

# **UNIVERSIDAD NACIONAL DE INGENIERÍA**

**FACULTAD DE INGENIERÍA ECONÓMICA Y CIENCIAS SOCIALES**



**“ESTUDIO DE LA DISTRIBUCIÓN DE AVES MARINAS EN RELACIÓN CON SUS  
PRESAS Y LA PROFUNDIDAD DEL LÍMITE SUPERIOR DE LA ZONA DE MÍNIMO  
OXÍGENO EN EL SISTEMA DE LA CORRIENTE DE HUMBOLDT UTILIZANDO  
ANÁLISIS DE PROCESOS PUNTUALES Y MODELOS LINEALES  
GENERALIZADOS”**

TESIS

PARA OPTAR POR EL TÍTULO PROFESIONAL DE  
INGENIERA ESTADÍSTICA

ZAIDA JESÚS QUIROZ CORNEJO

LIMA-PERÚ

2011

## **Agradecimientos**

*Me gustaría empezar agradeciéndole a la Universidad Nacional de Ingeniería - UNI, porque fue ahí donde inicio “este camino”. Gracias a todos aquellos buenos profesores que día a día nos brindan desprendidamente sus conocimientos y trabajan desde las aulas por un Perú mejor.*

*Gracias a IMARPE e IRD por ser instituciones forjadoras de investigadores. Gracias a IMARPE por brindarme los datos utilizados en esta tesis y al IRD por financiar mi estancia en IMARPE.*

*Mil gracias a Arnaud y Sophie Bertrand. Quiero agradecerles la oportunidad que me brindaron de realizar mi tesis con ustedes. Gracias por leer y releer “infinitas” veces esta tesis. Gracias por “todo” lo que me enseñaron durante esta etapa, por sus extraordinarias explicaciones y discusiones sobre ecología y estadística que me ayudaron a empezar, desarrollar y terminar la tesis. Gracias Arnaud por tus divertidas e interesantes clases de ecología marina y por tu constante preocupación en toda esta etapa. Gracias Sophie por incentivar mi interés por la ecología y la estadística espacial de las que poco sabía. Gracias a ambos por permitirme explorar e iniciarme en este mundo de la investigación, por brindarme nuevos horizontes, por sus consejos y por ser más que amigos.*

*Gracias a Nicolás Bez por sus comentarios y enseñanzas en mis inicios en la geoestadística. Gracias a Michael Ballón por sus enseñanzas en acústica. Un especial agradecimiento a Gildas Rondaut por sus enseñanzas en acústica y programación, por su disposición y paciencia, y por permitirme colaborar en la programación del Echopen en Matlab. También gracias a la profesora Magen Infante por sus sugerencias y recomendaciones en la etapa final de la tesis y a lo largo de la carrera.*

*Gracias a Rocío, Daniel, Ainboa, Rosmery, Gianinna, Claude, Alexandre y José porque con su interés y curiosidad me enseñaron que hay más por estudiar, gracias por los buenos momentos y sobre todo por su amistad. Quiero agradecerles especialmente a Rocío y Daniel, por la confianza y el apoyo que me brindaran en toda esta etapa, tanto profesional como emocional, por creer que es posible la investigación y la aplicación de la estadística en el Perú, por enseñarme a luchar por lo que uno quiere y hacer frente a las adversidades, y por permitirme ser parte de esta aventura. Gracias Ainboa por el tiempo que invertiste para ayudarme a comprender mejor la acústica y la biología marina, por las discusiones en geoestadística y en estadística en general que ampliaron mis conocimientos, gracias a Mathieu también por absolver mis dudas de geoestadística por el skype de Ainboa!, gracias Ainboa por tu preocupación como hermana y por tu alegría y entusiasmo. Y gracias Gianinna por tu amistad, por las charlas y almuerzos, por las caminatas, por tus consejos y por explicarme cada día más sobre biología e investigación.*

*También le agradezco infinitamente a mi mamá y hermana por su constante preocupación, consejos y apoyo incondicional en todas las etapas de mi vida, por comprenderme y apoyarme en todas mis decisiones, aunque siempre escoja lo más difícil. Gracias a mi papá y abuelo por su insistencia y motivación para que me gradúe. Gracias Viktor por tu comprensión, por tu permanente apoyo y por compartir mis alegrías. Y para finalizar quiero dar gracias a todos aquellos que hicieron posible la realización de esta tesis, y que el camino no termine aquí...*

## **Resumen**

Conocer la distribución de las aves marinas, sus presas y la profundidad del límite superior de la zona de mínimo oxígeno (ZMO) es esencial para garantizar la durabilidad del ecosistema, y en particular, entender la respuesta de las aves a la variabilidad en los procesos marinos. Actualmente a lo largo de la costa peruana se realizan cruceros acústicos que recogen información de las características físicas y biológicas en el Sistema de la Corriente de Humboldt (SCH). Por otro lado se cuenta con información de los desplazamientos de las aves marinas gracias al marcaje electrónico con GPS. Se proponen dos métodos estadísticos para estudiar la distribución de las aves, sus presas y la oxiclina (profundidad del límite superior de la ZMO): Como metodologías se aplican la función K de Ripley, que permite trabajar procesos puntuales y los Modelos Lineales Generalizados (MLG) que permiten trabajar datos continuos desde otra perspectiva, en ambos casos tomando explícitamente en cuenta escalas temporales y espaciales. En el primer capítulo se describe el contexto ecológico de este estudio, es decir, el ecosistema marino del SCH. En el segundo capítulo, se presenta la teoría estadística que enmarca la función K de Ripley como herramienta de estudio de los procesos puntuales y los MLG. Además se revisa brevemente la teoría de la geoestadística que se usara en el procesamiento de las bases de datos del siguiente capítulo. En el tercer capítulo se explica cómo se obtienen los datos y su procesamiento, como se combinaron las bases de datos usando escalas de tiempo y espacio, y las variables finales usadas en los dos métodos propuestos para cada escala. A continuación se presentan los resultados para las diferentes escalas de tiempo y espacio. Finalmente, se comparan y evalúan los resultados en términos estadísticos, así como su contribución en la ecología.

**Palabras clave:** Función K de Ripley, Modelos Lineales Generalizados, Distribución de aves marinas, anchoveta; *Engraulis ringens*, Escalas, Relaciones predador presa.

## **Abstract**

Knowing the association between seabirds, their prey and depth of the upper limit of the Oxygen Minimum Zone (OMZ) along different space-temporal scales is essential for guaranteeing the sustainability of the ecosystem, and in particular, to understand the response from seabirds to the variability in marine processes. Nowadays acoustic cruises are performed along the Peruvian coast which collects information about the physics and biologic features in the Humboldt Current System (HCS). On the other hand we have information on seabird's movement thanks to electronic tagging with GPS. Two methods are proposed to study associations between seabirds, their preys and the oxycline (Upper limit of the OMZ): As methodology applied consist in Ripley's K function, which allow us to work with point data, and Generalized Linear Models (GLM) which allow us to work with continuous data from another perspective, and in both cases taking into account explicitly spatial-temporal scales.

The first chapter describes the ecological context of this study, i.e. the HCS marine ecosystem. The second chapter presents the statistic theory underlying the Ripley's k function as study tool of punctual processes and GLM. The geostatistic theory is also briefly reviewed before being used in data base processing of the next chapter. The third chapter explains how is obtained the data and its processing, how are combined the databases using spatial-temporal scales, and the final variables used in the two methods proposed for each scale. Later, the results are presented for each spatial-temporal scale. Finally, the results are compared and evaluated statistically, as well as their contribution in ecology.

**Key words:** Ripley's K function, generalized linear models, Seabirds distribution, anchovy; *Engraulis ringens*, Scales, Predator-prey relationships.

# ÍNDICE

<i>Resumen</i>	1
<i>Abstract</i>	1
CAPITULO I	5
INTRODUCCIÓN	5
1. <i>Marco situacional</i>	6
1.1 <i>El sistema de la Corriente de Humboldt</i>	7
1.1.1 <i>Zona de mínimo oxígeno</i>	7
1.1.2 <i>Producción del sistema de la Corriente de Humboldt</i>	8
1.2 <i>Acústica</i>	9
1.3 <i>Las aves marinas</i>	10
1.4 <i>Las escalas espacio-temporales</i>	12
1.5 <i>Las relaciones predador-presa</i>	13
2. <i>Planteamiento del Problema</i>	15
3. <i>Objetivos</i>	15
4. <i>Pruebas de Hipótesis</i>	15
CAPITULO II	16
1. <i>Geoestadística</i>	17
1.1 <i>Variograma</i>	17
1.2 <i>Kriging</i>	20
2. <i>Procesos Puntuales</i>	21
2.1 <i>Definición y características de los Procesos Puntuales</i>	22
2.2 <i>Proceso Poisson Homogéneo (PPH)</i>	24
2.3 <i>La Función K-Ripley</i>	24
2.3.1 <i>Función K de Ripley teórica para un PPH</i>	24
2.3.2 <i>Estimación General de la función K-Ripley</i>	25
2.4 <i>La función L</i>	26
2.5 <i>Procesos Puntuales Multivariados</i>	27
<i>Función Doble K de Ripley teórica para PPH</i>	28
3. <i>Modelos Lineales Generalizados</i>	29
3.1 <i>Introducción</i>	29
3.2 <i>Definición de los MLG</i>	29
3.2.1 <i>Componente aleatorio</i>	29
3.2.2 <i>Componente sistemático</i>	32

3.2.3	<i>Función de enlace</i>	32
3.3	<i>Estimación de los parámetros</i>	33
3.4	<i>Evaluación del MLG</i>	36
3.4.1	<i>Pruebas de Bondad de Ajuste</i>	37
3.4.1.1	<i>Función de Desvianza</i>	37
3.4.1.2	<i>Estadística de Chi-cuadrado de Pearson</i>	38
3.4.2	<i>Pruebas de la significancia del modelo ajustado</i>	38
3.4.3	<i>Criterios de selección del modelo</i>	39
3.4.3.1	<i>Criterio de Información de Akaike (AIC)</i>	39
3.4.3.2	<i>Criterio de Información Bayesiana (BIC)</i>	39
3.4.3.3	<i>Desvianza explicada</i>	39
3.5	<i>Análisis de Residuales y Diagnostico del MLG</i>	39
3.5.1	<i>Diagnóstico de balanceo e influencia</i>	39
3.5.2	<i>Medida de apalancamiento (leverage)</i>	40
3.5.3	<i>Distancia de Cook</i>	40
3.5.4	<i>Residual ordinario</i>	41
3.5.5	<i>Residuales de Pearson</i>	41
3.5.6	<i>Residuales de Desvianza</i>	41
3.5.7	<i>Gráficos de diagnóstico</i>	42
CAPITULO III		43
1.	<i>Obtención de los datos</i>	45
1.1.	<i>Obtención de los datos acústicos</i>	45
1.2	<i>Obtención de los datos de aves marinas</i>	45
2.	<i>Pre-procesamiento de los datos</i>	46
2.1	<i>Pre-procesamiento de los datos acústicos</i>	46
2.2.	<i>Pre-procesamiento de los datos de aves marinas</i>	50
3.	<i>Procesamiento de las bases de datos</i>	51
3.1	<i>Malla espacio-temporal</i>	51
3.2.	<i>Bases de datos para análisis de procesos puntuales</i>	52
3.3.	<i>Bases de datos para MLG</i>	53
3.4	<i>Aplicación de la Función K-Ripley</i>	55
3.5	<i>Aplicación de geoestadística</i>	55
3.6	<i>Aplicación de MLG</i>	56
CAPITULO IV		57
4.1	<i>Resultados de procesos puntuales</i>	58

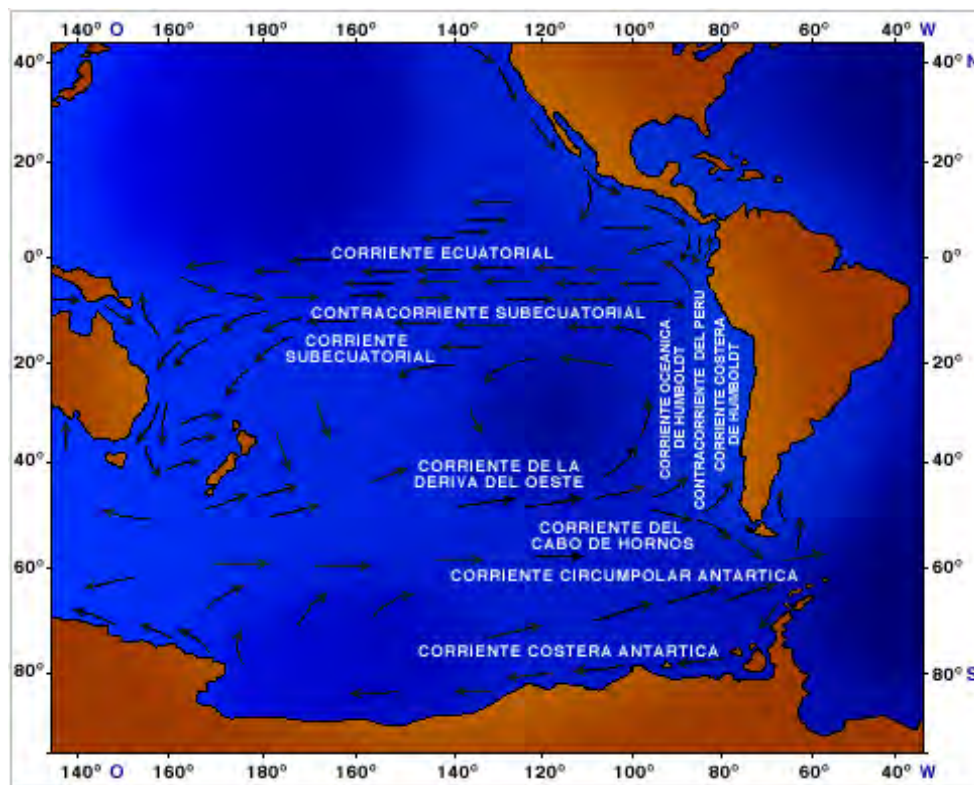
4.1.1	<i>Distribución espacial de las ARS</i>	58
4.1.2	<i>Distribución espacial de ARS y peces</i>	59
4.2	<i>Aplicación de geoestadística</i>	63
4.2.1	<i>Distribución espacial de la biomasa de peces</i>	63
4.2.2	<i>Distribución espacial de la profundidad del límite superior de la zona de mínimo oxígeno</i>	65
4.3	<i>Resultados de Modelos Lineales Generalizados</i>	67
4.3.1	<i>Resultados MLG Presencia – Ausencia (Probabilidad de presencia de ARS)</i>	67
4.3.2	<i>Resultados MLG conteos (Número de ARS - si está presente)</i>	74
CAPÍTULO V		79
CAPITULO VI		88
BIBLIOGRAFÍA		90
ANEXOS		96

# **CAPITULO I**

## **INTRODUCCIÓN**

## 1. Marco situacional

Un ecosistema marino contiene detritos<sup>1</sup> y miles de clases de organismos, incluyendo bacterias, fitoplancton, zooplancton, peces, aves marinas y mamíferos, junto con todos los componentes del medio ambiente abiótico<sup>2</sup> (Cury et al., 2001). Entender mejor el funcionamiento actual de los ecosistemas marinos es muy importante porque permite tomar en cuenta la biodiversidad marina y sobre todo proteger su sostenibilidad en el futuro mediante adecuados diagnósticos de su situación actual, estableciendo sistemas de evaluación y monitoreo, adoptando medidas y priorizando acciones para hacer frente a las variabilidades del ecosistema marino. El océano desde su formación, hace 4 000 millones de años, contiene la mayor parte de agua líquida de nuestro planeta. El Océano Pacífico es el mayor océano de la tierra, ocupa la tercera parte de su superficie. Se extiende aproximadamente de 15 000 km desde el mar de Bering y en el Ártico por el norte, hasta los márgenes congelados del mar de Ross en la Antártida por el sur. La circulación superficial de las aguas del Pacífico es generalmente en el sentido contrario de las agujas del reloj en el Hemisferio Sur. La Corriente Ecuatorial del Sur (Fig. 1) fluye hacia el oeste a lo largo del ecuador, gira hacia al sur, al este de Nueva Guinea, después gira al este a unos 50 grados de latitud sur, y se une a la corriente de circulación principal hacia el oeste del Pacífico Sur, que incluye la Corriente Circumpolar Antártica que da la vuelta a la Tierra. Al aproximarse a la costa de Chile, la Corriente Sur Ecuatorial se divide; una rama fluye alrededor del Cabo de Hornos y la otra gira al norte para formar la Corriente de Humboldt, ubicada en el Este del Océano Pacífico, frente a las costas peruanas y chilenas (Tomczak et al., 2005).



**Figura 1. Sistema general de corrientes superficiales del Océano Pacífico Sur. (Fuente: Ana Maria Duque Correa. <http://fluidos.eia.edu.co/hidraulica/articulos/interesantes/humboldt/humboldt.htm>.)**

<sup>1</sup> Los detritos son residuos, generalmente sólidos, que provienen de la descomposición de fuentes orgánicas (vegetales y animales). Es materia muerta. Aunque es materia orgánica en descomposición, hay numerosos seres vivos que se alimentan de ella degradándola aún más.

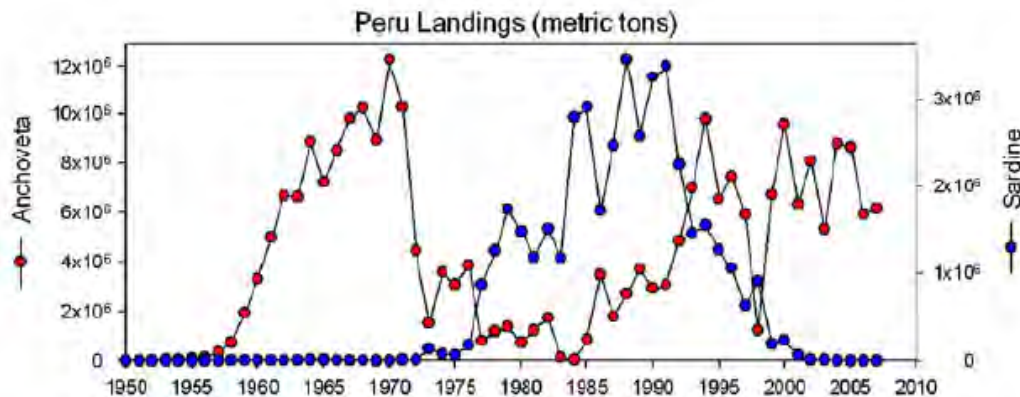
<sup>2</sup> En el ámbito de la biología y la ecología, el término abiótico designa para hacer referencia a aquellas estructuras, formaciones, elementos inertes en los que no es posible la vida, es decir que no forman parte o no son producto de los seres vivos (Tovar., 2000).



## 1.1 El sistema de la Corriente de Humboldt

El sistema de la Corriente de Humboldt (SCH) que abarca las costas de Chile y Perú, es uno de los más productivos sistemas costeros de surgencia del mundo. Produce más peces (ver Fig. 2 para los desembarques de anchoveta *Engraulis ringens* y sardina *Sardinops sagax*) por unidad de superficie que cualquier otro espacio marítimo, pues aun cuando representa menos del 0.1% de la superficie mundial de los océanos, en él se lleva a cabo más del 10 % de las capturas de peces del planeta, convirtiéndose en el ecosistema con mayor concentración de peces a nivel mundial (Chávez et al., 2008).

Esta producción biológica excepcional principalmente se debe a la presencia de zonas de surgencia o afloramiento que se producen por la acción del viento principalmente, el que desplaza grandes cantidades de aguas superficiales, creando un espacio que es llenado por el ascenso de aquellas que se encuentran en profundidades que fluctúan entre los 150 y 200 metros. Estas aguas son frías y muy ricas en nitratos y fosfatos, sustancias que provienen de la descomposición de los organismos del mar y/o de los residuos que llegan a él y que son básicas para el desarrollo de la vida. Las aguas ricas en nutrientes, al llegar a la superficie y por acción de la energía solar, facilitan la proliferación de las algas microscópicas que forman el fitoplancton, iniciándose así una cadena alimentaria. Este sistema tiene otras particularidades, algunas de las cuales han sido sugeridas como responsables de la alta producción (Bakun y Weeks., 2008; Chávez et al., 2008; Chaigneau et al., 2009; Ballón., 2010): (i) es el sistema de surgencia más cercano al Ecuador, (ii) tiene la más superficial e intensa zona de mínimo oxígeno (ZMO), (iii) tiene la más estable y menos intensa variabilidad física a meso-escala comparada con otros sistemas de surgencia, y (iv) está íntimamente ligado a las dinámicas del Pacífico ecuatorial y está sujeto a fluctuaciones largas interanuales y multi-décadas.



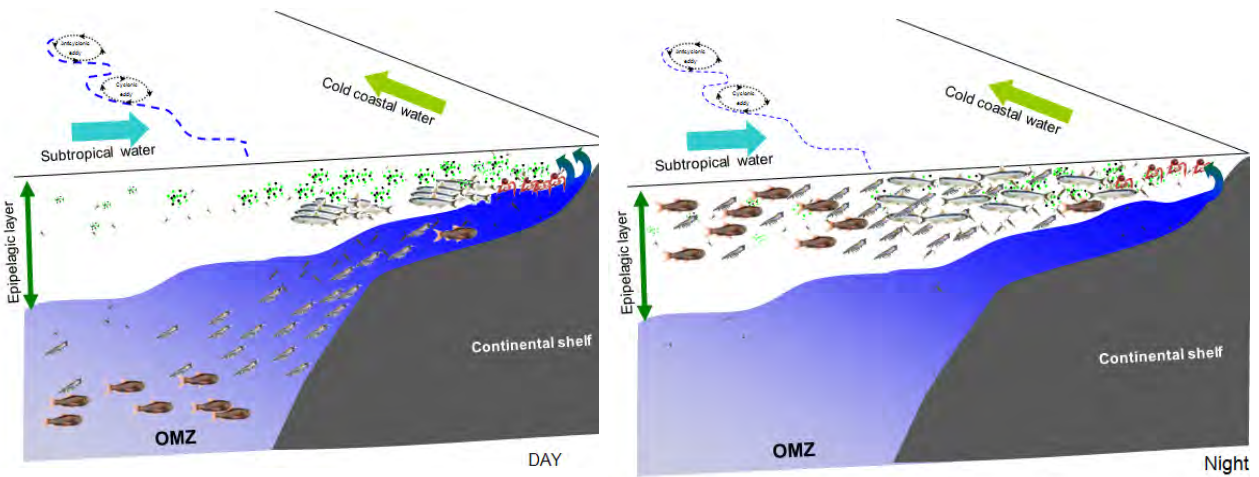
**Figura 2. Desembarques de anchoveta y sardina para el periodo de 1950 al presente. (Fuente: Chávez et al., 2008)**

El objetivo de esta sección es describir brevemente las características físicas y biológicas del Sistema de la corriente de Humboldt, para facilitar la interpretación de este estudio.

### 1.1.1 Zona de mínimo oxígeno

Los océanos incluyen extensas áreas llamadas zonas de mínimo oxígeno (ZMOs) que son zonas de contenido reducido de oxígeno situadas a nivel de los márgenes orientales de los océanos tropicales, a una profundidad de 20 a 800 metros. Las ZMOs se sitúan debajo de una capa superficial surtida de oxígeno disuelto. La frontera entre estos dos dominios forma la oxiclina (Bertrand et al., 2010a). La casi ausencia de oxígeno en la ZMO obliga a muchas especies a concentrarse en la capa superficial oxigenada, en ocasiones, teniendo tan sólo una decena de metros de altura (Fig. 3). Este fenómeno podría agravarse debido a que las ZMO se amplían bajo el efecto del cambio climático. Algunas especies de zooplancton, peces mesopelágicos y calamares han adaptado su metabolismo para habitar

temporalmente o permanentemente en la ZMO e incluso algunas la utilizan como un refugio a la depredación. Incidentalmente, la escasez de oxígeno en el fondo preserva los organismos muertos que se sedimentan y cuyo estudio permite reconstruir las condiciones climáticas y de productividad pasadas (Chávez et al., 2008).



**Figura 3. Visión esquemática del efecto de la Zona mínima de oxígeno en la distribución de las especies según día y noche (Fuente: M. Ballón y A. Bertrand).**

### 1.1.2 Producción del sistema de la Corriente de Humboldt

La producción primaria está relacionada con la cantidad de nutrientes. El fitoplancton es la principal base de la producción primaria en los océanos; son plantas que tienen capacidad fotosintética y que viven dispersas en el agua. En las zonas de afloramiento, el agua de surgencia rica en nutrientes, llega a la superficie, la luz del sol activa la producción de fitoplancton. Por lo tanto, la alta producción primaria es una de las características principales de la surgencia costera (Ryther., 1969).

Se denomina producción secundaria, la producción en organismos que se alimentan de fitoplancton, se trata principalmente del zooplancton, es decir, el plancton animal. En el SCH la producción secundaria está conformada principalmente por crustáceos, en particular copépodos, eufausidos, quetognatos y múnida. Espinoza y Bertrand (2008) y Espinoza et al. (2009) estudiaron el contenido estomacal de anchovetas y sardinas y encontraron que su principal alimento son los eufausidos y copépodos. Además en el SCH la mayoría del zooplancton migrante vive de día en la ZMO que constituye una zona de refugio donde pocos predadores pueden entrar, y migran a la superficie durante la noche para alimentarse del fitoplancton.

En el SCH, la producción terciaria (los organismos que comen zooplancton, en particular de peces como anchoveta, sardina o jurel *Trachurus murphyi*), es 10 veces más alta que en otros sistemas de surgencia comparables (Chávez et al., 2008). Además la mayor producción de peces consiste de una especie, la anchoveta, que está restringida a las masas de agua altamente productivas y en particular a las aguas costeras frías (Bertrand et al., 2004; Swartzman et al., 2008).

### La anchoveta Peruana

El ecosistema pelágico peruano está altamente dominado por la anchoveta (Bertrand et al., 2004). La anchoveta es un elemento crítico del SCH, manteniendo la mayor pesquería en el mundo (Fréon et al., 2008) con más de 6 millones de toneladas de anchoveta anuales capturadas durante los últimos cuatro años en el Perú (Fréon et al., 2010).

