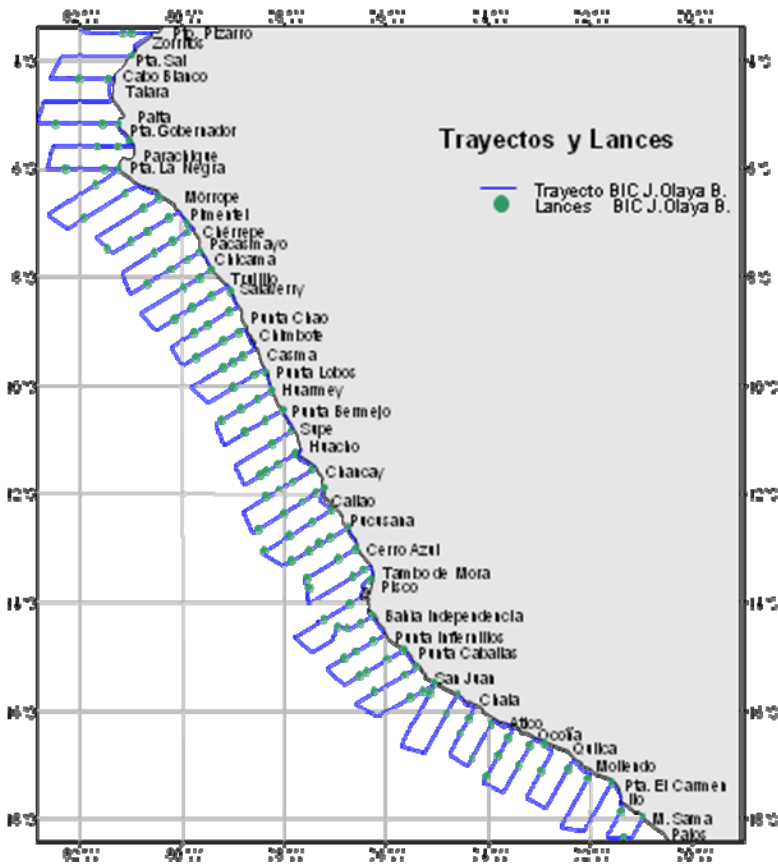


Figura 9. Crucero de evaluación acústica realizado por el IMARPE en febrero-abril 2005. La línea azul indica la trayectoria del BIC Olaya. Los puntos verdes indican las ubicaciones de los lances de pesca del BIC Olaya.

2.2 Datos Acústicos

Los datos acústicos fueron recolectados empleando una ecosonda digital SIMRAD EK-500 operando a 120 y 38 kHz. La selección y clasificación de los datos fueron llevadas a cabo por Michael Ballón (estudiante de doctorado del IRD) quien utilizó los programas de Echoview (SonarData Pty. Ltd., Hobart, Tasmania, Australia) y Matlab.

2.3 Pre-procesamiento

Para poder comparar los ecogramas provenientes de dos frecuencias distintas, se sincronizaron los números de emisiones acústicas o 'pings' (número y posición) entre los ecogramas (para esta tarea se utilizó Echoview). También se excluyó el ruido de la superficie (5 m de profundidad) y el

eco del fondo marino. Similarmente se limpiaron los ecogramas excluyendo las regiones que contenían ruido. Y se limpió el ruido generado por la absorción acústica¹.

Para este pre procesamiento se utilizaron los programas Ecoview y Matlab.

2.4 Discriminación de los grupos

Los peces tienen respuestas más semejantes y fuertes de energía de retrodispersión² a 38 y 120 kHz en comparación con el zooplancton (Simmonds y MacLennan, 2005). Utilizando esta propiedad se han generado ecogramas de 'peces' (fish) y 'no peces' en ambas frecuencias (38 y 120 kHz). Dado que los peces tienen una respuesta ligeramente superior a 38 que a 120, se procedió a refinar los ecogramas de peces (fish) manteniendo solo aquellos datos que cumplen con $120 \text{ Sv} - 38 \text{ Sv} < 2 \text{ dB}$ ³. Donde Sv es el volumen de la fuerza de retrodispersión.

La respuesta de los organismos gelatinosos son mayores en 120 kHz que en 38 kHz. Esta diferencia de respuesta en la frecuencia permite extraer los organismos gelatinosos ('blue noise') de la data acústica.

Se extraen los organismos gelatinosos del ecograma de "no peces" mediante la aplicación de una máscara booleana para seleccionar los datos con mayor retrodispersión a 120 que a 38 kHz. Como resultado, el ecograma de "no peces" se dividió en "ecogramas fluid-like" (lo que corresponde al zooplancton) y "ecogramas blue-noise" (Figura 10).

Los ecogramas obtenidos ('fish', 'blue noise' y 'fluid like') fueron exportados en formato Matlab. Estos datos consistieron en el volumen (Sv) por elemento de celda de una emisión acústica (la frecuencia de emisión era de 1 por segundo con el barco navegando a 10 nudos) de largo y 0.75 metros de alto, junto con las coordenadas de posición (latitud y longitud), el tiempo y la profundidad del fondo.

¹ Parte de la energía acústica que ha sido absorbida por la superficie

² Energía que es devuelta al receptor

³ Decibel es la unidad relativa logarítmica empleada en acústica y telecomunicaciones para expresar la relación entre dos magnitudes, acústicas o eléctricas, o entre la magnitud que se estudia y una magnitud de referencia

Las munidas (crustáceos competidores de la anchoveta, ver Gutiérrez et al., 2008) (i) son más grandes que cualquier otro organismo gelatinoso (fluid-like), (ii) tienen mayor retrodispersión que otros organismos gelatinosos (fluid-like), (iii) son siempre epipelágicos⁴ y realizan pequeñas migraciones verticales en la capa epipelágica (Gutiérrez et al., 2008). Utilizando las características anteriores se extrajeron los datos de munida del ecograma fluid-like mediante la aplicación de los siguientes umbrales en el fluid-like Sv: $-72 \text{ dB} < \text{Sv} < -53 \text{ dB}$. Los límites de los umbrales fueron determinados empíricamente de ecogramas donde la presencia o ausencia de munida fue confirmada por las redes de arrastre pelágicos.

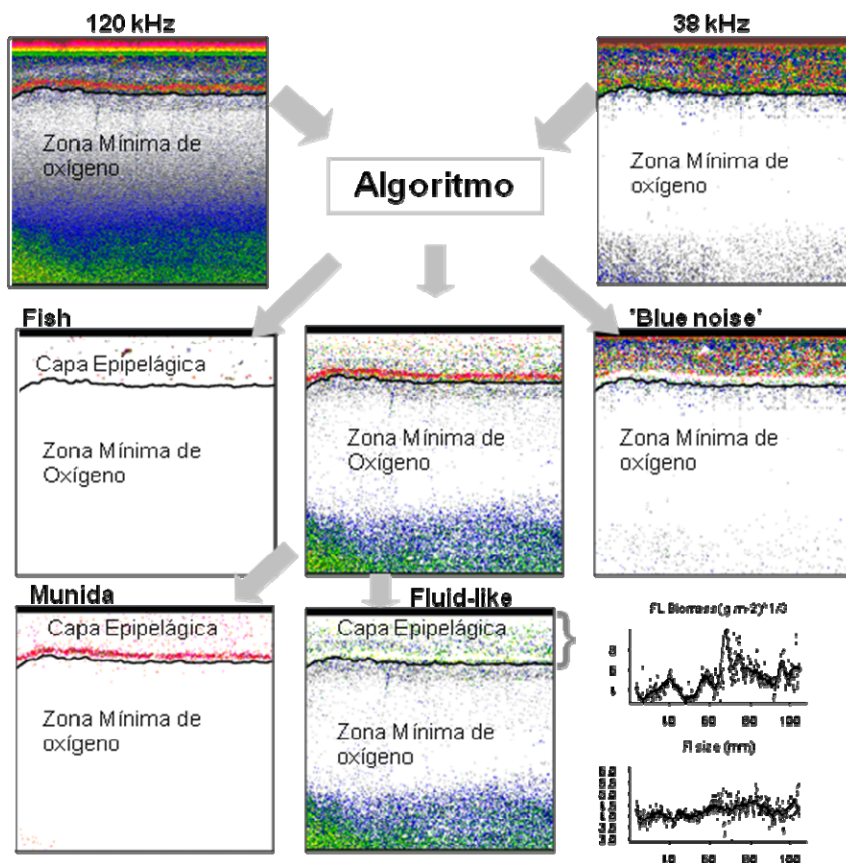


Figura 10. Esquema gráfico del proceso de separación de las bases de fish, blue noise, munida y fluid like.

⁴ Zona donde hay suficiente luz para realizar la fotosíntesis

Para reducir el tiempo de procesamiento los datos fueron transformados de Sv (Unidad de muestreo de volumen) a s_v (unidades) y horizontalmente muestreados mediante la estimación de la s_v mediana para cada 3 pings. Así, 3 pings era el tamaño de las UBM (unidad básica de muestreo). La resolución vertical del ecograma no se ha modificado en este proceso.

Finalmente hemos eliminado los datos que pertenecen al alba y al ocaso. Nos centramos únicamente en los períodos de día y de noche. Esto porque el comportamiento y la distribución del zooplancton son muy diferentes entre día y noche; los periodos crepusculares siendo periodos de transición entre día y noche y difícilmente interpretables (Fréon et al., 1993).

3 Descripción de las variables

3.1 Profundidad de la zona mínima de oxígeno

En el HCS la extensión vertical de la comunidad epipelágica de zooplancton es limitada por la presencia de la Zona Mínima de Oxígeno (ZMO) (Crales, et al., 2008; Ayón et al., 2008). Si la abundancia de la comunidad epipelágica es suficientemente grande para producir ecos acústicos, esta comunidad aparece en los ecogramas como una capa de dispersión superficial (Figura 11). Hemos definido el límite vertical de la comunidad epipelágica (VEEC) como la profundidad en que se alcanzó el 98% de la suma acumulada de eco acústico combinado (ecuación 50) de la comunidad epipelágica.

$$s_v \text{ combined echogram} = (s_v \text{ Fluid-like} \cdot 10^3 + s_v \text{ fish} + s_v \text{ blue noise} \cdot 10^3) / 3 \quad (50)$$

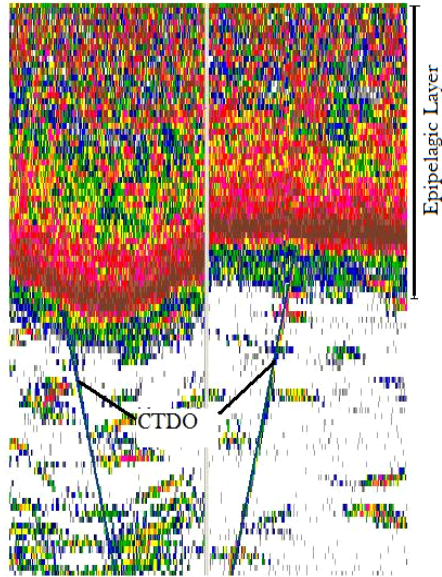


Figura 11. Ecograma acústico representando la distribución vertical de los organismos a lo largo de un transecto.

Este límite es una aproximación del límite superior de la zona mínima de oxígeno (Ballón et al., in prep; Bertrand et al., in prep).

Globalmente existe una tendencia de gran escala (cientos de km) costa mar afuera en la capa de la comunidad epipelágica (Bellier, 2007). Esta tendencia esconde parcialmente procesos a escala más pequeña que es el centro de interés de este trabajo. Entonces se procedió a quitar esta tendencia mediante regresiones robustas (anexo 1) aplicadas a cada transecto.

Finalmente en los análisis se utilizara los residuales de las regresiones (que son los valores sin tendencia). Estos residuales serán llamados como la variable física.

3.2 Estimación de biomásas

3.2.1 Biovolumen de fluid like

El biovolumen de fluid like y de munida se calcula multiplicando la densidad del organismo (ind.m^{-3}) por el volumen estimado de cada individuo "medio" (para mayor detalle de la estimación del biovolumen ver Mitson et al., 1996).

3.2.2 Abundancia de fish

Los ecogramas reducidos de fish fueron convertidos en coeficientes de millas náuticas (NASC), con un índice de abundancia de acuerdo a la siguiente fórmula.

$$\text{NASC (m}^2 \cdot \text{millas nauticas}^{-2}) = 4\pi \times 18522 \cdot sv \cdot T,$$

donde T fue la distancia vertical de integración de 0.75 m.

La biomasa integrada se calculó multiplicando los valores medios por la profundidad de la capa.

3.3 Base de datos

En síntesis para este trabajo contamos con datos de fish, fluid like, munida y blue noise cada emisión acústica (~ 5 m) a lo largo del recorrido del barco. Para reducir el tamaño de la base se calculó la mediana por cada 300 metros de distancia horizontal, esta reducción se realizó para atenuar el ruido de las distancias cortas (menores a 300 metros). Finalmente nuestra base de trabajo contó con 3490 observaciones para noche y 4300 observaciones para día.

Nuestra base de datos final estuvo compuesta por latitud, longitud, variable física, fluid like y fish. Para los datos de fluid like se utilizó una transformación log (fluid like) y para los datos de fish se utilizó la transformación log(fish +0.1). Estas transformaciones se realizaron para tener variables distribuidas simétricamente. Podemos observar una representación la base de datos en la tabla 4.

Tabla 4. Representación de la base de datos

Nº	Lat	Lon	Var. Física	Fluid Like	Fish
1	-6.53	-72.5	-5.35	x1	y1
2	-6.75	-72.7	-3.25	x2	y2
.
.
n				xn	yn

CAPITULO IV:
APLICACIÓN DE LOS MÉTODOS

1 Distribución del fluid like según periodo

La abundancia en zooplancton (fl) varió significativamente (test de Wilcoxon, $w=2820893$, $p < 2.2e^{-16}$) según el periodo nictimeral siendo más abundante en noche que en día (Figura 12).

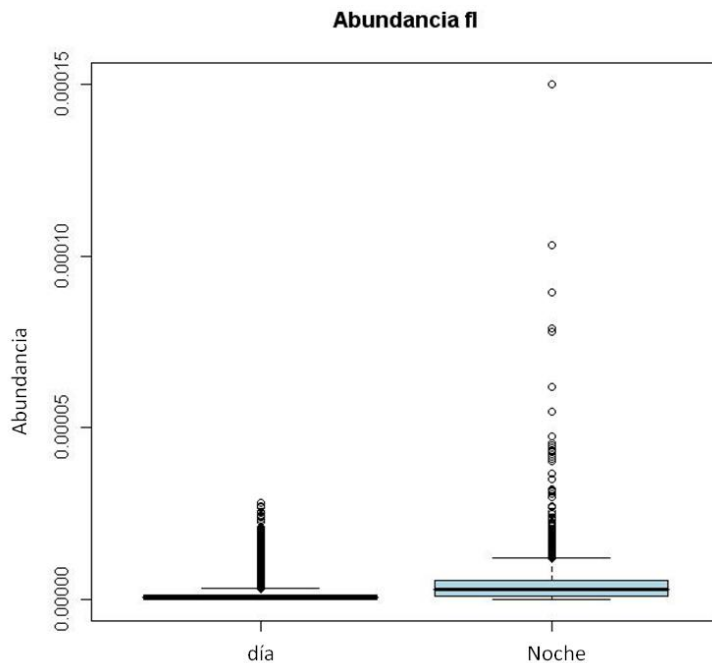


Figura 12. Distribución de la abundancia (g. m^{-3}) del fluid like según día y noche.

2 Aplicación del test de correlación con ajuste del tamaño de muestra

Se seleccionaron solamente aquellos transectos mayores a los 100 km. Para tener series igualmente espaciadas dentro de cada transecto se interpolaron los datos de la variable física y el log 'fluid-like'. La interpolación se realizó cada 300 metros.

Para estudiar la variabilidad a grandes distancias se tomó promedios cada 2 km en cada transecto. Luego se calculó la autocorrelación para cada serie de datos (variable física y log fluid like), de la autocorrelación presente en cada serie de datos se pudo obtener el tamaño de la estructura para cada serie. Finalmente se aplicó la correlación con ajuste del tamaño de muestra.

Para estudiar la variabilidad a pequeñas distancias se aplicó una suavización spline (anexo 2) para extraer la variación que ocurre a grandes distancias. Finalmente a las series sin efecto de las

grandes distancias (variable física y fluid like) se le aplicó el test de correlación con ajuste del tamaño de muestra.

Observando las ecuaciones (1) y (2) se realizó los análisis entre la **variable física y el log fluid like** (para datos de día y de noche) en cada transecto seleccionado. No se realizó el análisis de la correlación lineal con ajuste del tamaño de muestra para el log fish por la presencia de muchos ceros en esta serie. La presencia de ceros altera la estructura espacial de la variable.

2.1 Correlación con modificación del tamaño de muestra

Se seleccionaron 15 transectos (con distancias superiores a los 100 km) 7 de noche y 8 de día y se interpoló los datos cada 300 metros. Luego, para estudiar la relación a pequeñas distancias se calculó la correlación lineal entre la variable física y el log fluid like en cada transecto seleccionado (Tabla 7).

Un paso previo para aplicar el método de la corrección del tamaño de muestra efectivo era calcular la autocorrelación presente en cada variable dentro de cada transecto.

Luego de calcular la autocorrelación para la variable física y, también, para el log fluid like (datos de noche) se observó estructuras de diferente tamaño en cada transecto (Tabla 5). Para la variable física la mayor estructura observada fue en el transecto 32 con un tamaño de 24 km, mientras que para la variable fluid like la mayor estructura se observó en el transecto 54 con un tamaño de 30 km.

Tabla 5. Tamaño de las estructuras observadas los procesos de autocorrelación realizados para cada transecto datos de noche (en km).

Transecto	var. Física	log fluid like
16	6	21
26	18	27
32	24	24
48	6	9
54	6	30
78	18	6
82	15	6

De igual modo se calcularon las autocorrelaciones para la variable física y log fluid like dentro de cada transecto (datos de día). Se observó que en el transecto 18 se dio la mayor estructura para las dos variables, con un tamaño de 24 km para la variable física y de 54 km para el fluid like (Tabla 6).

Tabla 6. Tamaño de las estructuras observadas los procesos de autocorrelación realizados para cada transecto datos de día (en km).

Transecto	var. Física	log fluid like
18	24	54
24	6	15
46	15	33
58	15	4
64	21	21
72	21	6
80	12	9
88	7	9

Para los datos de noche se calcularon las correlaciones lineales entre la variable física y el log fluid like dentro de cada transecto (para datos de noche) (Tabla 7). Luego se procedió a calcular el factor de corrección con ajuste del tamaño de muestra. Se observó que de los 7 transectos seleccionados (para los datos de noche) se obtuvieron solamente 4 correlaciones lineales significativas ($p\text{-value} < 0.025$). El sentido de las correlaciones significativas varió según el transecto analizado (Tabla 7).

Tabla 7. Resultados de la correlación entre la variable física y el log fluid like (datos de noche) con ajuste del tamaño de muestra. Datos para series igualmente espaciadas cada 300 m.

Transecto	correlación	n	n'	pvalue
16	0.041902	415	35	6.06E-08
26	-0.192694	384	19	0.01261498
32	0.056055	441	12	0.08780894
48	0.114099	373	57	0.02893673
54	-0.220760	381	25	0.02257832
78	0.143089	330	40	1.49E-08
82	-0.106274	370	31	0.1616762

También se calculó las correlaciones entre la variable física y el log fluid like para los datos separados cada 2 km. Llegándose a observar luego del ajuste del tamaño de muestra que las correlaciones obtenidas para cada uno de los transectos no fueron significativas (Tabla 8).

Tabla 8. Resultados de la correlación entre la variable física y el log fluid like (datos de noche) con ajuste del tamaño de muestra. Datos para series igualmente espaciadas cada 2 km.

Transecto	correlación	n	n'	pvalue
16	-0.047906	62	25	0.40966014
26	-0.107700	56	17	0.34198254
32	0.161185	65	11	0.31759591
48	0.381241	54	26	0.02798541
54	-0.231891	56	21	0.15457783
78	-0.182866	48	23	0.1994469
82	-0.152607	54	24	0.23848855

Para los datos de día, se calculó la correlación entre la variable física y el log fluid like para cada transecto. Los resultados podemos observarlos en la Tabla 9. De las 8 correlaciones calculadas solamente 3 resultaron ser significativas (Tabla 9). Las correlaciones que resultaron ser significativas mostraron una asociación negativa entre la variable física y el log fluid like, esto ocurrió al estudiar la variabilidad a pequeñas distancias (datos separados cada 300 metros).

Tabla 9. Resultados de la correlación entre la variable física y el log fluid like (datos de día) con ajuste del tamaño de muestra. Datos para series igualmente espaciadas cada 300 m.

Transecto	correlación	n	n'	pvalue
18	-0.246419	538	14	0.09736664
24	-0.290136	433	44	0.10160212
46	-0.162328	508	32	0.36615851
58	-0.166707	360	32	0
64	-0.391263	404	9	0.03980275
72	-0.238894	427	45	0.06776685
80	-0.335630	428	20	0
88	-0.337723	350	17	0.01686535

Al realizar el cálculo de la correlación lineal en cada transecto seleccionado con los datos separados cada 2 km. Se observó una asociación negativa entre la variable física y el log fluid like. Luego de calcular el tamaño de muestra efectivo y al aplicarlo para calcular la significancia de estas correlaciones, se pudo observar que solamente 2 correlaciones fueron significativas (Tabla 10).

Tabla 10. Resultados de la correlación entre la variable física y el log fluid like (datos de día) con ajuste del tamaño de muestra. Datos para series igualmente espaciadas cada 2 km.