La anchoveta Peruana (Fig. 4) es un pez de la familia de los engraulidae; vive en aguas costeras frías, tiene una esperanza de vida máxima de 4 años y una longitud corporal máxima de 20 cm. La primera

madurez sexual ocurre alrededor de la edad de un año con una longitud aproximada de 12 cm. Los individuos de una talla menor a 12 cm son por consecuencia considerados como juveniles.



**Figura 4. La anchoveta Peruana – *Engraulis ringens***

Las anchovetas tienen una relativamente escasa capacidad migratoria, por ello no pueden eludir las condiciones desfavorables. También están restringidas a las aguas muy superficiales encima de la ZMO, por lo que son presas fáciles para los peces más grandes, los mamíferos, las aves y los pescadores (Gerlotto et al., 2006). ¿Cómo pueden entonces estos peces proliferar en tales magnitudes frente a las costas del Perú?

Los investigadores descubrieron que su secreto reside en parte en su flexibilidad biológica y de comportamiento, y existe una serie de características que pueden ayudar a explicar la aparente “paradoja” de la anchoveta peruana (Espinoza y Bertrand., 2008): dentro de las aguas costeras frías, la anchoveta puede forrajear eficientemente bajo una variedad de condiciones ambientales, utiliza comida de varios niveles tróficos<sup>3</sup>, contrariamente a la teoría según la cual su productividad estaría vinculada con una cadena alimentaria corta, las anchovetas no se alimentan principalmente de plancton vegetal (fitoplancton), seleccionan las presas más energéticas, esencialmente del plancton animal (zooplancton) de gran tamaño (Espinoza y Bertrand., 2008); además se concentra en áreas de refugio cuando las condiciones son adversas (Bertrand et al., 2004), puede adaptar su comportamiento reproductivo según las condiciones climáticas y distribuye su población a lo largo de un amplio rango de temperaturas (Gutiérrez et al., 2008).

## **1.2 Acústica**

Los métodos acústicos se basan en el estudio de la difusión del sonido, infrasonido<sup>4</sup> y ultrasonido<sup>5</sup>, es decir ondas mecánicas que se propagan a través de la materia (tanto sólida como líquida o gaseosa). En

---

<sup>3</sup> Se llama nivel trófico en ecología a cada uno de los conjuntos de especies, o de organismos, de un ecosistema que coinciden por el turno que ocupan en la circulación de energía y nutrientes, es decir, a los que ocupan un lugar equivalente en la cadena trófica, que es el proceso de transferencia de energía alimenticia a través de una serie de organismos, en el que cada uno se alimenta del precedente y es alimento del siguiente. También conocida como cadena alimenticia, es la corriente de energía y nutrientes que se establece entre las distintas especies de un ecosistema en relación con su nutrición. Así el fitoplancton tiene un nivel trófico de uno, el zooplancton que se alimenta de fitoplancton un nivel trófico de 2 mientras los que se alimentan de este zooplancton tienen un nivel trófico de 3.

<sup>4</sup> El infrasonido es una onda acústica o sonora cuya frecuencia está por debajo del espectro audible del oído humano (aproximadamente menor a 20 Hz), puede ser oído por las ballenas.

<sup>5</sup> Un ultrasonido es una onda acústica o sonora cuya frecuencia está por encima del espectro audible del oído humano (aproximadamente mayor a 20 000 Hz), puede ser escuchado por algunos animales como los delfines y los murciélagos.

la actualidad la tecnología acústica provee herramientas fundamentales para muchos investigadores de ciencias del mar como por ejemplo la ecosonda. La ecosonda tiene un transductor<sup>6</sup> el cual se instala generalmente en el casco del barco; este emite pulsos de sonido (“ping”) que se propagan en el mar y al encontrar objetos (peces, zooplancton, fondo, etc.) se refleja la onda de sonido, de tal forma que cierta energía regresa hacia el transductor, esta energía retro dispersada (eco) es detectada nuevamente por el transductor y convertida a señal eléctrica (Simmonds y MacLennan., 2005). Luego de un tiempo el transductor emite un nuevo pulso y se repite el proceso. Finalmente los sonidos medidos en decibeles se muestran en un ecograma. Gracias a esta tecnología se puede estudiar la composición del fondo marino, las localizaciones y biomasas de los peces y plancton, la distribución de las poblaciones marinas en el espacio y cómo cambian en el tiempo, entre otros.

En el Perú, desde 1983 el Instituto del Mar del Perú - IMARPE realiza cruceros sistemáticos haciendo uso de las técnicas acústicas (Gutiérrez et al., 2007; Simmonds et al., 2009). Actualmente, se realizan en promedio dos cruceros al año y gracias al reciente desarrollo de las herramientas acústicas es posible determinar cada segundo y a lo largo del trayecto de una nave, la distribución y abundancia de los recursos marinos. Además Ballón (2010) y Bertrand et al. (2010a) desarrollaron un método basado en datos acústicos para determinar el límite superior de la ZMO.

### 1.3 Las aves marinas

Las aves marinas pelágicas son predadores marinos altamente movibles y hábiles que forrajea<sup>7</sup> peces y zooplancton. Se pueden mover a grandes distancias a un relativo bajo costo energético en comparación con sus presas (Fauchald., 2009) y cuando sus presas son peces pelágicos gregarios forrajean en áreas muy pequeñas en concentraciones locales debido a la naturaleza de su presa (Duffy., 1983). Las aves con huevos o crías están limitadas en su rango de forrajeo porque tienen que retornar a las colonias regularmente, ya sea para alternar con su pareja para proteger sus huevos o para alimentar a sus crías (Weimerskirch., 2007).

El ecosistema de surgencia costero peruano alberga grandes poblaciones de aves marinas. La mayor presa de estas aves marinas es la anchoveta. Las dos especies que presentan los niveles poblacionales más altos son: el piquero peruano (*Sula variegata*) y el cormorán guanay (*Phalacrocorax bougainvillii*) (Fig. 5).



**Figura 5. Aves marinas del Perú con sus crías (piquero a la izquierda y guanayes a la derecha. Fuente: S. Bertrand et al., 2010b)**

El piquero se caracteriza por un aleteo bastante rápido y su forma de pescar en pique, entrecerrando las alas y cayendo al mar en un clavado casi vertical desde la altura, terminando eventualmente de perseguir la presa nadando brevemente con las alas. Aunque puede pescar solo o en grupos, también es frecuente

<sup>6</sup> Un transductor es un dispositivo capaz de transformar o convertir un determinado tipo de energía de entrada, en otra de diferente a la salida.

<sup>7</sup> Actividad que los animales realizan para conseguir alimento.

verlo en bandadas de cientos de individuos, realizando todos dicha maniobra en forma conjunta. Anida en cerros, pampas y acantilados (Mobley., 2009).

El guanay se alimenta de un conjunto de hasta 69 diferentes especies de peces (Jahncke., et al 1998), pero claramente la especie presa más importante para ellos es la anchoveta peruana. Nidifica principalmente en cerros y pampas, y lo hace en comunidades, llegando a formar colonias de anidamiento de cientos de miles de individuos. Y se alimenta principalmente por buceo, propulsado por las patas (Likoff., 2007).

El piquero y el guanay además son aves guaneras, así llamadas por el guano que producen, que es un fertilizante natural y completo, contiene todos los nutrientes que las plantas requieren para su normal crecimiento, desarrollo y producir buenas cosechas, es un producto ecológico, no contamina el ambiente, es biodegradable y mejora las propiedades físicas, químicas y biológicas del suelo. La abundancia global de aves guaneras fue estimada aproximadamente a más de 20 millones de individuos en la década de 1950 y de 2 millones en la década pasada (Weimerskirch et al., 2010). Sin embargo, la importancia en preservar las aves marinas del Perú, cuyo estado de conservación es vulnerable, no es únicamente por el guano que producen. Cury (2001) plantea que la eliminación de los predadores superiores en un ecosistema marino puede tener un fuerte impacto en los niveles tróficos menores.



**Figura 6. Ubicación de las Islas y Puntas Guaneras del Perú. (Fuente: AGRORURAL)**

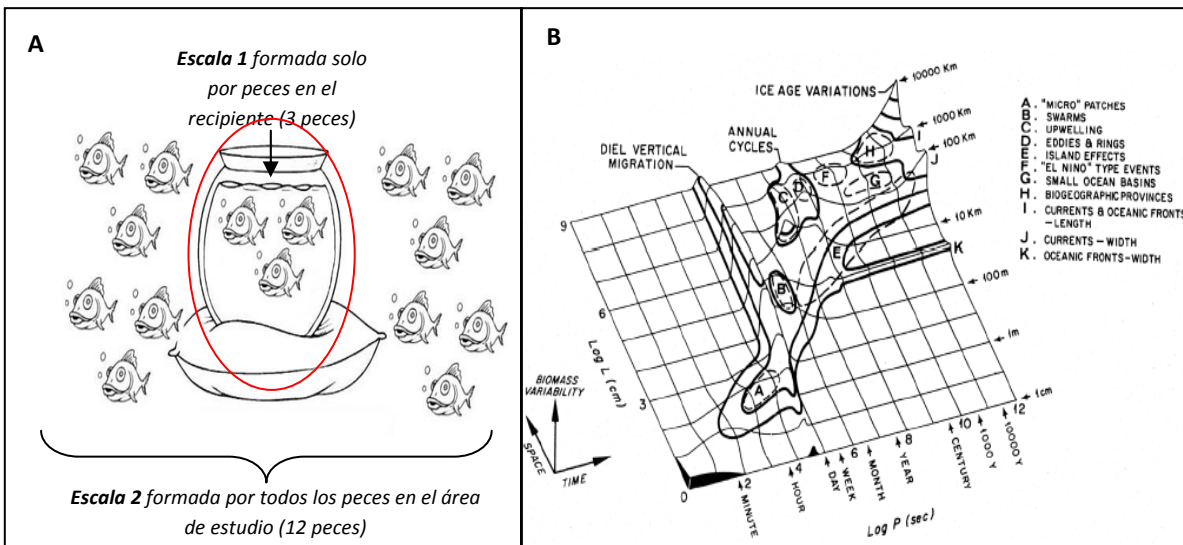
Como medidas de conservación estas especies se encuentran protegidas por el Estado mediante el D.S. No. 034-2004-AG. Además en julio del año 2006 se aprueba la Ley 28793-Ley de Protección, Conservación y Repoblamiento de Islas, Rocas y Puntas Guaneras, declara de interés nacional la protección, conservación y repoblamiento de 33 islas y puntas guaneras del país (Fig. 6). Las islas, islotes y puntas guaneras están protegidas también por la Ley Forestal y de Fauna Silvestre (Ley N° 27308) y su Reglamento (Decreto Supremo N° 014- 2001-AG). Mediante esta ley, se maneja y conserva las especies de fauna silvestre de las reservas costeras, islas y puntas, a través de los órganos competentes como AGRORURAL.

### 1.4 Las escalas espacio-temporales

Levin (1992) estima que a cada una de las escalas de análisis corresponde la descripción de un mecanismo dado, es por ello que la elección de la escala de estudio es primordial y debe ser explícita.

Describir un sistema biológico de manera pertinente necesita escoger una o más escalas de observación. Esta elección no es tan sencilla porque no existe una "correcta" escala a la cual podemos describir poblaciones o ecosistemas (Levin 1992., Wiens., 1989), y lo que se puede encontrar a una escala puede diferir mucho de lo encontrado en otra escala (Fig 7a).

Para un fenómeno complejo dado, hay tantos modelos y "teorías" del fenómeno como escalas de percepción y la variabilidad del sistema es condicionada a la escala de descripción (Haury et al., 1978; Fig 7b).



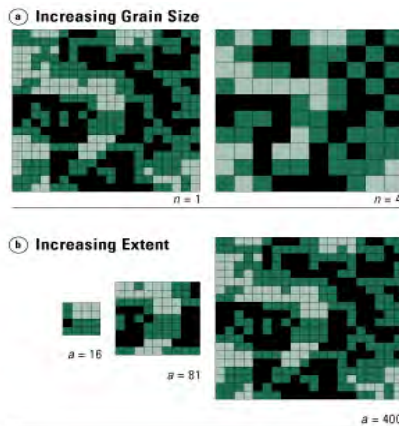
**Figura 7. A) Representación de las escalas espaciales de la variabilidad de abundancia de peces; B) Representación conceptual de las escalas espacio-temporales de la variabilidad de la biomasa en zooplancton y de los factores asociados a estas escalas, conocido como el diagrama de Stommel. (Fuente: Haury et al., 1978)**

La dinámica de poblaciones del predador y de sus presas, puede ser influenciada por factores a diferentes escalas. Intentar relacionar estas dinámicas directamente sin reconocer las diferencias en las escalas puede llevarnos a confusiones (Wiens, 1989). Entonces ¿Cómo tomar en cuenta las escalas de todos los componentes de un sistema que uno no conoce? Una posible solución es el diseño de estudios que incluyan un análisis de escalas múltiples (Cueto, 2006),

Las escalas espacio-temporales se pueden caracterizar de dos formas (Fig. 8):

- El grano que es el tamaño de las unidades individuales de observación, lo que los estadísticos llaman unidad básica de muestreo; y
- La extensión que es toda el área que abarca el estudio.

Ambos definen la resolución de un estudio. Cuando la escala de medida de una variable es cambiada, la varianza de dicha variable cambia y como ocurre esto depende de qué cambie, ya sea el grano o la extensión. Así cuando se mantiene constante la extensión y se incrementa la medida del grano, generalmente la variación espacial decrece (Wiens, 1989).



**Figura 8. Representación de la variación del grano (arriba) y de la extensión (abajo). Cuando se incrementa el tamaño del grano algunos parches pequeños de hábitats que originalmente podían ser diferenciados, ahora son incluidos dentro de la misma muestra. Cuando se aumenta la extensión del estudio se incorporan nuevos parches de hábitat que no estaban presentes en el área de estudio original.**

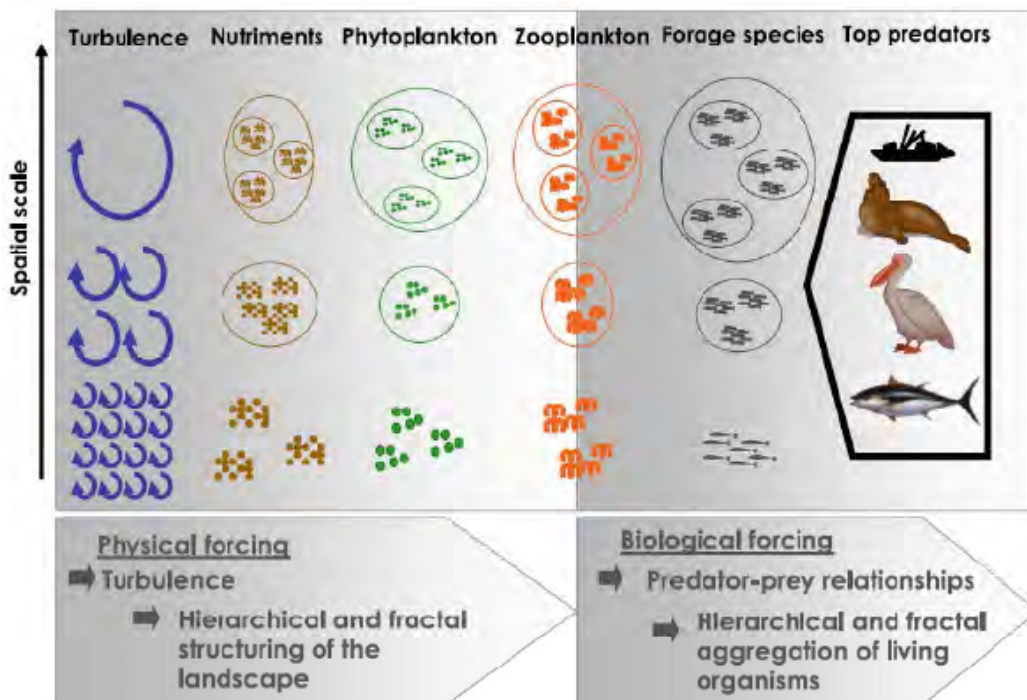
## 1.5 Las relaciones predador-presa

El funcionamiento del ecosistema marino depende de su estructura, diversidad e integridad. Para garantizar la durabilidad del ecosistema, las instancias de manejo de las actividades humanas (como la pesca) deben tomar en cuenta los diferentes componentes del sistema: las poblaciones explotadas pero también sus predadores, sus presas y sus competidores. Una alteración o disturbio de uno o algunos de los componentes de los ecosistemas marinos puede tener fuertes efectos en los niveles tróficos mayores o menores, dependiendo de si las cadenas alimenticias son controladas por los organismos de forrajeo o por los depredadores (Cury et al., 2001). Sin embargo, los ecosistemas marinos son por su naturaleza más difíciles de observar que los ecosistemas terrestres; "Nada es más difícil que estudiar un animal que por definición no podemos ver" (Tremblay et al., 2009).

Uno de los principales predadores superiores en los ecosistemas marinos son las aves marinas. Describir la densidad y distribución espacial de las aves marinas es un importante paso para entender su rol en el ecosistema marino y su vulnerabilidad a los cambios en estos ecosistemas. En particular, esta información es esencial para entender la respuesta de las aves a la variabilidad en los procesos marinos, físicos y biológicos, para evaluar los impactos probables y riesgos para las aves marinas de la sobrepesca de las presas (Burger et al., 2004). Estudios ecológicos sobre el comportamiento de forrajeo de las aves marinas son complicados primero por las grandes distancias recorridas por aves marinas forrajeras y segundo por la dificultad de observar las distribuciones de la presa. Como resultado, las relaciones funcionales entre el comportamiento de las aves marinas y sus ambientes bióticos y abióticos son difíciles de estudiar. Sin embargo los nuevos instrumentos tecnológicos de observación, nos proveen de poderosos "ojos" para observar el ecosistema (Tremblay et al., 2009). En el Perú, se está realizando

un esfuerzo para estudiar el comportamiento de forrajeo de las aves guaneras, así desde el 2007 se realizan experimentos de telemetría sobre piqueros y guanayes gracias a una colaboración de IMARPE-IRD<sup>8</sup>-CNRS<sup>9</sup>.

Una clave para entender la dinámica de la población de las aves marinas es la caracterización de las respuestas de comportamiento de las aves marinas frente a las fluctuaciones en la abundancia, distribución y calidad de la presa. Las relaciones predator-presa constituyen una importante clase de procesos ecológicos porque: 1) muchos de los organismos viven en ambientes donde las fuentes de comida están distribuidas heterogéneamente, 2) la distribución de las presas influye los movimientos de los consumidores a lo largo de vastas áreas, y 3) el uso de la presa por los predadores tiene importantes consecuencias sobre el flujo energético del ecosistema (Russell, et al., 1992). La fuerza de la relación espacial entre las aves marinas y su presa es escala-dependiente (Fig. 9). Es decir, los resultados obtenidos son dependientes tanto de la resolución (grano) como de la extensión del estudio (Swartzman y Hunt, 2000). Rose & Leggett (1990) plantean que para una misma extensión, a gran escala (grano) de estudio hay una correlación positiva entre la presencia de presas y de predadores. Cuando el grano disminuye y se acerca a la de las agregaciones, la correlación disminuye hasta volverse nula. Finalmente con un grano inferior a las agregaciones uno tiene una correlación negativa entre los predadores y las presas.



**Figura 9. Ilustración esquemática de los rasgos fundamentales de la ocupación del espacio en los ecosistemas marinos pelágicos. De un punto de vista estructural, todos los compartimientos del ecosistema están caracterizados por la presencia de estructuras jerárquicas, a lo largo de una amplia gama de escalas. Desde una perspectiva funcional, estas estructuras están generadas y mantenidas entre otros por: (i) la introducción de energía en la masa de agua por la turbulencia y (ii) la cadena de comportamientos generadas por la búsqueda de presas. (Fuente: Bertrand., et al., 2008b)**

<sup>8</sup> Institut de recherche pour le développement (Francia)

<sup>9</sup> Centre National de la Recherche Scientifique (Francia)



## **2. Planteamiento del Problema**

Varios investigadores han estudiado la distribución espacial de las aves y la distribución de sus presas o medio ambiente (Ejm. Burger et al., 2004; Swartzman y Hunt 2000; Fauchald, 2009; Tremblay et al., 2009; Vlietstra, 2005; Davoren et al., 2003; Weimerskirch, 2007; Zavalaga, et al 2010); sin embargo relacionar las variaciones del comportamiento de las aves con las del ecosistema no es sencillo, la resolución de los datos no siempre es consistente, la recolección de datos no siempre coinciden en espacio y tiempo, tiene que trabajarse a varias escalas, y las técnicas estadísticas son distintas a las clásicas, entre algunas de las limitantes. Lo cual nos lleva a plantearnos las siguientes preguntas:

- ¿Qué metodología estadística nos permite estudiar la distribución de las aves marinas, sus presas y el límite superior de la zona de mínimo oxígeno en el SCH?
- ¿Cuál es el comportamiento espacial de las aves marinas en el SCH?
- ¿Cómo es la distribución de las aves marinas relativamente a sus presas en el SCH?
- ¿Cómo es la distribución espacial de las aves marinas relativamente a la profundidad del límite superior de la zona de mínimo oxígeno en el SCH?

## **3. Objetivos**

### **General:**

Determinar mediante técnicas estadísticas en qué medida la presencia de aves marinas y su comportamiento espacial, reflejan las propiedades del ecosistema, tales como el hábitat físico (límite superior la zona de mínimo oxígeno) y los peces forraje frente al Perú.

### **Específicos:**

- Describir el comportamiento espacial de las aves marinas.
- Investigar las interacciones predador-presa a diferentes escalas, y
- Relacionar el hábitat de forrajeo de las aves, en particular la distribución de las aves marinas con la profundidad del límite superior de la zona de mínimo oxígeno y la distancia a la colonia (Isla Pescadores).

## **4. Pruebas de Hipótesis**

### **General:**

La presencia de aves marinas y su comportamiento espacial, reflejan las propiedades del ecosistema, tales como el hábitat físico (límite superior la zona de mínimo oxígeno) y los peces forraje frente al Perú.

### **Específicos:**

- El comportamiento espacial de las aves marinas es aleatorio.
- Existen interacciones predador-presa a diferentes escalas, y
- Existe relación entre distribución de las aves marinas con la profundidad del límite superior de la zona de mínimo oxígeno y la distancia a la colonia (Isla Pescadores).

# **CAPITULO II**

## **MARCO TEÓRICO**

En este capítulo, se revisa brevemente la teoría de la geoestadística; método que se usó para interpolar los datos correspondientes a la profundidad del límite superior de la zona de mínimo de oxígeno y la biomasa de peces del crucero acústico en el espacio. Estos datos interpolados se usarán en el procesamiento de datos para cruzar las bases de datos provenientes de acústica y aves. A continuación se presenta la teoría estadística que enmarca la función K de Ripley como herramienta de estudio de los procesos puntuales, que nos servirá para medir la asociación espacial entre zonas de búsqueda restringida de las aves y agregaciones de anchoveta. Finalmente se presenta la teoría estadística correspondiente a los Modelos Lineales Generalizados para estudiar la relación entre el comportamiento de las aves marinas, la estructura espacial de la anchoveta y otros factores abióticos.

## 1. Geoestadística

Los métodos clásicos de estimación estadística consideran los datos como realizaciones independientes de una misma variable aleatoria, suponiendo que no tienen relaciones entre sí y que se distribuyen igualmente. Sin embargo, cuando se tienen datos ubicados en el espacio los métodos clásicos son particularmente inadecuados ya que supone independencia entre las observaciones cuando en realidad las observaciones próximas tienen valores cercanos, mientras que aquellas que están más alejadas entre sí tienen una menor relación entre ellas, es decir existe autocorrelación espacial (Emery, 2008). La geoestadística brinda a la ecología nuevas herramientas para la interpretación de patrones espaciales de organismos y de numerosos componentes ambientales con los cuales interactúan (Rossi, 1992). La geoestadística permite estudiar, describir y caracterizar dicha continuidad espacial de variables autocorrelacionadas distribuidas en el espacio. Los métodos geoestadísticos tienen muchas aplicaciones en poblaciones de ecosistemas marinos recolectados por el método acústico (Ejm. Gerlotto et al, 1987; Simmonds et al., 1992; Bez, 1997; Petitgas, 1993; Bez et Rivoirard, 2000; Rivoirard et al., 2000).

En todo análisis espacial es muy útil hacer mapas que muestren información de los valores medidos, lo cual da una idea preliminar de la distribución y la continuidad espacial de estos. Se define el campo de estudio para posteriormente realizar variogramas experimentales y definir el modelo del variograma para realizar interpolación (Kriging), mapas, etc (Fig. 10).

### 1.1 Variograma

En geoestadística la autocorrelación de una distribución espacial está descrita por el variograma. En términos matemáticos es una función de distancia que mide en promedio la variabilidad entre todos los pares de puntos, en función de la distancia que los separa (Rivoirard et al., 2000).

El variograma experimental es un estimador insesgado del variograma teórico, representativo para una muestra de puntos, definido por la semivarianza  $\gamma^*(h)$  como:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{x_i - x_j \sim h} [z(X_i) - z(X_j)]^2$$

Donde:

- **Z(x)**: Se define como una variable regionalizada que es la realización de un proceso estocástico (aleatorio) definido en un espacio. Entonces su valor es la realización de variables aleatorias definidas en el punto  $x$  del espacio.
- **Lag (h)**: El 'lag' es la distancia  $h$  de separación entre los puntos, cualquier sea la dirección.
- **Tolerancia en las distancias (N(h))**: La tolerancia de distancias o 'Nlag' es el número de pares de puntos  $(x_i, x_j)$ , separados por un vector de distancia  $h$ .
- **Direcciones**: El variograma puede ser omnidireccional o bidireccional. Cuando el variograma es calculado para todos los pares de puntos en ciertas direcciones tales como norte-sur o este-

oeste, a veces presenta diferentes tipos de comportamientos en alguno de ellos, es decir presenta anisotropía, y el variograma debe ser bidireccional. En caso que no ocurra, el variograma depende únicamente de la magnitud de la distancia entre puntos  $h$  y se dice que es isotrópico y será omnidireccional (Armstrong, 1992).

- **Tolerancia en la dirección (tolerancia angular):** La tolerancia angular es un ángulo que controla el número de puntos del campo de estudio que están incluidos en el cálculo del variograma.

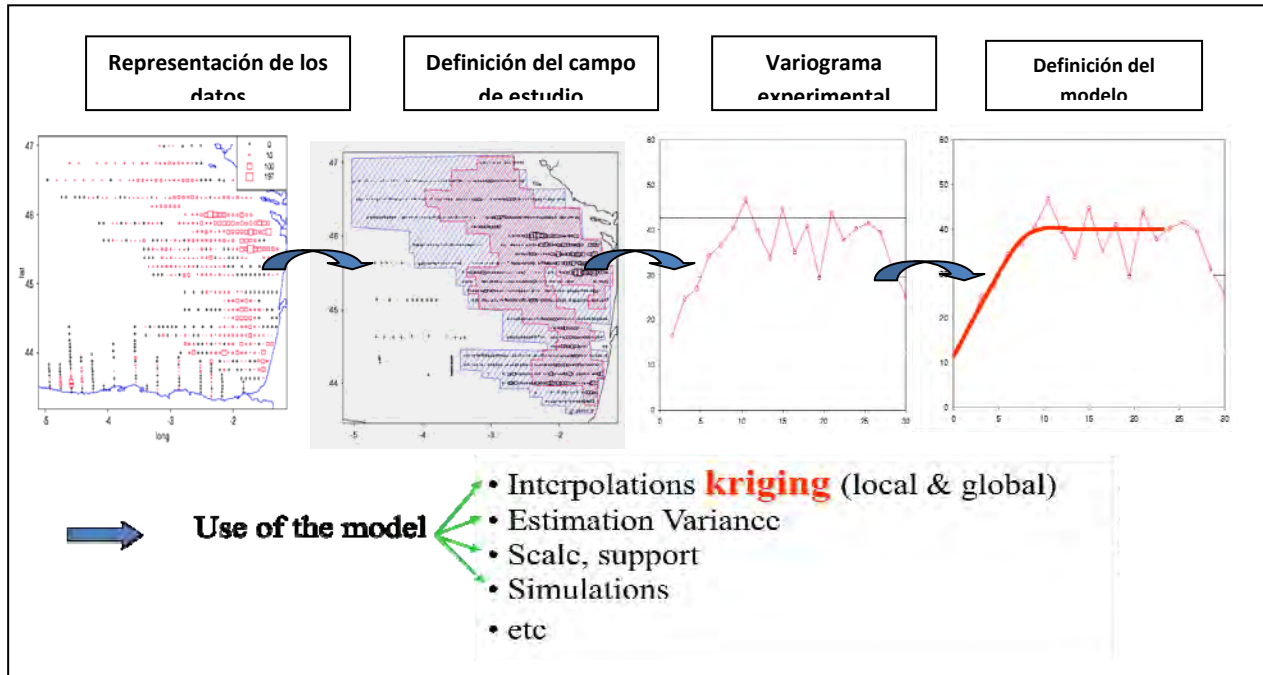


Figura 10. Esquema de metodología para el análisis geoestadística (Fuente: Nicolás Bez, 2009)

### Transformaciones

Muchas veces cuando la variable tiene una marcada asimetría, la estructura espacial de la variable definida con el variograma es más robusta si la variable es transformada, sin embargo al hacer la transformación se obtiene un estimador sesgado de la estructura original, por lo cual es necesario realizar una re-transformación (Rivoirard., et al 2000).

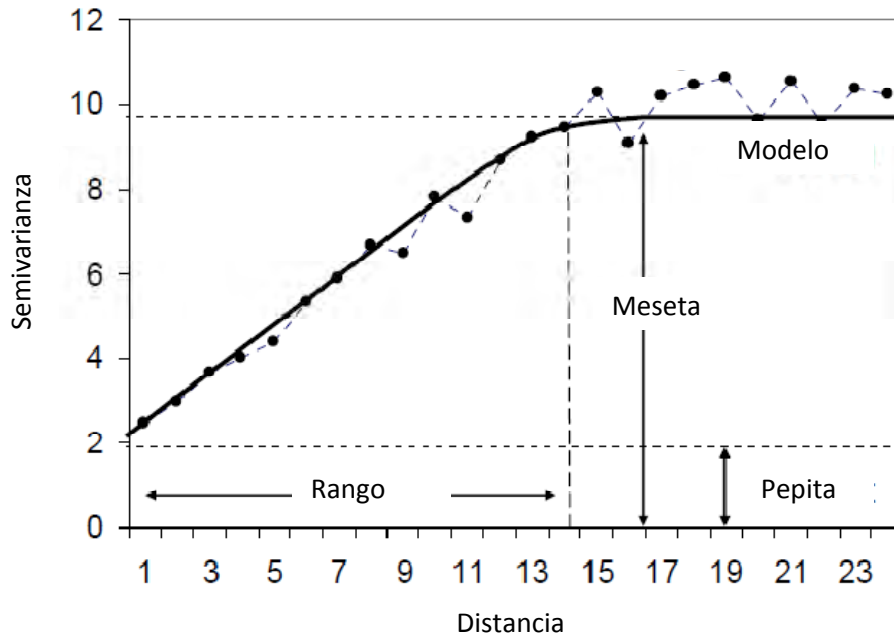
Para reducir la importancia de valores altos se puede realizar la siguiente transformación logarítmica:

$$L = \log\left(1 + \frac{Z}{b}\right); Z \text{ es la variable sin transformar y } b = E(Z)$$

### Modelos

El modelo del variograma es un modelo matemático que se ajusta al variograma experimental y en base a este modelo se estima la varianza a cualquier distancia, lo cual permitirá realizar posteriormente estimaciones. El modelo del variograma está definido por tres parámetros (Fig. 11):

- **Pepita (nugget):** es el comportamiento al origen, que detecta la variabilidad a muy pequeña escala, es decir representa micro estructuras que no son accesibles por la resolución muestral, por ello es conocido en la minería como efecto pepita.
- **Rango:** es la distancia a partir de la cual el variograma se estabiliza, es decir el valor de la semivarianza se vuelve constante y las muestras son espacialmente independientes unas de otras. El rango indica el tamaño de las estructuras.
- **Meseta:** es el valor máximo de las semi-varianza  $\gamma^*(h)$  obtenido en la muestra. Caracteriza el comportamiento del variograma a largas distancias.



**Figura 11. Parámetros del variograma experimental (línea punteada) y su modelo (línea sólida)**

Existe un gran número de modelos usados para ajustarse al variograma experimental:

$$\text{Nugget: } \gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ c & \text{otro caso} \end{cases}$$

$$\text{Esferico: } \gamma(h) = \begin{cases} c \cdot \left( 1.5 \left( \frac{h}{a} \right) - 0.5 \left( \frac{h}{a} \right)^3 \right) & \text{si } h \leq a \\ c & \text{otro caso} \end{cases}$$

$$\text{Exponencial: } \gamma(h) = c \left( 1 - \exp\left(\frac{-3h}{a}\right) \right)$$

$$\text{Gaussiano: } \gamma(h) = c \left( 1 - \exp\left(\frac{-3h^2}{a^2}\right) \right)$$

$$\text{Potencia: } \gamma(h) = c \cdot h^w \quad \text{con } 0 < w < 2$$

El modelo debe ser lo más simple posible, es preferible usar modelos con meseta verdadera como el esférico, ya que el modelo gaussiano o exponencial alcanzan la meseta de forma asintótica, o lo que es lo mismo, no la alcanzan nunca, lo cual no es muy realista. Además el variograma puede ser anidado, ello depende de las estructuras en el mismo.

Finalmente se debe tener en cuenta que el modelado del variograma varía según su uso, así Bez y Rivorard (2001) utilizan los parámetros del variograma para describir la estructura espacial (nivel de agregación e inercia de la distribución) e identificar las escalas asociadas a las agrupaciones de fitoplancton del Atlántico. Gerlotto (1993) utiliza los diferentes parámetros del variograma para definir las poblaciones acústicas. Bahri y Fréon (2000) calculan los variogramas para analizar el efecto de medio ambiente en la distribución de los peces del Mediterráneo. Finalmente, Gerlotto y Paramo (2003) ajustan los variogramas de las variables que describen la estructura interna de los cardúmenes para describir su estructura interna e identificar la existencia de 'núcleos'.

## 1.2 Kriging

El kriging es una técnica geoestadística que sirve para interpolar datos de una grilla (estimación local) y/o crear un mapa (estimación global). Su ventaja frente a otras técnicas de interpolación radica en incluir el comportamiento de la variable en el espacio. Las distancias usadas en las técnicas de kriging son distancias estadísticas a diferencia de las distancias geométricas utilizadas por ejemplo en el método del inverso de la distancia. Además el kriging pondera de forma distinta muestras que estén muy cerca entre sí y procedan de la misma región que muestras que estén por ejemplo al lado opuesto al punto que queremos interpolar (Gallardo, 2008). Los parámetros del semi-variograma elegido tienen una importancia relevante cuando se quiere interpolar. Así por ejemplo la pepita determinará como se reparten los diferentes pesos a las muestras que rodean un punto que queremos interpolar, el rango también influye, pues aquellos puntos que se encuentren más allá del rango tendrán la mínima capacidad predictiva. Respecto a la anisotropía, es decir aquellos puntos que se encuentren en el eje mayor de variación, se le asignarán coeficientes mayores a dichas muestras.

La vecindad del kriging se define como el espacio que contiene el sitio a estimar los datos utilizados en la estimación (Emery, 2008). La vecindad puede ser "vecindad única" cuando usa todas las muestras posibles, por ello se usa cuando no son tan numerosas, unos cientos a lo mas, también puede ser "vecindad móvil" en cuyo caso la vecindad debe ser escogida de tal forma que incluya las muestras que están cercanas, pero idealmente no del mismo lado que el punto objetivo, por ejemplo no todos del mismo transecto acústico (Rivoirard, 2000).

Existen diferentes tipos de kriging dependiendo de si la variable considerada es estacionaria, intrínseca o no-estacionaria. El tipo de kriging más común es el kriging ordinario, en el cual la media es estimada localmente. Cuando se conoce la media de la variable a estimar en el área de estudio se usa el simple kriging. Cuando se incluye la influencia de una tendencia espacial se usa el universal kriging.

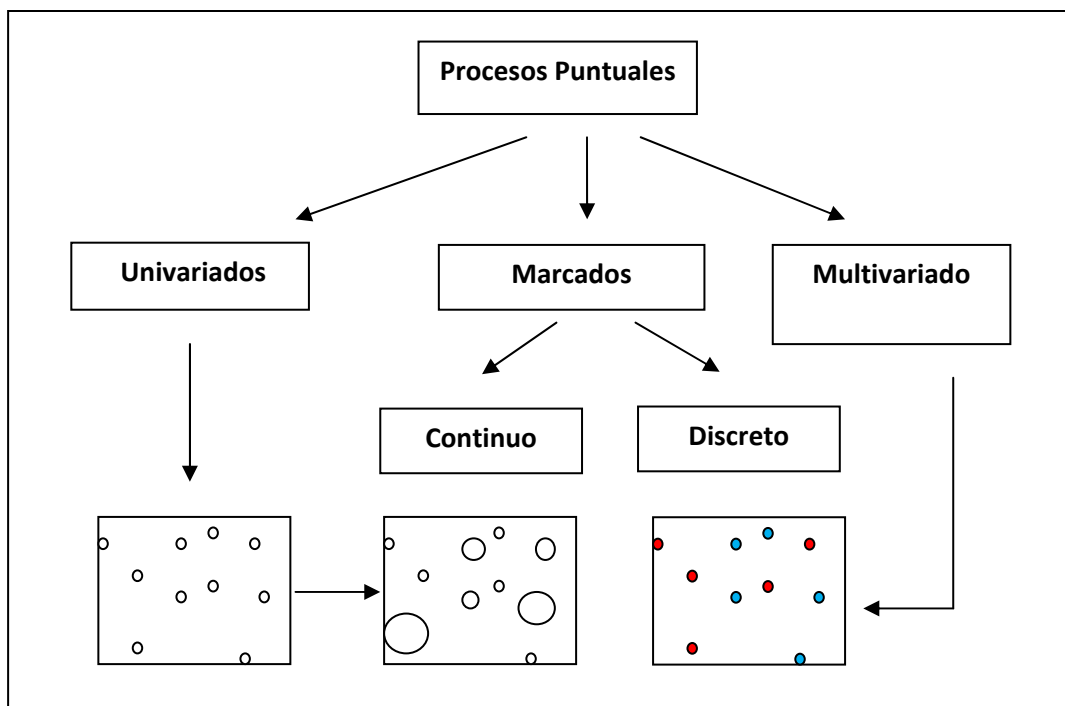
El kriging ordinario considera que la estimación de  $Z(V)$  está dada por  $\sum \lambda_i Z(x_i)$ , siendo  $V$  el dominio (posiciones cuyos valores se desea estimar). El kriging es el estimador óptimo, obtenido con las ponderaciones  $\lambda_i$  que minimizan la varianza. Dichas ponderaciones  $\lambda_i$  están basadas en el modelo del variograma, siendo la solución del sistema lineal:

$$\begin{cases} \sum_j \lambda_j \gamma(x_i - x_j) + \mu = \bar{\gamma}(x_i, V), & \text{para cualquier } i \\ \sum_i \lambda_i = 1, \end{cases}$$

donde  $\mu$  es el parámetro de Lagrange introducido para cumplir la condición  $\sum_i \lambda_i = 1$ .

## 2. Procesos Puntuales

Por más de 30 años los procesos puntuales han sido una mayor área de investigación en estadística espacial (Ripley, 1977, 1981; Cressie, 1993). La estadística espacial difiere de la estadística clásica, la cual típicamente analiza observaciones independientes y aplica conceptos tales como la media, la varianza muestral y la función de distribución. Gran parte de la teoría estadística clásica está basada en la asunción de la distribución normal. Basado en esto, las pruebas estadísticas e intervalos de confianza son las que se aplican en la mayoría de las áreas de estadística. Sin embargo los estadísticos de los procesos puntuales, toman en cuenta la correlación entre los patrones. La distancia relativa entre puntos muchas veces está correlacionada tanto como el número de puntos en regiones adyacentes, en adición, las características de los objetos representados por los puntos pueden ser espacialmente correlacionadas. De ahí que el análisis estadístico esta mucho más ligado con detectar y describir estas correlaciones. Además de la descripción de datos en el espacio, el estudio de los procesos puntuales espaciales proporciona una gran variedad de modelos de comportamientos particulares o procesos que pueden generar datos puntuales. Estos modelos son, generalmente, el resultado de considerar situaciones reales. Por ejemplo, en el caso biológico, la ubicación de animales depende de la existencia de depredadores y/o comida. Se puede concluir que la importancia de los procesos puntuales radica en que se pueden usar para responder a una variedad de objetivos dependiendo del área de aplicación. En primer lugar, se centra en la distribución espacial de los datos observados, en segundo lugar responde a la relación que existe entre distintos tipos de eventos y por último se centra en la intensidad de estos.



**Figura 12. Clasificación de los procesos puntuales según variables**

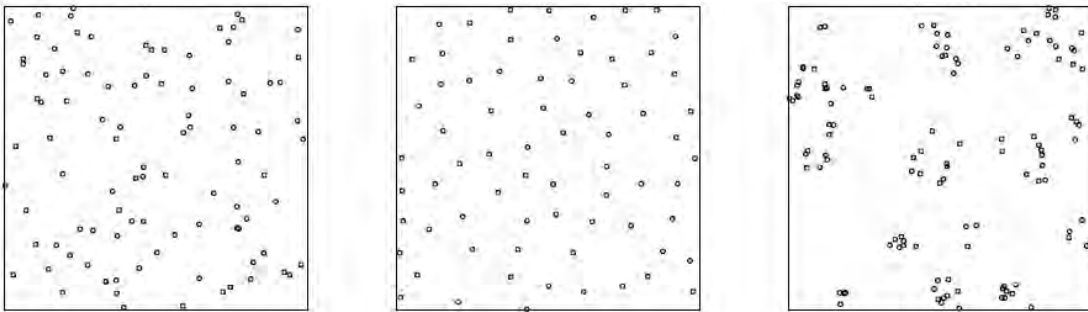
## 2.1 Definición y características de los Procesos Puntuales

El término ‘proceso’ implica que una serie de eventos<sup>10</sup> se desarrollan en un tiempo y espacio dado, sin embargo en la mayoría de los casos el tiempo es estudiado de forma independiente del espacio, es por ello que algunos autores prefieren usar el término ‘campo puntual’. El término puntual hace referencia a la colección de puntos en algún área. En la teoría de la probabilidad clásica una variable aleatoria describe todos los posibles resultados de un experimento. En términos matemáticos, un proceso puntual puede ser descrito por infinitas variables aleatorias, pues es un mapeo de las probabilidades espaciales dentro de algún espacio fijo. Se define un proceso puntual  $N$  como un modelo estocástico de una colección de eventos que caen en cierto espacio  $B$ , en este caso se considerara el espacio bidimensional  $B \subset \mathbb{R}^2$ .

Un proceso puntual se califica como marcado (Fig. 12) cuando cada punto en el patrón espacial tiene una información adicional llamada marca, la cual puede ser discreta, cuando contiene un factor es decir son diferentes tipos de puntos, o continua cuando cada punto contiene valores numéricos asociados. El primer caso sirve para evaluar la interacción entre dos o más procesos puntuales, mientras que el segundo sirve para evaluar si la diferencia numérica entre las marcas depende de la distancia de los puntos correspondientes, es decir, puntos vecinos tienden a tener marcas similares (Illian, et al 2008).

Un patrón puntual puede tratarse como un proceso puntual si dicho patrón es aleatorio. La realización de un proceso puntual es un conjunto desordenado de puntos. En ciertos casos los datos no son los datos originales, debido a que no forman un proceso puntual. En dichos casos, a veces se puede obtener un proceso puntual después de procesar los datos originales, por ejemplo cuando los datos originales son un conjunto de segmentos, y el proceso puntual está formado por los puntos finales, puntos medios, entre otros (Baddley et al., 2008).

Un patrón puntual de acuerdo a su distribución en el espacio puede ser (Fig. 13): (i) un espacio completamente aleatorio (CSR) si los eventos son igualmente probables de ocurrir en cualquier punto dentro del mismo (uniformes) y si no interactúan entre ellos (independientes), (ii) un espacio regular en caso los eventos interactúen repulsivamente, o (iii) un espacio agrupado en caso los eventos interactúen atractivamente. En general se piensa que un patrón “aleatorio” no debería ser caótico con pares de puntos cercanos y grandes espacios vacíos, por eso Ripley (1981) dice que “el ojo humano está absolutamente adaptado para clasificar patrones puntuales pero es también fácilmente engañado”.



**Figura 13. Patrones de puntos: aleatorio (izquierda), regular (centro), agrupado (derecha). (Fuente: Illian, et al 2008)**

---

<sup>10</sup> En estadística, un evento o suceso es un subconjunto de un espacio muestral, es decir, un conjunto de posibles resultados que se pueden dar en un experimento aleatorio.



## Medida de Momentos de Procesos Puntuales

Así como en la estadística clásica las variables aleatorias quedan definidas por los momentos de primer y segundo orden con los cuales se calcula la media y la varianza, los procesos puntuales están definidos por las medidas de los momentos (Moyal, 1962; Krickeberg, 1970; 1973), y que a diferencia de los momentos definidos para variables aleatorias como cantidades escalares, o los momentos definidos para vectores aleatorios como vectores, las medidas de momentos de procesos puntuales están definidos como medidas definidas en el espacio.

Sea un proceso puntual  $N$ , un conjunto Borel<sup>11</sup>  $B$  y el número de puntos  $N(B)$  en  $B$  es una variable aleatoria, entonces la medida del  $k$ -ésimo momento de  $N$  está dada por:

$$\mu_N^{(k)}(B_1 \times B_2 \times \dots \times B_k) = E(N(B_1)N(B_2) \dots N(B_k)) = \int_{\phi} \phi(B_1)\phi(B_2) \dots \phi(B_k) \prod_N d\phi$$

donde  $\prod_N d\phi = p(N \in d\phi)$  y  $\phi(B) = \sum_{i=1}^{\infty} I(s_i \in B)$

La medida del primer momento o medida promedio es:

$$\mu_N(B) = E(N(B)) = \int_B \phi(B) \prod_N d\phi$$

Si  $B \downarrow \{s\}$  donde  $s \in X$ ;  $X \subset \mathbb{R}^d$ :

$$\mu_N(B) = \int_B \lambda(s) ds$$

Luego la **intensidad de primer orden** es el número promedio de eventos por unidad de área en el punto  $s$  y está definida por:

$$\lambda(s) = \lim_{v(ds) \rightarrow 0} \frac{\mu_N(ds)}{v(ds)}$$

Donde  $v(\cdot)$  es la medida de Lebesgue ( $v(A) = |A|$  que es el volumen de  $A$ )

La **intensidad de segundo orden** describe la covarianza (o correlación) entre valores de procesos de diferentes regiones en el espacio:

$$\lambda_2(s, u) = \lim_{\substack{v(ds) \rightarrow 0 \\ v(du) \rightarrow 0}} \frac{\mu_N^{(2)}(ds \times du)}{v(ds)v(du)}$$

La medida de momento de segundo orden  $\lambda_2(s, u)$  sirve también para determinar el momento de segundo orden reducido  $\mathcal{K}_N$  que está definido como la función  $K$  de Ripley.

Un proceso puntual puede ser estacionario cuando no es afectado por la traslación (cuando  $\lambda$  es constante). También puede ser isotrópico cuando no es afectado por la rotación o invariante cuando es estacionario e isotrópico.

---

<sup>11</sup> Un conjunto Borel es un conjunto en un espacio topológico que puede formarse a partir de conjuntos abiertos (o, equivalentemente, a partir de conjuntos cerrados) a través de las operaciones contables de la unión, la intersección contables, y complemento.

## 2.2 Proceso Poisson Homogéneo (PPH)

Un proceso homogéneo de Poisson es sinónimo de un proceso espacial completamente aleatorio (Cressie, 1993), este proceso se caracteriza por dos propiedades fundamentales:

- i. Uniformidad: implica que el número de eventos por unidad de área sea constante e igual a  $\lambda$ , por lo cual cumple la asunción de estacionariedad. Entonces para un proceso homogéneo de Poisson:

$$\mu_N(ds) = \lambda(s)v(ds) = \lambda v(ds)$$

$$\hat{\lambda} = \frac{n}{A}$$

donde  $n$  es el número de eventos observados y  $A$  es el área de la región de estudio.

Por lo tanto un proceso homogéneo de Poisson es automáticamente también isotrópico.

- ii. Independencia: que implica que la ocurrencia de un evento es independiente de la ocurrencia de otros eventos, a esta propiedad también se le conoce como "completamente aleatorio". Por lo cual la medida de segundo orden queda definida como:

$$\lambda_2(s, u) = \lim_{\substack{v(ds) \rightarrow 0 \\ v(du) \rightarrow 0}} \frac{\mu_N^{(2)}(ds \times du)}{v(ds)v(du)} = \lim_{\substack{v(ds) \rightarrow 0 \\ v(du) \rightarrow 0}} \frac{\lambda(s)v(ds)\lambda(u)v(du)}{v(ds)v(du)} = \lambda(s)\lambda(u) = \lambda^2$$

Un proceso homogéneo de Poisson al ser estacionario e isotrópico es por lo tanto invariante.

## 2.3 La Función K-Ripley

La función K función se conoce como función K de Ripley porque fue Ripley (1976) quien escribió toda la teoría de las propiedades de los procesos puntuales de segundo orden, tomando en cuenta ya no solo el vecino más cercano si no todos los vecinos. Debido a ello su importancia radica en describir la dependencia espacial a lo largo de una variedad de escalas de un proceso puntual. Esta función ha sido usada desde entonces y se ha convertido en una medida estándar presentada en la bibliografía de procesos puntuales (Ripley, 1981; Cressie, 1993).

La función K de Ripley está definida por  $\mathcal{K}_N$  :

$$\mathcal{K}_N = \lambda^{-1}(\mathbf{h})E(N_0(\mathbf{h}))$$

donde  $E(N_0(h))$  es el número esperado de eventos dentro de una distancia  $h$  de un evento aleatorio  $N$ .

Los procesos homogéneos de Poisson son usados de forma práctica para calcular la función teórica de la función K de Ripley.

### 2.3.1 Función K de Ripley teórica para un PPH

Para un proceso homogéneo de Poisson la función K de Ripley está definida como:

$$\mathcal{K}_N = \lambda^{-1} \int_0^h \frac{a}{\lambda} \lambda_2^\circ(u) du$$

donde  $a$  es el área de una esfera de radio  $h$  en el espacio bidimensional ( $d=2$ ):

$$a = \frac{d\pi^{d/2}h^{d-1}}{\Gamma(1 + \frac{d}{2})} = 2\pi h$$

En un proceso homogéneo de Poisson  $\lambda_2^\circ = \lambda^2$ , entonces la función de K-Ripley para un proceso homogéneo de Poisson es:

$$\mathcal{K}_N = \pi h^2$$

### **2.3.2 Estimación General de la función K-Ripley:**

La función K de Ripley está definida por  $\mathcal{K}_N$ :

$$\mathcal{K}_N = \lambda^{-1}(\mathbf{h})E(N_0(\mathbf{h}))$$

En la práctica se puede estimar  $E(N_0(\mathbf{h}))$  en base a las distancias entre eventos (Diggle et al., 1983):

$$\hat{E}(N_0(\mathbf{h})) = n^{-1} \sum_{i=1}^n \sum_{j \neq i} I(\mathbf{u}_{ij} \leq \mathbf{h})$$

donde  $u_{ij} = \|x_i - x_j\|$  es la distancia entre los eventos  $x_i$  y  $x_j$ ;

$$I = \begin{cases} 1; & \text{si } u_{ij} \leq h \\ 0; & \text{si } u_{ij} > h \end{cases}$$

Entonces la función de K de Ripley estimada es:

$$\hat{\mathcal{K}}_N = \lambda^{-1}(\mathbf{h})n^{-1} \sum_{i=1}^n \sum_{j \neq i} I(\mathbf{u}_{ij} \leq \mathbf{h})$$

Entonces la función K de Ripley (Fig. 14a) consiste en tomar un evento fijo e ir contando cuantos puntos lo rodean a diferentes distancias  $h$ , lo mismo que equivale a estudiar la distribución a distintas escalas espaciales, de igual forma se procede para los demás eventos.