Transecto	correlación	n	n'	pvalue
18	-0.436839	80	12	0.07338999
24	-0.339405	63	34	0.02523203
46	-0.126608	75	18	0.30525833
58	-0.560591	53	24	0.00198605
64	-0.517123	59	8	0.08501061
72	-0.111067	62	33	0.26986095
80	-0.468791	63	16	0.02997299
88	-0.641893	51	13	0.00754659

3 Aplicación del método geoestadístico

Para la aplicación de la geoestadística se utilizó el software R versión 2.8.1, utilizando la librería gstat. Para modelar la estructura espacial de la variable física (también del log fluid like y log fish) se calculó un variograma robusto (**ecuación 30**) omnidireccional con una separación de 1 km ($h=1$ km). El variograma omnidireccional fue generado en todas las direcciones. Para cada estructura observada en el variograma se probaron diferentes modelos teóricos, quedándonos con los que modelaron mejor la estructura presente en el variograma (variograma anidado). Luego de determinar los parámetros (meseta, rango, pepita) correspondientes a cada modelo teórico, se procedió a calcular la interpolación espacial mediante el método del kriging, esto se realizó solamente para mostrar una imagen de la variable en estudio no siendo relevante para identificar las estructuras espaciales. Para esta interpolación se utilizó el modelo ajustado al variograma omnidireccional. La interpolación fue realizada mediante un kriging ordinario, para lo cual se utilizó

una grilla que abarcó desde los 5 hasta los 19 grados de latitud sur y de los 70 hasta los 82 grados de longitud oeste.

Para cuantificar la dependencia espacial entre dos variables (log fluid like - variable física, log fish - variable física y log fish - log fluid like) se ha calculado el variograma cruzado (**ecuación 33**) e identificado las estructuras presentes en el variograma. Los análisis fueron realizados para los datos de día y datos de noche.

3.1 Modelamiento geostatístico

3.1.1 Análisis de estructuras espaciales de la variable física

Luego de calcular un variograma robusto para la variable física – noche (Figura 13) y día (Figura 14), se ajustó un modelo exponencial a las estructuras espaciales observadas en los variogramas respectivos. Llegándose a identificar en los dos casos un rango de 30 km. Se observó una estructura similar de la variable física para los datos de día como para los datos e noche.

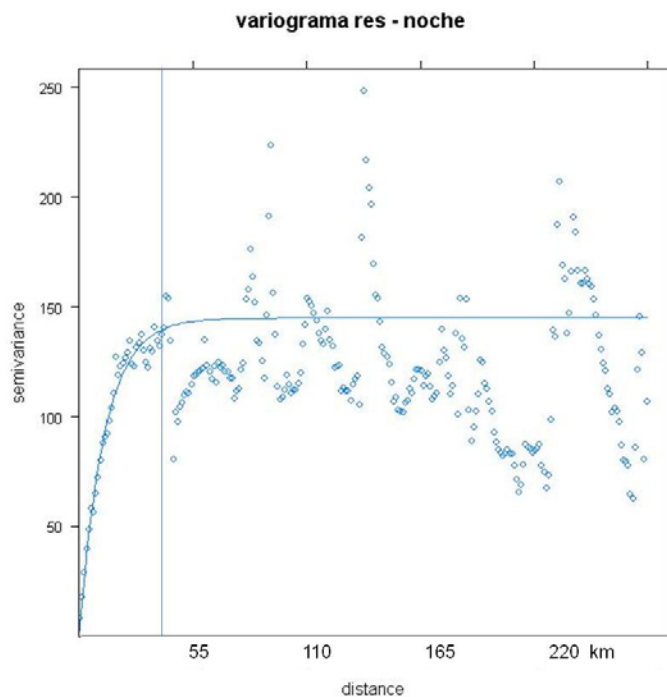


Figura 13. Variograma robusto de variable física - noche. Estructura modelada mediante un modelo exponencial, con un rango de 30 km.

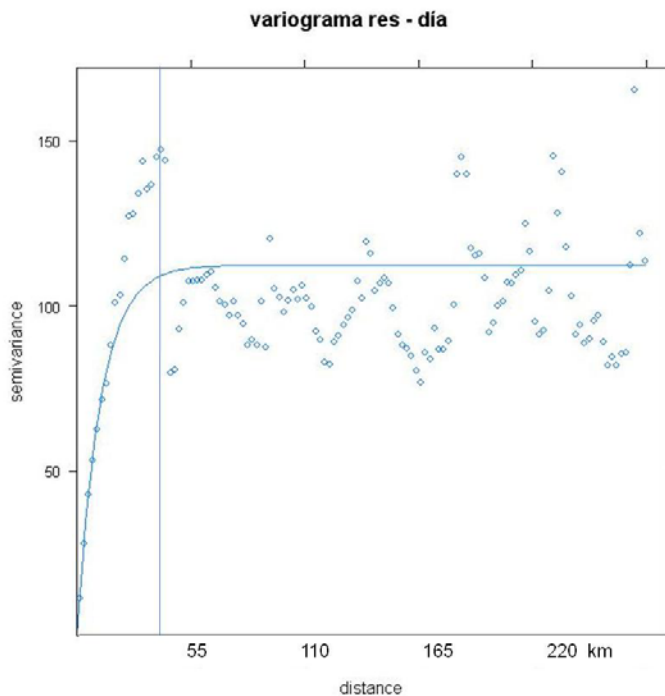


Figura 14. Variograma de la variable física – día. Estructura espacial modelada mediante un modelo exponencial con un rango de 30 km.

Tabla 11. Modelos ajustados al variograma del log fl noche.

	modelo	meseta	rango	pepita
variable física noche	Exponencial	140	30	0
variable física día	Exponencial	120	30	0

3.1.2 Análisis de estructuras espaciales del log fluid like

Se calculó el variograma robusto para los datos de log fluid like noche, se observó tres estructuras espaciales en el variograma. A la primera estructura espacial (Figura 15a) se le ajustó un modelo exponencial con un rango de 15 km; a la segunda estructura espacial se le ajustó un modelo materno con un rango de 70 km (Figura 15 b). Finalmente la tercera estructura espacial fue modelada mediante un modelo exponencial con un rango de 130 km (Figura 15c).

Para los datos de fluid like día, de igual modo, se observó tres estructuras espaciales. Siendo la primera estructura modelada mediante un modelo exponencial teniendo un rango de 20 km (Figura

16a). La segunda estructura mediante un modelo exponencial con un rango de 75 km (Figura 16b), Finalmente la tercera estructura fue modelada por un modelo Bessel, la cual tenía un rango de 120 km (Figura 16c). Los rangos de la pequeña estructura variaron según los datos de día y noche.

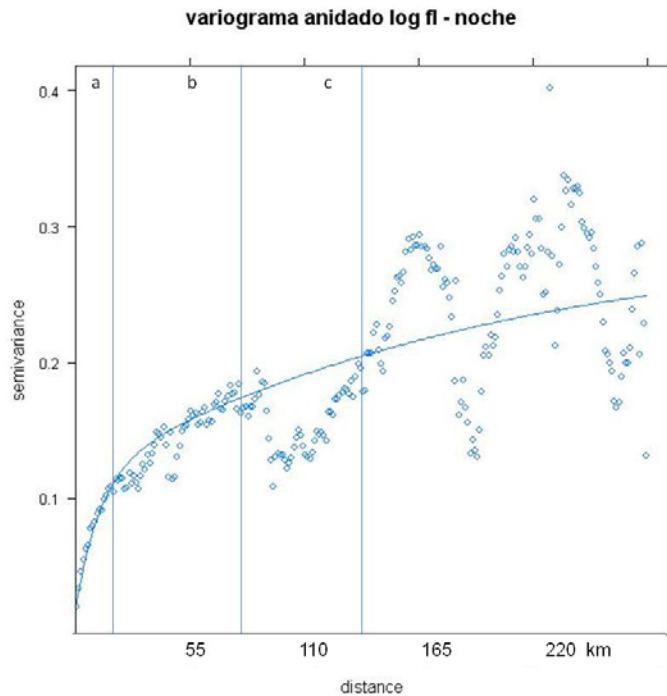


Figura 15. Variograma anidado del log fl noche; a) Modelo Exponencial con rango de 15 km; b) Modelo Materno con rango de 70 km; c) Modelo Exponencial con rango de 130 km.

Tabla 12. Modelos ajustados al variograma del log fl noche.

Modelo	Meseta	Rango (km)	Pepita
Exponencial	0.12	15	0.02
Materno	0.17	70	0.02
Exponencial	0.22	130	0.02

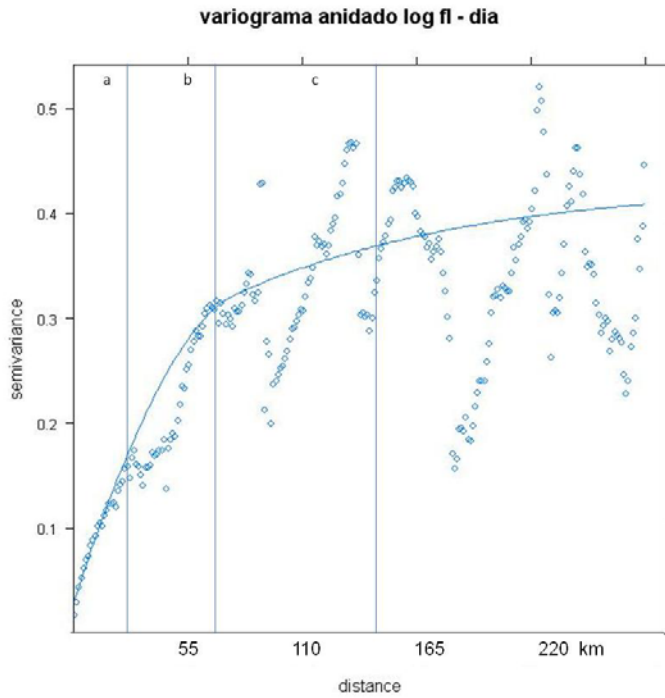


Figura 16. Variograma anidado del log fl día; a) Modelo Exponencial con rango de 20 km; b) Modelo Materno con rango de 70 km; c) Modelo Exponencial con rango de 120 km.

Tabla 13. Tres modelos ajustados al variograma robusto del log fl día.

Modelo	Meseta	Rango (km)	Pepita
Exponencial	0.17	20	0.02
Esférico	0.3	70	0.02
Bessel	0.35	120	0.02

Finalmente, utilizando los modelos y parámetros del variograma anidado (Tabla 6 y 7) del log fluid like se procedió a realizar la interpolación espacial mediante el kriging ordinario. Los mapas interpolados para el log fluid like son apreciados en las Figuras 17 (para noche) y 18 (para día).

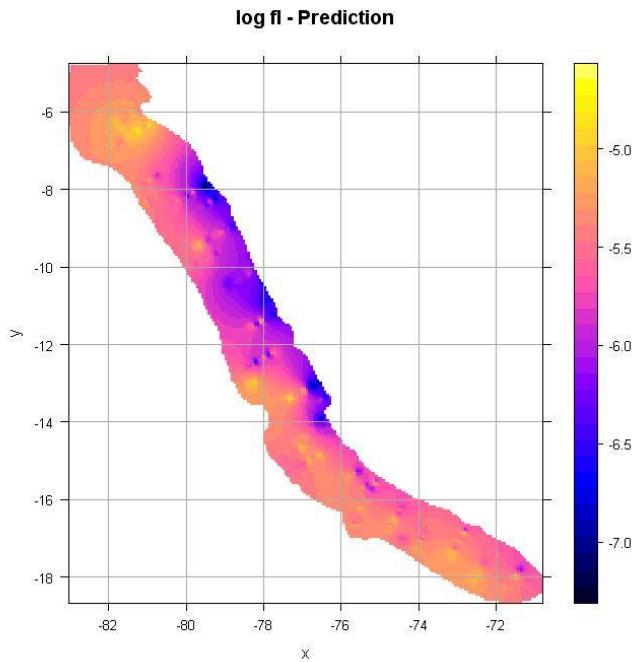


Figura 17. Interpolación espacial de la distribución del log fluid like para datos de noche. Interpolación mediante el método kriging ordinario utilizando el variograma anidado (Tabla 12).

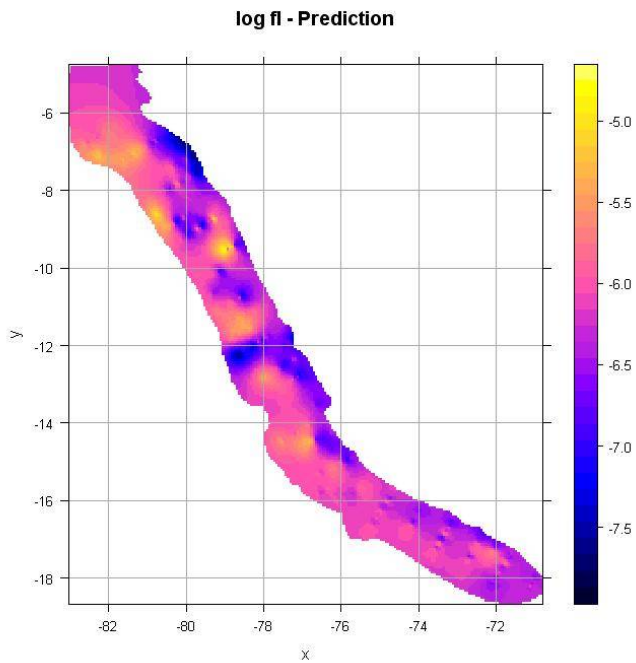


Figura 18. Interpolación espacial de la distribución del log fluid like para datos de día. Interpolación mediante el método kriging ordinario utilizando el variograma anidado (Tabla 13).

3.1.3 Análisis de estructuras espaciales del log fish

Luego de calcular los variogramas a los datos de log fish se observaron la presencia de tres estructuras (para ambos variogramas). Luego de ajustar los modelos a cada estructura para los datos de noche y día (Tablas 9 y 10) se observó que la pequeña estructura del log fish noche tuvo un rango de 25 km (Figura 19a), siendo diferente para los datos de día con un rango de 20 km (Figura 20a). La segunda estructura presenta un rango de 55 km para el día (Figura 20b) y de 70 km para la noche (Figura 19b). Finalmente la tercera estructura de igual rango, 120 km, para los dos variogramas.

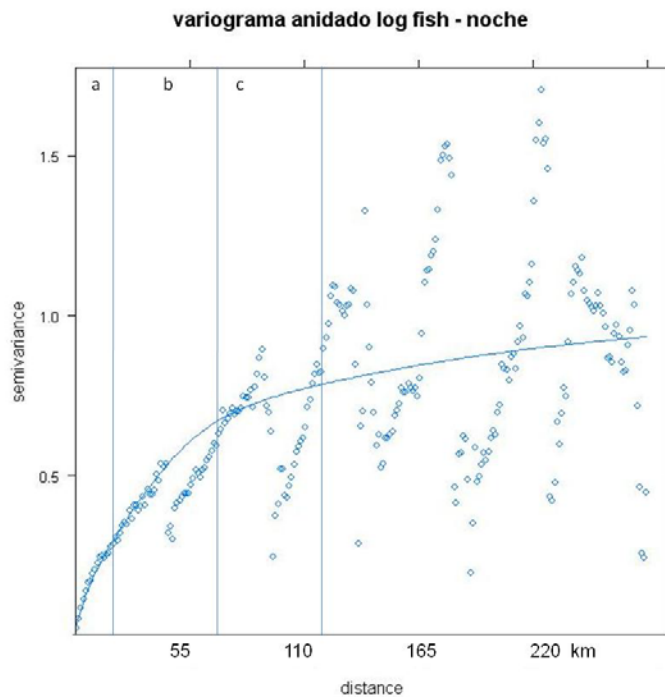


Figura 19. Variograma del log fish noche; a) Modelo Exponencial con rango de 25 km; b) Modelo Esférico con rango de 70 km; c) Modelo Bessel con rango de 120 km.

Tabla 14. Modelos ajustados al variograma robusto del log fish noche.

Modelo	Meseta	Rango (km)	Pepita
Exponencial	0.5	25	0.05
Materno	0.8	70	0.05
Exponencial	0.18	120	0.05

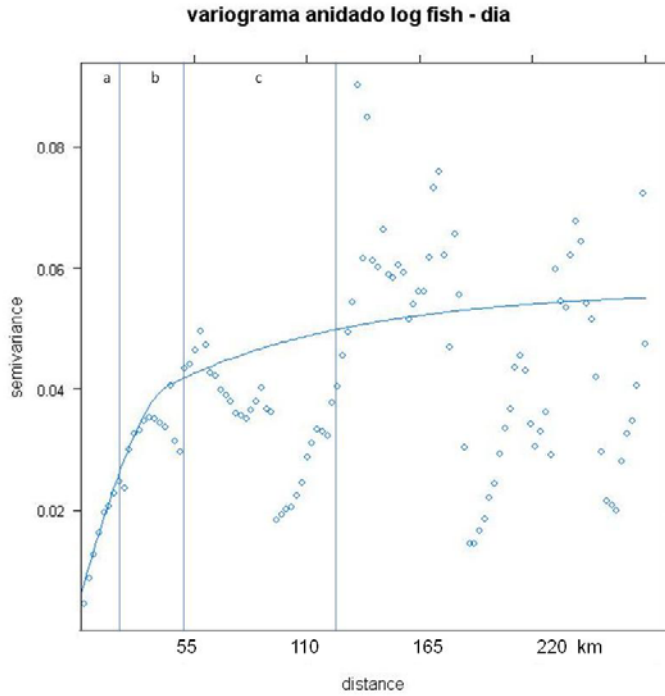


Figura 20. Variograma robusto del log fish día. Dicho variograma presenta tres estructuras espaciales. a) Modelo exponencial con rango de 20 km. b) modelo materno con rango de 55 km. c) Modelo exponencial con rango de 120 km.

Tabla 15. Tres modelos ajustados al variograma robusto del log fish día.

Modelo	Meseta	Rango	Pepita
Exponencial	0.025	20	0.008
Materno	0.045	55	0.008
Exponencial	0.06	120	0.008

Utilizando los modelos y los parámetros de la Tabla 14 y 15 se procedió a interpolar mediante el método de kriging ordinario la distribución espacial del log fish noche y día (Figuras 21 y 22).

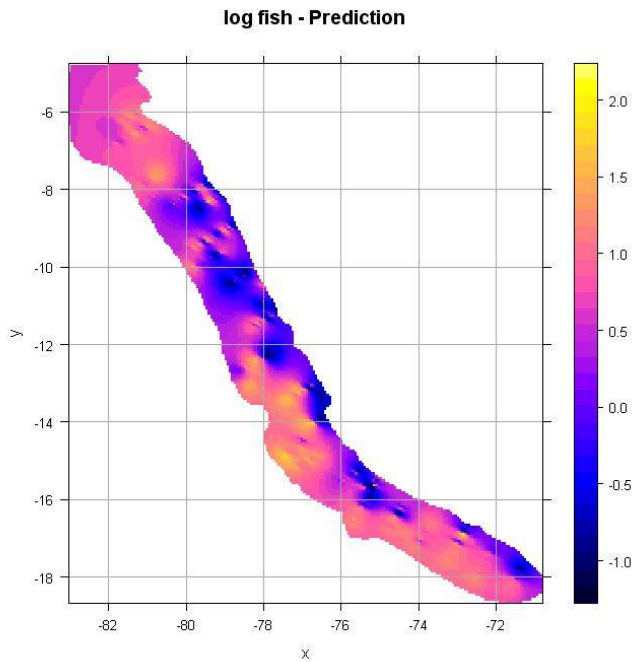


Figura 21. Interpolación espacial de la distribución del log fish parara datos de noche. Interpolación mediante el método kriging ordinario utilizando el variograma anidado (Tabla 14).

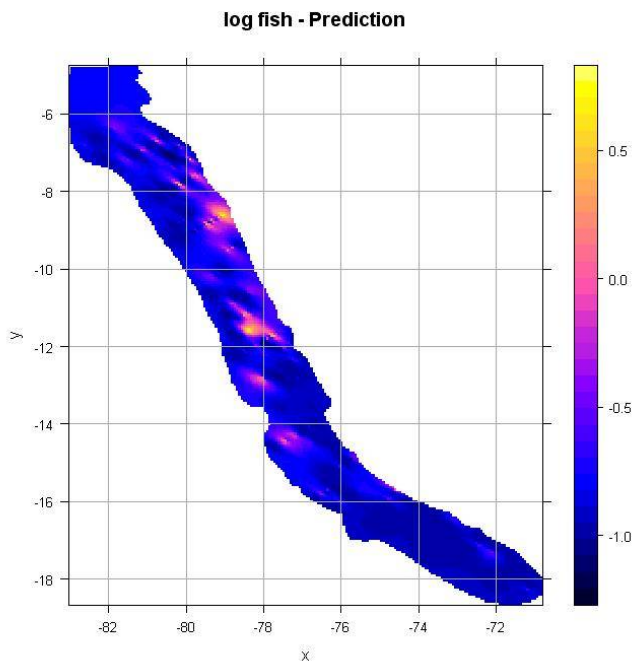


Figura 22. Interpolación espacial de la distribución del log fish parara datos de día. Interpolación mediante el método kriging ordinario utilizando el variograma anidado (Tabla 15).

3.2 Analizando la estructura de coregionalización

Luego de haber identificado las estructuras espaciales para la variable física, el log fluid like y el log fish (para datos de día y noche) se estudió el comportamiento en conjunto (dos a dos) de las variables en función al espacio.

3.2.1 Coregionalización entre variable física y log fluid like

El variograma cruzado permite estudiar el comportamiento de la covarianza a través del espacio entre dos variables, en nuestro caso entre el log fl y la variable física (Figura 23) para datos de noche.

El variograma cruzado entre el log fl y la variable física para los datos de noche (Figura 23) presenta una covarianza negativa, con un valor absoluto de la covarianza que aumenta desde 0 hasta los 30 km (Figura 23a). Esta covariación disminuye de los 30 hasta los 70 km aproximadamente (Figura 23b). Nuevamente la covariación aumenta (en valor absoluto) hasta aproximadamente los 130 km.