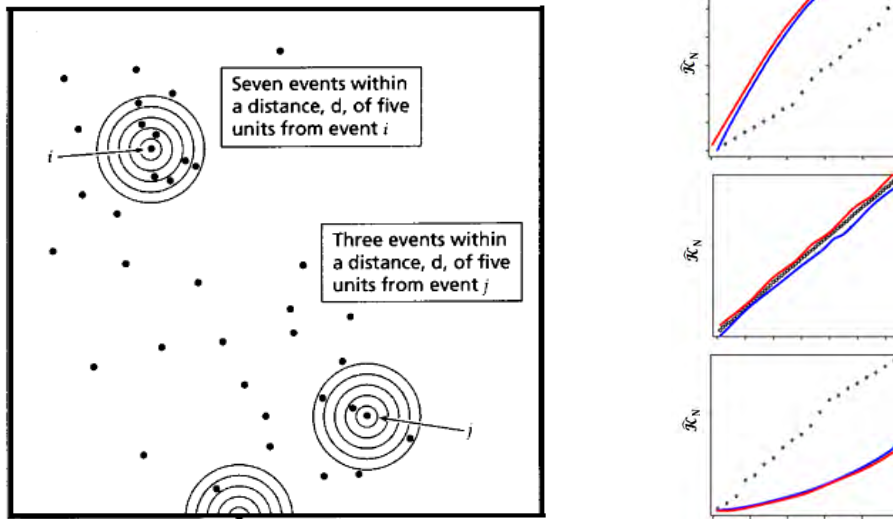
### **Estimación de la función K-Ripley para un PPH**

Para estimar la función K-Ripley bajo el supuesto de un proceso homogéneo de Poisson se considera la intensidad constante, entonces simplemente la función K de Ripley estimada es:

$$\hat{\mathcal{K}}_N(\mathbf{h}) = \lambda^{-1}n^{-1} \sum_{i=1}^n \sum_{j \neq i} I(\mathbf{u}_{ij} \leq \mathbf{h}) = A \cdot n^{-2} \sum_{i=1}^n \sum_{j \neq i} I(\mathbf{u}_{ij} \leq \mathbf{h})$$

Considerando la función K de Ripley teórica y la estimada se busca probar la hipótesis nula que la distribución espacial de un proceso es completamente aleatoria, lo que implica que la distribución de los eventos observados consiste en un proceso de Poisson Homogéneo y que por lo tanto  $\hat{\mathcal{K}}_N(h) = \pi h^2$ . En otro caso se puede pensar que si  $\hat{\mathcal{K}}_N(h) > \pi h^2$  el proceso puntual es agregado, o si  $\hat{\mathcal{K}}_N(h) < \pi h^2$  el proceso puntual es regular. Pero para aceptar que la función  $\hat{\mathcal{K}}_N(h)$  estimada es realmente significativa se usan técnicas de simulación para calcular los intervalos de aceptación de CSR. Para eso se hace  $m$  simulaciones independientes de  $n$  eventos en la región de estudio (donde  $m$  puede ser por ejemplo 99). Para cada patrón puntual simulado, se puede estimar  $\hat{\mathcal{K}}_N(h)$  y usar el cuantil 0.05 y cuantil 0.95 de estas funciones estimadas para definir una aproximación de la envoltura del intervalo simulado teórico.

Entonces si el valor estimado de  $\widehat{\mathcal{K}}_N(h)$  está entre los dos intervalos se puede concluir que se trata de un CSR, pero si cae por encima del intervalo mayor entonces se trata de un proceso agregado, mientras si cae por debajo del intervalo menor, hay evidencia de inhibición espacial o regularidad entre los eventos (Fig. 14b).



**Figura 14. (a) Izquierda: Estimación de una función K de Ripley para eventos dentro de una distancia d (Fuente: Gatrell et al., 1996). (b) Derecha: Función  $\widehat{\mathcal{K}}_N(h)$  de los patrones uniforme o regular (arriba), aleatorio (centro) y agregado (abajo); la línea punteada negra representa el valor empírico de la función K de Ripley, y las líneas roja y azul los intervalos de aceptación de CSR (Fuente: Betrand S, 2006).**

## 2.4 La función L

La función L propuesta por Besag (Ripley., 1977), y es una función lineal muy usada en la práctica porque gráficamente es más fácil de interpretar que la función K, además que la varianza de L es aproximadamente constante bajo CSR (Ripley et al., 1979). Se define la función L estimada como:

$$\widehat{L}(h) = [\widehat{\mathcal{K}}_N(h)/\pi]^{1/2} - h$$

En el caso de CSR la función L teórica viene dada por el valor teórico de la función K de Ripley, entonces:

$$L(h) = [\pi h^2/\pi]^{1/2} - h = 0$$

Entonces comparando el valor de la función L teórica con el valor de la función L estimado, aceptamos la hipótesis nula que el proceso es CSR si  $\widehat{L}(h) - t = 0$ , o caso contrario se puede pensar que si  $\widehat{L}(h) - t > 0$  el proceso puntual es agregado, o si  $\widehat{L}(h) - t < 0$  el proceso puntual es regular (Fig. 15).

Pero para aceptar que la función  $\widehat{L}(h) - t$  estimada es realmente significativa se usan técnicas de simulación para calcular los intervalos de aceptación de CSR. Para eso se hacen  $m$  simulaciones independientes de  $n$  eventos en la región de estudio (donde  $m$  puede ser por ejemplo 99). Para cada patrón puntual simulado, se puede estimar  $\widehat{L}(h) - t$  y usar el cuantil 0.05 y cuantil 0.95 de estas funciones estimadas para definir una aproximación de la envoltura del intervalo simulado teórico. Entonces si el valor estimado de  $\widehat{L}(h) - t$  está entre los dos intervalos se puede concluir que se trata de un CSR, pero si cae arriba del intervalo mayor entonces se trata de un proceso agregado, mientras si cae por debajo del intervalo menor, hay evidencia de inhibición espacial o regularidad entre los eventos.

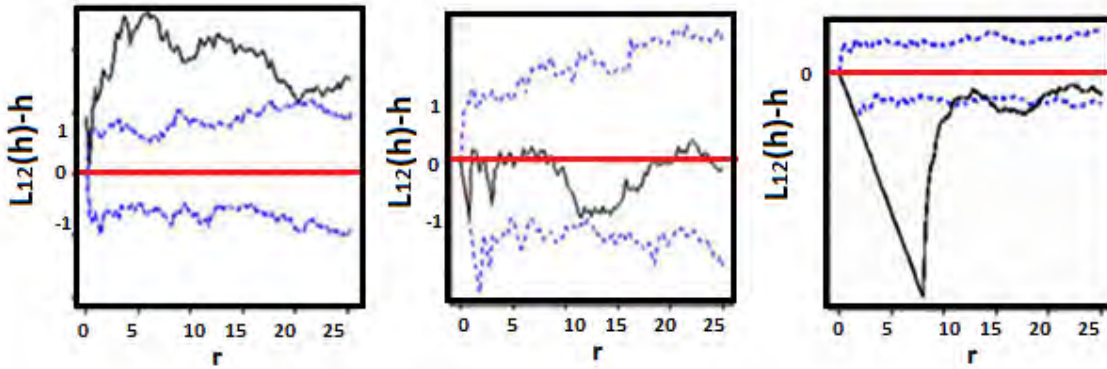


Figura 15. Función  $\hat{L}(h)$  de los patrones aleatorio (izquierda), agrupado (centro) y uniforme (derecha); la línea roja representa el valor teórico de la función  $L$ , la línea negra el valor estimado de la función  $L$  y las líneas azules los intervalos de aceptación de CSR.

## 2.5 Procesos Puntuales Multivariados

La estimación de la función  $K$  de Ripley para un único conjunto de eventos no siempre es muy informativa. Por eso surge la necesidad de considerar un proceso puntual multivariado que es un caso especial de un proceso puntual marcado, donde las marcas son variables categóricas. Estas marcas son importantes porque permiten describir un proceso puntual que incluye información adicional de las ubicaciones de cada evento, de tal forma que el proceso puntual se clasifica según las marcas en dos o más tipos (según especies, casos/control, etc.).

Formalmente la medida de momentos para un proceso puntual bivariado pueden ser definida de igual forma que para el caso univariado. Por consiguiente sea el proceso puntual  $N=(N_1, N_2)$ , los conjunto de Borel  $B_i$  y el número de puntos  $N(B_i)$  en  $B_i$  variables aleatorias, la medida del segundo momento cruzado está dada por:

$$\mu_{N_1 N_2}(B_1 \times B_2) = E(N_1(B_1)N_2(B_2)) = \int_{\phi} \phi_1(B_1)\phi_2(B_2) \prod_N d\phi$$

donde  $\prod_N d\phi = p(N \in d\phi)$  y  $\phi(B) = \sum_{i=1}^{\infty} I(s_i \in B)$

Por otro lado la intensidad de primer orden se define igual que en el caso univariado. Mientras que la intensidad de segundo orden está definida por:

$$\lambda_{12}(s, u) = \lim_{\substack{v(ds) \rightarrow 0 \\ v(du) \rightarrow 0}} \frac{\mu_{N_1 N_2}(ds \times du)}{v(ds)v(du)}$$

La generalización de la función  $K$  de Ripley para dos tipos de eventos  $j=1,2$  conocida como **dobles K de Ripley** es:

$$\mathcal{K}_{ij}(h) = \lambda_j^{-1}(h)E(N_{0ij}(h))$$

donde  $N_0(h)$  es el número de eventos de tipo  $j$  dentro de una distancia  $h$  de un evento aleatorio de tipo  $i$ , y  $\lambda_j$  es la intensidad del tipo de evento  $j$ .

Las relaciones entre dos procesos puntuales pueden ser planteadas mediante hipótesis distintas:

- Etiquetado aleatorio: dadas las ubicaciones, las marcas son condicionalmente independientes e idénticamente distribuidas.
- Independencia de componentes: los subprocesos de eventos de cada marca, son procesos puntuales independientes.

La primera hipótesis responde a si existe una relación entre las distribuciones relativas de los eventos de los dos procesos puntuales, es decir si existe interacción entre los dos procesos y la segunda hipótesis responde a si existe una relación entre las distribuciones de las densidades asociadas a los eventos de cada proceso.

### **Función Doble K de Ripley teórica para PPH**

Para procesos homogéneos de Poisson la función doble K de Ripley está definida como:

$$\mathcal{K}^{12}(h) = \lambda_1^{-1} \lambda_2^{-1} \int_0^h a \lambda_{12}^\circ(u) du$$

donde  $a$  es el área de una esfera de radio  $h$  en el espacio bidimensional ( $d=2$ ):

$$a = \frac{d\pi^{d/2} h^{d-1}}{\Gamma(1 + \frac{d}{2})} = 2\pi h$$

Asumiendo que  $N_1$  y  $N_2$  son procesos independientes  $\lambda_{12}^\circ = \lambda_1(s)\lambda_2(u)$ , entonces la función doble K de Ripley es:

$$\mathcal{K}^{12}(h) = \pi h^2$$

### **Estimación General de la función Doble K de Ripley:**

Para estimar la función doble K de Ripley bajo el supuesto de un proceso homogéneo de Poisson se considera la intensidad constante, entonces simplemente la función K de Ripley estimada es:

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

donde:  $u_{ik,jl} = \|x_{ik} - x_{jl}\|$  es la distancia entre la  $k$ -ésima ubicación del evento de tipo  $i$  y la  $l$ -ésima ubicación del evento de tipo  $j$  y  $A$  es el área de la región de estudio.

Si los eventos de tipo 1 y 2 son generados por procesos independientes entonces  $\hat{\mathcal{K}}_{12}(h) = \pi h^2$ . Si los eventos de tipo 1 y 2 son generados por etiquetado aleatorio y los tipos son asignados por los resultados de una secuencia de ensayos independientes de Bernoulli, entonces  $\hat{\mathcal{K}}_{11}(h) = \hat{\mathcal{K}}_{22}(h) = \hat{\mathcal{K}}_{12}(h)$ . Y en particular la independencia y el etiquetado aleatorio son equivalentes si y solo si los componentes del proceso son procesos homogéneos de Poisson (Diggle et al., 2006).

Pero para aceptar que la función  $\hat{\mathcal{K}}_{12}(h)$  estimada es realmente significativa se usan técnicas de simulación para calcular los intervalos de aceptación de CSR como en el caso univariado. Si el valor estimado de  $\hat{\mathcal{K}}_{12}(h)$  está entre los dos intervalos se puede concluir que no existe interacción entre los procesos, pero si cae por encima del intervalo mayor entonces se trata procesos agregados, mientras si cae por debajo del intervalo menor, hay evidencia de inhibición espacial o regularidad entre los procesos. Además se puede usar de forma similar al caso univariado la función L, definida como:

$$\hat{L}_{12}(h) = [\hat{\mathcal{K}}_{12}(h)/\pi]^{1/2} - h$$

Y cuyo valor teórico es  $L_{12} = 0$ .

### 3. Modelos Lineales Generalizados

#### 3.1 Introducción

En el análisis univariado el modelo lineal general surge de la necesidad de cuantificar las relaciones entre un conjunto de variables, en la que una de ellas se denomina variable respuesta o dependiente y las restantes son las variables explicativas o independientes. Este modelo asume que la variable dependiente sigue una distribución normal y homocedasticidad (McCulloch y Searle, 2001), supuestos que no siempre se cumplen, surgiendo la necesidad de estudiar otros modelos. Por ejemplo, es muy común en ecología que a medida que aumenta la media de la muestra, aumente también su varianza. Estos problemas se pueden llegar a solucionar mediante la transformación de la variable respuesta (por ejemplo tomando logaritmos). Sin embargo estas transformaciones no siempre consiguen corregir la falta de normalidad, la heterocedasticidad (varianza no constante) o la no linealidad de nuestros datos (Cayuela., 2009). Una alternativa a la transformación de la variable respuesta y a la falta de normalidad es el uso de los modelos lineales generalizados.

Los modelos lineales generalizados (MLG) son una extensión del clásico Modelo Lineal a una familia más general, involucran una variedad de distribuciones seleccionadas de una familia exponencial especial (McCulloch y Searle, 2001; Lindsay, 1997). Hace más de un siglo se estudian métodos estadísticos que son casos especiales de GLM, su historia se puede remontar muy brevemente en la Tabla 1, el término fue por primera vez introducido por Nelder y Wedderburn (1972) y progresivamente se han aplicado en diversas disciplinas como agricultura, demografía, ecología, economía, ingeniería, geografía, geología, historia, medicina, ciencias políticas, psicología y sociología (Lindsay, 1997).

Tabla 1.

<b>AÑO</b>	<b>ESTUDIOS</b>	<b>AUTORES</b>
1900	· <i>Regresión Lineal Múltiple</i>	<i>Legendre-Gauss</i>
1920-1935	· <i>ANOVA</i>	<i>Fisher</i>
1922	· <i>Función de enlace</i>	<i>Fisher</i>
1922	· <i>Ensayos de dilución</i>	<i>Fisher</i>
1934	· <i>Familia Exponencial</i>	<i>Fisher</i>
1935	· <i>Análisis Probit</i>	<i>Bliss</i>
1944-1952	· <i>Proporciones para logit</i>	<i>Berkson-Dyke y Patterson</i>
1960	· <i>Análisis Item</i>	<i>Rasch</i>
1963	· <i>Modelos de conteo log lineales</i>	<i>Birch</i>
1965-1966-1967	· <i>Modelos de regresión de supervivencia</i>	<i>Feigl y Zelen, Zippin y Armitage, Glasser</i>
1966	· <i>Polinomios inversos</i>	<i>Nelder</i>
1972	· <b><i>Modelos Lineales Generalizados</i></b>	<b><i>Nelder y Wedderburn</i></b>

#### 3.2. Definición de los MLG

Los MLG están formados por tres componentes descritos a continuación.

##### 3.2.1 Componente aleatorio:

Sea el vector aleatorio  $Y = (y_1, y_2, \dots, y_n)'$  cuyos elementos son independientes e idénticamente distribuidos. La función de distribución de este vector pertenece a la familia exponencial y tiene la siguiente forma:

$$f(y, \theta, \phi) = \exp\left\{\frac{1}{a(\phi)}(y\theta - b(\theta)) + c(y, \phi)\right\}; \phi > 0$$

donde  $\theta$  es un parámetro canónico,  $\phi$  es el parámetro de dispersión y  $a(\cdot)$ ,  $b(\cdot)$  y  $c(\cdot)$  son funciones conocidas.

Por ejemplo la función de densidad de la distribución normal puede ser reescrita como:

$$f(y; \mu) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right]$$

$$f(y; \mu, \sigma) = \exp\left[\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2}\left(\frac{y^2}{\sigma^2} - \log(2\pi\sigma^2)\right)\right]$$

Donde:

$$a(\phi) = \sigma^2; b(\theta) = \frac{\mu^2}{2}; c(y, \phi) = \frac{1}{2}\left(\frac{y^2}{\sigma^2} - \log(2\pi\sigma^2)\right);$$

Una distribución normal es, por definición, continua. En ocasiones, sin embargo, la variable dependiente sigue una distribución que no es continua y, por tanto, los valores estimados por el modelo han de seguir el mismo tipo de distribución que los datos de partida. Para detectar si nuestros datos son o no normales es conveniente: (1) conocer el tipo de variable respuesta y su naturaleza; y (2) analizar los residuos del modelo una vez ajustado el modelo (ya sea un modelo lineal o un GLM con una distribución de errores no normal). Son parte de la familia exponencial las funciones de distribución: Normal, Binomial, Poisson, Gamma y Binomial Negativa, ya que pueden reescribirse en la forma  $f(y, \theta, \phi)$  como se muestra en la Tabla 2.

El empleo de una u otra distribución depende de la naturaleza de la variable respuesta, así por ejemplo se puede usar la distribución:

- **Binomial:** para proporciones y datos de presencia/ausencia (por ejemplo tasas de mortalidad; tasas de infección; porcentaje de parasitismo; porcentaje de éxito reproductivo; presencia o ausencia de una determinada especie).
- **Poisson o Binomial Negativa:** para conteos de casos (por ejemplo abundancia de una especie, número de muertos por accidentes de tráfico; número de días con heladas en el mes de enero; número de colonias de bacterias).
- **Gamma:** con datos continuos que muestran un coeficiente de variación constante, esto es, en donde la varianza aumenta según aumenta la media de la muestra de manera constante (por ejemplo número de presas comidas por un predador en función del número de presas disponibles).

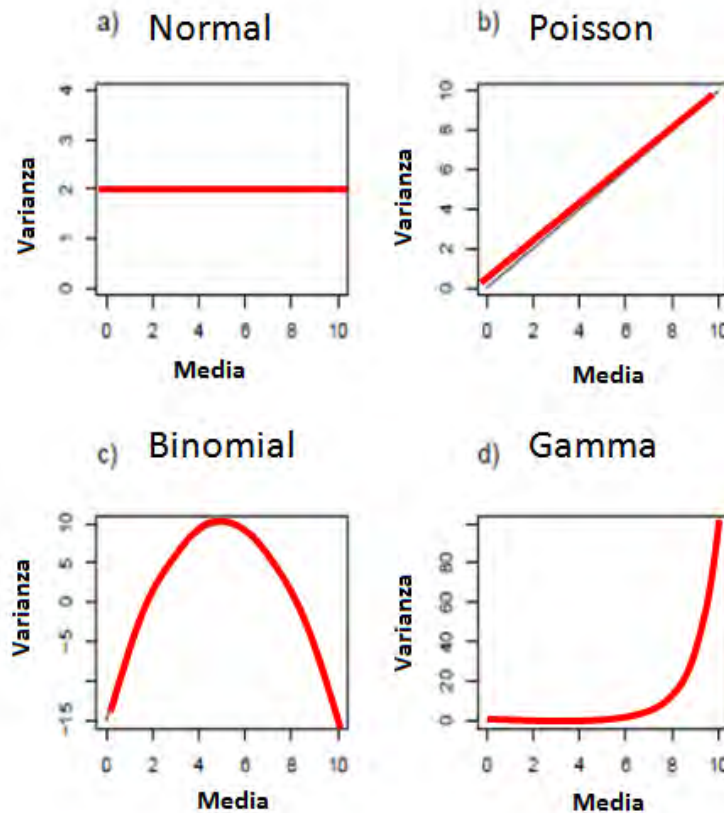
Tabla 2. Algunas distribuciones de la Familia Exponencial

Distribución	$a(\phi)$	$\theta$	$b(\theta)$	$c(y, \phi)$	$\mu(\theta)$	$V(\mu)$
Normal: $N(\mu, \sigma^2)$	$\sigma^2$	$\mu$	$\frac{\theta^2}{2}$	$-\frac{1}{2}\left[\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\right]$	$\theta$	1
Poisson: $P(\mu)$	1	$\log(\mu)$	$e^\theta$	$-\log(y!)$	$e^\theta$	$\mu$
Binomial: $B(m, \pi)$	1	$\log\left(\frac{\mu}{m - \mu}\right)$	$m \log(1 + e^\theta)$	$\log\binom{m}{y}$	$\frac{me^\theta}{1 + e^\theta}$	$\frac{\mu}{m}(m - \mu)$
Binomial Negativa: $BN(\mu, k)$	1	$\log\left(\frac{\mu}{\mu + k}\right)$	$-k \log(1 - e^\theta)$	$\log\left(\frac{\Gamma(k + y)}{\Gamma(k)y!}\right)$	$\frac{ke^\theta}{1 - e^\theta}$	$\mu\left(\frac{\mu}{k} + 1\right)$
Gamma: $G(\mu, v)$	$v^{-1}$	$-\frac{1}{\mu}$	$-\log(-\theta)$	$v \log(vy) - \log(y) - \log\Gamma(v)$	$-\frac{1}{\theta}$	$\mu^2$
Normal Inversa: $IG(\mu, \sigma^2)$	$\sigma^2$	$-\frac{1}{2\mu^2}$	$-(-2\theta)^{1/2}$	$-\frac{1}{2}\left[\log(2\pi\sigma^2) + \frac{1}{\sigma^2 y}\right]$	$(-2\theta)^{-1/2}$	$\mu^3$



### Media y Varianza

El supuesto central en los modelos lineales es que la varianza es constante (Fig. 16a). En el caso de los conteos, sin embargo, donde la variable respuesta esta expresada en números enteros y en donde hay a menudo muchos ceros en los datos, la varianza podrá incrementar linealmente con la media (Fig. 16b). Con proporciones, donde hay un conteo del número de fallos de un evento, así como del número de éxitos, la varianza tendrá una forma de U invertida en relación a la media (Fig. 16c). Cuando la variable respuesta siga una distribución Gamma, entonces la varianza incrementa de una manera no lineal con la media (Fig. 16d).



**Figura 16: Relación entre la media y la varianza de los datos bajo distintos supuestos; (a) la varianza es constante; (b) la varianza incrementa con la media; (c) la varianza tiene forma de U invertida en relación a la media; (d) la varianza se incrementa de manera no lineal con la media. (Fuente: Cayuela, 2009)**

Para calcular la media de  $y_i$  en los GLM se usa la función de máxima verosimilitud:

$$l(y_i, \theta, \phi) = \log f_Y(y_i, \theta, \phi);$$

$$l(y_i, \theta) = \frac{1}{a(\phi)} (y_i \theta - b(\theta)) + c(y_i, \phi);$$

$$E\left(\frac{dl}{d\theta}\right) = 0; \frac{dl}{d\theta} = \frac{\{y_i - b'(\theta)\}}{a(\phi)}; \text{entonces } E\left(\frac{dl}{d\theta}\right) = E\left(\frac{\{y_i - b'(\theta)\}}{a(\phi)}\right) = 0;$$

$$E\left(\frac{dl}{d\theta}\right) = \frac{\{\mu_i - b'(\theta)\}}{a(\phi)} = 0;$$

$$E(y_i) = \mu_i = b'(\theta)$$

Para calcular la varianza se procede de forma similar:

$$E\left(\frac{d^2l}{d\theta^2}\right) + E\left(\frac{dl}{d\theta}\right)^2 = 0;$$

$$\frac{d^2l}{d\theta^2} = \frac{b''(\theta)}{a(\phi)};$$

$$E\left(\frac{b''(\theta)}{a(\phi)}\right) = -\frac{b''(\theta)}{a(\phi)};$$

$$E\left(\frac{d^2l}{d\theta^2}\right) + E\left(\frac{dl}{d\theta}\right)^2 = -\frac{b''(\theta)}{a(\phi)} + \left[\frac{\{\mu_i - b'(\theta)\}}{a(\phi)}\right]^2 = -\frac{b''(\theta)}{a(\phi)} + \frac{Var(Y)}{(a(\phi))^2} = 0;$$

$$Var(y_i) = a(\phi)b''(\theta)$$

### 3.2.2 Componente sistemático:

Como en el caso de los modelos lineales, el componente sistemático está dado por el predictor lineal.

$$\eta = \sum_{j=0}^n x_j \beta_j$$

donde las  $x_j$  representan las covariable o variables explicativas del modelo que no pueden estar altamente correlacionadas y los  $\beta_j$  son parámetros cuyos valores son desconocidos y necesitan ser estimados.

### 3.2.3 Función de enlace:

La función de enlace o cadena es útil en los MLG porque permite restringir los valores de  $\mu$ , puesto que tiene en cuenta otras distribuciones a parte de la normal. Por ejemplo si la variable respuesta sigue una distribución de Poisson, la media toma valores positivos mientras que  $\eta$  no necesariamente, en este caso una función de enlace como el logaritmo asegura que la media sea positiva (Fig. 17). La función de enlace  $g(\mu)$ , relaciona el componente aleatorio y el componente sistemático vinculando el predictor lineal  $\eta$  con el valor esperado de la variable respuesta,  $E(Y/X) = \mu$ , a través de la función

$$g(\mu) = X\beta$$

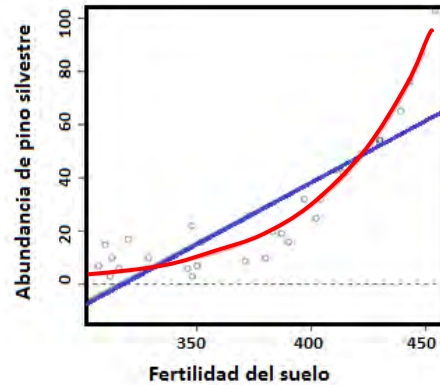
$$\eta = g(\mu); \mu = g^{-1}(\eta)$$

donde  $g$  es una función diferenciable, monótona e invertible.

Las tres principales funciones de enlace son:

- Función logit:  $\eta = \log \left\{ \frac{\mu}{1-\mu} \right\}$
- Función Probit:  $\eta = \Phi^{-1}(\mu)$
- Función Log-log complementaria:  $\eta = \log \{-\log(1 - \mu)\}$

Otras funciones de enlace según la distribución se muestran en la Tabla 3.



**Fig. 17. Relación entre la abundancia de una especie (por ejemplo, número de pies de pino silvestre en una parcela forestal) y la fertilidad del suelo. Los valores quedan acotados por encima de 0 cuando utilizamos una función de enlace de tipo logarítmica (curva roja). La línea azul representa el ajuste del modelo sin usar una función de enlace. (Fuente: Cayuela, 2009)**

**Tabla 3. Familias y función de enlace.  
(Función de enlace por defecto en el paquete R es D).**

Función de enlace	Binomial	Gamma	Gaussiana	Poisson	Binomial Negativa	Gaussiana Inversa
Logit	D					
Probit	x					
Cloglog	x					
Identidad		D	D	x	x	
Inversa		x				
Log		x		D	D	
Raíz				x	x	
$1/\mu^2$						D

### 3.3 Estimación de los parámetros

Los GLM utilizan varios métodos para estimar los parámetros, el óptimo método es por máxima verosimilitud, el cual tiene las propiedades de consistencia y eficiencia asintótica (Cordeiro y Demetrio, 2000). Las estimaciones de  $\beta$  por medio de este método son obtenidas vía el algoritmo conocido como proceso de mínimos cuadrados ponderados iterativos (Iteratively reweighted least squares - IWLS; McCullagh y Nelder, 1983).

$$f(y_1, \dots, y_N; \theta_1, \dots, \theta_N) = \prod_{i=1}^N \exp \left\{ \frac{1}{a(\phi)} (y_i \theta - b(\theta)) + c(y_i, \phi) \right\}$$

$$f(y_1, \dots, y_N; \theta_1, \dots, \theta_N) = \exp \left[ \sum_{i=1}^N \frac{1}{a(\phi)} y_i \theta + \sum_{i=1}^N \frac{1}{a(\phi)} b(\theta) + \sum_{i=1}^N c(y_i, \phi) \right];$$

$$l = \sum_{i=1}^N \frac{1}{a(\phi)} [y_i \theta + b(\theta)] + \sum_{i=1}^N c(y_i, \phi)$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i \frac{d\theta}{d\beta} - \frac{db(\theta)d\theta}{d\theta d\beta}] = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \frac{d\theta}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \frac{d\theta d\mu_i}{d\mu_i d\beta}$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \frac{d\theta d\mu_i d(g(\mu_i))}{d\mu_i d(g(\mu_i)) d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \frac{d\theta}{d\mu_i} \left[ \frac{d(g(\mu_i))}{d\mu_i} \right]^{-1} \frac{dx'_i \beta}{d\beta}$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \frac{d\theta}{d\mu_i} \left[ \frac{d(g(\mu_i))}{d\mu_i} \right]^{-1} x'_i$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \left[ \frac{d\mu_i}{d\theta} \right]^{-1} \left[ \frac{d(g(\mu_i))}{d\mu_i} \right]^{-1} x'_i$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \left[ \frac{d\mu_i}{d\theta} \right]^{-1} \left[ \frac{d(g(\mu_i))}{d\mu_i} \right]^{-1} x'_i$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] \left[ \frac{d^2 b_i}{d\theta^2} \right]^{-1} \left[ \frac{d(g(\mu_i))}{d\mu_i} \right]^{-1} x'_i$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} \sum_{i=1}^N [y_i - \mu_i] [b''(\theta_i)]^{-1} \left[ \frac{d(g(\mu_i))}{d\mu_i} \right]^{-1} x'_i$$

$$\frac{dl}{d\beta} = \frac{1}{a(\phi)} [Y - \mu] W \Delta X' = \mathbf{0} \dots (i)$$

$$\text{donde } W = \left[ b''(\theta_i) \left( \frac{d(g(\mu_i))}{d\mu_i} \right)^2 \right]^{-1}; \Delta = \frac{d(g(\mu_i))}{d\mu_i}$$

$$Y W \Delta X' = \mu W \Delta X' \dots (1)$$

Los parámetros  $\mu$ ,  $W$  y  $\Delta$  son funciones no lineales, por lo tanto no pueden ser resueltas analíticamente, es necesario usar procesos numéricos iterativos para estimar  $\beta$ .

El método de Newton-Raphson es un método numérico en el cual una solución de una ecuación  $f(x) = 0$  está basada en una aproximación de Taylor para una función  $f(x)$  en la vecindad del punto  $x_0$ , es decir:

$$f(x) = f(x_0) + (x - x_0)f'(x_0) = 0$$

$$\text{obteniéndose } x = x_0 - \frac{f(x_0)}{f'(x_0)}, \text{ o de una forma más general } x^{(m+1)} = x^{(m)} - \frac{f(x^{(m)})}{f'(x^{(m)})}$$

Ahora considerando que se desea obtener una solución del sistema de ecuaciones de  $\frac{dl}{d\beta} = 0$  y usando una versión multivariada del método de Newton-Raphson, se tiene que  $\beta^{(m+1)} = \beta^{(m)} + [I(\beta^{(m)})]^{-1} \frac{\partial l}{\partial \beta} \Big|_{\beta=\beta^{(m)}}$  donde  $I(\beta^{(m)})$  es la inversa de las derivadas de segundo orden de  $l(\beta)$ .

Cuando las derivadas parciales de segundo orden son fácilmente halladas el método de Newton-Raphson es bastante útil. Otro método muy usado es el método de puntuación de Fisher, en general, más sencillo (que coincide con el método de Newton-Raphson en el caso de las funciones de enlace canónico). Este método implica una sustitución de la matriz de derivadas parciales de segundo orden por la matriz de valores esperados de derivadas parciales, es decir, la matriz de información observada por la matriz de información esperada de Fisher.

La información de Fisher mide que tanta información da una variable aleatoria  $X = (x_1, \dots, x_n)$  sobre un parámetro desconocido  $\theta$ . Si el logaritmo de  $f(X, \theta)$  es doblemente diferenciable respecto a  $\theta$ , la información de Fisher puede ser escrita como

$$I(\theta) = -E \left[ \frac{\partial^2 l(Y, \theta)}{\partial \theta \partial \theta'} \right] \dots (*)$$

Usando (\*) calculamos la Información de Fisher de  $\beta$ :

$$\frac{\partial^2 l(Y, \beta)}{\partial \beta \partial \beta'} = \frac{\partial}{\partial \beta'} \left( \frac{dl}{d\beta} \right) = \frac{\partial}{\partial \beta'} \left( \frac{1}{a(\phi)} [Y - \mu] W \Delta X' \right)$$

$$\frac{\partial^2 l(Y, \beta)}{\partial \beta \partial \beta'} = \frac{1}{a(\phi)} \left\{ -W \Delta X' \frac{\partial \mu}{\partial \beta'} + [Y - \mu] \frac{\partial W \Delta}{\partial \beta'} X' \right\}$$

$$-E \left[ \frac{\partial^2 l(Y, \beta)}{\partial \beta \partial \beta'} \right] = -E \left[ \frac{1}{a(\phi)} \left\{ -W \Delta X' \frac{\partial \mu}{\partial \beta'} + [Y - \mu] \frac{\partial W \Delta}{\partial \beta'} X' \right\} \right] = \frac{1}{a(\phi)} W \Delta X' \frac{\partial \mu}{\partial \beta'} + 0$$

$$-E \left[ \frac{\partial^2 l(Y, \beta)}{\partial \beta \partial \beta'} \right] = \frac{1}{a(\phi)} W \Delta X' \frac{\partial \mu}{\partial g(\mu)} \frac{\partial g(\mu)}{\partial \beta'} = \frac{1}{a(\phi)} W \Delta X' \left( \frac{\partial g(\mu)}{\partial \mu} \right)^{-1} \frac{\partial X' \beta}{\partial \beta'}$$

$$-E \left[ \frac{\partial^2 l(Y, \beta)}{\partial \beta \partial \beta'} \right] = \frac{1}{a(\phi)} X' W \Delta \left( \frac{\partial g(\mu_i)}{\partial \mu} \right)^{-1} X = \frac{1}{a(\phi)} X' W \Delta \Delta^{-1} X = \frac{1}{a(\phi)} X' W X$$

La matriz de información de Fisher de  $\beta$  es consistentemente estimada por:

$$I(\beta) = \frac{1}{a(\phi)} X' W X \dots (ii)$$

Entonces como:

$$\beta^{(m+1)} = \beta^{(m)} + [I(\beta^{(m)})]^{-1} \left. \frac{\partial l}{\partial \beta} \right|_{\beta=\beta^{(m)}}$$

De (i) y (ii) tenemos que:

$$\beta^{(m+1)} = \beta^{(m)} + \left[ \frac{1}{a(\phi)} X' W X \right]^{-1} \frac{1}{a(\phi)} [Y - \mu] W \Delta X'$$

$$\beta^{(m+1)} = \beta^{(m)} + [X' W X]^{-1} [Y - \mu] W \frac{d(g(\mu_i))}{d\mu_i} X'$$

$$\beta^{(m+1)} = [X' W^{(m)} X]^{-1} W^{(m)} X' z^{(m)}$$

$$\text{donde } W^{(m)} = \left[ b''(\theta_i) \left( \frac{d(g(\mu_i))}{d\mu_i} \right)^2 \right]^{-1} ; \eta_i^{(m)} = \sum_{r=1}^p x_{ir} \beta_r^{(m)}$$

$$z^{(m)} = \eta_i^{(m)} + [y_i - \mu_i^{(m)}] g'(\mu_i^{(m)})$$

Un método usual para iniciar el proceso iterativo es especificar una estimación inicial, así para calcular  $\beta^{(2)}$  cada observación puede ser considerada como una estimación de su valor medio, es decir:

$$\mu_i^{(1)} = y_i \text{ y así se calcula } \eta_i = g(\mu_i^{(1)}) = g(y_i), \text{ y } w_i^{(1)} = \left[ b''(\theta_i) \left( \frac{d(g(y_i))}{dy_i} \right)^2 \right]^{-1}$$

donde  $\theta_i$  esta en función de  $\mu_i$  y por lo tanto de  $y_i$ .

$$\beta^{(2)} = [X'W^{(1)}X]^{-1}W^{(1)}X' \eta_i^{(1)}$$

Se usa la función  $\beta^{(m+1)} = [X'W^{(m)}X]^{-1}W^{(m)}X'z^{(m)}$  para  $m=3, \dots, k$ , siendo  $k-1$  el número necesario de iteraciones para la convergencia, de tal forma que el parámetro  $\hat{\beta}$  no cambie significativamente ya que  $\beta^{(k+1)}$  se aproxima a  $\hat{\beta}$  cuando  $k$  es lo suficientemente grande.

Sea  $\hat{\beta}$  el valor final de los coeficientes de regresión y sin son correctas las hipótesis del modelo, incluyendo la elección de la función de enlace se puede demostrar que asintóticamente:

(i) El estimador de  $\beta$  es asintóticamente estimado, es decir para muestras grandes:  $E(\hat{\beta}) = \beta$

(ii) La varianza del estimador de  $\beta$  es:  $Var(\hat{\beta}) = a(\phi)(X'V^{-1}X)^{-1}$

Siendo  $V$  una matriz diagonal formada por las varianzas de los parámetros estimados en el predictor lineal, aparte de  $a(\phi)$ .

### 3.4 Evaluación del MLG

La aplicación de MLG incluye una correcta especificación de la función de enlace, un adecuado número de variables explicativas así como una correcta escala de medida de las mismas y la no influencia indebida de ciertas observaciones. Así por ejemplo si se admite una combinación satisfactoria de la variable respuesta y la función de enlace, el primer objetivo será determinar cuántos términos son necesarios en la estructura lineal para una descripción razonable de los datos. Pues un número grande de variables explicativas puede conducir a un modelo que explique bien los datos pero con un alto grado de complejidad en la interpretación. Por otro lado, un pequeño número de variables explicativas puede conducir a un modelo de interpretación fácil pero que se ajuste pobremente a los datos.

El *modelo nulo* es el modelo más simple, el cual se utiliza como modelo de referencia. Contiene como único parámetro al valor esperado  $\mu$ , para todas las observaciones. Habitualmente es incapaz de representar adecuadamente la estructura de los datos, asume un efecto nulo de las variables explicativas. En el *modelo saturado* el número de parámetros estimados es igual al número de observaciones. En datos individuales, utilizar este modelo implicaría estimar un número de parámetros igual al tamaño muestral. En la práctica el modelo nulo es demasiado simple y el modelo saturado no es informativo, existen otros dos modelos no tan extremos: el modelo minimizado que incluye el mínimo número de variables para un ajuste bueno, y el modelo maximizado que incluye el máximo número de variables que pueden ser consideradas. Cualquier modelo entre estos dos últimos se le llama modelo ajustado o modelo corriente (Cordeiro y Demetrio, 2008; Lindsay, 1997).

El problema principal de la selección de variables explicativas es determinar la utilidad de la inclusión de una variable extra en el modelo ajustado (en estudio), o se debe verificar la falta de ajuste inducida por

la omisión de esta variable. Con la finalidad de discriminar entre modelos, las medidas de discrepancia deben ser introducidas para medir el ajuste de un modelo.

### 3.4.1 Pruebas de Bondad de Ajuste

#### 3.4.1.1 Función de Desviación

La desviación del modelo fue propuesta por Nelder y Wedderburn (1972) y sirve para evaluar el ajuste del modelo, así como comparar modelos. La desviación residual muchas veces llamada simplemente como desviación compara el logaritmo de la verosimilitud del *modelo saturado* con el logaritmo de la verosimilitud de un *modelo ajustado*. El valor de la función de logaritmo de verosimilitud, para el modelo ajustado, nunca puede ser mayor que el del modelo saturado, porque el modelo ajustado contiene menos parámetros, entonces la desviación del modelo es siempre mayor o igual a cero y se define como:

$$D_p(\mu) = 2 [\ln L(\text{modelo saturado}) - \ln L(\hat{\mu})] = 2[l(\text{modelo saturado}) - l(\hat{\mu})]$$

En la tabla 4 se muestran funciones de desviación para algunas distribuciones.

Un modelo bien ajustado a los datos, con una verosimilitud grande tiene un pequeño valor de desviación, indicando que con un número menor de parámetros se obtiene un ajuste tan bueno como con el modelo saturado. Por lo tanto, la desviación es una medida de distancia de los valores ajustados en relación con los datos observados.

La desviación escalonada esta expresada como un múltiplo del parámetro de dispersión (McCullagh y Nelder, 1991).

$$S_p(\mu) = \phi^{-1} D_p(\mu)$$

Tabla 4. Funciones de desviación de algunos modelos

Modelo	Desviación
Normal	$D_p = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$
Binomial	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \log \left( \frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right]$
Poisson	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (y_i - \hat{\mu}_i) \right]$
Binomial Negativo	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (y_i + k) \log \left( \frac{\hat{\mu}_i + k}{y_i + k} \right) \right]$
Gamma	$D_p = 2 \sum_{i=1}^n \left[ y_i \log \left( \frac{\hat{\mu}_i}{y_i} \right) + \left( \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right) \right]$
Normal Inverso	$D_p = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}$

Cuando el parámetro de dispersión es igual a uno (modelos Binomial y Poisson) la desviación escalonada es igual a la desviación clásica.

Es bastante usada como "regla de oro" que el modelo no tiene un buen ajuste si la desviación escalonada no es significativamente mayor que el número de grados de libertad de la desviación residual  $n-p$ .

Según los criterios formales de esta prueba la función desviación escalonada tiene aproximadamente una distribución  $\chi^2_{n-p}$  (McCullagh y Nelder, 1991) y las pruebas de hipótesis son:

$H_0$ : El modelo ajustado es adecuado. (Se acepta si  $S_p(\mu) \leq \chi^2_{(\alpha, n-p)}$  )

$H_a$ : El modelo ajustado no es adecuado. (Se acepta si  $S_p(\mu) > \chi^2_{(\alpha, n-p)}$  )

Aunque el residual de la desviación es una medida de ajuste muy usada, McCullagh and Nelder (1989) argumentan que (al menos en el modelo MLG Binomial) valores grandes de desviación no siempre son evidencia de un ajuste pobre.

### **3.4.1.2 Estadística de Chi-cuadrado de Pearson:**

Como medida de "falta de ajuste", la  $X^2$  es frecuentemente recomendada, se calcula como la suma de los residuales de Pearson al cuadrado:

$$X^2 \text{ Pearson} = \sum_{i=1}^n \left( \frac{y_i - \mu}{\sqrt{V(\mu)}} \right)^2$$

Para muestras grandes se aproxima a  $\chi^2_{n-p}$ . Las pruebas de hipótesis son las mismas que para la función de desviación, y el modelo ajustado es adecuado si  $X^2 \text{ Pearson} \leq \chi^2_{(\alpha, n-p)}$ .

### **3.4.2 Pruebas de la significancia del modelo ajustado**

Para evaluar la significancia general del modelo ajustado se usan la prueba de razón de verosimilitud y la prueba de Wald, que miden qué tanto mejor el modelo ajustado se ajusta a los datos que el modelo nulo.

Las hipótesis a probar son:

$$H_0: B_1 = B_2 = \dots = B_p = 0$$

$H_a$ : Al menos un  $B_i \neq 0$

#### **Prueba de razón de verosimilitud (LRT)**

Conocido también como prueba "G" compara el logaritmo de la verosimilitud del modelo nulo con el logaritmo de verosimilitud del modelo ajustado.

$$LRT = -2 \ln \left( \frac{L(\text{modelo nulo})}{L(\text{modelo ajustado})} \right)$$

$$LRT = -2[\ln L(\text{modelo nulo}) - \ln L(\text{modelo ajustado})] \sim \chi^2_p$$

Se rechaza  $H_0$  si  $LRT \geq \chi^2_{(\alpha, p)}$ .

En el caso de las distribuciones Binomial y Poisson, el valor de LRT es igual que el de las desviaciones respectivas, pues  $\phi = 1$ .

#### **Prueba de Wald ( $W^*$ )**

Un método alternativo de probar la significancia del modelo es usar la prueba de Wald que está definida por:

$$W^* = [\hat{\beta}_{\text{ajustado}} - \beta_{\text{nulo}}]^T \hat{\text{Var}}^{-1}(\hat{\beta}) [\hat{\beta}_{\text{ajustado}} - \beta_{\text{nulo}}]$$

donde  $\hat{\text{Var}}(\hat{\beta})$  es la estimación de la matriz varianza-covarianza asintótica de  $\hat{\beta}$ , entonces:



$$W^* = \frac{1}{\phi} [\hat{\beta}_{ajustado} - \beta_{nulo}]^T (X^T \widehat{W} X) [\hat{\beta}_{ajustado} - \beta_{nulo}]$$

La cual tiene se aproxima a una distribución  $\chi^2$  con grados de libertad igual a la dimensión de  $\hat{\beta}_{ajustado}$  y se rechaza  $H_0$  si  $W \geq \chi^2_{(\alpha,p)}$ .

Para muestras grandes LRT y la prueba de Wald dan resultados similares, aunque no los mismos, sin embargo para muestras pequeñas, pueden diferir, investigaciones han demostrado que el uso de LRT da aproximaciones más confiables para pequeñas muestras que con la prueba de Wald (McCulloch et al., 2001; Agresti., 1996) debido a que para muestras grandes la varianza se infla.

### 3.4.3 Criterios de selección del modelo

#### 3.4.3.1 Criterio de Información de Akaike (AIC)

Otra medida usada con frecuencia para discriminar entre modelos es el Criterio de Información de Akaike conocido como AIC, definido como:

$$AIC = -2 \ln L + 2 * (p)$$

donde  $p$  es el número de parámetros.

Este índice evalúa tanto el ajuste del modelo a los datos como la complejidad del modelo. Mientras menor sea el valor de AIC el ajuste del modelo será mejor.

#### 3.4.3.2 Criterio de Información Bayesiana (BIC)

Es muy similar al AIC pero incorpora el tamaño de muestra, penaliza la adición de parámetros en mayor medida que el AIC. Se recomienda su uso cuando el tamaño de muestra es grande y el número de parámetros pequeño. Está definido como:

$$BIC = -2 \ln L + \ln(n) * (p)$$

#### 3.4.3.3 Desviación explicada

Como en MLG no se tiene un coeficiente de determinación, un estadístico similar es la desviación explicada, llamado también como pseudo  $R^2$  (Dobson, 2002), que nos da una idea de la variabilidad de los datos y se calcula como:

$$Desviación explicada = 100 \times \frac{desviación\ nula - desviación\ residual}{desviación\ nula} \%$$

donde la desviación nula se calcula igual que la desviación residual pero en vez del modelo ajustado utiliza el modelo nulo.

### 3.5 Análisis de Residuales y Diagnóstico del MLG

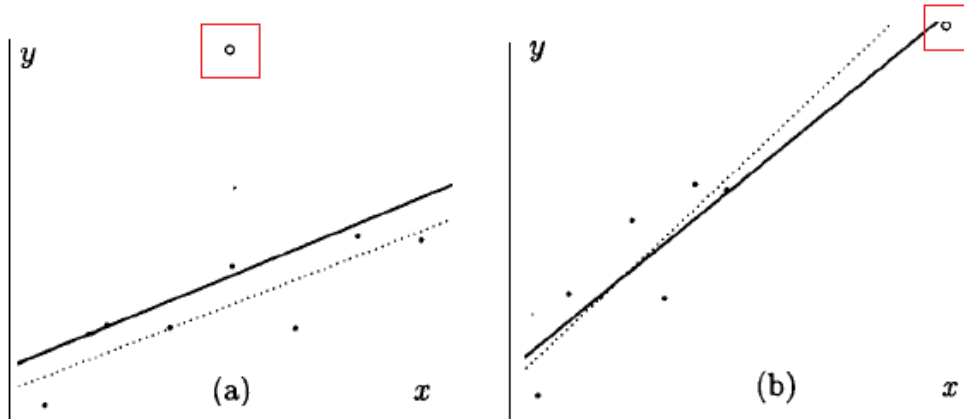
Como en cualquier procedimiento de ajuste de modelo, es importante el análisis de residuales para ajustar el MLG. Los residuales pueden guiar sobre la adecuación del modelo y dar una indicación sobre la adecuación de la función de enlace seleccionada. Además los residuales pueden también indicar la presencia de valores atípicos que requieren mayor investigación. Las técnicas usadas para el análisis de residuos y diagnóstico en los GLM son muy similares a las usadas en los modelos lineales clásicos de regresión, con algunas adaptaciones para que sean aplicables a todas las distribuciones.

#### 3.5.1 Diagnóstico de balanceo e influencia:

Un *punto de balanceo* (Fig. 18a) es un valor desacostumbrado de  $x$ ,  $y$  que puede controlar ciertas propiedades del modelo, aunque estos puntos no afectan las estimaciones de los coeficientes de

regresión si tienen un efecto marcado en las estadísticas del modelo como el pseudo R-cuadrado y sobre los errores estándar de las estimaciones de los coeficientes de regresión. Un *punto de influencia* (Fig. 18b) tiene un impacto notable sobre los coeficientes del modelo, porque “jala” al modelo en su dirección ya que tanto x como la variable respuesta se alejan de los demás puntos.

En el MLG se busca que el modelo sea representativo de todas las observaciones de la muestra. En consecuencia, se debe localizar esos puntos influyentes y evaluar su impacto sobre el modelo. Si esos puntos en realidad son malos se debe evaluar la posibilidad de eliminarlos de la muestra. Por otro lado, puede que no haya nada malo en ellos pero si controlan las propiedades del modelo, es bueno conocerlo, porque podría afectar al final el MLG.



**Figura 18. Gráficos que muestran el efecto de puntos extremos en relación con los demás puntos. a) Punto de influencia b) Punto de balanceo**

### 3.5.2 Medida de apalancamiento (leverage):

El diagnóstico del modelo ajustado se realiza a través de un análisis de las observaciones porque queremos verificar si es que algunas de ellas son puntos de balanceo.

La matriz de proyección (matriz de sombrero) para los MLG se define como:

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$$

donde  $H$  es simétrica e idempotente y depende de las variables explicativas, la función de enlace y la función de varianza, siendo más difícil su interpretación de medida de “leverage”. Se demuestra que  $V^{-1/2}(\hat{\mu} - \mu) \cong H V^{-1/2}(Y - \mu)$  donde  $V = \text{diag}\{V(\mu_i)\}$ , demostrando los elementos de  $h_{ij}$  pueden ser vistos como la cantidad de balanceo o palanqueo ejercido por la  $i$ -ésima observación  $y_i$  sobre el  $i$ -ésimo valor ajustado  $\mu_i$ .

Con frecuencia la atención se dirige hacia los elementos diagonales  $h_{ii}$  de la matriz  $H$ . La  $\text{Traza}(H) = \sum h_{ii} = p$ , y trabajando con la forma estandarizada  $h'_{ii} = n h_{ii} / p$  entonces  $\sum h'_{ii} = n$ .