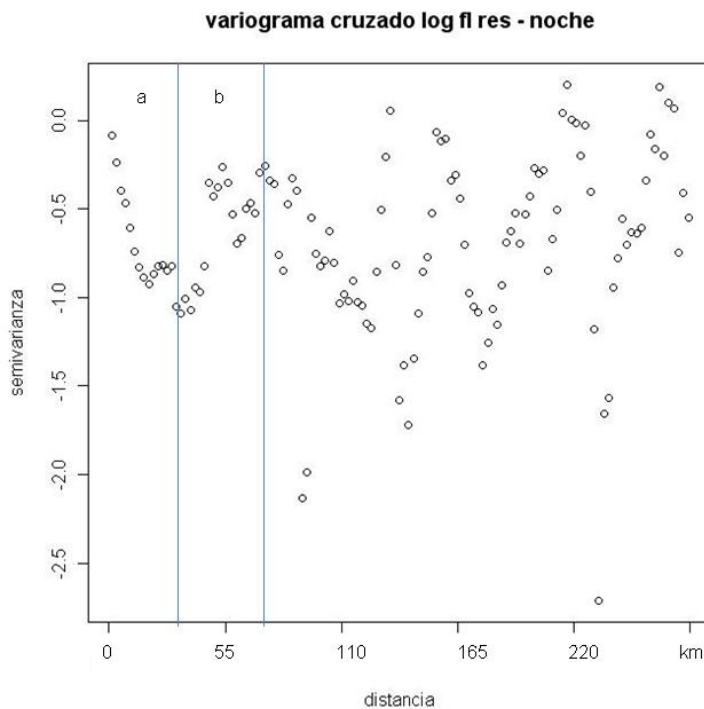


Figura 23. Variograma cruzado entre el log fl y la variable física para datos de noche.

De igual forma se calculó el variograma cruzado para la variable física y el log fluid like para datos de día (Figura 24). En el variograma cruzado se observó la presencia de dos estructuras espaciales. También se observó que la covariación espacial entre el log fluid like y la variable física fue negativa. En valor absoluto, la covarianza espacial aumentó desde 0 hasta los 40 km aproximadamente (Figura 24a), para decrecer de los 40 hasta los 65 km (Figura 24b).

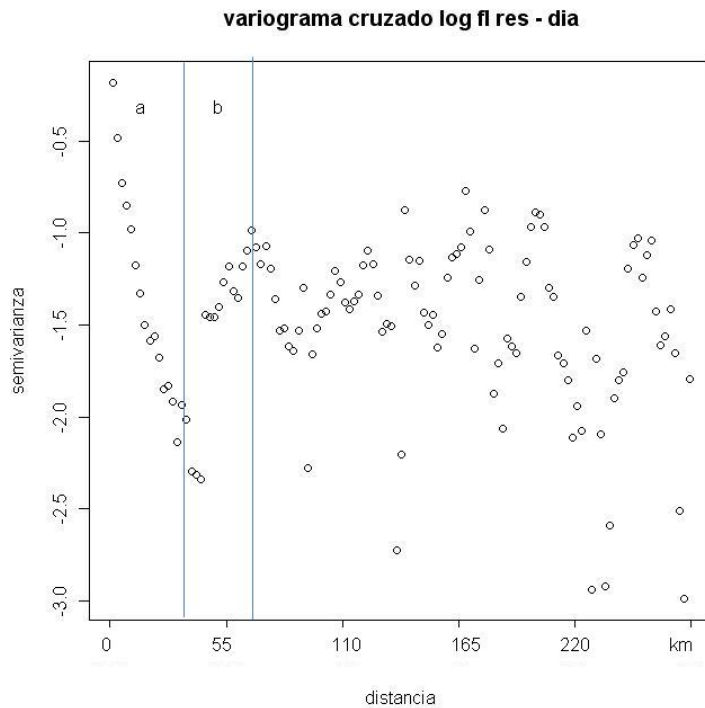


Figura 24. Variograma cruzado entre el log fl y la variable física para datos de día.

3.2.2 Coregionalización entre variable física y log fish

El variograma cruzado entre la variable física y el log fish para los datos de noche, reveló la presencia de tres estructuras. Se observó que existe una relación negativa en la covarianza de la variable física y el log fish, la cual cambia a través del espacio. La primera estructura ocurre de forma creciente (en valor absoluto) de los 0 hasta los 25 km aproximadamente (Figura 25a), para luego decrecer desde los 25 hasta los 50 km (Figura 25b). Finalmente se observa que la covariación aumenta (en valor absoluto) desde los 50 km hasta los 80 km (Figura 25c).

El variograma cruzado para la variable física y el log fish (datos de día) (Figura 26) Reveló la existencia de tres estructuras. Se observó que la relación negativa de la covarianza entre la variable física y el log fish. En valor absoluto la covarianza aumenta desde los 0 hasta los 30 km (Figura 26a), para luego decrecer desde los 30 km hasta los 60 km (Figura 26b) y finalmente nuevamente aumenta desde los 60 km hasta los 110 km (Figura 26c).

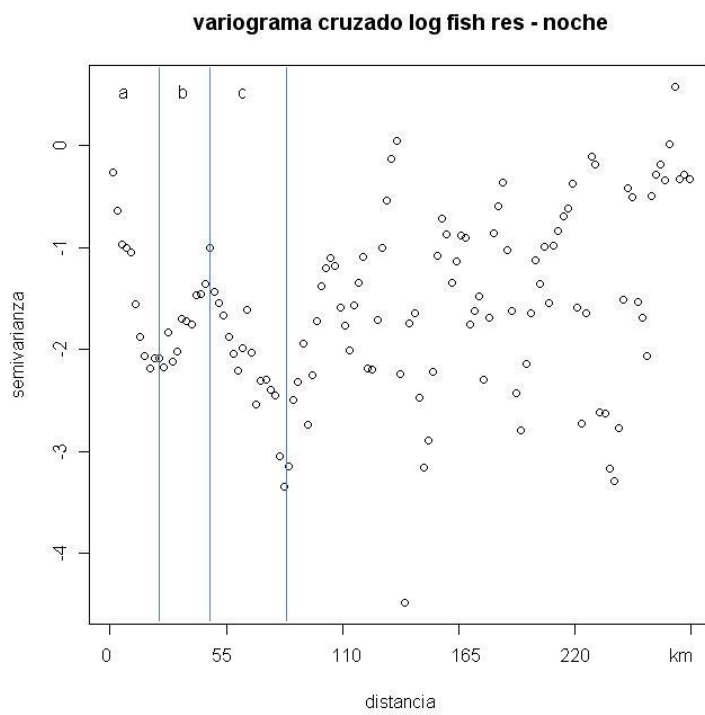


Figura 25. Variograma cruzado de la variable física y el log fish para datos de noche.

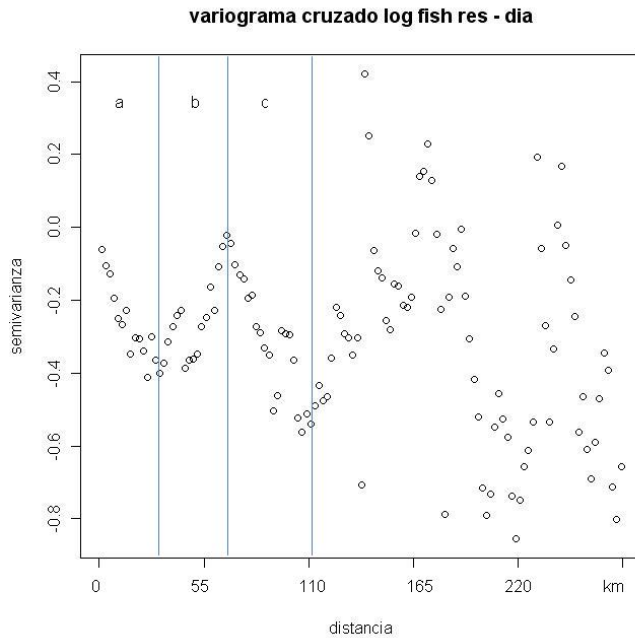


Figura 26. Variograma cruzado entre la variable física y el log fish para datos de día.

3.2.3 Coregionalización entre log fluid like y log fish

Luego de calcular el variograma cruzado entre el log fish y el log fluid like (para datos de noche), se observó una covariación positiva entre el log fish y log fluid like. Este variograma cruzado tuvo tres estructuras (Figura 27), la primera abarcó de los 0 hasta los 25 km (Figura 27a), luego de los 25 a los 55 km (Figura 27b) y finalmente de los 55 hasta los 90 km (Figura 27c).

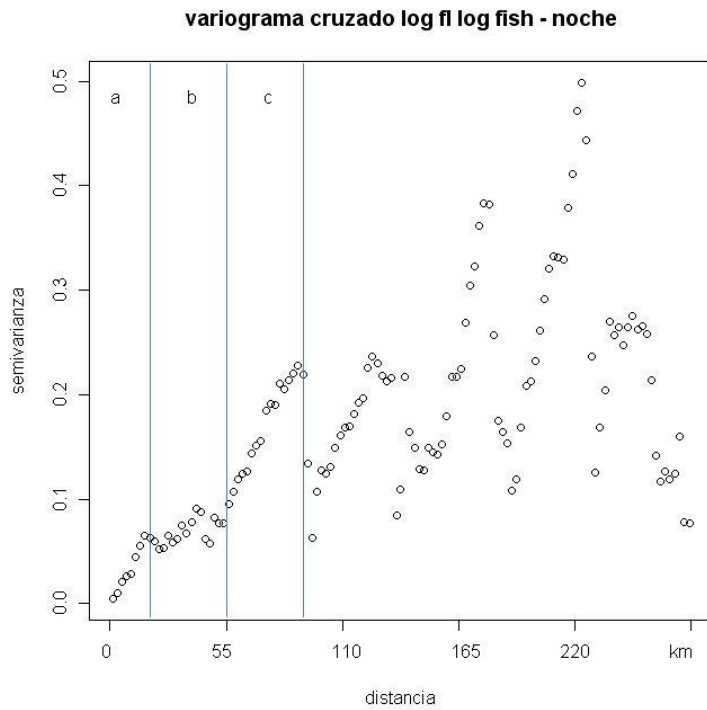


Figura 27. Variograma cruzado entre el log fish y el log fluid like para datos de noche.

En el variograma cruzado entre el log fish y el log fluid like (para datos de día) se observó tres estructuras, siendo la primera de los 0 hasta los 25 km (Figura 28 a), la segunda estructura fue de los 25 hasta los 55 km (Figura 28 b), para finalmente observarse otra desde los 55 hasta los 120 km (Figura 28c). En este último rango se observó un decrecimiento de la covarianza (sin llegar a cero).

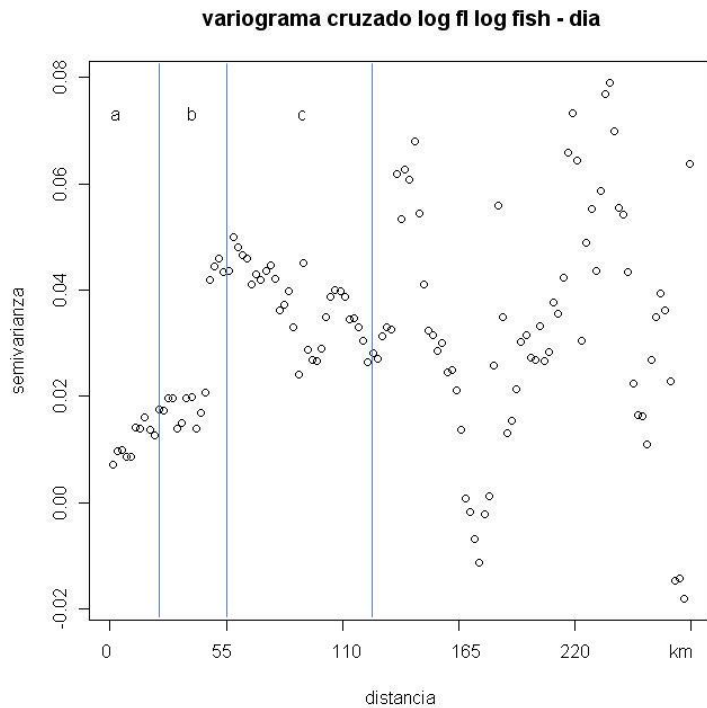


Figura 28. Variograma cruzado entre el log fish y el log fluid like para datos de día.

4 Aplicación de los PCNM

Para realizar los análisis hemos utilizado el software R (versión 2.8.1), en especial la librería PCNM desarrollado por Pierre Legendre, Daniel Borcard, Stephane Dray y Miquel de Cáceres.

Se probó la significancia de los vectores propios (asociados a los valores propios positivos) mediante el test de moran utilizando 999 permutaciones. Seleccionándose aquellos vectores propios con una correlación espacial altamente significativa ($p\text{-value} < 0.01$) para usarlos como variables explicativas en regresiones lineales múltiples con test de permutación.

Entonces para las regresiones lineales múltiples se tomó cada variable (**variable física, log fluid like y log fish**) como variable respuesta y como variables explicativas los vectores propios significativos. Estas regresiones nos mostrarán aquellos vectores propios que son relevantes para explicar la distribución espacial de la variable respuesta (variable física, log fluid like y log fish respectivamente).

Para cada PCNM variable se calculó un variograma y se procedió a ajustar un modelo teórico (para identificar el rango), luego se generó un gráfico entre los rangos encontrados y los valores propios que generaron los PCNM variables. Con ayuda del gráfico nosotros elegimos el número y rango de los submodelos. Dentro de cada submodelo se identificaron los PCNM variables que se tienen rangos dentro de los límites del submodelo.

Lo anteriormente descrito se realizó para las PCNM variables generados por la variable física, luego para las PCNM variable generadas por el log fluid like y finalmente para las generadas por el log fish.

Para estudiar la relación que existe entre la **variable física y el log fluid like** se aplicó regresiones múltiples dentro de los submodelos obtenidos para el log fluid like. Estas regresiones tuvieron como variable respuesta el log fluid like y como variables explicativas a los PCNM variables pertenecientes a cada submodelo y, también, a la variable física.

De igual forma para conocer la relación que existe entre el **log fish con el log fluid like y la variable física** se plantearon regresiones múltiples para cada submodelo obtenido para el log fish. Para estas regresiones se tomó como variable respuesta al log fish y como variables explicativas a los PCNM variables pertenecientes a cada submodelo en conjunto con la variable física y el log fluid like. Los análisis se realizaron tanto para los datos de día como para los datos de noche.

4.1 Descomposición del espacio

Después de calcular las distancias euclidianas entre todas las posiciones geográficas de las observaciones se truncó la matriz con un umbral de 80 km para los datos de noche. Mediante el análisis de coordenadas principales aplicada a la matriz de distancias truncadas generada por las coordenadas de los datos de noche obtuvimos 720 vectores propios asociados a valores propios positivos (Tabla 16). Mediante el test de Moran (con 999 permutaciones) se determinó que 307 de estos vectores propios tenían autocorrelación significativa ($p < 0.01$) (Figura 29).

De manera similar procedimos para la base de día. Luego de calcular las distancias entre todas las coordenadas de las observaciones se truncó la matriz de distancias con un umbral de 85 km. Mediante el análisis de coordenadas principales se procedió a descomponer la matriz de distancias truncadas. Teniendo 1263 vectores propios asociados a valores propios positivos. El análisis de autocorrelación mediante el test de Moran con permutación indicó que de los 1263 valores propios positivos 518 fueron significativos ($p < 0.001$) (Figura 30).

Tabla 16. Número de vectores propios significativos.

Base	Umbral	Vectores propios asociados a valores propios positivos	Vectores propios significativos (test de Moran)
Noche	80 km	720	307
Día	85 km	1263	518

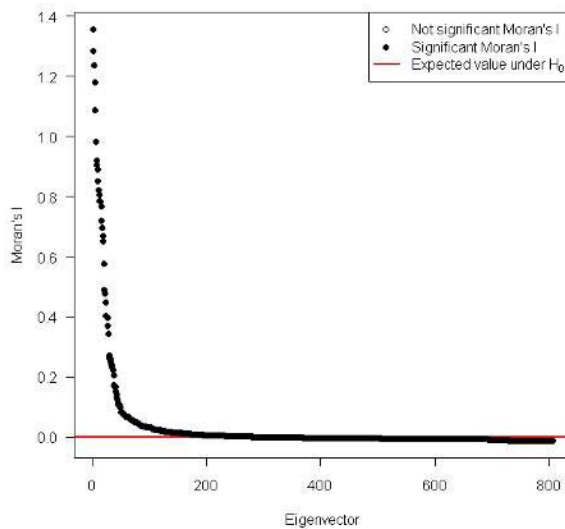


Figura 29. Plot de los valores propios después de la diagonalización de la matriz de distancias truncada para observaciones de noche. La significancia de estos índices fue calculado mediante los índices de Moran.

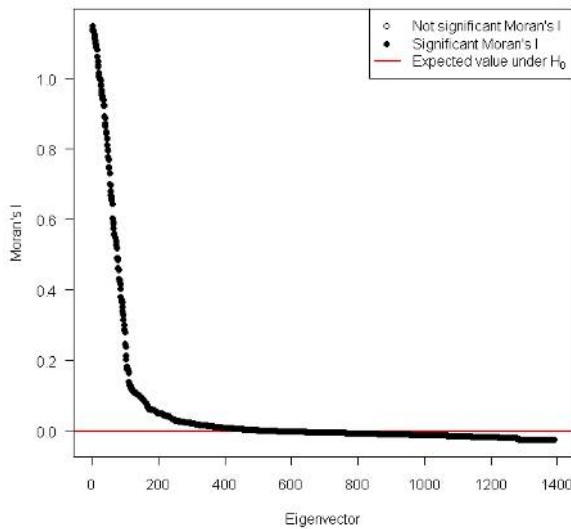


Figura 30. Plot de los valores propios después de la diagonalización de la matriz de distancias truncada para observaciones de día. La significancia de estos índices fue calculado mediante los índices de Moran

4.2 Análisis de estructuras

4.2.1 Estructuras espaciales para la variable física

Luego de identificar el número de vectores propios significativos (para los datos de noche), se procedió a realizar regresiones lineales múltiples (con 999 permutaciones). Para tales regresiones se tomó como variable respuesta a la variable física (para datos de noche) y como variables explicativas a los 307 vectores que resultaron ser significativos (Tabla 16). Los resultados de las regresiones indicaron que de los 307 vectores solamente 41 resultaron ser relevantes para explicar la variabilidad de la variable física. Estos 41 vectores llegaron a explicar el 31% de la variabilidad de la variable física ($R^2_{adj}=0.3132$).

Luego para cada uno de los 41 vectores se le ajustó un modelo de variograma (esto con el objetivo de identificar el rango de la estructura en cada vector). Después de calcular los rangos de los 41 vectores propios y con ayuda visual de la Figura 31 se agrupó en tres submodelos, el primer submodelo abarca desde los 0 hasta los 12 km (Figura 31a), seguido por el segundo submodelo que abarcó desde los 12 hasta los 66 km (Figura 31b). Finalmente el tercer submodelo abarco a aquellos vectores con rangos superiores a los 66 km (Figura 31c).

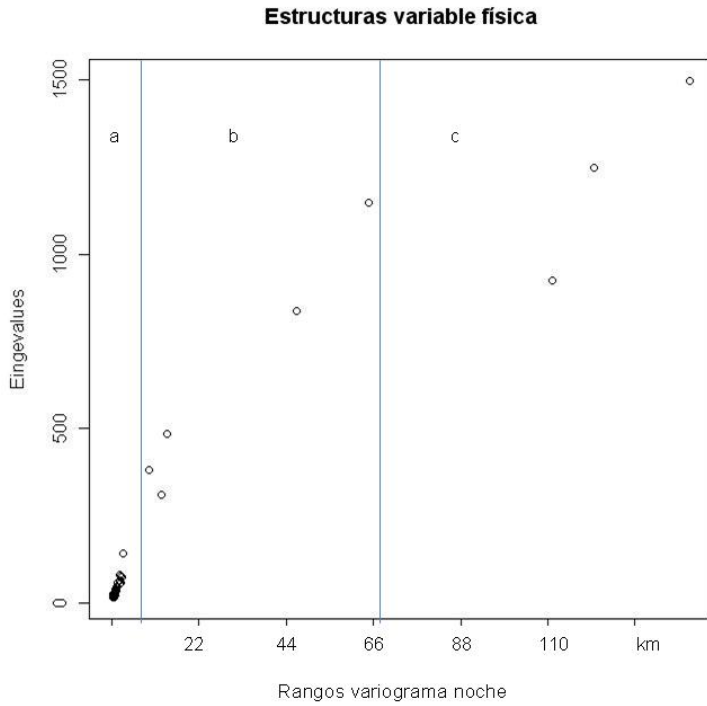


Figura 31. Rangos de las 41 PCNM variables significativas para explicar la variabilidad de la variable física noche.

Para los datos de día, se procedió a realizar regresiones lineales múltiples (con 999 permutaciones) tomando como variable respuesta la variable física (datos de día) y como variables explicativas a los 518 vectores que resultaron ser significativos (Tabla 16). Los resultados de las regresiones indicaron que de los 518 vectores solamente 38 resultaron ser relevantes para el modelo. Estos 38 vectores explicaron el 32% de la variabilidad de la variable física día ($R^2_{adj}=0.3241$).

Luego para cada uno de los 38 vectores se ajustó un modelo de variograma (esto con el objetivo de identificar el rango de cada vector). Después de calcular los 38 rangos (de los 38 vectores) y con ayuda visual de la Figura 32 se agrupó en tres submodelos. El primer submodelo abarcó desde los 0 hasta los 15 km (Figura 32a), seguido por el segundo submodelo que abarcó desde los 15 hasta los 60km (Figura 32b), finalmente el tercer submodelo abarco a aquellos vectores con rangos superiores a los 60 km (Figura 32c).