Se sugiere el uso de  $h'_{ii} > \frac{2p}{n}$ , es decir  $h'_{ii} > 2$  para indicar puntos de alto apalancamiento. Una gráfica de  $h'_{ii}$  versus los índices con un límite  $h'_{ii} = 2$  es generalmente una herramienta útil para buscar puntos de apalancamiento. Cuando las observaciones con elementos diagonales grandes también posean residuales grandes, estas observaciones probablemente serán influyentes.

### 3.5.3 Distancia de Cook:

Los puntos influyentes se detectan mediante el análogo del estadístico de Cook de los modelos lineales clásicos. La influencia puede ser medida a través del cambio en la estimación de los parámetros cuando una  $i$ -ésima observación es retirada.

De esta manera, se evalúa  $\hat{\beta}_{(i)} - \hat{\beta}$  donde  $\hat{\beta}_{(i)}$  denota la estimación eliminando la  $i$ -ésima observación, y  $\hat{\beta}$  aquella con este valor incluido. En definitiva es una medida de distancia entre  $\hat{\beta}$  y  $\hat{\beta}_{(i)}$ . La estadística de Cook  $LD_i$  para uso de los modelos lineales generalizados propuesta según McCullagh y Nelder (1991) es:

$$LD_i = (\hat{\beta}_{(i)} - \hat{\beta})(X^T W X)(\hat{\beta}_{(i)} - \hat{\beta})/p\hat{\phi}$$

Los puntos con grandes valores  $LD_i$  tienen una gran influencia en el estimado de  $\hat{\beta}$ .

### 3.5.4 Residual ordinario

Los residuales originales u ordinarios del modelo lineal generalizado son las diferencias entre las observaciones y los valores ajustados:

$$e_i = y_i - \hat{y}_i = y - \mu$$

En los MLG los residuales no siempre tienen media cero, varianza constante y distribución simétrica, por lo cual surge la necesidad de nuevos tipos de residuales.

### 3.5.5 Residuales de Pearson

Según Johnson (2006) el residual de Pearson es el residual más lógico e intuitivo. Estos residuales corrigen la heterocedasticidad (Cameron y Travedi, 1998) debido a que incorporan la varianza de  $\mu$ , sin embargo, una desventaja es que su distribución es bastante asimétrica para modelos no normales (Cordeiro y Demetrio, 2008). En general se recomienda que su uso solo para fines de diagnóstico en general, pues para detectar valores atípicos u observaciones influyentes es preferible utilizar los residuos de desvianza. El residual de Pearson está definido como:

$$r_p = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

Esto residuales son importantes porque son una forma de estimar  $\phi$  usando el promedio de los residuales:

$$\phi = \frac{1}{N-p} \sum_{i=1}^N \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

El Residual de Pearson Estudentizado es recomendable cuando se tienen muestras pequeñas, ya que incorpora el  $i$ ésimo ( $h_{ii}$ ) elemento de la matriz sombrero (H). Está definido como:

$$r_i^{D*} = \frac{y - \mu}{\sqrt{V(\mu)(1 - h_{ii})}}$$

### 3.5.6 Residuales de Desvianza

El residual de desvianza representa la distancia de la observación  $y_i$  a su valor ajustado  $\hat{\mu}_i$  medida en la escala logaritmo de la función de verosimilitud. El  $i$ -ésimo residual de desvianza se define como la raíz cuadrada de la contribución de la  $i$ -ésima observación a la desvianza, multiplicada por el signo del residual original:

$$r_i^D = \sqrt{d_i} \text{ signo } (y_i - \hat{\mu}_i); \quad \sum (r_i^D)^2 = D$$

donde  $d_i$  es llamado componente de la desvianza.

El residual desvianza tiene una distribución exacta normal en el caso de la regresión normal y en el caso de la regresión normal inversa.

El Residual de Desviación estudentizado está definido como:

$$r_i^{D*} = \frac{r_i^D}{\sqrt{1 - h_{ii}}}$$

donde  $h_{ii}$  es el  $i$ -ésimo elemento de la matriz  $H$ .

### 3.5.7 Gráficos de diagnóstico:

Las técnicas gráficas más usadas para analizar la adecuación de los Modelos Lineales Generalizados son las siguientes:

- **Residuos vs. valores ajustados:** Es recomendado, por ejemplo el gráfico de algún tipo de residuo estudentizado versus  $\hat{\eta}$ . El comportamiento estándar de este gráfico es una distribución de los residuos en torno de cero con una amplitud constante. Este gráfico sirve para verificar la función de enlace. No tiene significado para datos binarios.
- **Gráfico de probabilidad normal o semi-normal de los residuos:** Son construidos de la misma forma que para los modelos clásicos de regresión usando una distribución pertinente. Mediante este gráfico se puede observar la existencia de observaciones atípicas y la adecuación del modelo.

Dado que la distribución de los residuales no es conocida, los “gráficos semi-normales” con envoltura simulada son herramientas de diagnóstico muy útiles (Atkinson et al., 1985). La idea es usar el clásico gráfico semi-normal mediante la adición de una envoltura simulada que se puede utilizar para decidir si los residuales observados son consistentes con el modelo ajustado, de tal forma que un modelo es correcto si los residuos calculados a partir de las observaciones caen dentro de la envoltura. Se construyen de la misma forma que para el modelo de regresión clásica, utilizando, sin embargo, la distribución correspondiente.

Para generar estos gráficos se propone seguir los siguientes pasos:

- a) Ajustar un determinado modelo a un conjunto de datos y generar una muestra simulada de  $n$  observaciones usando el modelo ajustado como si fuera el modelo verdadero.
- b) Ajustar el modelo a la muestra generada, y calcular los valores absolutos ordenados de los residuales.
- c) Repetir los pasos (a) y (b)  $k$  veces.
- d) Considere  $n$  conjuntos de los  $k$  estadísticos ordenados, y para cada uno calcular el promedio, mínimo y máximo valor.
- e) Graficar estos residuales ordenados de la muestra original respecto a los scores semi-normales  $\Phi^{-1}((t + n - 1)/8)/(2n + 1/2)$ .

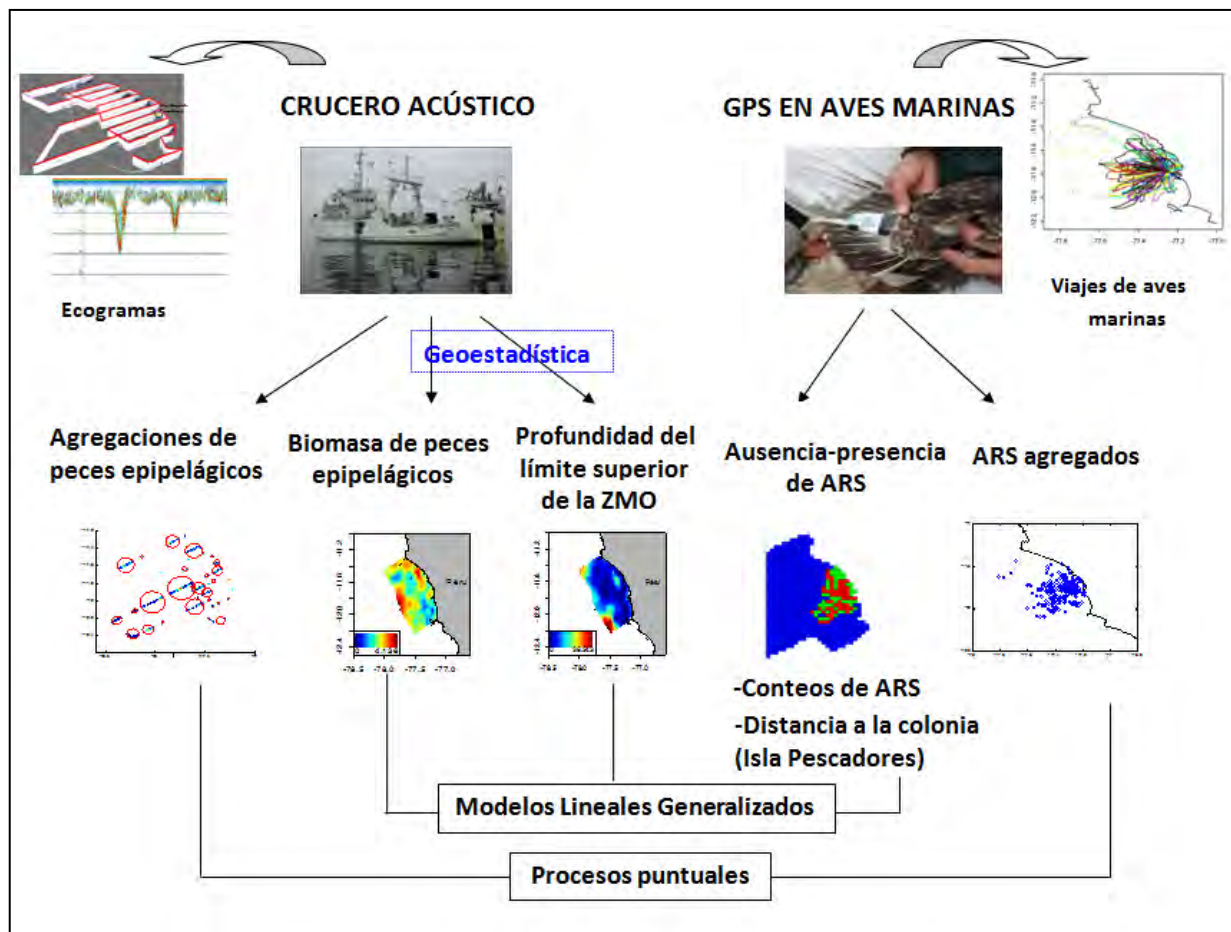
Atkinson (1985) sugiere usar  $k=19$ , tal que la probabilidad que el valor absoluto de un residual se encuentre fuera de la envoltura sea aproximadamente igual a  $1/20=0.05$ .

- **Valores absolutos de residuos estudentizados vs. valores ajustados:** Sirven para verificar la función de varianza. El patrón nulo es una distribución aleatoria de media cero y amplitud constante. Una función de varianza escogida erradamente mostrará una tendencia en la media. En general, la no adecuación de la función de varianza será tratada como sobredispersión, tema tratado más adelante.
- **Gráficos de  $h_i$ ,  $LD_i$  vs. orden de las observaciones:** sirven para localizar observaciones con residuo, leverage ( $h_{ii}$ ) o distancia de Cook grandes (Cordeiro y Demetrio, 2008), es decir para la visualización de los puntos discordantes e influyentes.

# **CAPITULO III**

## **MATERIALES Y MÉTODOS**

Como se ha mencionado en la parte introductoria, en este estudio se utilizaron datos de prospección acústica y telemetría de aves marinas. Como se aplicaron dos técnicas estadísticas distintas, se ve la necesidad de crear dos tipos de bases de datos finales para cada escala. Para la base de datos de procesos puntuales, se utilizaron los datos acústicos para encontrar agregaciones de peces epipelágicos, y en el caso de los datos de aves marinas se trabajó con las ARS (Area Restricted Searches), es decir la zonas de las trayectorias en los cuales las aves realizan bucles, lo que se considera como un marcador de la detección de presas (Fauchald y Tveraa, 2003). La base de datos para los MLG estuvo formada por una malla, como los datos acústicos provienen de transectos se realizó un kriging para cada variable de interés, y los datos de aves se ubicaron según sus latitudes y longitudes dentro de cada malla. Finalmente ambas bases de datos tuvieron en cuenta las ya mencionadas escalas espacio-temporales, por lo tanto se crearon mallas espacio-temporales para poder combinar bases de acústica con bases de datos de aves (Fig. 19).



**Figura 19. Esquema para obtener las bases de datos finales**

## 1. Obtención de los datos

### 1.1. Obtención de los datos acústicos

Los datos acústicos utilizados en este estudio provienen de un crucero pelágico realizado entre noviembre y diciembre del 2008 por el Instituto del Mar del Perú a bordo del buque de investigación científica José Olaya Balandra (Fig. 20). Los datos acústicos fueron recolectados con una ecosonda científica Simrad EK-60 en dos frecuencias, 38 y 120 kHz. El área de estudio total cubrió el litoral entre Paita (05°00'S) e Ilo (17°40'S). Y en particular se estudió el área "Topineme"<sup>12</sup>, porque los datos de aves fueron recolectados en los alrededores de la Isla Pescadores (77° 15' 52 S y 11° 46' 28 W).

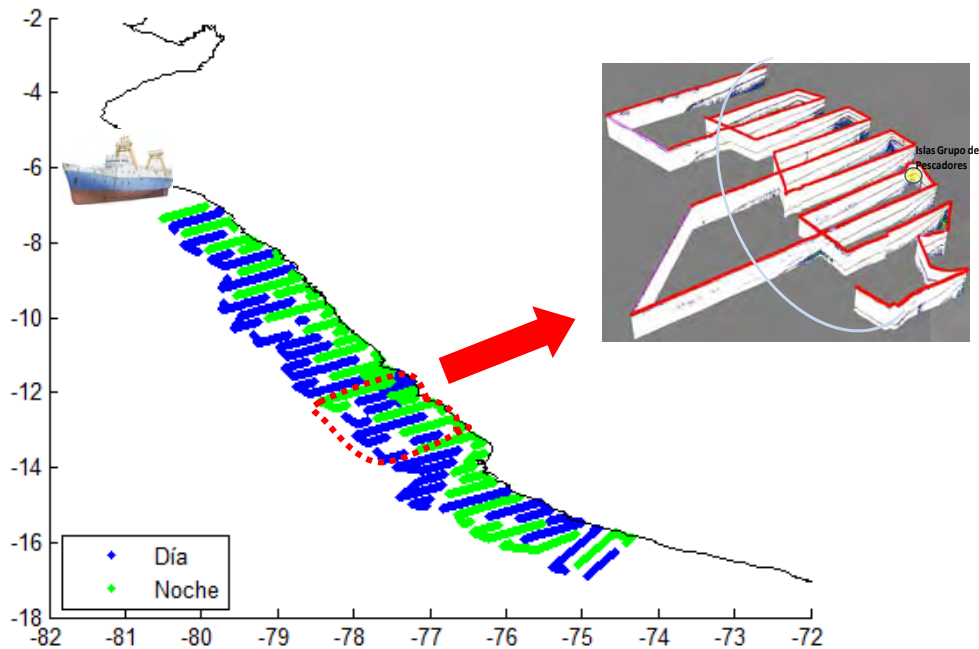


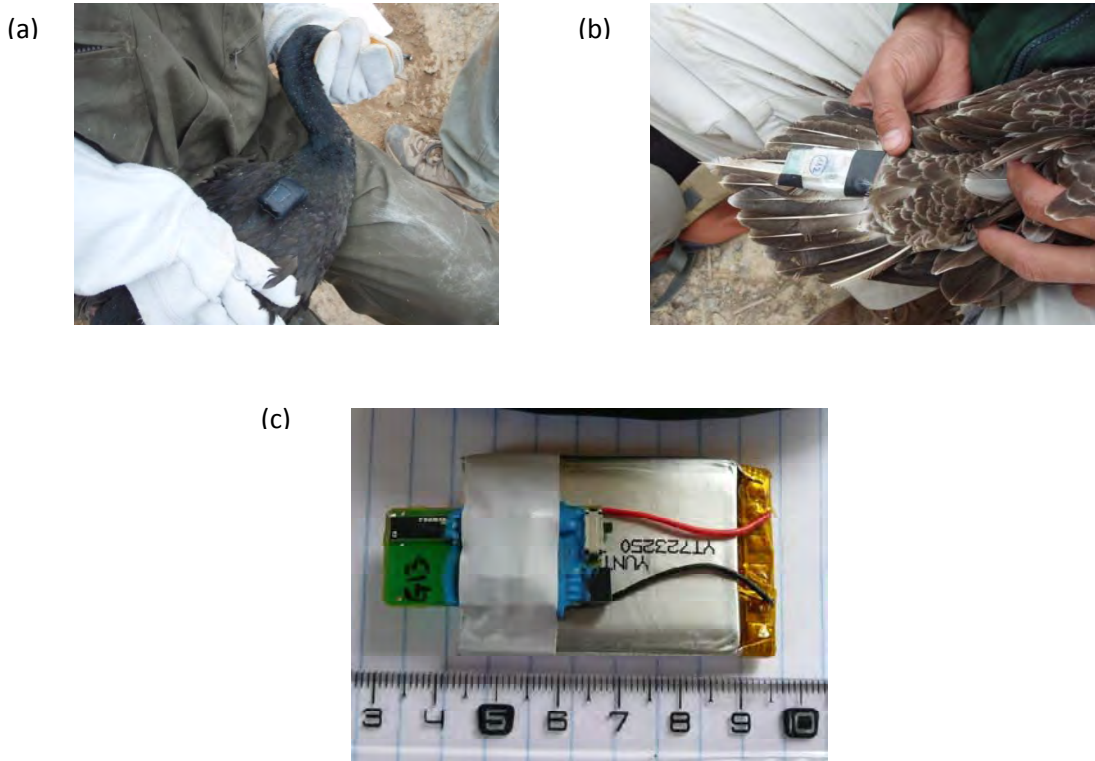
Figura 20. Área Muestreada total y Área TOPINEME

### 1.2 Obtención de los datos de aves marinas

Los datos recolectados consisten en datos de seguimiento de aves marinas con GPS (sistema de posicionamiento global) de tipo "GiPsy Technosmart" (Fig. 21). Los GPS fueron colocados en guanajes y piqueros que tenían crías, para asegurar que vuelvan a su nido y recuperar los GPS (estas GPS graban y no emiten; luego para obtener los datos de posición, hace falta recuperar el aparato). Los GPS grabaron, cada segundo, las coordenadas (latitud, longitud en grados, minutos y segundos) de la posición del ave, así como el tiempo en el que se capturaron las señales (fecha, hora, minutos y segundos). Se trabajó con

<sup>12</sup> El proyecto "Topineme" es manejado por un grupo multi-disciplinario de investigadores del IMARPE, IRD, y CNRS y tiene como objetivo determinar en qué medida los predadores superiores (aves marinas, lobos y pescadores), su abundancia, su dieta y su comportamiento espacial reflejan las propiedades de los principales componentes del ecosistema, desde el hábitat físico, pasando por las comunidades planctónicas y hasta las poblaciones de peces forraje frente al Perú.

aves equipadas en la Isla Pescadores, una pequeña isla ubicada a 7.5 Km de la costa central del Perú, entre noviembre y diciembre del 2008.



**Figura 21. Marcaje de aves a) Piquero b) Guanay c) GPS: '28 GiPsy Technosmart'**

## **2. Pre-procesamiento de los datos**

### **2.1 Pre-procesamiento de los datos acústicos**

Para discriminar entre peces y otros tipos de organismos se aplicó un análisis bi-frecuencia, se sincronizó el número de emisión (pin) y la posición entre ecogramas usando un paquete especializado en tratamiento de datos acústicos llamado Echoview, que permite el procesamiento de datos obtenidos mediante ecosondas.

Se sabe que en muchas mediciones existe un factor no controlable. En el caso de los datos acústicos, unos de los factores no controlables más importantes son los ruidos (Simmonds y MacLennan, 2005) provocados por el barco o el ruido provocado por la reflexión del sonido en el fondo del mar. Estos ruidos pueden causar distorsión en la medición como sobreestimaciones (Simmonds y MacLennan, 2005), por ello se excluyó el ruido de la superficie (5 m de profundidad) y las demás regiones con ruido fueron eliminadas de los ecogramas usando un algoritmo basado en la función de ruido con la forma:

$$20*\log(R)+2*a *R + offset$$

Donde  $R$  es el rango (en m),  $a$  es la frecuencia del coeficiente de absorción (en  $\text{dB.m}^{-1}$ ) y el valor de desplazamiento (en dB) es el ruido inicial asumido en el primer metro. Los valores de  $a$  y el desplazamiento fueron determinados usando un programa escrito en Matlab por Paul Fernandes (Marine lab., Aberdeen, UK), y fueron usados en un algoritmo lineal del Echoview. A continuación, se hizo uso de



un algoritmo (Ballón et al., aceptado), para clasificar toda esta información en peces, fluid like<sup>13</sup> y blue noise<sup>14</sup> (Fig. 22).

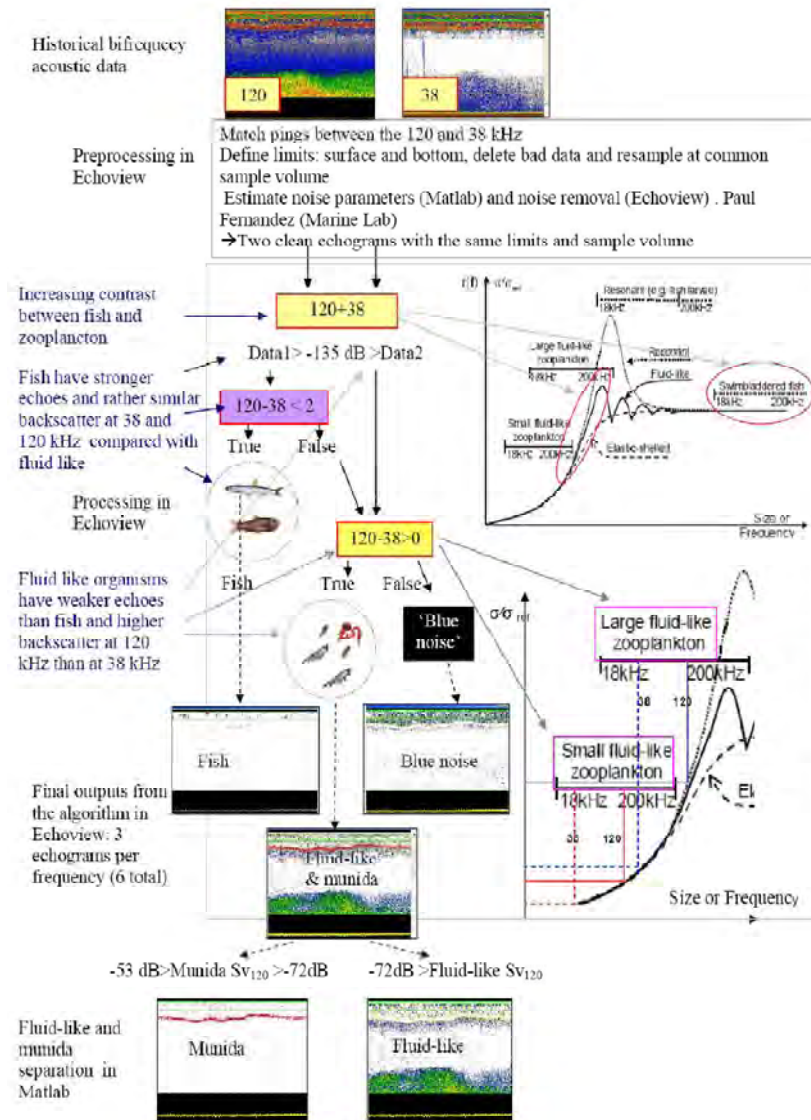


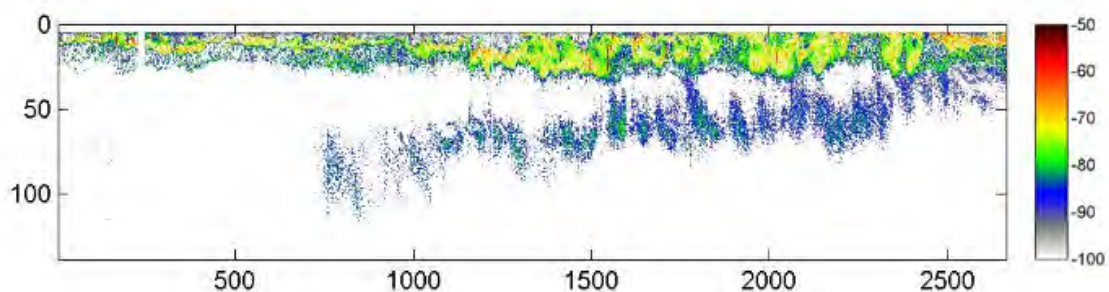
Figura 22. Discriminación Acústica por el algoritmo bi-frecuencia desarrollado por Ballón (2010).

<sup>13</sup> El 'fluid-like' es una de las clase de plancton según la acústica; el cuerpo de los fluid-like es más o menos cilíndrico con una longitud mayor que el ancho y todo el cuerpo contribuye al eco, a este grupo pertenece la mayoría de zooplankton, en particular los crustáceos (ej. copépodos, eufausidos).

<sup>14</sup> El blue noise incluye todos aquellos organismos que según su fuerza de retrodispersión no fueron clasificados como fluid-like o como peces adultos. El blue noise esta principalmente constituido por pequeños organismos como larvas de peces con vejiga natatoria y organismos gelatinosos.

Los peces con vejiga natatoria<sup>15</sup>, tales como la anchoveta, tiene ligeramente mayor fuerza de retrodispersión (Sv) a 38 que a 120 kHz, mientras que el zooplancton tiene mayor retrodispersión a 120 KHz. Esta diferente respuesta de frecuencia hace posible usar las frecuencias de 38 y 120 KHz para extraer los organismos de tipo fluid-like y peces de los datos acústicos. Para asegurar una buena discriminación entre peces y otras especies, se incremento el contraste entre los diferentes grupos sumando la respuesta de retro dispersión a ambas frecuencias. Luego, en base a las observaciones, se escogió un valor de umbral de -135 dB y que se uso para crear ecogramas de peces y no peces para cada frecuencia. Se refinó el ecograma de peces manteniendo únicamente aquellos que cumplían  $Sv_{120} - Sv_{38} < +2$  dB. Los seis ecogramas (fish, fluid-like y blue noise) en ambas frecuencias fueron exportados y procesados en Matlab. Estos datos consistían del volumen de la fuerza de retrodispersión (Sv) por celdas elementales de un pin de longitud por segundo y 0.5 m de altura, juntos con las coordenadas posicionales (latitud y longitud), tiempo y profundidad.

Cuando la zona mínima de oxígeno (ZMO, zona con bajas concentraciones de oxígeno disuelto) está presente como en el caso del litoral peruano, la extensión vertical de la comunidad epipelágica (zooplancton, organismos gelatinosos y pescado que viven en las capas superficiales) es restringida a la parte oxigenadas encima de la ZMO. Al anochecer la comunidad mesopelágica que se refugia en la ZMO durante el día migra hacia la superficie y se mezcla con la comunidad epipelágica. Al amanecer estos migran verticalmente y se refugian de nuevo en la ZMO (Fig. 23). Debido a ello solo los periodos de día y noche fueron considerados por separado, pues no es posible determinar la profundidad de la extensión vertical de la comunidad epipelágica cuando los organismos mesopelágicos migran hasta el límite superior de la zona mínima de oxígeno.



**Figura 23. Ciclo nictimeral, se muestra cómo las comunidades mesopelágicas migran hacia la superficie al anochecer.**

Para eliminar ecos de ruidos potencialmente altos, se aplicaron los siguientes umbrales en estos datos: -21dB para peces (38 kHz), -56 dB para blue-noise (38 kHz), y -53 dB para fluid-like (120 kHz). También ecos de blue-noise bajo el umbral de -100 dB fueron excluidos del análisis.

A continuación se separo múnida<sup>16</sup> *Pleuoncodes monodon* (Fig. 24) de otros organismos fluid-like por varias razones, pues su abundancia se incremento dramáticamente en el sistema desde mediados de 1996 y ocupa un nicho ecológico específico (Gutiérrez et al., 2008), la múnida es más grande que cualquier otro zooplancton fluid-like en el ecosistema y no juega el mismo rol ecológico (es decir la múnida no es alimento de pequeños peces forrajeros), la múnida tiene una mayor retrodispersión que otros organismo fluid-like presentes, la múnida siempre es epipelágica y realiza pocas migraciones verticales dentro de la capa epipelágica (Gutiérrez et al., 2008).

<sup>15</sup> La vejiga natatoria es un órgano de flotación que poseen muchos peces. Se trata de una bolsa de paredes flexibles, llena (generalmente) de gas, situada dorsalmente por debajo de la columna vertebral y por encima del tubo digestivo. Controla la flotabilidad mediante un complejo sistema de intercambio gaseoso con la sangre, y permite al pez ascender o descender en el agua sin necesidad de utilizar la musculatura.

<sup>16</sup> La múnida es un crustáceo grande tipo langostino.



**Figura 24. Múnida**

Para extraer la múnida de otros fluid-like se aplicaron los siguiente umbrales en el fluid-like Sv120:  $-72\text{dB} < \text{Sv} < -53\text{dB}$ . Los limites de los umbrales fueron considerados empíricamente usando ecogramas donde la presencia o ausencia de múnida fue confirmada por redes de arrastre epipelágicos.

Con la finalidad de reducir el tiempo del procesamiento y asegurar la representatividad de los datos, se estimó la mediana cada tres pines, luego se procedió a calcular y estimar la variables de interés.

La base de datos de primer orden con datos acústicos para este estudio estuvo compuesta por las siguientes variables continuas:

**La profundidad del límite superior de la zona mínima de oxígeno- oxiclina:** Se estima que el límite superior de la ZMO corresponde a la profundidad (en m) de la extensión vertical de la comunidad epipelágica (VEEC; Bertrand et al., 2010a). Se define la profundidad de la VEEC ( $Z_{VEEC}$ ) como la profundidad estimada en la cual el 98% de la suma acumulada de ecos acústicos de la comunidad epipelágica se alcanza (Ballón, 2010; Bertrand et al., 2010a). El eco acústico de blue noise y fluid-like se multiplicó por un factor  $10^3$ , pues tienen menor fuerza de retrodispersión que los peces, entonces la suma acumulada es:

$$(\text{Sv fluid like} * 10^3 + \text{Sv Blue-noise} * 10^3 + \text{Sv peces})/3$$

**Biomasa de peces epipelágicos:** La fuerza de retrodispersión (Sv) de peces fue convertida en el coeficiente de dispersión de área náutica (NASC en  $\text{m}^2 \text{n.mi.}^{-2}$ ), un índice de biomasa, acorde a la fórmula de (MacLennan et al, 2002):

$$\text{NASC} = 4\pi(1852)^2 * \text{Sv peces} * T$$

donde T es la distancia de integración vertical de 0.5 m. La biomasa integrada se calculó multiplicando los valores medios por la profundidad de la capa. Finalmente se usó los lances de pesca de arrastre realizados a lo largo del crucero para clasificar los peces en epipelágicos y mesopelágicos.

**Latitud acústica:** La latitud fue convertida de grados a km usando una transformación de la variable original.

$$\text{Latitud\_Km\_acustica} = \text{latitud\_grados\_acustica} * 60 * 1.852$$

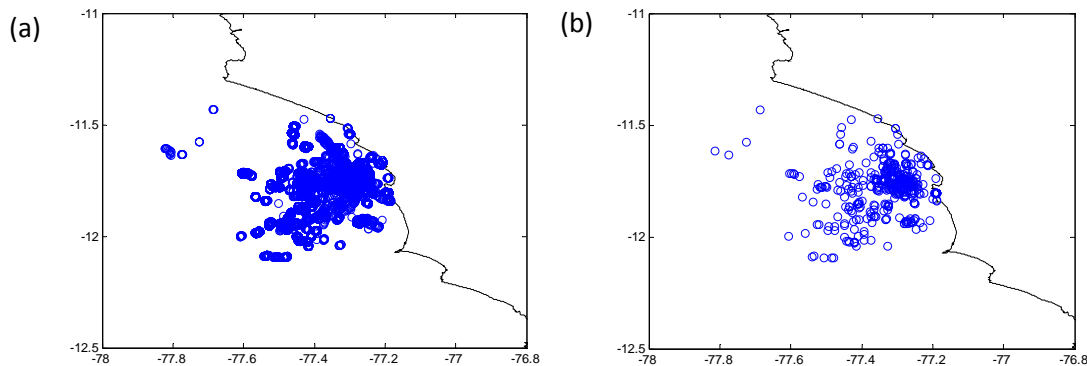
**Longitud acústica:** La longitud fue convertida de grados a km usando una transformación de la longitud en base al coseno de la latitud para tomar en cuenta la redondez del planeta:

$$\text{Longitud\_Km\_acustica} = \text{longitud\_grados\_acustica} * 60 * \cos((\text{latitud\_grados\_acustica} * \pi) / 180) * 1.852$$

**Tiempo acústica:** Es el tiempo promedio en que el crucero pasó por determinada posición. Está formada por el día, mes, hora, minutos y segundos.

## 2.2. Pre-procesamiento de los datos de aves marinas

Los viajes de forrajeo de las aves fueron reconstruidos con los datos grabados por el GPS. En efecto algunos datos estaban incompletos debido a limitaciones de la batería o problemas de posicionamiento; por ello estos se removieron para que no afecten los resultados. Para cada viaje se identificó áreas de búsqueda restringida (ARS). El ARS es una manera de caracterizar la estrategia de búsqueda de alimento utilizada por muchos animales, la cual implica que un animal que esta forrajeando debe aumentar su esfuerzo de búsqueda en las zonas donde los recursos son abundantes. Por ejemplo, un animal debe aumentar su ritmo de giro y / o reducir su velocidad como una respuesta a los aumentos en la tasa de encuentro de presa, adoptando así el llamado comportamiento de ARS (Pinaud y Weimerskirch, 2007). Para calcular las ARS se usó el método First Passage Time method (FPT) que se define como el tiempo necesario para que un animal cruce un círculo con un radio determinado. Es una medida de cuánto tiempo pasa un animal dentro de un área determinada (Fauchald y Tveraa, 2003). En los registros de GPS, se determinó para cada posición si pertenecía a una zona de ARS o no. Una ARS duró más que el tiempo de una grabación GPS, cada ARS correspondió entonces a una serie continua de posiciones etiquetadas como 'en ARS'. Para evitar luego que una ARS sea representada por un número variable de puntos de posiciones, generando un problema de *autocorrelación artefactual*, se consideró que una ARS se ubicaba (asociaba a la posición GPS) solamente en el punto promedio de una serie continua de posiciones determinadas como 'en ARS' (Fig.25). Este procedimiento se realizó para todos los viajes. Y para cada agrupación se calculó la posición promedio (longitud, latitud), el tiempo promedio.



**Figura 25. a) Total de ARS. b) ARS agrupados.**

La *base de datos de primer orden* con datos de aves para este estudio estuvo compuesta por las siguientes variables:

**ARS:** Identificó si se puede considerar la posición del viaje de forrajeo del ave como un área de búsqueda restringida. Es una variable indicadora: Si es ARS (1) y si no es ARS (0).

**Latitud ARS:** Es una variable continua en grados.

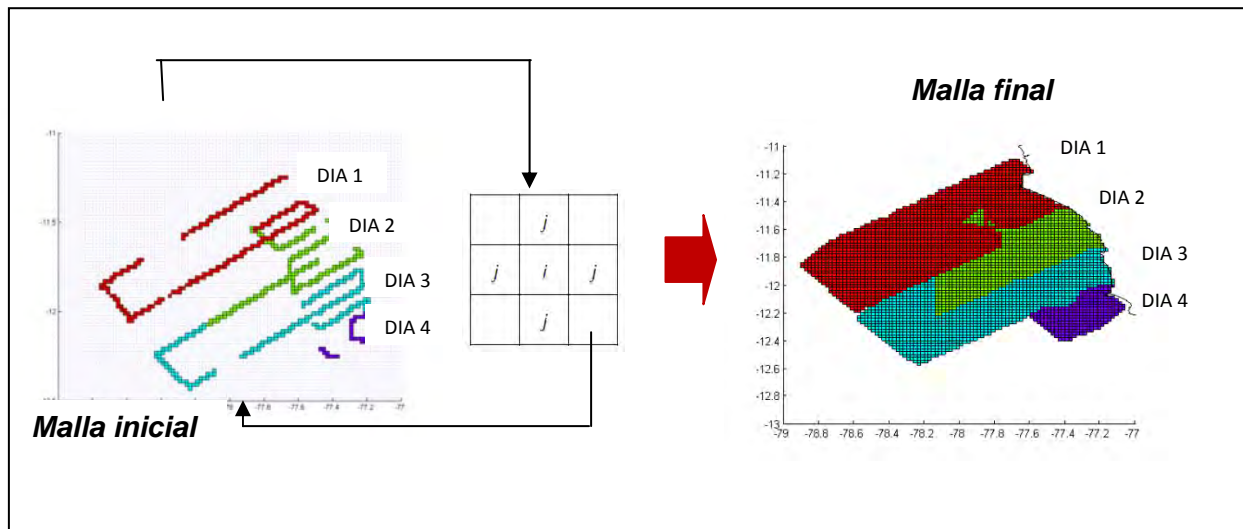
**Longitud ARS:** Es una variable continua en grados.

**Tiempo ARS:** Es el tiempo en el que se emitió la señal del GPS para cada posición. Es una variable continua formada por el día, mes, hora, minutos y segundos.

### 3. Procesamiento de las bases de datos

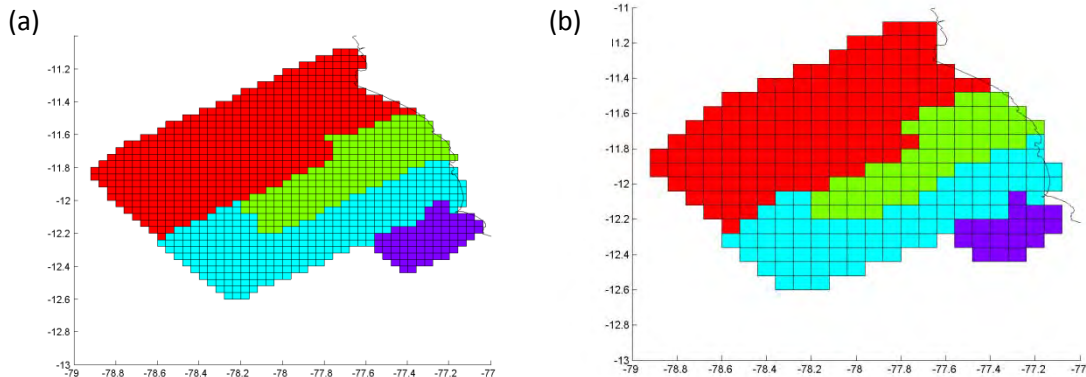
#### 3.1 Malla espacio-temporal

En la introducción se explicó la importancia de las escalas espaciales y temporales. Para cruzar las bases de datos de acústica y de aves se creó una malla espacio-temporal base. Como los transectos del crucero acústico se realizaron ordenadamente se tomaron como base para crear cada malla espacio-temporal. Se creó una malla en blanco (con determinada escala espacial, por ejemplo celdas con longitud de 1.1km), y se rellenó primero las celdas por donde paso el crucero acústico, es decir, las celdas por donde pasó tomaron “como valor” la fecha correspondiente al crucero. Luego, se consideraron las celdas más próximas alrededor de las ya rellenadas (Fig. 26), y se contó la mayor cantidad de celdas continuas con determinada fecha para rellenar la celda, se realizó este proceso sucesivamente hasta rellenar completamente la malla.



**Figura 26. Método utilizado para crear la malla espacio temporal (Escala espacial 1.1km)**

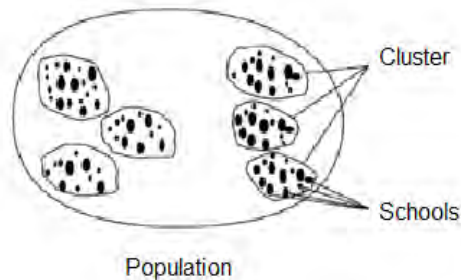
Se realizó el mismo procedimiento para las siguientes escalas espaciales: 8.9km (0.08°, Fig. 27a) ,6.7km (0.06°), 4.4km (0.04°, Fig. 27b), 2.2km (0.02°) y 1.1km (0.01°).



**Figura 27. Ejemplo de mallas espacio temporales que muestran la variación del grano: a) Escala espacial 4.4km. b) Escala espacial 8.9km.**

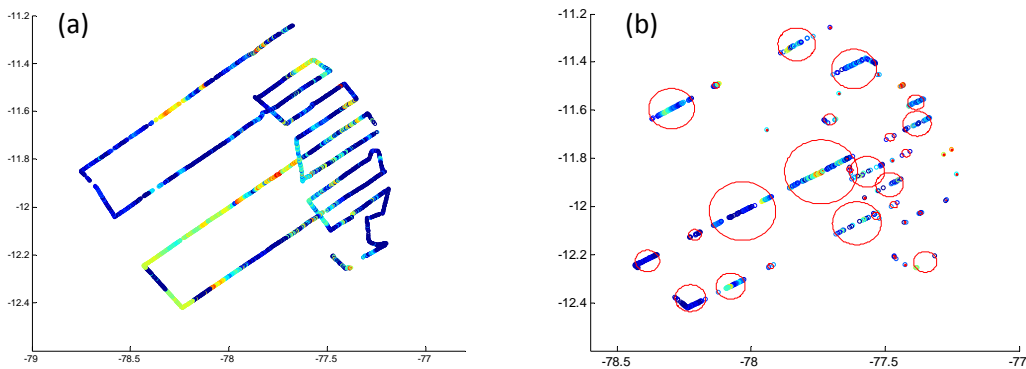
### 3.2. Bases de datos para análisis de procesos puntuales

Las agregaciones (Fig. 28) se consideran como eventos discretos de un proceso puntual. Usando la base de datos acústica de primer orden obtenida en el pre-procesamiento de datos acústicos (Fig. 29a) se agruparon las unidades básicas de muestreo cada 200 metros, para facilitar la identificación de las agregaciones de peces epipelágicos.



**Figura 28. Nivel de organización clásicos: cardúmenes (“Schools”), agregaciones (“Clusters”), sub-población o poblaciones, Fuente: Fréon et al. (2005a).**

Para definir las agregaciones/parches se utilizó la definición de un parche según la cual son regiones donde la abundancia es superior a lo que se puede esperar en el caso de una distribución aleatoria (Haury y McGowan, 1998). El promedio de la biomasa de peces epipelágicos (NASC) fue igual a  $460.5 \text{ m}^2 \text{ mn}^{-2}$ . Nos quedamos con las partes de los transectos (agrupadas cada 200 m) con valores de NASC mayores al promedio. Luego se calculó la distancia ortodrómica<sup>17</sup> entre las posiciones consecutivas de las partes de los transectos mayores al promedio. Las posiciones que ocurrieron a una distancia menor a 5 km formaron parte de la misma agregación, caso contrario se consideró como una nueva agregación (Fig. 29b). Para cada agregación de peces definida, se calculó la biomasa promedio, el radio de la agregación y la posición promedio (longitud y latitud).



**Figura 29. a) Distribución espacial de la biomasa de peces epipelágicos correspondiente a la base de acústica de primer orden, cada 3 pines. b) Biomasa de peces epipelágicos agrupados cada 200 metros e identificación de agregaciones (círculos rojos).**

Para el análisis de procesos puntuales se consideró únicamente aquellas posiciones de la base de datos de aves obtenida en el pre-procesamiento de datos donde se identificó presencia de ARS. A continuación, se usó la malla espacio-temporal de 0.01km obtenida en la sección 3.1, para determinar qué ARS puede o no incluirse en cada base de datos final según las escalas temporales que se estudiaron (Tabla 5).

<sup>17</sup> Es la distancia entre dos posiciones en grados sexagesimales  $(x_1, y_1)$  y  $(x_2, y_2)$  de la superficie terrestre, en km está definida como:

$$d = (1.852 * 60 * 180 / \pi) * \text{acos}(\sin(y_1 * \pi / 180) * \sin(y_2 * \pi / 180) + (\cos(y_1 * \pi / 180) * \cos(y_2 * \pi / 180) * \cos(x_2 * \pi / 180 - x_1 * \pi / 180)))$$

Tabla 5

<b>Escala temporal (tolerancia)</b>
Más menos tres días
Más menos dos días
Más menos un día
El mismo día

Las variables finales (Tabla 6) para cada proceso puntual según las cuatro escalas propuestas son las siguientes:

- **ARS:** es una variable categórica. Cuando es ARS es igual a uno, caso contrario es igual a cero.
- **Agregaciones\_peces:** es una variable categórica. Cuando es Agregación de peces epipelágicos es igual a uno, caso contrario es igual a cero.
- **Latitud:** es la latitud promedio de la agregación o de ARS. Variable continua en grados.
- **Longitud:** es la longitud promedio de la agregación o de ARS en grados.

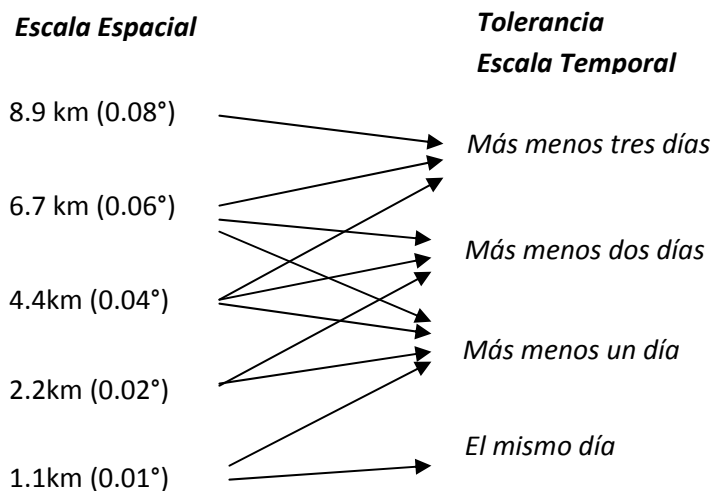
Tabla 6.

Longitud	Latitud	ARS	Agregaciones_peces
-77.53	-11.45	0	1
...	...	...	...
-78.13	-11.5	0	1
-77.29	-11.73	1	0
...	...	...	...
-77.26	-11.94	1	0

### 3.3. Bases de datos para MLG

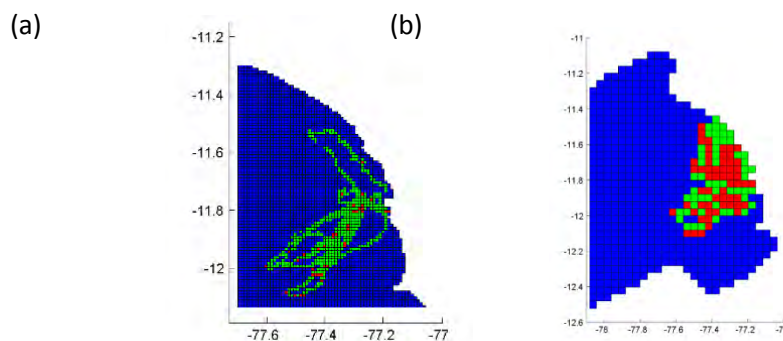
En la base de datos acústica de primer orden, obtenida en el pre-procesamiento de datos acústicos solo tenemos valores para los transectos (sección 2.1). Se realizó una interpolación mediante el uso de geostatística, de tal forma obtuvimos una malla con celdas de 1.1 km con los valores estimados de la profundidad del límite superior de la ZMO y la abundancia de peces epipelágicos.

A continuación, se usaron las cinco mallas espacio-temporales (sección 3.1) para incluir tolerancias temporales. Una malla con escala temporal de más menos un día de tolerancia, implicó que para incluir el dato en la malla (a la escala espacial fijada) debió ocurrir como máximo un día antes o el día posterior a la "fecha" de la celda que le corresponde dicho dato. El objetivo fue crear doce combinaciones de mallas de diferentes escalas de tiempo y espacio, según las mostradas en la Figura 30.



**Figura 30. Combinaciones de mallas a diferentes escalas temporales y espaciales para MLG.**

Cada combinación de malla a diferentes escalas se cruzó con las variables de aves obtenidas en el pre-procesamiento de datos (Fig. 31). Se calculó el valor promedio de los valores interpolados de abundancia de peces y la profundidad del límite superior de la ZMO según su latitud y longitud de cada celda.



**Figura 31. Mallas cruzadas con ARS a diferentes escalas. Rojo (ARS=1) y verde (ARS=0). a) Escala 1.1km b) Escala 4.4km.**

Las variables finales (Tabla 7) de las doce combinaciones de mallas espacio-temporales son las siguientes:

- **Abundancia de ARS total:** se calculó el número total de ARS.
- **Presencia – Ausencia de ARS:** se consideró presencia si por lo menos existió un ARS en la celda, caso contrario si paso un ave pero no se considero ningún ARS se considero como ausencia.

$$Presencia\ ausencia\ (ARS) = \begin{cases} 1 & \text{Al menos (número ARS} \geq 1) \\ 0 & \text{número ARS} = 0 \end{cases}$$

- **Biomasa de peces epipelágicos:** se usaron los valores estimados con geoestadística y se calculó el promedio de la biomasa de peces en cada celda, Variable continua, NASC en  $m^2 \cdot mn^{-2}$ .
- **Profundidad del límite superior de la ZMO:** se usaron los valores estimados con geoestadística y se calculó el promedio del límite superior de la zona de mínimo oxígeno en cada celda, Variable continua en metros.
- **Longitud\_inicial y Longitud\_final:** son las longitudes que determinan la mínima y máxima longitud de la celda. Variable continua en grados.
- **Latitud\_inicial y Latitud\_final:** son las latitudes que determinan la mínima y máxima latitud de la celda. Variable continua en grados.



- **Distancia a la isla:** se calculó la distancia ortodrómica (que es la distancia más corta entre dos puntos de una esfera (la Tierra).) entre la posición del ARS (número y presencia-ausencia) y la Isla Pescadores. Variable continua en km.

$$A = \left[ \text{sen} \left( \frac{\text{lat. } i * \pi}{180} \right) * \text{sen} \left( \frac{\text{lat} * \pi}{180} \right) \right]$$

$$B = \left[ \text{cos} \left( \frac{\text{lat. } i * \pi}{180} \right) * \text{cos} \left( \frac{\text{lat} * \pi}{180} \right) * \text{cos} \left( \left( \frac{\text{lon. } i * \pi}{180} \right) - \left( \frac{\text{lon} * \pi}{180} \right) \right) \right]$$

donde: posición ARS = (lon, lat) y posición de la isla = (lon. i, lat. i)

$$\text{Dist}_{\text{isla}} = 60 * 1.852 * \left[ \frac{180 * \text{acos}\{A + B\}}{\pi} \right]$$

Tabla 7. Muestra de la base de datos final para el MLG para la una malla con escala espacial 1.1 km

Longitud inicial	Longitud final	Latitud inicial	Latitud final	Biomasa de peces epipelágicos	Profundidad del límite superior de ZMO	Abundancia de ARS	P/A
-77.37	-77.36	-11.89	-11.88	59.502	-11.038	1	1
-77.35	-77.34	-11.75	-11.74	48.178	-8.5946	NA	0
-77.34	-77.33	-11.93	-11.92	22.973	-10.97	2	1
-77.34	-77.33	-11.92	-11.91	28.964	-10.423	1	1
-77.33	-77.32	-11.87	-11.86	60.14	-8.7743	NA	0

### 3.4 Aplicación de la Función K-Ripley

Para calcular la función K de Ripley univariada solo se tomo en cuenta la variable ARS cuando es igual a uno (Tabla 6). Se calculo la función K de Ripley univariada para cada escala temporal (día a día, más menos un día, más menos dos días, más menos tres días) hasta los 20 km de radio(r), que es la cuarta parte de la longitud del menor lado del área (80km), debido a que después de este umbral los radios cubren casi por completo el área incrementando la varianza de los estimadores dramáticamente conforme crece el radio (Dessard et al., 2004; Baddley et al., 2010). Asimismo se graficó la función L(r) – empírica y teórica, debido a que es más fácil visualizar gráficamente los resultados, sobre todo a pequeñas escalas espaciales.

Para calcular la función K de Ripley bivariada se tomo en cuenta la variable ARS y la variable agregaciones de peces. Se calculó la función K de Ripley bivariada tomando como puntos fijos las agregaciones de peces, por ello en este caso se cuentan los puntos de ARS alrededor de cada agregación de peces. También se considero un radio máximo de 20 km y se grafico la función L(r) – empírica y teórica para cada escala temporal respectivamente.