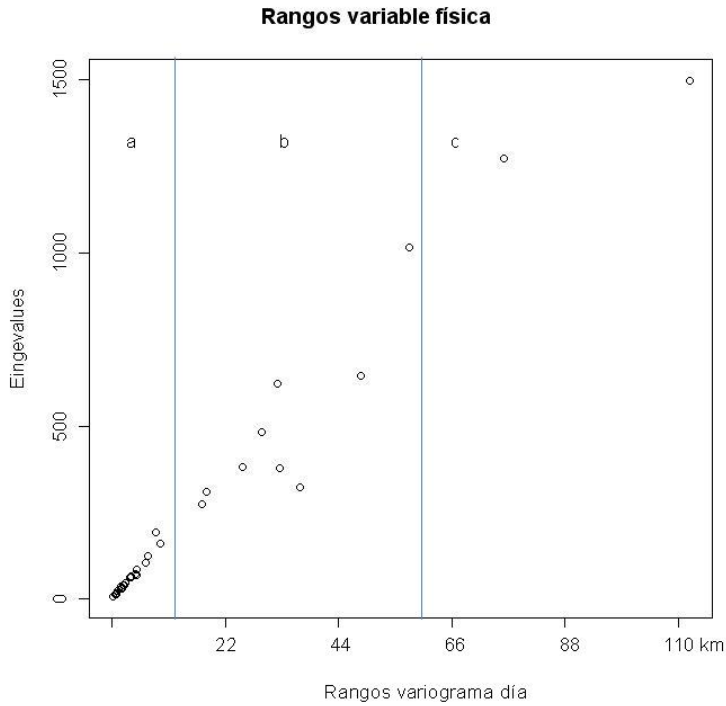


Figura 32. Rangos de las 38 PCNM variables significativas para explicar la variabilidad de la variable física día.

4.2.2 Estructuras espaciales para el log fluid like

Luego de haber determinado el número de vectores propios significativos para los datos de noche (Tabla 16), se realizó regresiones lineales múltiples con test de permutación (999 permutaciones). Teniendo como variable respuesta el log fl noche y como variables explicativas los 307 vectores propios significativos. De estas regresiones se obtuvieron como resultado que solamente 36 vectores resultaron ser relevantes para el modelo. Estos 36 vectores propios llegaron a explicar el 53% de la variabilidad del log fl noche ($R^2_{ajus}=0.5338$)

Mediante ajustes de variogramas se procedió a determinar los rangos para cada uno de los 36 vectores propios (Figura 33). Luego con ayuda visual de la Figura 33 se decidió particionar en tres submodelos. EL primer submodelo abarcó hasta los primeros 13 km (Figura 33a). El segundo submodelo abarcó desde los 13 hasta los 60 km (Figura 33b). Finalmente el tercer submodelo abarcó a aquellos vectores propios con rangos superiores a los 60 km (Figura 33c).

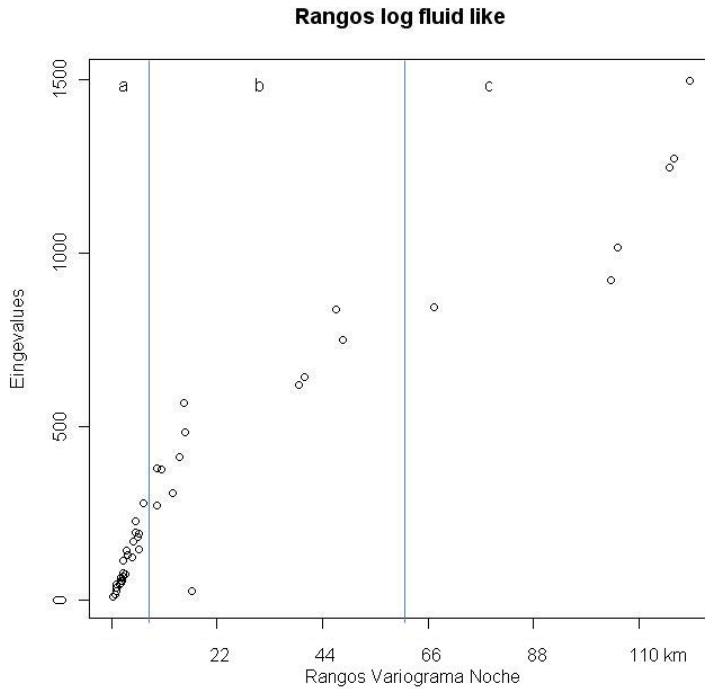


Figura 33. Rangos de las 36 PCNM variables significativas que explican la variabilidad de la variable log fluid like noche.

De igual manera que en el caso de noche, hemos realizado múltiples regresiones lineales entre el log fl día y los 518 vectores propios significativos (Tabla 16). Los resultados de las regresiones múltiples indicaron que existen 80 vectores propios relevantes para el modelo, los cuales llegan a explicar el 58% de la variabilidad del log fl ($R^2_{ajus}=0.5769$).

Mediante el ajuste de variogramas a cada uno de los 80 vectores propios se identificó el rango para cada vector propio. Nuevamente observando la Figura 34 se decidió separar en tres submodelos. Siendo el primer submodelo compuesto por los vectores propios que tuvieron rangos inferiores a los 15 km (Figura 34a). El segundo submodelo estuvo compuesto por aquellos vectores con rangos comprendidos entre los 15 km y los 50 km (Figura 34b). El tercer submodelo estuvo compuesto por aquellos vectores que resultaron tener rangos superiores a los 50 km (Figura 34c).

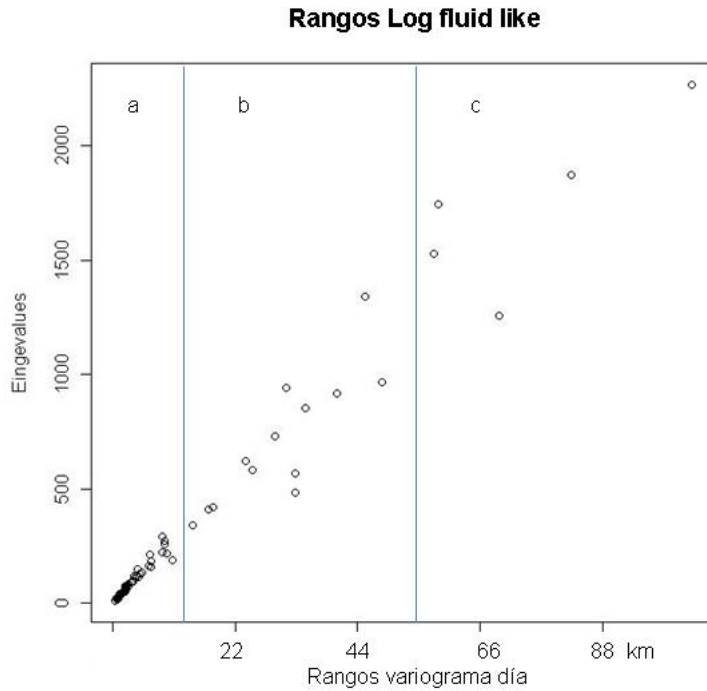


Figura 34. Rangos de las 80 PCNM variables significativas que explican la variabilidad de la variable log fluid like día.

4.2.3 Estructuras espaciales para el log fish

Para los datos de log fish noche se procedió a realizar regresiones múltiples con test de permutación (999 permutaciones). Se utilizó el log fish noche como variable respuesta y los 307 vectores propios significativos como variables explicativas (Tabla 16). Teniendo como resultado que 37 vectores propios fueron relevantes para el modelo. Estos 37 vectores propios explican el 42% de la variabilidad del log fish noche ($R^2_{\text{ajus}}=0.4210$).

Luego a cada uno de los 37 vectores propios se le ajustó un modelo de variograma con el fin de identificar los rangos para cada vector. Con ayuda de la Figura 35 se decidió separar en tres submodelos. El primer submodelo estuvo compuesto por aquellos vectores que tuvieron rangos inferiores a los 20 km (Figura 35a). El segundo submodelo estuvo compuesto por aquellos vectores que tuvieron rangos entre los 20 y los 60 km (Figura 35b). Finalmente el tercer submodelo estuvo

compuesto por aquellos vectores que tuvieron rangos superiores a los 60 km (Figura 35c).

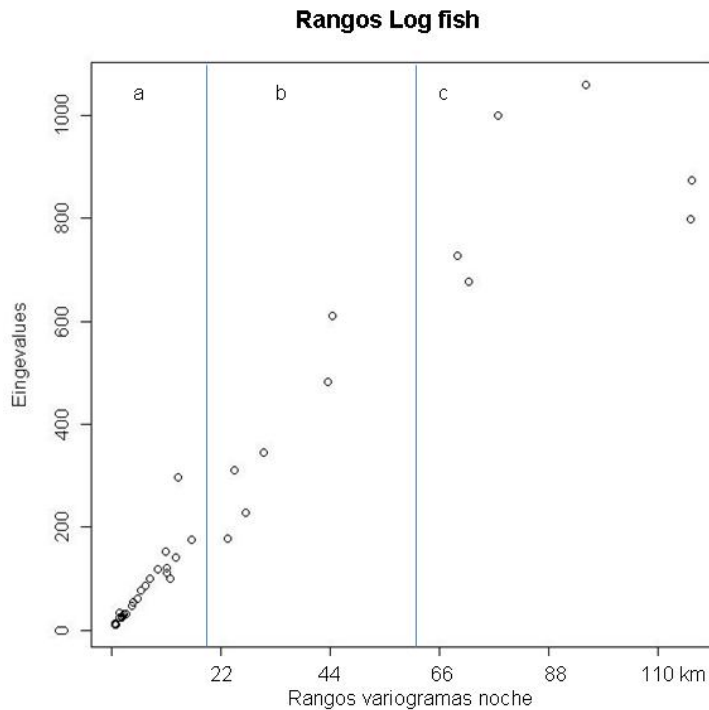


Figura 35. Rangos de las 37 PCNM variables significativas que explican la variabilidad de la variable log fish noche.

Para los datos de día se utilizó en regresiones lineales múltiples como variable respuesta el log fish día y como variables explicativas los 518 vectores propios significativos. Teniendo como resultado que 43 vectores propios resultaron ser relevantes para el modelo planteado. Estos 43 vectores propios llegaron a explicar el 33% de la variabilidad del log fish día.

Se ajustó un variograma para cada uno de los 43 vectores propios (con el objetivo de identificar el rango del vector). Luego observando la figura se decidió dividir en tres submodelos. El primer submodelo estuvo compuesto por aquellos vectores que tuvieron rangos inferiores a los 12 km (Figura 36a). El segundo submodelo estuvo compuesto por aquellos vectores que tuvieron rangos comprendidos entre los 12 y los 50 km (Figura 36b). El tercer submodelo estuvo compuesto por aquellos vectores con rangos superiores a los 50 km (Figura 36c).

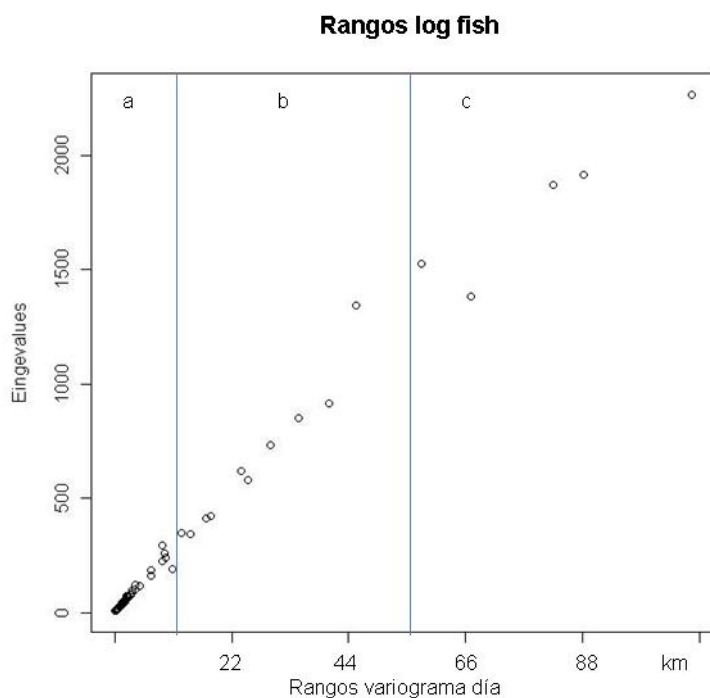


Figura 36. Rangos de las 43 PCNM variables significativas que explican la variabilidad de la variable log fish día.

Tabla 17. Resumen de los submodelos para noche y día

		Submodelo 1	Submodelo 2	Submodelo 3
Noche	Variable física	12 km	12 - 66 km	> 66 km
	Log fluid like	13 km	13 - 60 km	> 60 km
	Log fish	20 km	20 - 60 km	> 60 km
Día	Variable física	15 km	15 - 60 km	> 60 km
	Log fluid like	15 km	15 - 50 km	> 50 km
	Log fish	12 km	12 - 50 km	> 50 km

En la Tabla 17 se observó que los rangos de los submodelos variaron según el periodo (día o noche); esta variación fue más fuerte en los primeros submodelos y más estable para los otros dos submodelos.

4.3 Análisis de los submodelos (relaciones entre variables)

4.3.1 Relación entre el log fluid like y la variable física

Luego de haber definido los submodelos para el log fluid like para día y noche (Tabla 17). Con el objetivo de conocer la relación entre el log fluid like y la variable física se realizó regresiones lineales múltiples para cada submodelo identificado del log fluid like.

Tomando como variable respuesta el log fluid like noche y como variables explicativas los vectores correspondientes a cada submodelo planteado para el log fluid like noche y, también, a la variable física. Las tres regresiones planteadas resultaron ser significativas con un p-value < 0.01.

En el primer submodelo se llegó a explicar el 10 % de la variabilidad del log fluid like. En este submodelo la variable física resultó ser significativa, teniendo una relación negativa con el log fluid like pues tuvo un coeficiente negativo (Tabla 18).

En la regresión para el segundo submodelo, se llegó a explicar el 8% de la variabilidad del log fluid like. Nuevamente la variable física resultó ser significativa para el modelo. La variable física tuvo una relación negativa (coef=-0.0039) con el log fluid like (Tabla 18).

Finalmente en el tercer submodelo, se llegó a explicar el 13% de la variabilidad del log fluid like, siendo nuevamente la variable física significativa para el modelo y teniendo la variable física una relación negativa (coef=-0.0045) con el log fluid like (Tabla 18).

Tabla 18. Resultados de la regresión entre el log fluid like y las pcnm variables en cada submodelo determinado. Se presenta el R^2_{adj} de la regresión y el coeficiente y la significancia de la variable física. Datos para noche

	Submodelo 1		Submodelo 2		Submodelo 3	
R^2_{adj}	0.1046		0.0811		0.1337	
	coef	P	coef	p	coef	p
var física	-0.0031	0.001	-0.0039	0.001	-0.0045	0.001

Ahora para los datos de día, tomando como variable respuesta el log fluid like día y como variables explicativas los vectores correspondientes a cada submodelo planteado para el log fluid like día

(Tabla 17) y, también, a la variable física. Las tres regresiones planteadas resultaron ser significativas con un p-value < 0.01. Se observó que para las tres regresiones planteadas la variable física fue relevante para explicar la variabilidad del log fluid-like día (Tabla 19). Se observó en los tres submodelos una relación inversa entre la variable física y el log fluid like (Tabla 19). Los coeficientes de correlación ajustados variaron para cada submolelo, llegando a ser mayor en el primer submodelo (hasta los 15 km) (Tabla 13).

Tabla 19. Resultados de la regresión entre el log fluid like y las pcnm variables en cada submodelo determinado. Se presenta el R^2_{adj} de la regresión y el coeficiente y la significancia de la variable física. Datos para día

	Submodelo 1		Submodelo 2		Submodelo 3	
R^2_{adj}	0.3158		0.2601		0.1567	
	coef	p	coef	p	coef	p
var física	-0.0128	0.001	-0.0129	0.001	-0.01316	0.001

4.3.2 Relación entre el log fish, el log fluid like y la variable física

Utilizando cada uno de los submodelos del log fish noche (Tabla 17) se procedió a realizar regresiones múltiples. Para esto se utilizó como variable respuesta el log fish y como variables explicativas los vectores pertenecientes a cada submodelo en conjunto con la variable física y el log fluid like (para datos de noche).

Se observó que los tres modelos de regresión fueron significativos. Para los tres modelos resultaron ser significativos la variable física y el log fluid like. Los coeficiente de la variable física y el log fluid like revelaron que hubo una relación positiva entre el log fluid like y log fish para los tres submodelos y una relación negativa entre la variable física y el log fish en los tres submodelos (Tabla 20). La variable física y el log fl resultaron ser significativos en los tres modelos.

Tabla 20. Coeficientes y p-value calculados en las regresiones para cada uno de los submodelos del log fish noche.

	Submodelo 1		Submodelo 2		Submodelo 3	
R^2_{adj}	0.269998		0.2132		0.2452	
	coef	p	Coef	p	coef	p
var física	-0.0053	0.001	-0.0049	0.001	-0.0050	0.001
log fl	0.6455	0.001	0.6858	0.001	0.7438	0.001

Para los resultados de día se utilizó los submodelos del log fish día (Tabla 17). Se procedió a calcular regresiones múltiples utilizando como variable respuesta el log fish y como variables explicativas los vectores pertenecientes a cada submodelo en conjunto con la variable física y el log fluid like (para datos de día). Se observó que los tres modelos planteados fueron significativos. Los resultados de la regresión mostraron que el log fluid like fue significativo para los submodelos 2 y 3, teniendo este una relación directa con el log fish. Esta relación directa se dio en los tres submodelos (Tabla 21). A diferencia de la variable física, que fue significativa solamente para el primer y tercer submodelo (p -value < 0.01). Se observó una relación inversa entre la variable física y el log fish, esto se dio en los tres submodelos.

Los coeficientes de correlación ajustados llegaron a explicar solamente el 14%, 7% y 4% de la variabilidad del log fish (para el submodelo uno, dos y tres respectivamente) (Tabla 21).

Tabla 21. Coeficientes y p-value calculados para cada uno de los submodelos del log fish día.

	Submodelo 1		Submodelo 2		Submodelo 3	
R^2_{adj}	0.1497		0.0658		0.0433	
	coef	p	coef	p	coef	P
Var física	-0.00296	0.001	-0.00061	0.363	-0.00184	0.008
Log fl	0.03581	0.014	0.11338	0.010	0.09740	0.001

5 Aplicación de los análisis wavelets

Para el estudio de los wavelets se utilizó el toolbox wavelets desarrollado en plataforma matlab por Torrence y Compo.

Para poder aplicar los wavelets es necesario tener datos igualmente espaciados, por tal motivo se realizó una interpolación espacial cada 300 m dentro de cada transecto. La interpolación se realizó para la **variable física, el log fluid like y el log fish**. Luego se realizó los análisis wavelets para cada una de las tres variables dentro de cada transecto. También se aplicó los wavelets cruzados y el análisis de fase para la **variable física, el log fluid like y el log fish**, esto nuevamente para cada transecto.

Los análisis se realizaron por separado para los datos de día y para los datos de noche. Utilizando la transformada de Morlet se realizó el análisis wavelets para cada transecto, tanto para datos de día como para datos de noche. El nivel de confianza del 95 % (líneas a rayas en la Figura 32d y e) se calculó mediante 1000 bootstrap de la serie analizada (log fl, log fish y var física).

5.1 Análisis wavelets

Utilizando la transformada de Morlet se aplicó los wavelets a cada uno de los transectos (para noche y día). Los wavelets fueron aplicados a la variable física, log fluid like y log fish.

Se realizó una revisión de los resultados de los wavelets para cada transecto (datos de noche), llegándose a observar para la variable física noche estructuras a partir de los 3 km y en otros transectos estructuras del tamaño de 25 km (anexo 3). Estas estructuras se desarrollaron en diferentes ubicaciones del transecto (ver espectro local de la potencia de anexo 3). De igual forma se observó en la variable física día resultados semejantes a los encontrados para la variable física noche.

Para el log fluid like noche se observó estructuras desde muy pequeños tamaños (500 m) observado en transectos de log fluid like noche (Figuras fl anexos noche). También se observó

estructuras del tamaño de 20 km en otros transectos del log fluid like noche. Estos resultados nuevamente se observó en día, con estructuras del tamaño de 2 km en unos transectos del fluid like día y de 20 km en otros transectos de fluid like día.

Los resultados del análisis wavelets para el log fish noche mostraron que existen diferentes tamaños de estructura presentes en la variable. Variando desde los 2 km hasta los 20 km aproximadamente (ver espectro local de la potencia de wavelets para fish en anexo 3). Se encontró que las mayores estructuras (de 10 a 20 km) se dieron mayormente para los datos de noche. Para los datos de día se observó mayormente estructuras del tamaño de 6 km aproximadamente (anexo 3).

5.2 Análisis del Wavelets cruzado.

El análisis de wavelets cruzado entre el log fl y la variable física reveló que la asociación entre estas dos variables varió para cada transecto. También no se evidencio una asociación significativa entre las dos variables a lo largo de todo el transecto. En la Tabla 22 se presenta un resumen de las frecuencias de todos los transectos (noche) de la asociación entre el log fl y la variable física. Se observó que existe una mayor asociación entre las dos variables del tamaño de 8 km, estos resultados fueron similares tanto de día como de noche (Figura 37c).

Analizando el wavelets cruzado para el log fish y la variable física se observó que la asociación entre estas se desarrolla principalmente entre 8 y 16 km en los datos de noche (Tabla 22) y hasta 8 km para los datos de día (Tabla 23). Y la asociación entre el log fl y el log fish rebelaron que ocurre una mayor asociación hasta los 8 km (Tabla 22 y Tabla 23), resultados similares se encontraron entre los datos de día y de noche. Esta asociación no se desarrolló continuamente a los largo de todo el transecto (Figura 37c y 38c).

Tabla 22. Tabla resumen de las frecuencias de estructuras observadas en el wavelets cruzado para datos de noche

	Var física-log fl	Var física-log fish	Log fl-log fish
0-8 km	14	7	15
8-16 km	6	11	6
16-32 km	7	9	6

Tabla 23. Tabla resumen de las frecuencias observadas en el wavelets cruzado para datos de día

	Var física-log fl	Var física-log fish	Log fl-log fish
0-8 km	23	25	30
8-16 km	8	10	5
16-32 km	6	2	2

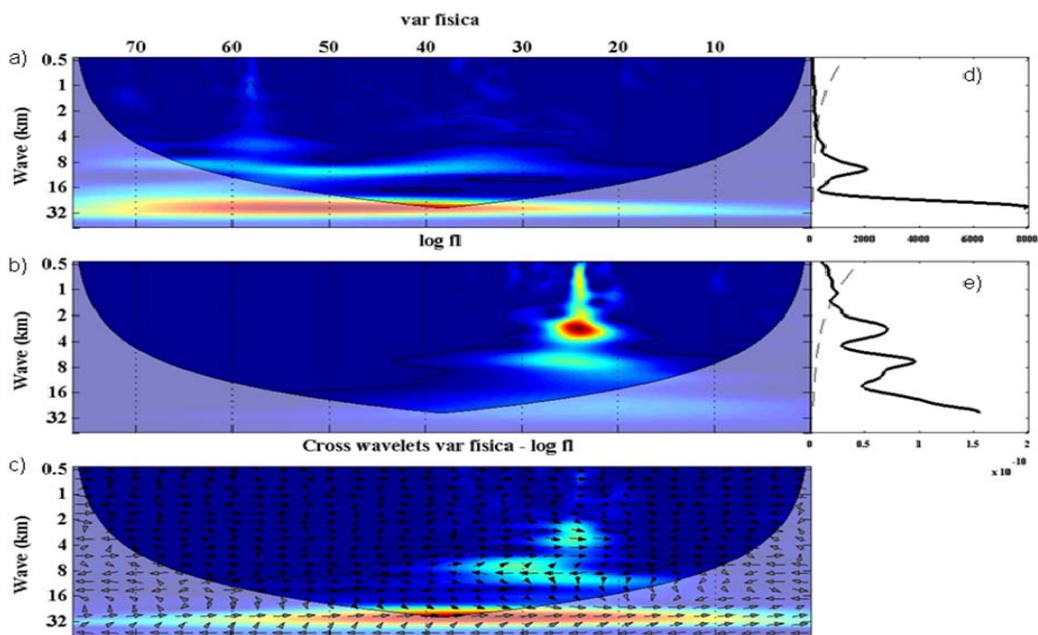


Figura 37. Análisis Wavelets para el transecto 44 noche; a) Espectro local de la potencia de wavelets para var física, b) Espectro local de la potencia de wavelets para log fl, y c) cross wavelets entre var física y log fl; d) espectro global de la potencia de wavelets para var física, e) espectro global de la potencia de wavelets para log fl.

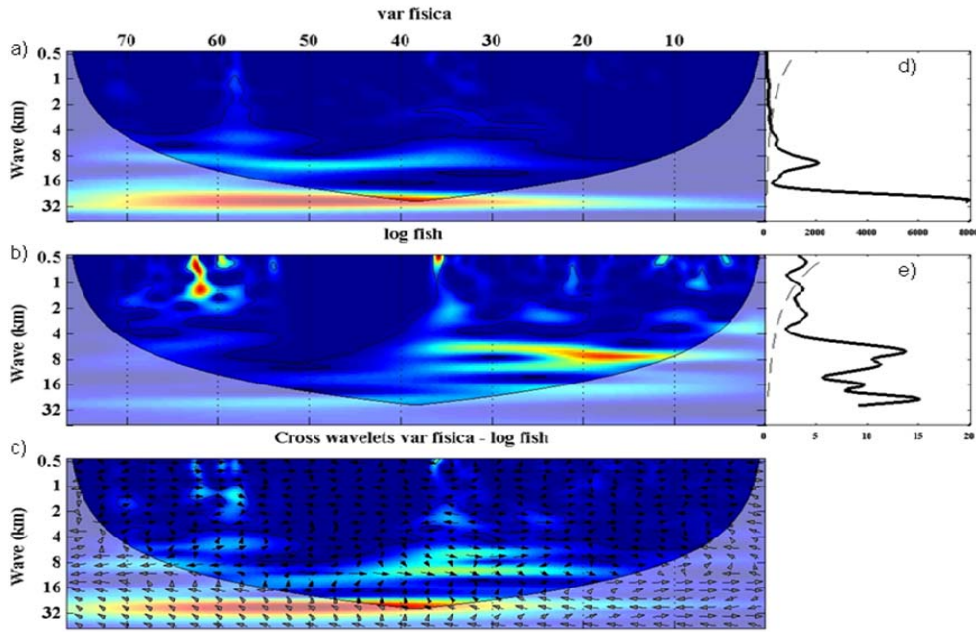


Figura 38. Análisis Wavelets para el transecto 44 noche; a) Espectro local de la potencia de wavelets para var física, b) Espectro local de la potencia de wavelets para log fish, y c) cross wavelets entre var física y log fish; d) espectro global de la potencia de wavelets para var física, e) espectro global de la potencia de wavelets para log fish. Análisis para datos de noche.

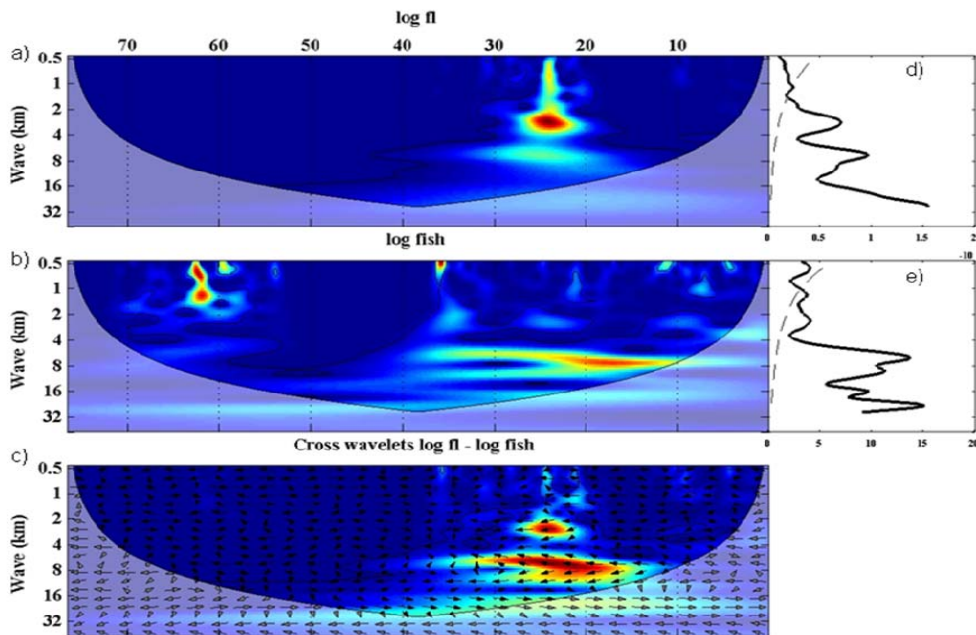


Figura 39. Análisis Wavelets para el transecto 44 noche; a) Espectro local de la potencia de wavelets para log fl, b) Espectro local de la potencia de wavelets para log fish, y c) cross wavelets entre log fl y log fish; d) espectro global de la potencia de wavelets para log fl, e) espectro global de la potencia de wavelets para log fish. Análisis para datos de noche. Análisis para datos de noche.

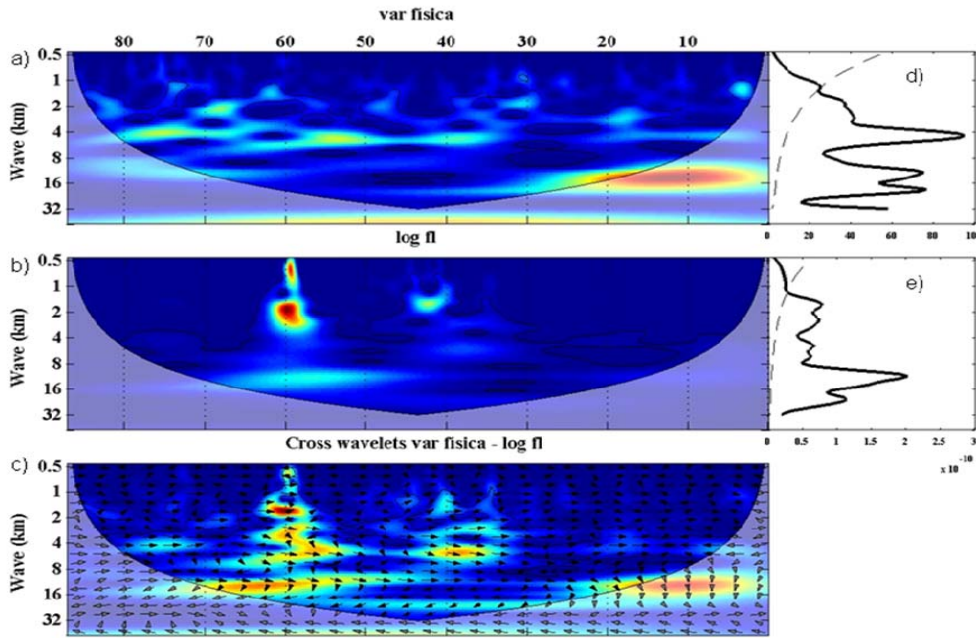


Figura 40. Análisis wavelets para el transecto 42 día; a) Espectro local de la potencia de wavelets para var física, b) Espectro local de la potencia de wavelets para log fl, y c) cross wavelets entre var física y log fl; d) espectro global de la potencia de wavelets para var física, e) espectro global de la potencia de wavelets para log fl. Análisis para datos de día.

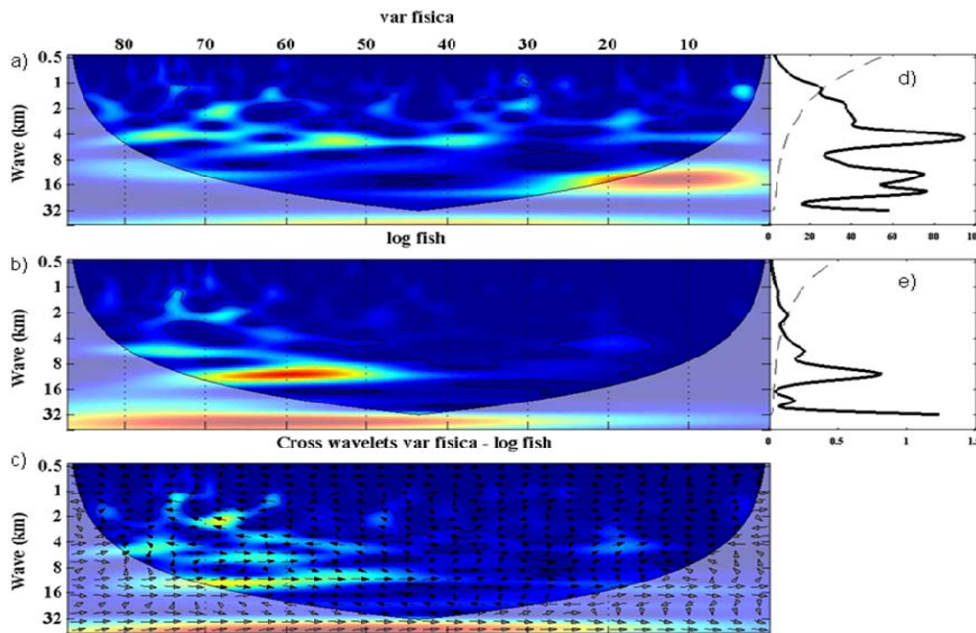


Figura 41. Análisis wavelets para el transecto 42 día; a) Espectro local de la potencia de wavelets para var física, b) Espectro local de la potencia de wavelets para log fish, y c) cross wavelets entre var física y log fish; d) espectro global de la potencia de wavelets para var física, e) espectro global de la potencia de wavelets para log fish. Análisis para datos de día.

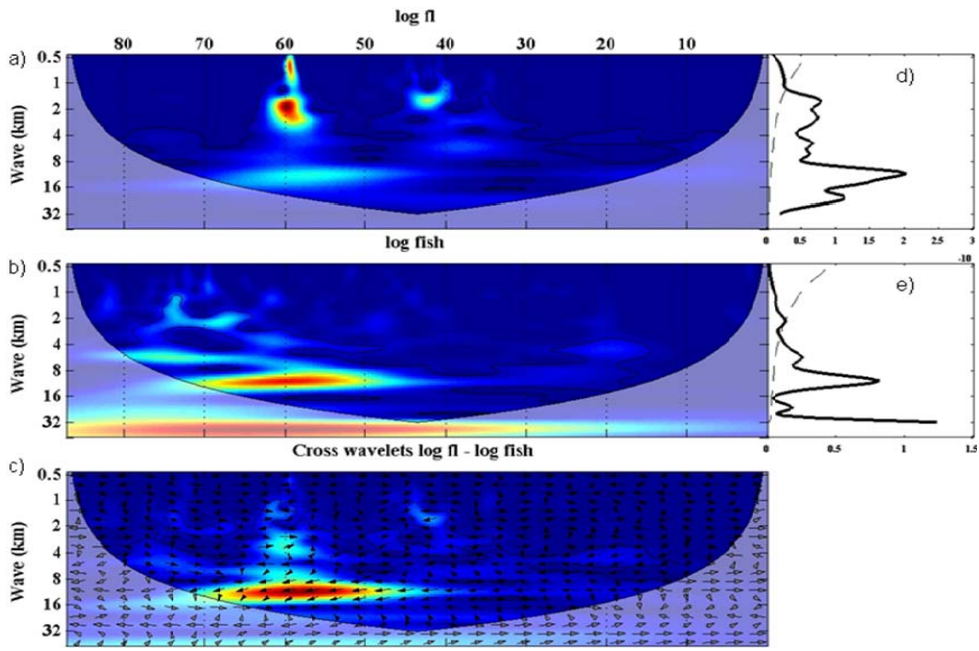


Figura 42. Análisis wavelets para el transecto 42 día; a) Espectro local de la potencia de wavelets para log fl, b) Espectro local de la potencia de wavelets para log fish, y c) cross wavelets entre log fl y log fish; d) espectro global de la potencia de wavelets para log fl, e) espectro global de la potencia de wavelets para log fish. Análisis para datos de día

CAPITULO IV:

DISCUSIÓN

1 De las metodologías

Las metodologías estudiadas en esta sección nos permiten un análisis del comportamiento de los procesos ecológicos a través el espacio. La Tabla 24 presenta una comparación entre los métodos.

Tabla 24. Descripción de las metodologías utilizadas.

	Correlación con ajuste de tamaño de muestra	Geoestadística	PCNM	Wavelets
Objetivos	Estudiar la correlación lineal entre dos variables	Describir y modelar la estructura espacial	- Describir la estructura espacial dominante en un proceso. - Describir la relación espacial de un proceso con covariables	- Descripción de la estructura espacial del proceso - Describir la relación entre dos procesos a cada lugar del espacio
Supuestos	Estacionariedad	Estacionariedad	Estacionariedad	No estacionariedad
Variables	Continuas	Continuas	Continuas	Continuas/discretas
Estructura Espacial	Autocorrelación	Variograma robusto	Variogramas	Espectro global/local de wavelets
Interacción entre dos variables	Si	Si - variograma cruzado	Si - regresión	Si - wavelets cruzado
Interacción a más de dos variables	No	No	Si - regresión	No
Tipos de datos	Unidimensional	Bidimensional	Bidimensional	Unidimensional
Muestreo	Regular	Regular-irregular	Regular-irregular	Regular
Ventajas	Facilidad de cálculo	Descompone la varianza en diferentes escalas espaciales	Analiza un gran rango de escalas espaciales (Brind'Amour et al., 2005)	Descomposición espacializada de la variabilidad de la variable. No requiere variables estacionarias
Desventajas	No obtenemos un valor espacializado de la relación entre las dos variable. Análisis unidimensional	Pérdida de la información a distancias menores a h	Subjetividad en la elección de los submodelos.	Análisis unidimensional de la serie de datos (análisis por transectos en nuestro caso)

La correlación lineal es muy usada por su simplicidad en la interpretación y facilidad del cálculo. Este método es inadecuado aplicar a series temporales o espaciales cuando hay presencia de autocorrelación entre los datos. Para solucionar este problema hemos usado el ajuste de Clifford et al. (1989) para el tamaño de muestra efectivo. Este ajuste está basado en el cálculo previo de las autocorrelaciones para cada serie de datos. Estas autocorrelaciones son calculadas para datos igualmente espaciados, por ello es necesario tener datos distribuidos unidimensionalmente. Otra ventaja de la correlación con ajuste del tamaño de muestra ocurre en el poco tiempo computacional que requiere el método. Sin embargo, como se observó en la ecuación 4, el cálculo del tamaño de muestra ajustado depende de las autocorrelaciones (de las dos variables a correlacionar). Entonces cuando la autocorrelación no disminuye con la distancia, la corrección del tamaño de muestra no es fiable (Dale y Fortin, 2009). El coeficiente de correlación calculado es un estadístico global de la asociación de las variables en estudio (Clifford et al., 1989). Por ejemplo, si existe una relación negativa a pequeñas distancias y una asociación positiva a grandes distancias, el resultado del coeficiente de correlación podría dar como resultado global una asociación positiva o negativa dependiendo de la estructura de la matriz de datos. Por tal razón en el presente trabajo se separaron los efectos a pequeñas distancias y los efectos a grandes distancias.

Los métodos geoestadísticos permiten cuantificar la variación espacial mediante el variograma. En este trabajo hemos utilizado el variograma robusto propuesto por Cressie y Hawkins (1980). Este variograma robusto no es afectado por la presencia de observaciones atípicas en la serie de datos. Un elemento importante para calcular un variograma (y del variograma cruzado) es la elección del tamaño de h (distancia de separación entre observaciones) ya que de acuerdo a este tamaño se cuantificará la variación espacial presente en la variable. Siendo así que las relaciones que se dan a distancias inferiores a h no son cuantificadas. En nuestro caso se eligió un $h=1$ km motivo por el cual no se pudo estudiar las relaciones que ocurrían a menos de 1 km. Nosotros hemos utilizado diferentes modelos para ajustar las estructuras presentes en los variogramas calculados. Esta sucesión de ajustes de modelos se le llama variograma anidado. Una debilidad de los variogramas anidados radica en la elección del número de modelos a utilizar (Bellier et al., 2007). La elección del número de modelos a utilizar (y por ende el número de estructuras

espaciales) fue realizada observando el variograma de los datos. Llegándose a observar tres estructuras. La elección del número de modelos es muy importante pues a partir de estos modelos se explicará el comportamiento de las variables en estudio. Así como la interacción de estas variables con algunas covariables como es el caso de la variable física.

La metodología PCNM propuesta por Borcard y Legendre (2002) viene siendo utilizada satisfactoriamente para describir la estructura espacial de las comunidades de plantas (Jones et al., 2008; Bellier et al., 2007) y de las comunidades de peces marinos (Brind'Amour et al., 2005). Una debilidad (hasta el momento) de los PCNM se presenta en la elección de las estructuras (submodelos). Por el momento, la elección de las estructuras es de manera subjetiva. Para la elección de los submodelos nosotros hemos aplicado la propuesta de Bellier et al. (2007), quienes proponen ajustar variogramas para cada PCNM variable, luego identificar el rango en cada variograma y con ayuda gráfica de los rangos elegir el número de submodelos. Sin embargo esta propuesta sigue siendo subjetiva. El estudio de los submodelos permite cuantificar la relación entre una variable explicativa y una variable respuesta, así como la significancia de la variable explicativa. Los futuros trabajos relacionados a los métodos PCNM deben de estar direccionados a desarrollar técnicas para poder separar estos submodelos. Nosotros creemos que este método es muy útil, pues descompone la variación espacial en un conjunto de variables ortogonales (variables no correlacionadas) los que permiten estudiar a diferentes estructuras la variable de interés.

En principio los métodos wavelets fueron desarrollados para el análisis de señales (en el campo de la física), teniendo hasta el momento pocas aplicaciones al campo de la ecología (Ménard et al., 2007). La ventaja de los métodos wavelets radica en la descomposición espacializada de la variabilidad presente en la variable de estudio. Con esta descomposición se puede describir el lugar de la ocurrencia (de estructuras) dentro de la serie. Un inconveniente de este método radica en la necesidad de realizar el estudio para series unidimensionales. En nuestro caso, ante la necesidad de hacer análisis por transectos solamente se pudo detectar estructuras pequeñas (hasta 20 km aproximadamente). El hecho de realizar un estudio por transectos nos permitió hacer, mediante los análisis wavelets, una descripción detallada de lo que ocurre dentro de cada

transecto, observándose señales desde los 2 km y ubicados en diferentes posiciones dentro de los transectos.

Una diferencia entre una serie temporal y una serie espacial es que para el primero existe una autocorrelación unidireccional y observaciones a intervalos iguales (Dale y Fortín, 2009) mientras para el segundo, la autocorrelación se desarrolla en todas las direcciones. Es así que los métodos desarrollados inicialmente para series temporales presentan deficiencias cuando se aplican a series espaciales.

Se ha podido observar que los métodos wavelets proporcionan información de la ubicación en el espacio y del tamaño de las estructuras del fenómeno en estudio. Pero desde el momento en que se hace la semejanza con la escala temporal, tenemos la necesidad de hacer nuestro estudio de manera unidimensional. Esto conlleva a perder información del tipo espacial. Ante tal situación es importante buscar una alternativa para poder utilizar los métodos wavelets de manera bidimensional. Una aproximación de este tipo fue hecho por Álvarez y Sansó (2008), que desde el enfoque bayesiano realizan una regresión wavelets para estimar las reservas de petróleo. Creemos que es importante tratar de implementar este enfoque wavelets a las investigaciones de ecología.

Al igual que la propuesta de Ávares y Sansó (2008), desde la estadística bayesiana surge una nueva metodología desarrollada en los últimos 10 años. Esta metodología nos permite evaluar el cambio de un proceso (físico) a través del tiempo y espacio. Este método se basa en implementar modelos dinámicos espacio-temporales, los cuales vienen siendo aplicados con mucho éxito en el campo de la climatología y la epidemiología (Ej. Stroud et al., 2001; Nobre et al., 2005; Fuentes et al., 2006; Chen et al., 2006). Dichos modelos toman en consideración los fenómenos estacionales que ocurren a través del tiempo y los fenómenos de estructura espacial que ocurren en las series espaciales, teniendo como principal ventaja el considerar en sus metodologías la correlación que ocurre entre el espacio y tiempo. En nuestro caso no fue posible trabajar con ésta metodología por la carencia de tener datos distribuidas tanto espacial como temporalmente (a lo largo del tiempo).

Existen otras metodologías que incorporan la autocorrelación (Anexo 3), pero nosotros no hemos utilizado dichas metodologías porque evalúan globalmente la asociación entre las series estudiadas (para el caso de los modelos SAR y test de mantel), y la razón por la que no se usó el test parcial de mantel fue porque la interpretación de este test no es fácil (Legendre y Legendre, 1998).

En este trabajo se realizó una comparación teórica de los métodos utilizados, No se pudo realizar una comparación cuantitativa de los mismos pues no se identificó una prueba estadística adecuada para cuantificar la eficiencia de los métodos utilizados. Una razón importante para no poder tener esta prueba es que los métodos de la correlación como también los métodos wavelets fueron usados para cada transecto y a diferencia de los métodos geoestadísticos y wavelets que se aplicaron para todos los transectos (el conjunto de datos). Para hacer comparables los cuatro métodos había la posibilidad de aplicar los cuatro métodos a cada transecto y así poder obtener una prueba estadística de eficiencia de los métodos, nosotros no hemos aplicado esa propuesta porque un objetivo en esta investigación fue identificar las estructuras espaciales que ocurren entre los procesos físicos (ZMO) y biológicos (peces y zooplancton), Entonces si nos limitáramos a hacer una comparación transecto por transecto perdíamos información proporcionado por procesos espaciales. Cabe mencionar que es interesante hacer una comparación de estos métodos de una manera cuantitativa, es así que queda abierta la posibilidad de futuros trabajos para cuantificar la eficiencia de estos métodos.

2 Identificación de las estructuras espaciales

A partir de esta sección nos referiremos a la variable física como la zona mínima de oxígeno (ZMO), al log fluid like como zooplancton y al log fish simplemente como fish.

Los cuatro métodos mostraron estructuras para la ZMO de, aproximadamente, 500 m hasta 30 km (a excepción de los PCNM que mostraron estructuras desde los 500 m hasta los 120 km). Esto quiere decir que existen estructuras de diferentes tamaños dentro de estos límites (500 m, 2 km, 5 km, 30 km como ejemplos). Dentro de este intervalo (500 m a 120 km) se observó la ocurrencia de

tres rangos de estructuras resaltantes, siendo la pequeña desde los 500 m hasta los 30 km, la mediana de los 30 km hasta los 60 km y la grande que abarca de los 60 km a los 120 km. Estas estructuras observadas corresponden a procesos (física) de sub-mesoescala (ejemplo ondas internas y pequeños remolinos de 100s m a kms) y de mesoescala (ejemplo frentes, remolinos 10s km) (Haury et al., 1978; Bertrand et al., 2008) (Figura 43). Mediante los análisis wavelets se observó que estas estructuras ocurren en diferentes ubicaciones dentro de cada transecto, no habiendo diferencia en ellos durante el periodo día y noche. Este ultimo resultado es similar a lo encontrado por Ciales et al. (2008), quienes, mediante las pruebas de *U de Mann – Whitney*, no encontraron diferencias entre los datos de día y de noche en la profundidad de la Zona Mínima de Oxígeno.

De igual forma para el zooplancton se observó la existencia de estructuras entre los 500 m hasta los 120 km, identificándose tres rangos de estructuras dentro de estas. La pequeña desde los 500 m hasta los 25 km, seguidos de otra desde los 25 km hasta los 60 km y finalmente la mayor estructura que corresponde desde los 60 hasta los 120 km aproximadamente. Se observó variaciones del tamaño de las estructuras (menores a los 25 km) para los datos de zooplancton de día y para los de noche, esta diferencia puede deberse a la migración vertical del zooplancton en los periodos nictimerales (Escribano et al., 2009; Bertrand et al., 2008) y al tamaño de los parches que estos forman dentro de estos periodos.

En el caso de los datos de fish, se observó nuevamente la presencia de tres estructuras. La pequeña estructura hasta los 25 km, la mediana estructura de los 25 km hasta los 70 km, y la estructura mayor que abarcó desde los 70 hasta los 130 km. Existiendo variación según el ciclo nictimeral (día o noche) en la pequeña estructura. Estas estructuras corresponden a escalas espaciales descritas en la literatura. En efecto se sabe que los peces se distribuyen en cardumenes (metros a 10s m), 'clusters' de cardumenes (kms), clusters de clusters (10s km), y poblaciones (100s km) (Gerlotto y Paramo, 2003; Fréon et al., 2005; Bertrand et al., 2008).

En síntesis; cualquier sea el parámetro analizado las grandes estructuras corresponden a procesos de mesoescala (10s km – 100s km) y la pequeña estructura (500 m hasta los 20 o 30 km) corresponde a la presencia de la sub-mesoescala (Figura 43).

3 Relación entre Zooplancton y Zona Mínima de Oxígeno (ZMO)

Los resultados evidenciaron que existe una relación negativa entre el zooplancton y la profundidad de la ZMO. Esto ocurrió en las dos escalas de estudio (mesoescala y sub-mesoescala). Esto significa que hay mayor abundancia de zooplancton en las zonas más profundas (la profundidad ZMO se tomó con signo negativo).

Mediante los wavelets en la sub-mesoescala se pudo observar diferentes tamaños de escalas, donde el proceso físico (ondas internas) concentra el zooplancton en parches.

Estos resultados son coherentes con la teoría, en efecto una mayor profundidad de la ZMO refleja la presencia de una onda interna (500 m a kms), de un remolino anticiclónico de sub-mesoescala (kms) o de mesoescala (10s km). Estas estructuras son zonas de convergencia de las corrientes que acumulan las partículas (Bakun, 1996; Farmer y Armi 1999, Haury *et al.* 1978; Pineda 1999; Franks 2005). Estos resultados evidencian que el parámetro físico que hemos usado, la profundidad de la ZMO, es un buen proxy de los procesos físicos y que la distribución del zooplancton está determinada por la estructuración física a lo largo de un gran rango espacial.

4 Relación entre fish con covariables el zooplancton y la ZMO

Los tres métodos mostraron una relación positiva y significativa (en la gran mayoría de los casos) entre la abundancia de peces y la disponibilidad en presas (zooplancton). Este resultado no dependió de la escala, sin embargo, parece que la relación se fortalece con la escala.

Contradictoriamente con la teoría (Rose y Leggett, 1990), no se observó relación negativa entre la presencia de predadores y presas a pequeña escala. Sin embargo cabe recordar que (a la excepción de los análisis wavelet) no se estudio relaciones a escala inferior a 1 km donde se espera relaciones potencialmente negativas. También se observa que en la sub-mesoescala la

relación predador-presa no es significativa para los datos de día, esto es debido a la migración vertical que realiza el zooplancton en la ZMO, el zooplancton utiliza la ZMO como refugio ante sus predadores (Escribano et al., 2009).

Las relaciones positivas observadas confirman el rol del zooplancton en la ecología de la anchoveta (Espinoza y Bertrand, 2008). Los resultados enseñan que el patrón de distribución de la anchoveta depende de la de sus presas.

La relación entre profundidad de la ZMO y peces es negativa (de noche) tal como lo observado en el caso del zooplancton pero de manera más débil. Este resultado es lógico. En efecto los procesos físicos determinan la distribución de los organismos pasivos. Luego los predadores buscan las agregaciones de presas (Figura 44) pero la relación entre la física y los peces es indirecta en este caso (la relación se hace a través de un nivel intermediario, el zooplancton).

También se ha detectado que la relación era más baja de día que de noche. Este resultado puede estar relacionado a las migraciones nictemerales de los organismos. En efecto durante el día gran parte del zooplancton migra debajo de la oxiclina y no está más accesible a su predador (Bertrand et al., 2008; Escribano et al., 2009). Las estructuras físicas superficiales siguen concentrando el zooplancton que se queda en la superficie y forman parches que son menos poblados ya que la mayoría de la biomasa se distribuye en la profundidad (bajo la ZMO). Parches superficiales pocos densos son menos atractivos para la anchoveta lo que explicaría la baja relación (Bertrand et al., 2008).

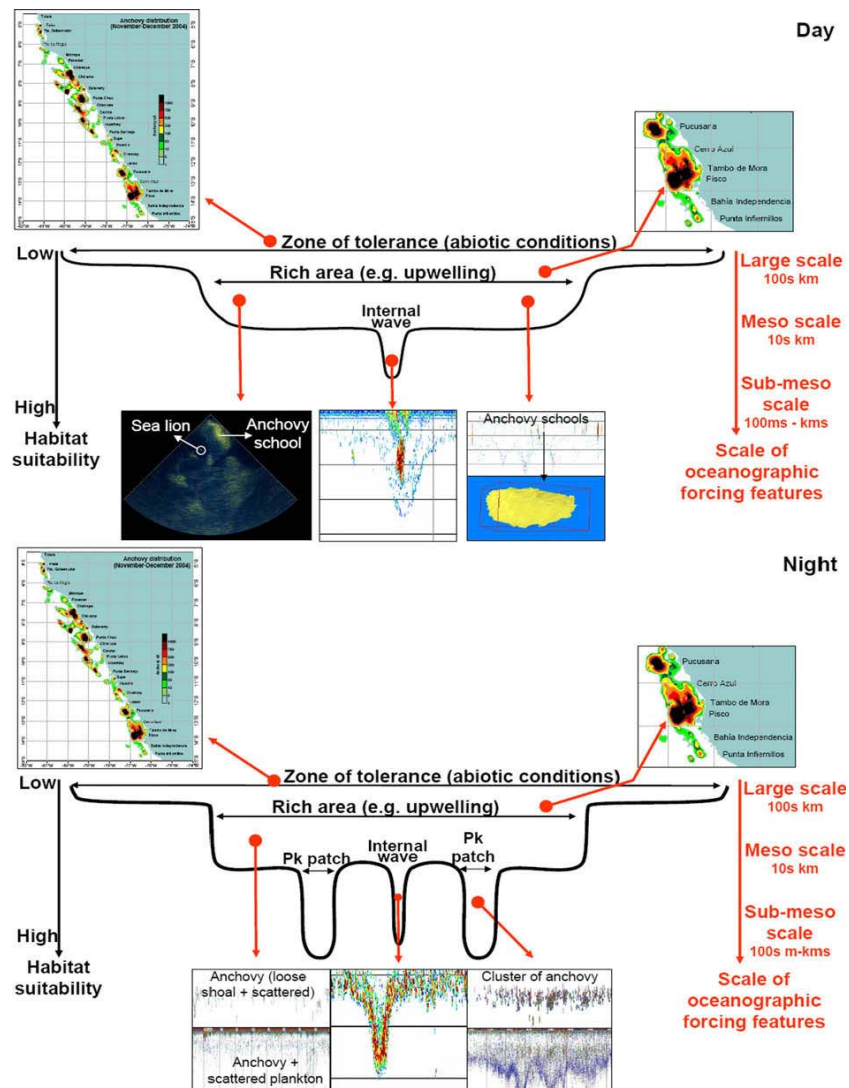


Figura 43. Síntesis de la variación día – noche de la clase de hábitat anchoveta a través de las escalas y sus consecuencias sobre la distribución espacial 3D de la anchoveta, adaptado de 'habitat-based basin model framework' de MacCall (1990). La profundidad del cardumen aumenta con las condiciones oceanográficas (p.ej. la temperatura, la profundidad de la ZMO) y biótico (p.ej. la depredación, la abundancia de presa, y la organización espacial). Dentro de esta zona general, la calidad del hábitat aumenta en áreas ricas en presa que rodea una zona de upwelling por ejemplo. Dentro de las áreas ricas, estructuras de sub-mesoscala (p. ej. ondas internas solitarias, remolinos) pueden concentrar a la presa, la calidad de hábitat aumenta mucho más. Durante el día (Figura superior), la profundidad del cardumen es más baja que durante la noche ya que la presa es menos accesible (una parte del plancton ha migrado debajo de la ZMO) y la depredación por depredadores superiores (p.ej. las aves marinas) son más altas. Anchoveta forma cardúmenes lo que es atacado por depredadores. Durante la noche (figura inferior) la profundidad de la cuenca aumenta como la presa se hace más disponible y reducen la depredación. Fuente: Bertrand et al., 2008.

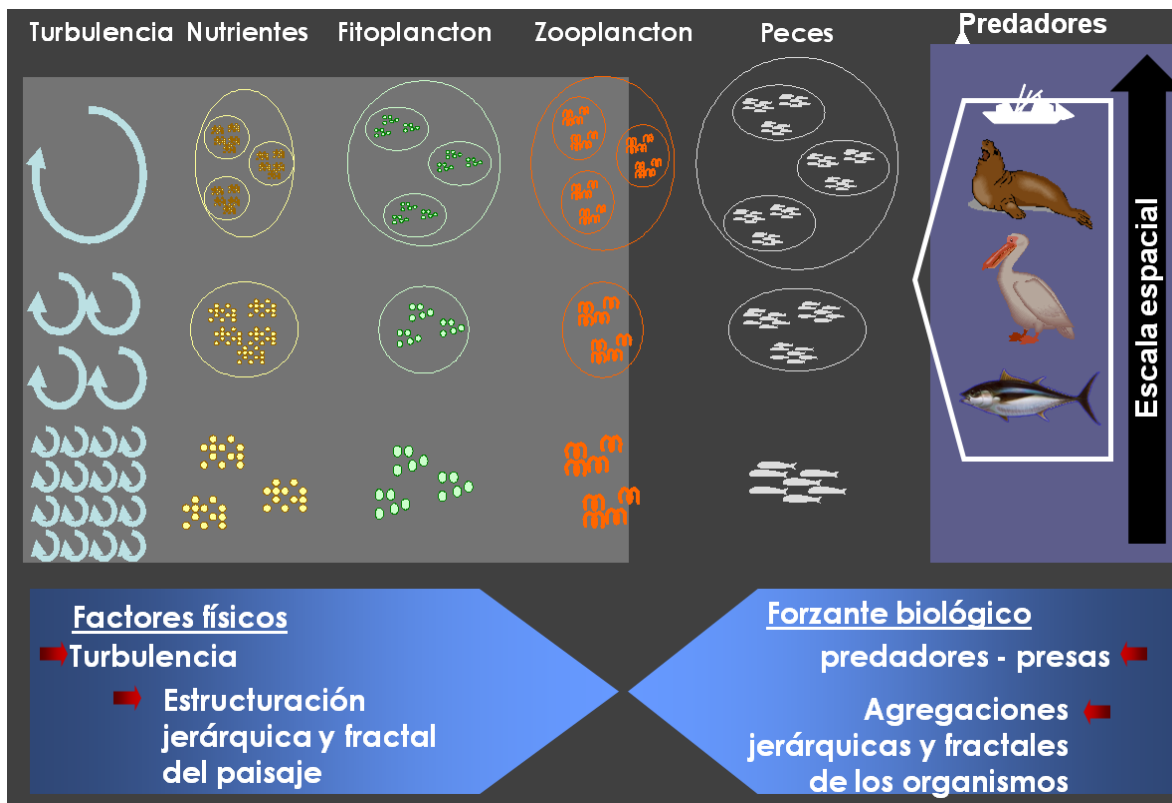


Figura 44. Representación esquemática del concepto de transferencia bottom-up de comportamiento y estructuración espacial. El forzante físico estructura a los organismos pelágicos por la turbulencia en las masas de agua. La disipación de esta turbulencia es fractal en la naturaleza y genera una estructuración jerárquica de las masas de agua. Las partículas inertes (nutrientes) y parte de organismos vivos (el fitoplancton y la mayor parte del zooplancton) pasivamente son organizadas en el espacio por este forzante físico. Entonces, las interacciones biológicas como relaciones de predador-presa ocurren hasta cierto punto de esta estructuración espacial a lo largo de la cadena de trófica. Fuente: Bertrand, S. et al. (2008).

CAPITULO IV:
CONCLUSIONES

En este trabajo hemos presentado la aplicación de 4 metodologías estadísticas que nos permiten identificar las estructuras espaciales dentro de los procesos (físicos, biológicos). También nos permite cuantificar las relaciones existentes entre estos procesos. Los cuatro métodos son: la correlación con ajuste del tamaño de muestra, la geoestadística, los PCNM y los wavelets.

Luego de aplicar estas metodologías nosotros observamos que:

1. con el objetivo de identificar las estructuras presentes, observamos que si el estudio es unidimensional, son los métodos wavelets los más adecuados a usar. Los wavelets son usados para datos estacionarios y no estacionarios, también nos permite identificar el rango de la estructura espacial y, además, la ubicación de esta estructura. También nos permite estudiar las correlaciones entre dos procesos.

Al contrario, si el estudio es bidimensional, son los métodos PCNM los más adecuados a utilizar. Estos métodos nos permiten estudiar los procesos a diferentes escalas, también permiten conocer que estructuras (y el tamaño de estas estructuras) son las más importantes para el proceso. Además es posible estudiar la interacción de los procesos ecológicos con los procesos medioambientales mediante regresiones multivariadas.

2. Los 4 métodos mostraron resultados similares para los procesos físicos y los procesos biológicos (zooplancton y peces). Estos resultados muestran la existencia de tres estructuras, siendo la primera estructura de los 500 m hasta los 30 km, la segunda estructura es de los 30 km hasta los 60 o 70 km y, la tercera estructura, de los 60 km o 70 km hasta los 130 km.

3. Estas tres estructuras corresponden a dos escalas espaciales ya descritas en la literatura. Estas escalas son: la sub-meso escala (100s m a kms) y la mesoescala (10s km a 100s km). En la sub-mesoescala existen diferencias entre los periodos nictimerales (día y noche) para los procesos biológicos. Estos cambios son debido a la migración vertical del zooplancton. Al contrario no se observó cambios significativos en los periodos nictimerales en los procesos de mesoescala.

5. Las relaciones observadas entre las variables son acordes a la teoría, pues se observó relaciones positivas (directas) entre la profundidad de la zona mínima de oxígeno (ZMO) y el

zooplancton. Luego también observamos una relación directa entre el zooplancton y los peces, lo cual muestra que la distribución de los peces es influenciado por la distribución del zooplancton. Por último la relación positiva entre la profundidad de la ZMO y los peces no fue de muy fuerte debido a que la distribución de los peces es forzada por la distribución del zooplancton.

Finalmente cabe mencionar que existe una necesidad latente, en el Perú, de una interacción directa entre los investigadores en estadística y los investigadores en ecología con el fin de aplicar métodos estadísticos más complejos al entendimiento de los procesos ecológicos. Si se llegan a aplicar métodos estadísticos que nos permitan valorar de mejor manera la información existente (que proviene de sistemas ecológicos), los científicos podrán comprender mejor el funcionamiento de los ecosistemas, entonces sus recomendaciones serán más adecuadas y si los responsables de tomar las decisiones escuchan estas recomendaciones podremos conservar los ecosistemas.

BIBLIOGRAFÍA

- Alheit, J. & Niquen, M., 2004. Regime shifts in the Humboldt Current ecosystem. *Progress in Oceanography*, 60: 201–222.
- Álvarez, G. & Sansó, B., 2008. Bayesian Wavelet Regression for Spatial Estimation. *Journal of Data Science*, 6: 219-229.
- Anderson, M.J. & Robinson, J., 2001. Permutation Test for Linear Models. *Australian and New Zealand Journal of Statistics*, 43(1): 75-88.
- Ayón, P., Swartzman, G., Bertrand, A., Gutiérrez, M. & Bertrand, S., 2008. Zooplankton and forage fish species off Peru: Large-scale bottom-up forcing and local-scale depletion. *Progress in Oceanography*, 79: 208-214.
- Bakun, A., 1996. Patterns in the Ocean: Ocean Processes and Marine Population Dynamics, University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico, 323 pp.
- Bellier, E., Monestiez, P., Durbec, J.-P. & Candau, J.-N., 2007. Identifying spatial relationships at multiple scales: principal coordinates of neighbour matrices (PCNM) and geostatistical approaches. *Ecography*, 30: 385-399.
- Bertrand, A. et al., 2004a. Diel vertical behaviour, predator-prey relationships, and occupation of space by jack mackerel (*Trachurus murphyi*) off Chile. *Marine Science*, 61: 1105-1112.
- Bertrand, A. et al., 2008. Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. *Progress in Oceanography*, 79: 264-277.
- Bertrand, A., Josse, E., Bach, P. & Dagorn, L., 2003. Acoustics for ecosystem research: lessons and perspectives from a scientific programme focusing on tuna-environment relationships. *Aquatic Living Resources*, 16: 197-203.
- Bertrand, A., Segura, M., Gutiérrez, M. & Vásquez, L., 2004b. From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. *Fish and Fisheries*, 5: 296-316.
- Bertrand, S., Dewitte, B., Tam, J., Díaz, E. & Bertrand, A., 2008. Impacts of Kelvin wave forcing in the Peru Humboldt Current system: Scenarios of spatial reorganizations from physics to fishers. *Progress in Oceanography*, 79: 278-289.
- Borcard, D. & Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, 153: 51-68.
- Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H., 2004. Dissecting The Spatial Structure of Ecological Data at Multiple Scales.
- Brind'Amour, A., Boisclair, D., Legendre, P. & Borcard, D., 2005. Multiscale spatial distribution of a littoral fish community in relation to environmental variables. *Oceanography*, 50(2): 465-479.
- Cazelles, B. et al., 2008. Wavelet analysis of ecological time series. *Oecologia*, 156: 287-304.
- Cazelles, B., Chavez, M., McMichael, A. & Hales, S., 2005. Nonstationary Influence of El Niño on the Synchronous Dengue Epidemics in Thailand. *PLoS medicine*, 2(4): 313-318.
- Chavez, F.P., Bertrand, A., Guevara-Carrasco, R., Soler, P. & Csirke, J., 2008. The northern Humboldt Current System: Brief history, present status and a view towards the future. *Progress in Oceanography*, 79: 95-105.
- Chen, L., Fuentes, M. & Davis, J.M., 2006. Spatial-temporal statistical modeling and prediction of environmental processes. In: O.U. Press (Editor), *Hierarchical Modelling For The Environmental Sciences : Statistical Methods And Applications*.
- Cliff, A.D. & Ord, K., 1970. Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography*, 46: 269-292.
- Cliff, A.D. & Ord, K., 1981. Distribution theory for the joint-count, I, and C statistics, *Spatial Processes, Models and Applications*., London, pp. 34-67.
- Clifford, P., Richardson, S. & Hemón, D., 1989. Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45(1): 123-134.
- Cressie, N., 1993. *Statistics for Spatial Data*. Jhon Wiley & Sons, Inc., 887 pp.
- Cressie, N. & Hawkins, D.M., 1980. Robust estimation of the variogram: I. *Mathematical Geology*, 12(2): 115-125.
- Criales-Hernández, M. et al., 2008. Zooplankton vertical distribution and migration off Central Peru in relation to the oxygen minimum layer. *Helgoland Marine Research*, 62: 85-100.
- Cueto, V., 2006. Escalas en Ecología: su importancia para estudios de la selección de hábitat de aves. *Hornero*, 21: 1-13.

- Dale, M.R.T. et al., 2002. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25: 558-577.
- Dale, M.R.T. & Fortin, M.-J., 2002. Spatial Autocorrelation and Statistical test in Ecology *Ecoscience*, 9 (2): 162-167.
- Dale, M.R.T. & Fortin, M.-J., 2009. Spatial Autocorrelation and Statistical Tests: Some Solutions. *Journal of Agricultural, Biological and Environmental Statistics*, 14: 188-206.
- Dale, M.R.T. & Mah, M., 1998. The use of wavelets for spatial pattern analysis in ecology. *Vegetation Science*, 9: 805-814.
- Diggle, P.J. & Ribeiro Jr, P.J., 2007. *Model-based Geostatistics*. Springer Series in Statistics, New York, 227 pp.
- Dormann, C.F. et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30: 609-628.
- Dray, S., Legendre, L. & Peres-Neto, P.R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 96: 483-493.
- Escribano, R., Hidalgo, P. & Krautz, C., 2009. Zooplankton associated with the oxygen minimum zone system in the northern upwelling region of Chile during March 2000. *Deep-Sea Research II*, 56: 1049-1060.
- Espinoza, P. & Bertrand, A., 2008. Revisiting Peruvian anchovy (*Engraulis ringens*) trophodynamics provides a new vision of the Humboldt Current system. *Progress in Oceanography*: 13.
- Farmer, D. & Armi, L., 1999. The Generation and Trapping of Internal Solitary Waves over Topography. *Science*, 283(5398): 188-190.
- Fortin, M.-J. & Dale, M.R.T., 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, 358 pp.
- Fortin, M.-J., Drapeau, P. & Legendre, P., 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio*, 83: 209-222.
- Franks, P.J.S., 2005. Plankton patchiness, turbulent transport and spatial spectra. *Marine Ecology Progress Series*, 294: 295-309.
- Fréon, P., Cury, P., Shannon, L. & Roy, C., 2005. Sustainable exploitation of small pelagic fish stocks challenged by environmental and ecosystem changes: A review. *Bulletin of Marine Science*, 76(2): 385-462.
- Fréon, P., Soria, M., Mullon, C. & Gerlotto, F., 1993. Diurnal variation in fish density estimate during acoustic surveys in relation to spatial distribution and avoidance reaction. *Aquatic Living Resources*, 6: 221-234.
- Fuentes, M., Kittel, T. & Nychka, D., 2006. Sensitivity of ecological models to their climate drivers: Statistical ensembles for forcing. *Ecological Applications*, 16(1): 99-116.
- Genton, M., 1998. Highly Robust Variogram Estimation. *Mathematical Geology*, 30(2): 213-221.
- Georgakarakos, S. & Kitsiou, D., 2008. Mapping abundance distribution of small pelagic species applying hydroacoustics and Co-Kriging techniques, *Hydrobiologia*, pp. 155-169.
- Gerlotto, F. & Paramo, J., 2003. The three-dimensional morphology and internal structure of clupeid schools as observed using vertical scanning multibeam sonar. *Aquatic Living Resources*, 16: 113-122.
- Giannoulaki, M. et al., 2003. Effect of coastal topography on the spatial structure of the populations of small pelagic fish. *marine Ecology Progress Series*, 265: 243-253.
- Gower, J.C., 1966. Some distance properties of Latent root and vector methods used in multivariate analysis. *Biometrika*, 53: 325-338.
- Grinsted, A., Moore, J.C. & Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11: 561-566.
- Gutiérrez, D. et al., 2008. Oxygenation episodes on the continental shelf of central Peru: Remote forcing and benthic ecosystem response. *Progress in Oceanography*, 79: 177-189.
- Hauray, L.R., McGowan, J.A. & Wiebe, P.H., 1978. Patterns and processes in the time-space scales of plankton distributions. In: J.H. Steele (Editor), *Spatial Pattern in Plankton Communities*, New York, USA, pp. 277-327.
- Jombart, T., Dray, S. & Dufour, A.B., 2009. Finding essential scales of spatial variation in ecological data: a multivariate approach. *Ecography*, 32: 161-168.
- Jones, M.M. et al., 2008. Explaining variation in tropical plant community composition: influence of environmental and spatial data quality. *Oecologia*, 155: 593-604.

- Keitt, T.H., Bjørnstad, O.N., Dixon, P.M. & Citron-Pousty, S., 2002. Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*, 25: 616-625.
- Legendre, P., Borcard, D. & Peres-Neto, P.R., 2005. Analysing beta diversity: Partitioning the spatial variation of community composition data. *Ecological Monographs*, 75(4): 435–450.
- Legendre, P. & Fortin, M.-J., 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80: 107-138.
- Legendre, P. & Legendre, L., 1998. *Numerical Ecology*. Elsevier, 870 pp.
- Legendre, P. & Troussellier, M., 1988. Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation. *Limnology and Oceanography*, 33(5): 1055-1067.
- Levin, S.A., 1992. The Problem of Pattern and Scale in Ecology. *Ecology*, 73(6): 1943-1967.
- Lichstein, J.W., 2006. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, 188: 117-131.
- Link, J., 2002. Ecological Considerations in Fisheries Management: When Does It Matter? *Fisheries*, 27: 10-17.
- Liu, Y., Liang, S. & Weisberg, R., 2007. Rectification of the Bias in the Wavelet Power Spectrum. *Journal of Atmospheric and Oceanic Technology*, 24: 2093-2102.
- López de Luzuriaga & Olano, J.M., 2006. Con los pies en el suelo: incluyendo la estructura espacial de los datos en los análisis multivariantes. *Ecosistemas*, 15 (3): 59-67.
- Mantel, N., 1967. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27: 209-220.
- Maronna, R.A., Martin, R.D. & Yohai, V.J., 2006. *Robust Statistics Theory and Methods*. John Wiley & Sons, 410 pp.
- Ménard, F., Marsac, F., Bellier, E. & Cazelles, B., 2007. Climatic oscillations and tuna catch rates in the Indian ocean: a wavelets approach to time series analysis. *Fisheries Oceanography*, 16:1: 95-104.
- Moran, P.A.P., 1948. The Interpretation of Statistical Maps. *JSTOR*, 10-2: 243-251.
- Nobre, A.A., Schmidt, A.M. & Lopes, H.F., 2005. Spatio-temporal models for mapping the incidence of malaria in Pará. *Environmetrics*, 16(3): 291 - 304.
- Pebesma, E., 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*, 30: 683-691.
- Petitgas, P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries*, 2: 231-249.
- Pikitch, E.K. et al., 2004. *Ecosystem-Based Fishery Management*.
- Pineda, J., 1999. Circulation and larval distribution in internal tidal bore warm fronts. *Limnology and Oceanography*, 44(6): 1400–1414.
- Rose, G.A. & Leggett, W.C., 1990. The Importance of Scale to Predator-Prey Spatial Correlations: An Example of Atlantic Fishes. *Ecology*, 71(1): 33-43.
- Rousseeuw, P.J., 1984. Least Median of Squares Regression. *American Statistical Association*, 79(388): 871-880.
- Rousseeuw, P.J. & Van Driessen, K., 2006. Computing LTS Regression for Large Data Sets. *Data Mining and Knowledge Discovery*, 12: 29–45.
- Simmonds, J. & MacLennan, D., 2005. *Fisheries Acoustics Theory and Practice*. Blackwell Publishing company, 429 pp.
- Sousa da Silva, A., 2006. *Modelos gaussianos geoestadísticos espacio-temporales y aplicaciones*, Universidad de Sao Paulo, Sao Paulo, 70 pp.
- Stroud, J.R., Müller, P. & Sansó, B., 2001. Dynamic models for spatiotemporal data. *Royal Statistical Society*, 63(4): 673-689.
- Torrence, C. & Compo, G.P., 1997. *A Practical Guide to Wavelet Analysis*, University of Colorado, Colorado.
- Vivar-Rojas, J.C., 2004. Uma nova classe de modelos espaço-temporais para dados de área. Tesis de Estadística Thesis, Universidade Federal do Rio de Janeiro – UFRJ, Rio de Janeiro, 100 pp.
- Wagner, H., 2003. Spatial Covariance in Plant Communities: Integrating Ordination, Geostatistics, and Variance Testing. *Ecology*, 84(4): 1045-1057.
- Webster, R. & Oliver, M., 2007. *Geostatistics for Environmental Scientists*. John Wiley & Sons Ltd, 309 pp.
- West, M. & Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics, New York, 671 pp.

Wiens, J.A., 1989. Spatial scaling in ecology, *Functional Ecology*, pp. 385-397.

Anexos

Anexo 1:
Regresión Robusta

Sea la ecuación $y_i = \beta^T x_i + \varepsilon_i$, $i = 1, \dots, n$, una regresión lineal. Frecuentemente los estimadores de β son estimados mediante la minimización de la suma de cuadrados de los errores.

Esta forma de estimar los parámetros β es fuertemente influenciada por la presencia de valores atípicos (outliers). Desde la estadística robusta surgen alternativas para estimar los parámetros (β). Esta estimación no es afectada por la presencia de outliers en la serie de datos. El estimador Least Trimmed Squares (LTS) es un método que no toma en consideración los puntos atípicos.

El estimador LTS de los parámetros está basado en la minimización de la suma de cuadrados de los residuales truncados por h . El estimador LTS de β viene dado por:

$$\beta^* = \min \sum_{i=1}^h r_{[i]}^2(\beta)$$

donde $r_{[i]}^2(\beta)$ representa el estadístico de orden i de los $r_1^2(\beta)$, $r_2^2(\beta)$, $r_3^2(\beta)$, \dots , $r_n^2(\beta)$ con $r_i(\beta) = y_i - \beta^T x_i$. Donde la constante h tiene que cumplir la condición $n/2 < h \leq n$, esta constante implica la ruptura del estimador LTS. Lo cual quiere decir que las observaciones $n-h$ residuales más grandes no afectarán al estimados LTS.

Anexo 2:

Suavización por Spline

Un spline de orden p con k nudos, t_1, \dots, t_k en el intervalo $[a,b]$ es una función que se obtiene dividiendo primero el intervalo $[a,b]$ en los subintervalos $[x_0,x_1), [x_1,x_2), \dots, [x_k,x_{k+1}]$, con $x_0=a$ y $x_{k+1}=b$ y usando luego un polinomio de grado menor o igual que p en cada uno de los subintervalos.

El suavizador por spline se obtiene minimizando

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int [f''(t)]^2 dt$$

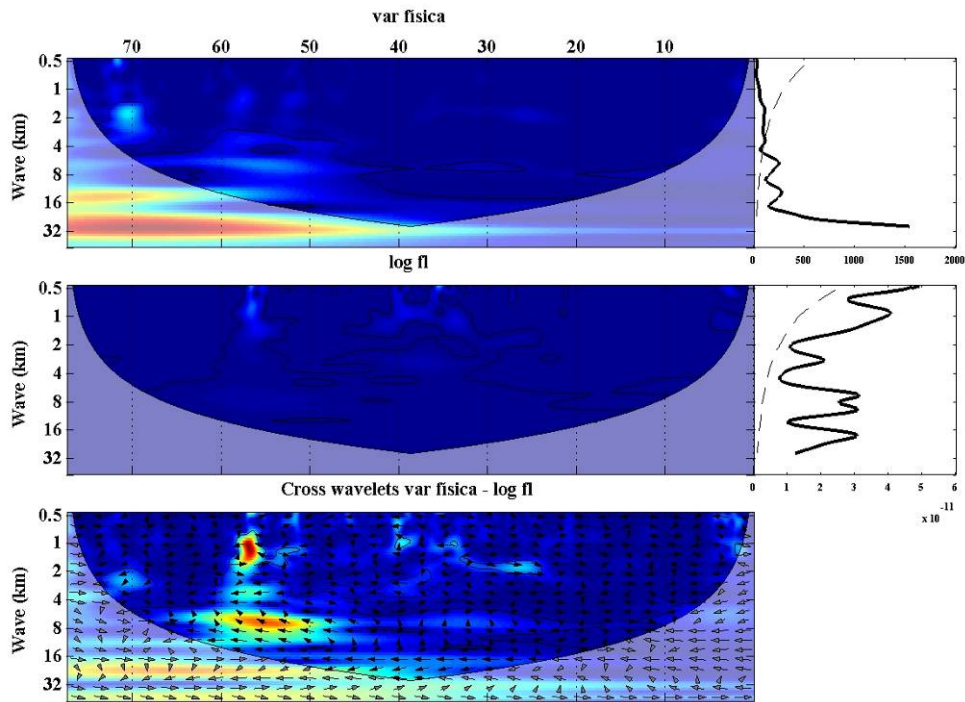
donde y y x es el par de observaciones (para $i=1, \dots, n$).

El primer término es una medida de la bondad de ajuste del modelo y el segundo término es una medida del grado de suavidad. El parámetro de suavidad λ es positivo y gobierna el intercambio entre la suavidad y la bondad de ajuste del suavizador.

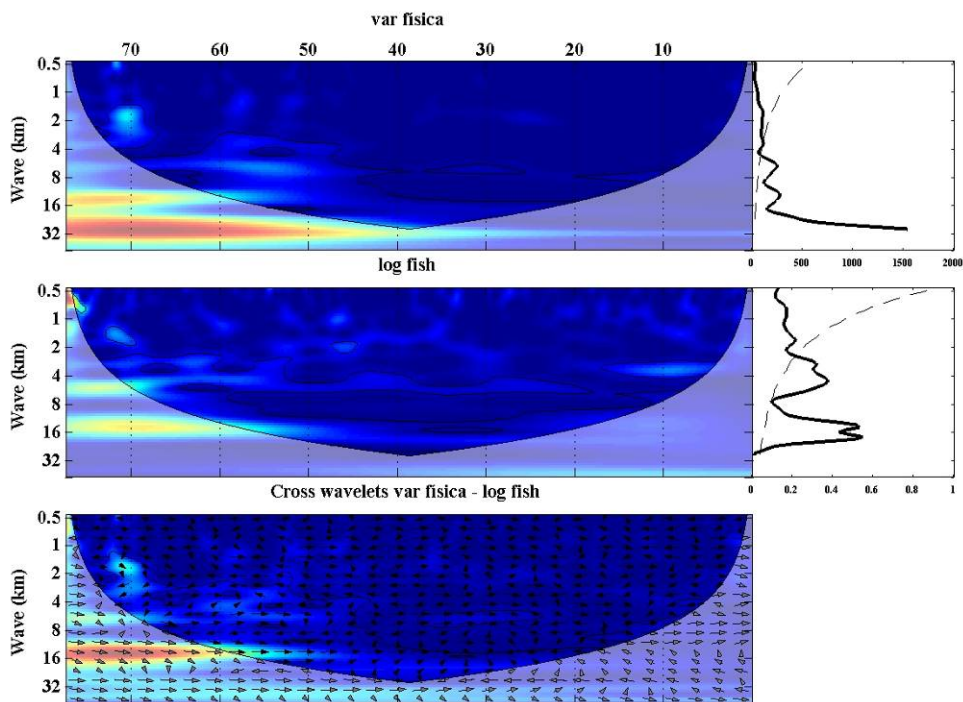
Anexo 3:
Resultados Wavelets

Anexo resultados de los wavelets

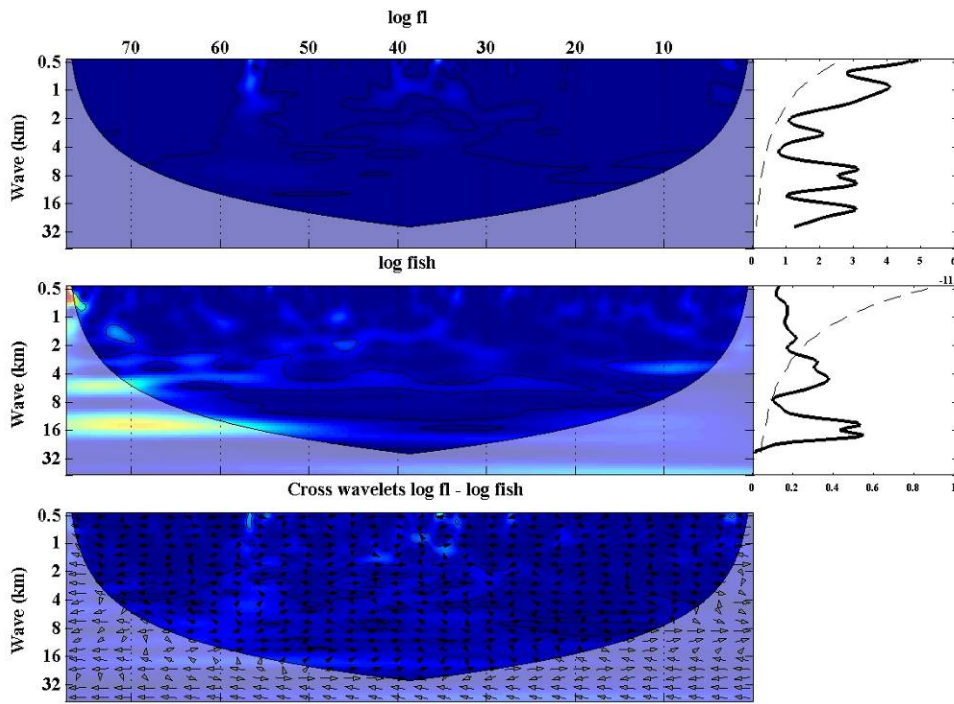
Transecto 22 noche. Análisis wavelets para variable física, log fluid like



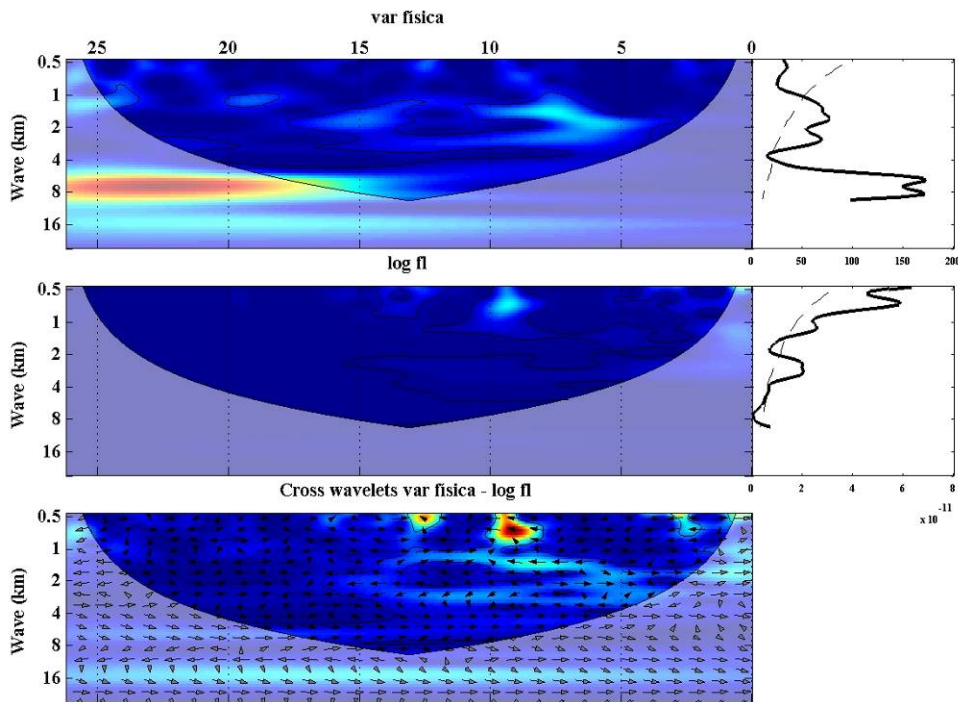
Análisis wavelets para variable física, log fish



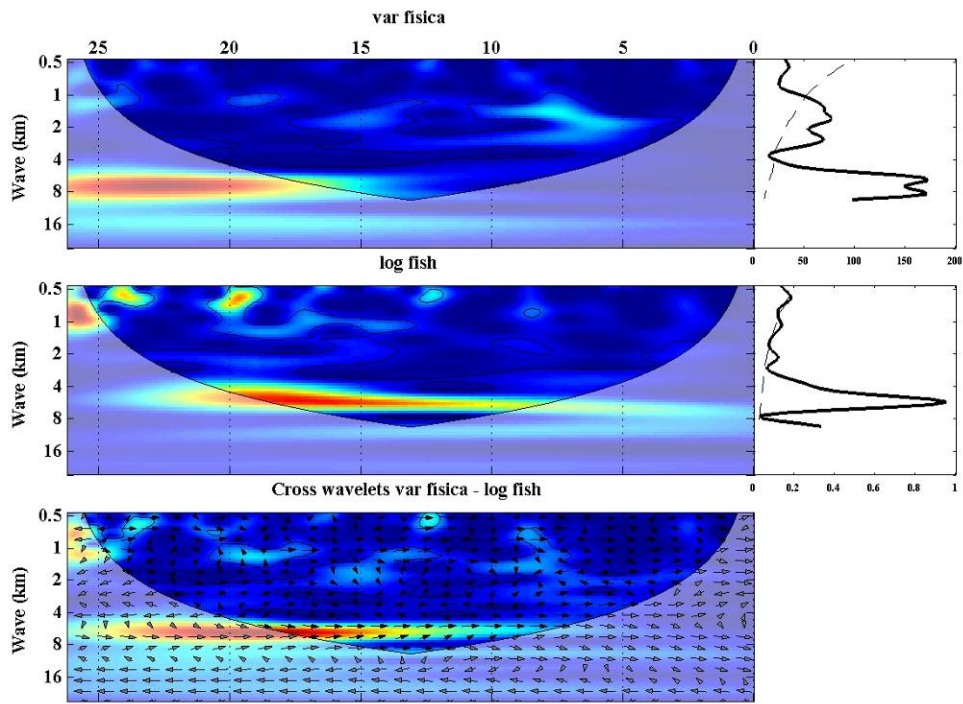
Análisis wavelets para log fluid like y log fish



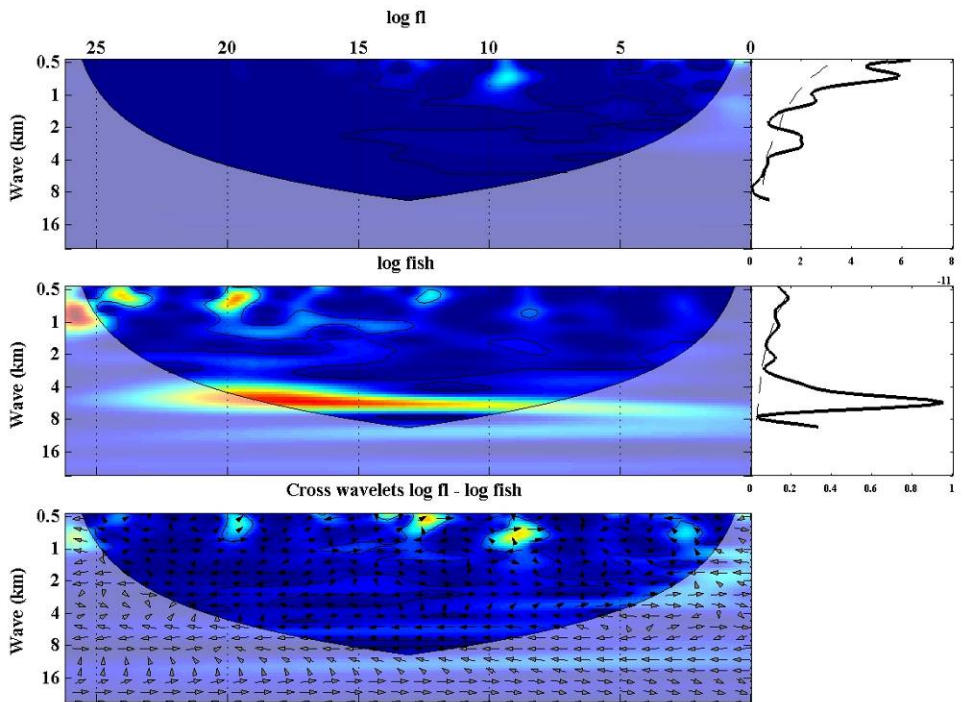
Transecto 24 noche. Análisis wavelets para variable fisica, log fluid like



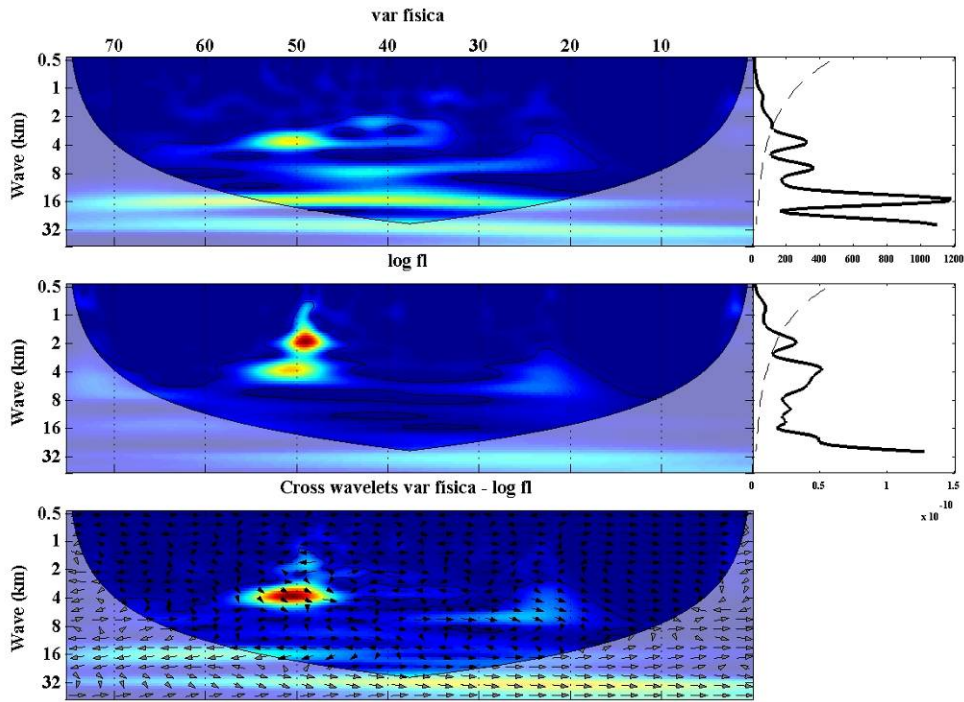
Análisis wavelets para variable física, log fish



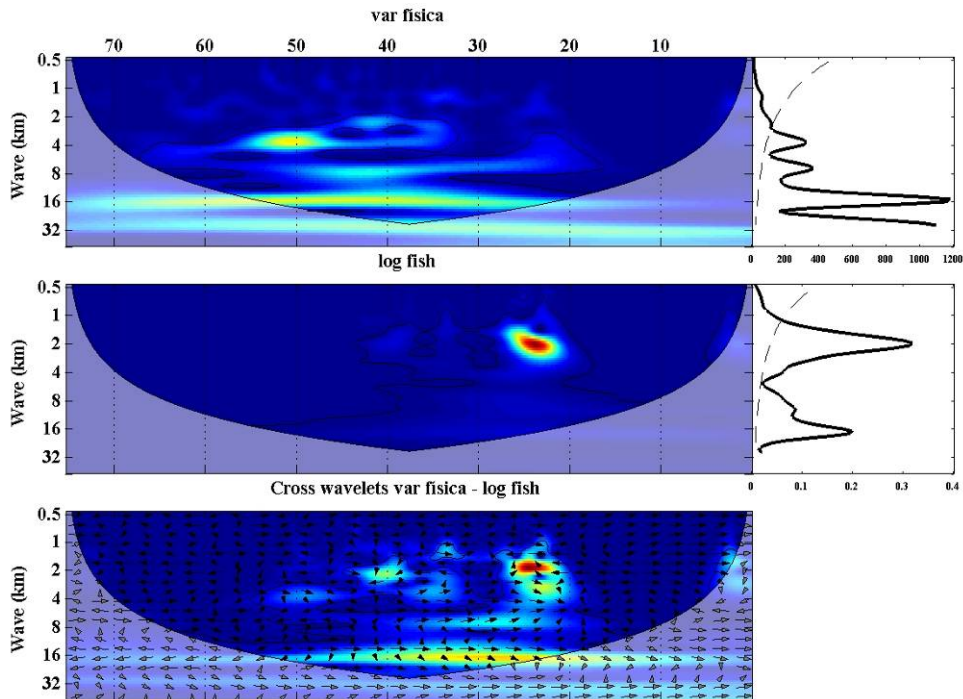
Análisis wavelets para log fluid like y log fish



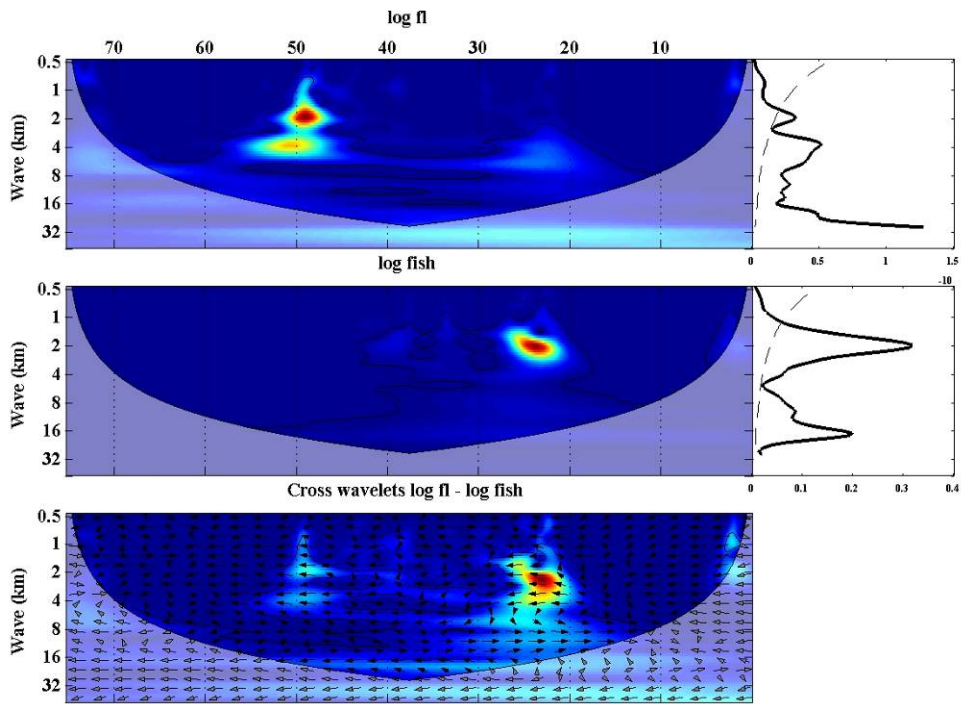
Transecto 14 día. Análisis wavelets para variable física, log fluid like



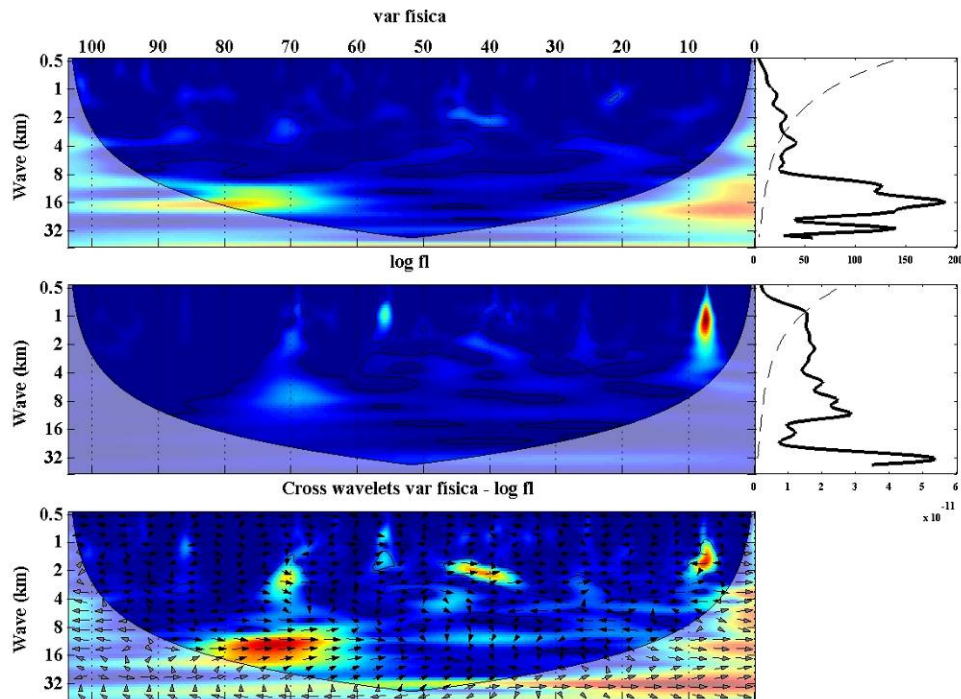
Análisis wavelets para variable física, log fish



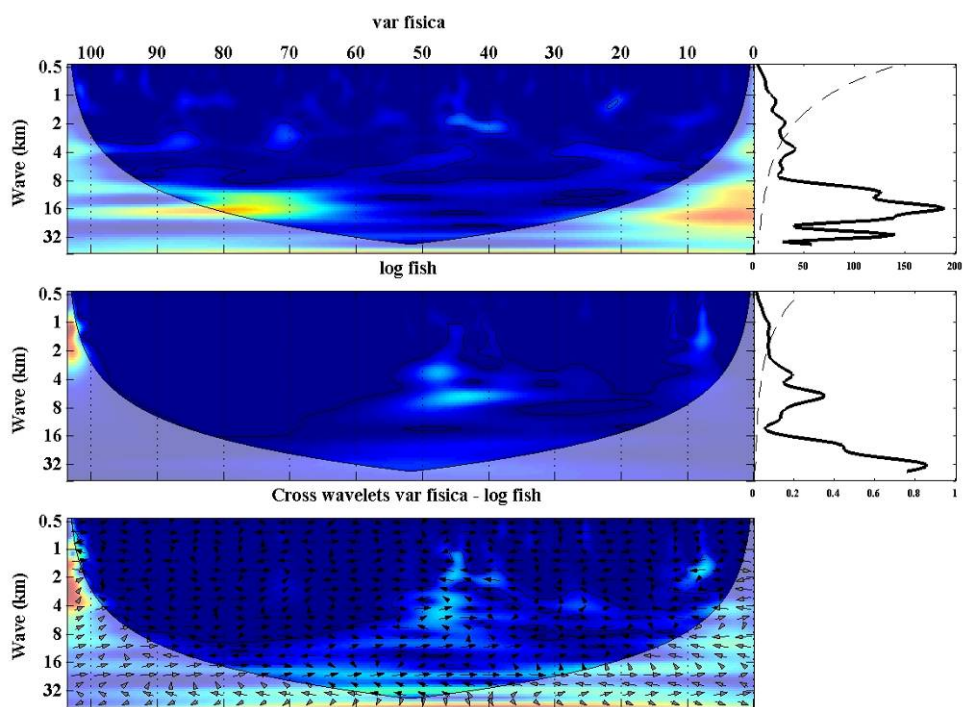
Análisis wavelets para log fluid like y log fish



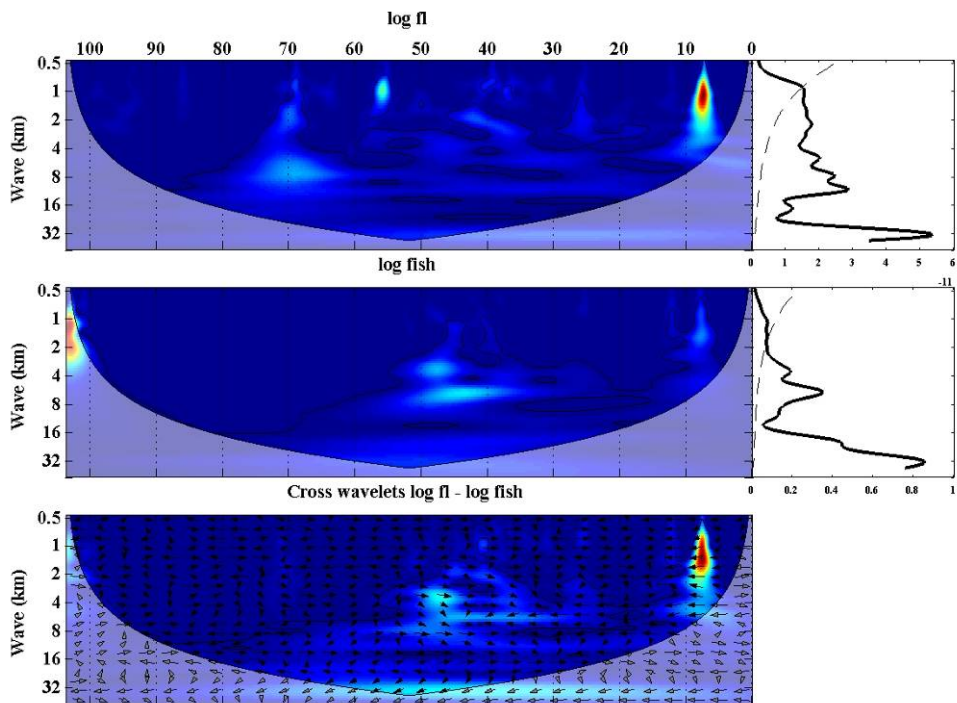
Transecto 40 día. Análisis wavelets para variable física y log fluid like



Análisis wavelets para variable física y log fish



Análisis wavelets para log fluid like y log fish



Anexo 4:

Otros métodos para tratar datos con autocorrelación espacial

1 Revisión de otros métodos para tratar la autocorrelación espacial

Los métodos descritos a continuación no fueron considerados en el análisis porque tienen como resultado una evaluación global de la serie o por tener una muy difícil interpretación como es el caso de los test parciales de mantel.

Todos los métodos que se mencionarán a continuación tienen como supuesto básico estacionariedad espacial (Dormann et al. 2007) el cual ya fue revisado anteriormente.

2 Test de Mantel

El test de Mantel (1967) compara dos matrices de distancias o similitudes. Este test se calcula mediante la suma del producto cruzado de similitudes entre las dos matrices en la cual se excluye los valores de la diagonal. Estos valores se excluyen ya que resultan del producto de las similitudes consigo mismas (ver Legendre y Legendre 1998). El test de Mantel tiene como hipótesis nula que las distancias (o similitudes) en la matriz A son linealmente independientes de las distancias (o similitudes) de la matriz B, (A y B son matrices de distancias o similitudes).

Hay una aproximación del test de Mantel. Primero se debe de estandarizar los valores en cada uno de los vectores de la matriz. Por ser una matriz simétrica el número de elementos a tomar será $d = n(n-1)/2 - 1$. El valor del test de Mantel se encontrará entre -1 y 1.

El test estandarizado de Mantel es de la siguiente forma:

$$r_M = \frac{1}{(d-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{A_{ij} - \bar{A}}{S_A} \right) \left(\frac{B_{ij} - \bar{B}}{S_B} \right)$$

2.1 Test parcial de Mantel

Sean las matrices A, B y C. El test parcial de Mantel $r_M(A * B.C)$ calcula la correlación de A y B controlando el efecto de C. Para el cálculo se estandarizan las matrices A, B y C y la forma del test parcial de Mantel es:

$$r_M(A * B.C) = \frac{r_M(AB) - r_M(AC) * r_M(BC)}{\sqrt{1 - r_M(AC)^2} \sqrt{1 - r_M(BC)^2}}$$

Los resultados del test parcial de mantel son muy difíciles de interpretar (Legendre y Legendre, 1998)

3 Regresiones Múltiples de Matrices de Distancias (MRM)

Los MRM presentan dos ventajas sobre el test parcial de Mantel (Lichstein 2006). El primero es que se pueden utilizar métodos de regresiones polinomiales, no lineales y GAM. Y la segunda es que el aumento de las matrices explicativas permite que cada variable (aumentada) ambiental esté representada por su propia matriz de distancias. Esto proporciona una mejor correlación especie-ambiente (Lichstein 2006).

La forma de representar los MRM es $Y = X\beta + \varepsilon$ donde X es la matriz de distancias euclidianas provenientes de las posiciones de cada observación. En resumen el método MRM hace los siguientes pasos.

Descompone en k intervalos de distancias d_1, d_2, \dots, d_k a la Matriz X , luego generamos una matriz con la condición:

$$D = \begin{cases} 1 & \text{si } X_{ij} \in d_i \\ 0 & \text{otro caso} \end{cases}$$

Después de la reparametrización tenemos la siguiente estructura matricial.

$$\begin{pmatrix} d_{1,2} \\ d_{1,3} \\ d_{1,4} \\ d_{1,5} \\ d_{2,3} \\ d_{2,4} \\ d_{2,5} \\ d_{3,4} \\ d_{3,5} \\ d_{4,5} \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \\ y_{41} \\ y_{42} \\ y_{51} \\ y_{52} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{51} \\ \varepsilon_{52} \end{pmatrix}$$

Estructura para 5 intervalos de distancias (Lichstein 2006).

Con la condición de $\sum \tau_i = 0$ y reconstruyendo el modelo con la condición dada tenemos

$Y = X * \beta^* + \varepsilon$ donde los coeficientes en β^* representan los índices de correlación, y graficando estos índices se obtiene el correlograma.

4 Modelos Autorregresivos Simultaneos (SAR)

Los modelos Autorregresivos Simultaneos (SAR) asumen estacionariedad espacial y normalidad de los errores (Dormann et al. 2007). Los modelos SAR se presentan de tres maneras las cuales dependen del proceso de autocorrelación.

El primer modelo SAR asume que el proceso autorregresivo ocurre solamente en la variable respuesta. Se incluye en el modelo el componente autorregresivo de la siguiente manera.

$$Y = X\beta + \rho WY + \varepsilon$$

donde ρ es el parámetro de autocorrelación y W es la matriz de pesos espaciales.

El segundo modelo SAR asume que la autocorrelación afecta a X y Y , entonces en el modelo anterior también se incluye el término $WX\gamma$ donde γ describe los coeficientes del modelo del conjunto de predictores espaciales WX . El modelo tiene la forma

$$Y = \rho WY + X\beta + WX\gamma + \varepsilon$$

Finalmente el tercer modelo asume que el proceso proviene solamente de ε . El modelo tiene es:

$$Y = X\beta + \lambda W\mu + \varepsilon$$

Anexo 5:

Pruebas de permutación

Test de Aleatorización (o permutación)

Las pruebas de permutación permiten realizar los test estadísticos cuando los datos no cumplen los supuestos básicos (como por ejemplo la normalidad).

Las pruebas de asignación al azar asumen que las observaciones pueden ser intercambiables lo cual garantiza que cualquier arreglo es igualmente probable de generar (Fortín and Dale 2006)

El método de permutación, también llamada de ordenación aleatoria, es un acercamiento muy general a pruebas de hipótesis estadísticas. La permutación y la ordenación aleatoria son consideradas sinónimas.

El método consiste en tener 2 variables x y y .

1) Calcular el estadístico requerido (ejm. correlación) entre las dos variables, este estadístico será el valor a probar (r_0).

2) Permutar la variables x

3) Calcular nuevamente la correlación entre y y x^* (x^* variable permutada)

4) Generar los pasos 2 y 3 iterativamente

5) Contrastar el estadístico original (r_0) con los estadísticos generados en el paso 3, para este contraste se puede utilizar la prueba t-student.

Luego en base a los resultados del paso 5 se decide si el estadístico inicial (r_0) es significativo o no (para mayor información ver Legendre y Legendre, 1998. Cap I).