### 3.5 Aplicación de geoestadística

Primero se evaluó la asimetría de la variable biomasa de peces epipelágicos y la profundidad del límite superior de la zona de mínimo oxígeno mediante histogramas, gráficos de probabilidad normal y la prueba de normalidad Shapiro-Wilks. Cuando se comprobó la asimetría de la variable se procedió a la transformación logarítmica respectiva. A continuación se realizaron variogramas bidireccionales para cada variable debido a que la costa peruana tiene una inclinación de 29° en la dirección este-oeste y 129° en la dirección norte-sur. En caso los variogramas bidireccionales resultaron ser distintos se

ajustaron modelos para cada variograma caso contrario se ajustó un modelo al variograma omnidireccional. Posteriormente se interpolaron los datos mediante un kriging ordinario usando una vecindad móvil para lo cual se usó una grilla con celdas de 1.1km.

### 3.6 Aplicación de MLG

Previo a la aplicación de los MLG se realizó un análisis exploratorio mediante histogramas de frecuencias de las covariables biomasa de anchoveta (*NASC*), Distancia a la isla, Profundidad del límite superior de la ZMO y latitud para conocer la naturaleza de los datos. Luego las covariables fueron exploradas mediante diagramas de cajas para determinar la existencia de valores atípicos. Se consideran como valores atípicos los datos menores a  $Q1 - 1.5 \times (Q3 - Q1)$  o mayores a  $Q3 + 1.5 \times (Q3 - Q1)$ , donde  $Q3$  es el tercer cuartil (75%) de los datos,  $Q1$  es el primer cuartil (25%) de los datos y  $(Q3 - Q1)$  es el rango intercuartil (Zuur et al., 2009). Dichos valores atípicos fueron eliminados de la base de datos. Se evaluó la normalidad de las covariables mediante la prueba de Anderson-Darling. Para determinar la relación entre las covariables, se utilizó el coeficiente de correlación de Spearman ( $\rho$ ) empleado para medir el grado de asociación entre dos variables en estadística no paramétrica (Siegel y Castellan., 2003). El  $\rho$  oscila entre +1 y -1, indicando asociaciones fuertes cuando el valor absoluto  $\rho$  es muy cercano a la unidad ( $\rho > 0.7$ ), provocando multicolinealidad entre las covariables.

En este estudio se realizó MLG para determinar la relación entre la distribución de las aves y la biomasa de peces, profundidad del límite superior de la ZMO, distancia a la isla y latitud, a diferentes escalas.

En primer lugar se utilizaron los datos de presencia y ausencia de ARS (P/A). Como en este caso la variable respuesta es binaria se distribuye como una Binomial y las funciones de enlace posibles son *logit*, *probit* o *cloglog*. Esto permite modelar altas variabilidades así como los ceros frecuentes que corresponden a la ausencia de ARS, debido posiblemente a comportamientos similares de las aves frente a variables ambientales, o la condición gregaria de su alimento, entre otras causas. Se evaluó la bondad de ajuste mediante la función de desviación y el estadístico Chi-cuadrado de Pearson, si los p-valores son muy pequeños ( $p\text{-valor} < 0.05$ ), proveerán evidencia de un pobre ajuste del modelo (Agresti, 1996). En caso la bondad de ajuste sea adecuada concluiremos que la distribución Binomial se ajustó adecuadamente a la variable respuesta y se evaluó la significancia del modelo (pruebas Wald y razón de verosimilitud). Caso contrario la falta de ajuste puede deberse a una mala especificación de la función de enlace o a la existencia de outliers, se evaluaron ambos casos mediante análisis de residuales y diagnóstico de MLG.

En segundo lugar se modela la abundancia de ARS, en este caso la variable respuesta son conteos de ARS (solo cuando está presente), son valores positivos ( $\geq 1$ ) por lo tanto sería inadecuado asumir distribuciones normal o Gamma, se puede asumir una distribución Poisson. Sin embargo esta asume que la media y varianza son iguales, y si la varianza es mucho mayor que la media puede que exista sobredispersión, se realizó un gráfico de la media respecto de la varianza estimada de cada observación para comprobarlo. Y luego se evaluó la bondad de ajuste mediante la función de desviación y el estadístico de Chi-cuadrado de Pearson. Se evalúa la significancia del modelo (pruebas de Wald y razón de verosimilitud), las funciones de enlace (*log*, *raíz* o *identidad*), outliers, residuales, hasta ajustar un modelo adecuado y significativo. En caso la bondad de ajuste no sea adecuada concluiremos que la distribución Poisson no es adecuada. Entonces otra opción fue asumir una distribución Binomial negativa, la cual acepta cierto grado de sobredispersión. Y nuevamente se evaluó el modelo, las funciones de enlace, outliers, residuales, hasta ajustar un modelo adecuado y significativo.

Finalmente se compararon todos los MLG con una adecuada bondad de ajuste (función de desviación y el estadístico de Chi-cuadrado  $p\text{-valor} > 0.05$ ) y aceptación de la significancia de los modelos (pruebas de Wald y razón de verosimilitud  $p\text{-valor} < 0.05$ ) mediante la evaluación del menor AIC y BIC para cada escala. Se aplicó una eliminación Stepwise basada en el AIC, donde el mejor modelo tiene menor AIC. Y se determinó cuáles de las covariables se debían incluir en cada MLG y se escogió el modelo final para cada escala.

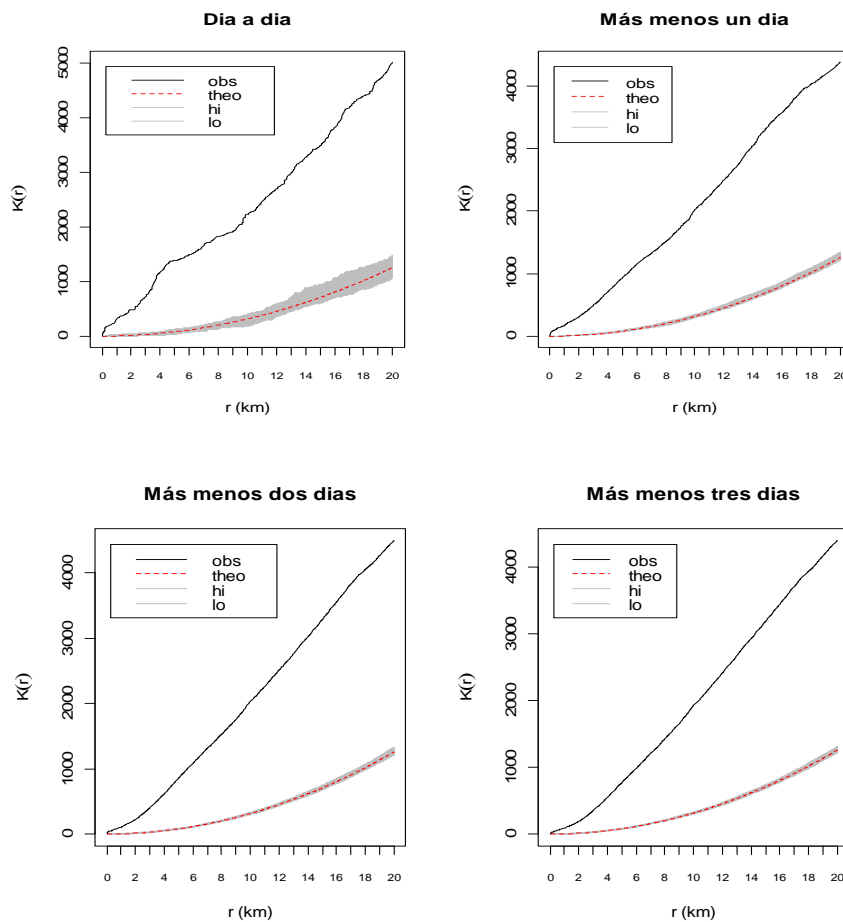
# **CAPITULO IV**

## **RESULTADOS**

## 4.1 Resultados de procesos puntuales

### 4.1.1 Distribución espacial de las ARS

La aplicación de la función K de Ripley para cada escala temporal (Fig. 33) muestra que la función K de Ripley observada cae fuera de los límites de la envoltura que define la distribución teórica de una Poisson homogénea. Este patrón indica que globalmente existe evidencia estadística que las ARS no se distribuyen de forma independiente. Además se muestra que en todas las escalas temporales la función K de Ripley observada es mayor que el límite superior de la envoltura ( $>hi$ ). Por lo tanto existe evidencia estadística significativa que las ARS están agrupadas en el espacio, formando parches en algunos lugares de las trayectorias. Sin embargo probablemente el proceso sea atractivo solo de forma indirecta porque no hay justificación para que varias ARS se encuentren cerca a menos que reflejen otro proceso que si lo sea como es la distribución de los peces. También se calculó y graficó la función  $L(r)-r$  (Anexo 1), los resultados fueron similares a los obtenidos con la función K de Ripley.



**Figura 33. Función K de Ripley (obs) de ARS para cada escala temporal. Función teórica de k de Ripley de Proceso Homogéneo de Poisson de k de Ripley (theo). Envoltura de la distribución teórica ( $<hi$  y  $>lo$ ). Radio en Km.**

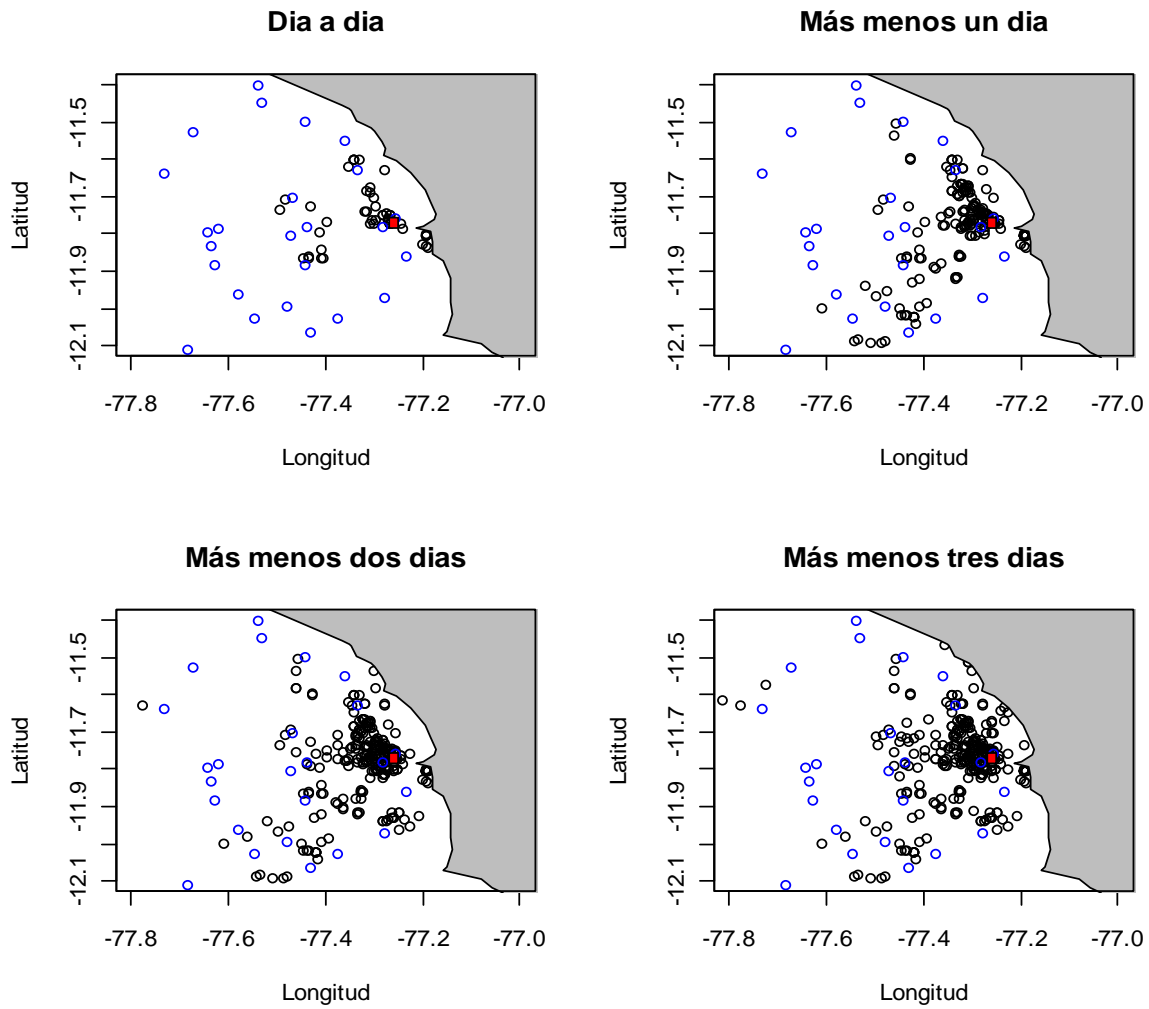
### 4.1.2 Distribución espacial de ARS y peces

Se realizaron gráficos exploratorios de la distribución de las ARS y agregaciones de peces epipelágicos en la zona Topineme en todas las escalas temporales (Fig. 34). Se observaron agregaciones dispersas de peces epipelágicos que cubren los alrededores de la Isla Pescadores así como un elevado número de ARS cercanas a dichas agregaciones.

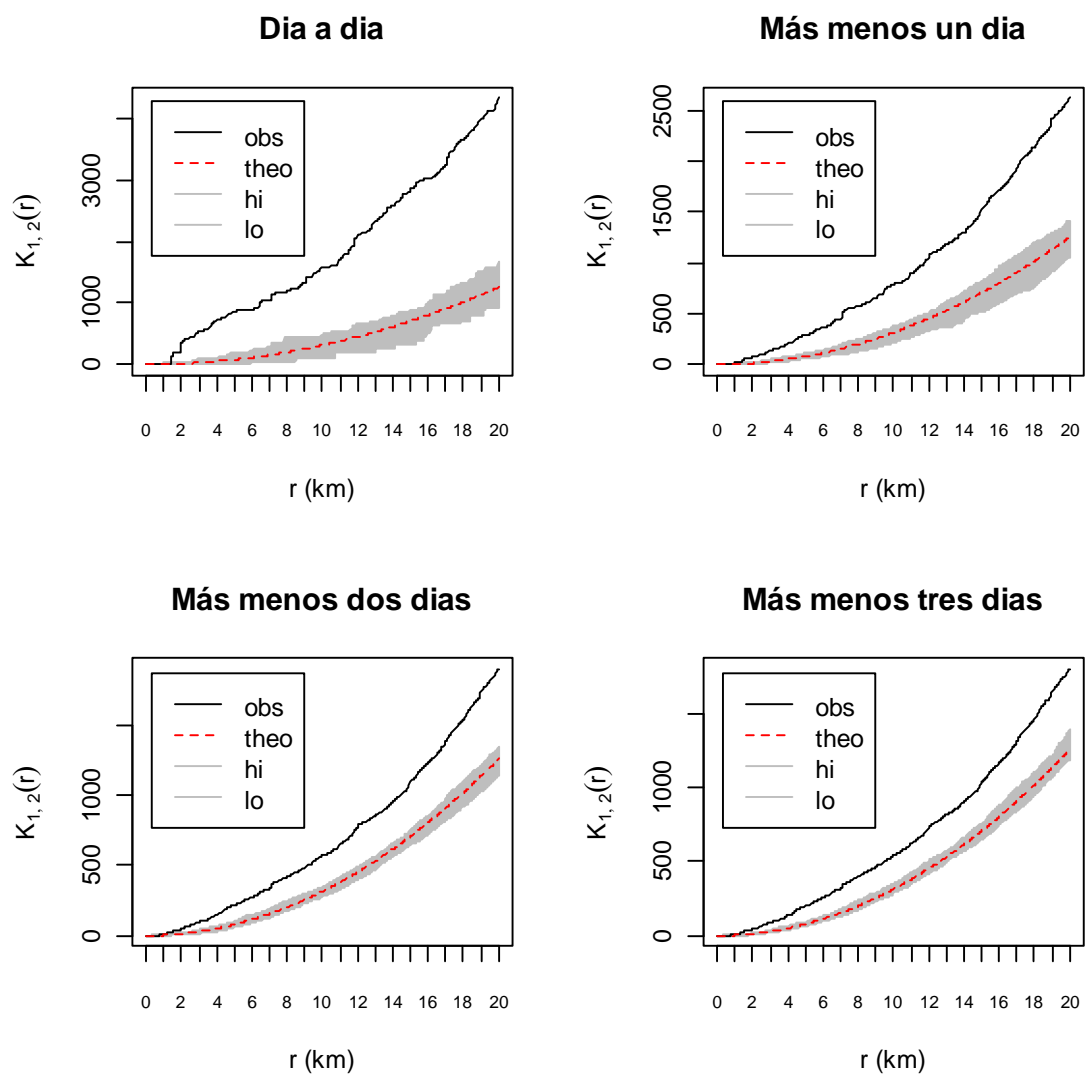
La aplicación de la función K de Ripley bivariada para cada escala temporal (Fig. 35) muestra que globalmente, la función K de Ripley bivariada observada es mayor que el límite superior de la envoltura que define la distribución teórica de una Poisson Homogénea. Por lo tanto las ARS y agregaciones de peces epipelágicos no se distribuyen de forma independiente y existe evidencia significativa que existe asociación espacial entre las ARS y las agregaciones de peces en todas las escalas.

Los gráficos de la función  $L(r)-r$  (Fig. 36) muestran que:

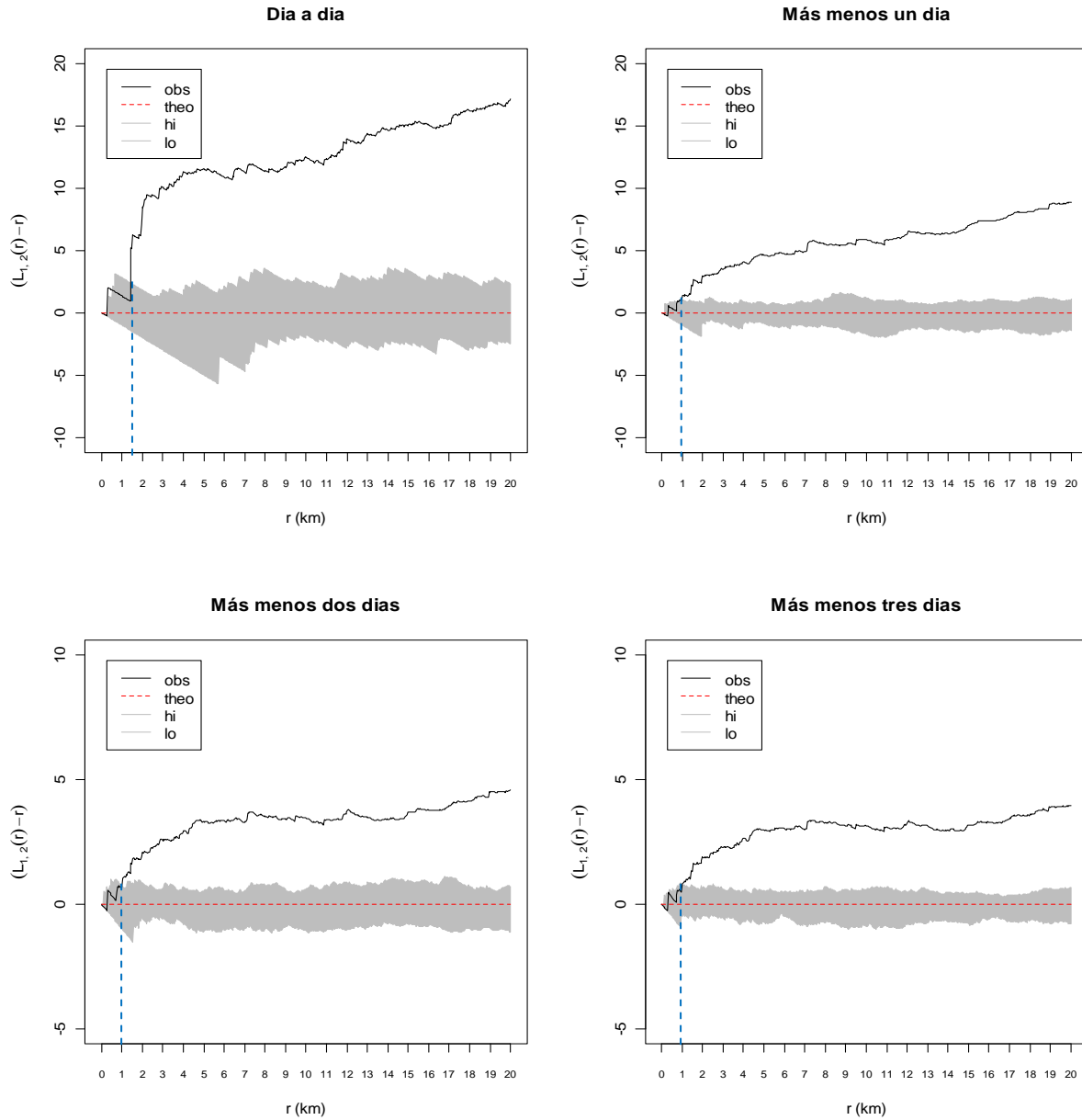
- En la escala temporal día-día hay evidencia que cuando la distancia entre las agregaciones de peces y las ARS es menor a 1.5 km aproximadamente existe cierto grado de atracción o agregación entre las ARS y agregaciones de peces. Sin embargo no llega a ser estadísticamente significativa debido a que la función  $L(r)-r$  empírica no es mayor que el límite superior de la envoltura ( $>h$ ). Por lo tanto se considera que a escala inferior a 1.5 km, con nuestros resultados las ARS y agregaciones de peces parecen distribuirse de forma independiente es decir que no existe interacción entre las ARS y las agregaciones de peces, sin embargo dicha asociación negativa puede deberse al límite de la resolución de nuestros datos, a muy pocos datos, y no realmente al proceso biológico. Cuando la distancia máxima entre las agregaciones de peces y las ARS es superior a 1.5 km, hay evidencia de que las ARS y agregaciones de peces no se distribuyen de forma independiente, es decir existe interacción de atracción o agregación entre ambos procesos puntuales (Fig. 36a).
- Cuando se incrementa la escala temporal con un límite de tolerancia de un día, dos días, o tres días hay evidencia que cuando la distancia entre las agregaciones de peces y las ARS es menor a 1 km existe cierto grado de atracción o agregación entre las ARS y agregaciones de peces. Sin embargo no llega a ser estadísticamente significativa. Por lo tanto se considera que a escala inferior a 1 km, con nuestros resultados las ARS y agregaciones de peces parecen distribuirse de forma independiente es decir que no existe interacción entre las ARS y las agregaciones de peces, relación que puede deberse al límite de la resolución de nuestros datos y no realmente al proceso biológico mismo. Cuando la distancia entre las agregaciones de peces y las ARS es superior a 1 km hay evidencia de que las ARS y agregaciones de peces no se distribuyen de forma independiente, es decir existe interacción de atracción o agregación entre ambos procesos puntuales (Fig. 36 b, c y d).



**Figura 34. Distribución espacial de ARS (burbujas negras) y agregaciones de peces epipelágicos (burbujas azules). Ubicación de la Isla Pescadores (cuadrado rojo).**



**Figura 35.** Función K de Ripley bivariado (obs) de ARS y agregación de peces epipelágicos para cada escala temporal. Función teórica de k de Ripley bivariado de Proceso Homogéneo de Poisson (theo). Envoltura de la distribución teórica de k de Ripley bivariado (<hi >lo). Radio en Km.



**Figura 36. Función  $L(r) - r$  de ARS y agregaciones de peces para cada escala temporal. Función teórica de  $L(r) - r = 0$  de Proceso Homógeno de Poisson (theo). Envoltura de la distribución teórica  $L(r) - r$  ( $<hi$  y  $>lo$ ). Línea punteada azul representa el radio( $r$ ) a partir del cual las ARS y agregaciones de peces interactúan atractivamente. Radio  $r$  en Km.**

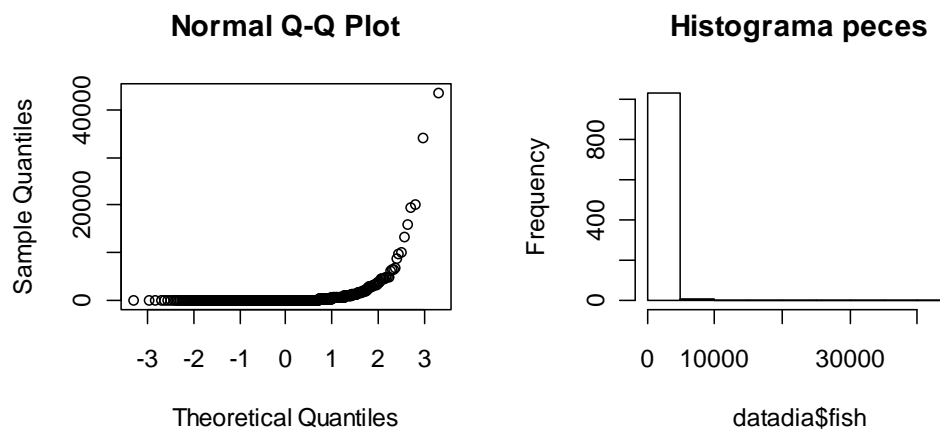


## 4.2 Aplicación de geoestadística

### 4.2.1 Distribución espacial de la biomasa de peces

Se procedió a evaluar la normalidad (Fig. 37) de la variable biomasa de peces epipelágicos. La distribución no resultó ser normal, es positivamente sesgada con sobredispersión, pues tiene un gran número de valores cero que no deben eliminarse cuando son ceros verdaderos y válidos. También se realizó la prueba de normalidad Shapiro-Wilk y se rechazó la normalidad de la variable biomasa de peces epipelágicos ( $W=0.17$  y  $p\text{-valor}= 2.2e-16 < 0.05$ ) con el 95% de confianza. Estos resultados nos indican que realizar el variograma con la variable biomasa de peces epipelágicos transformada es viable.

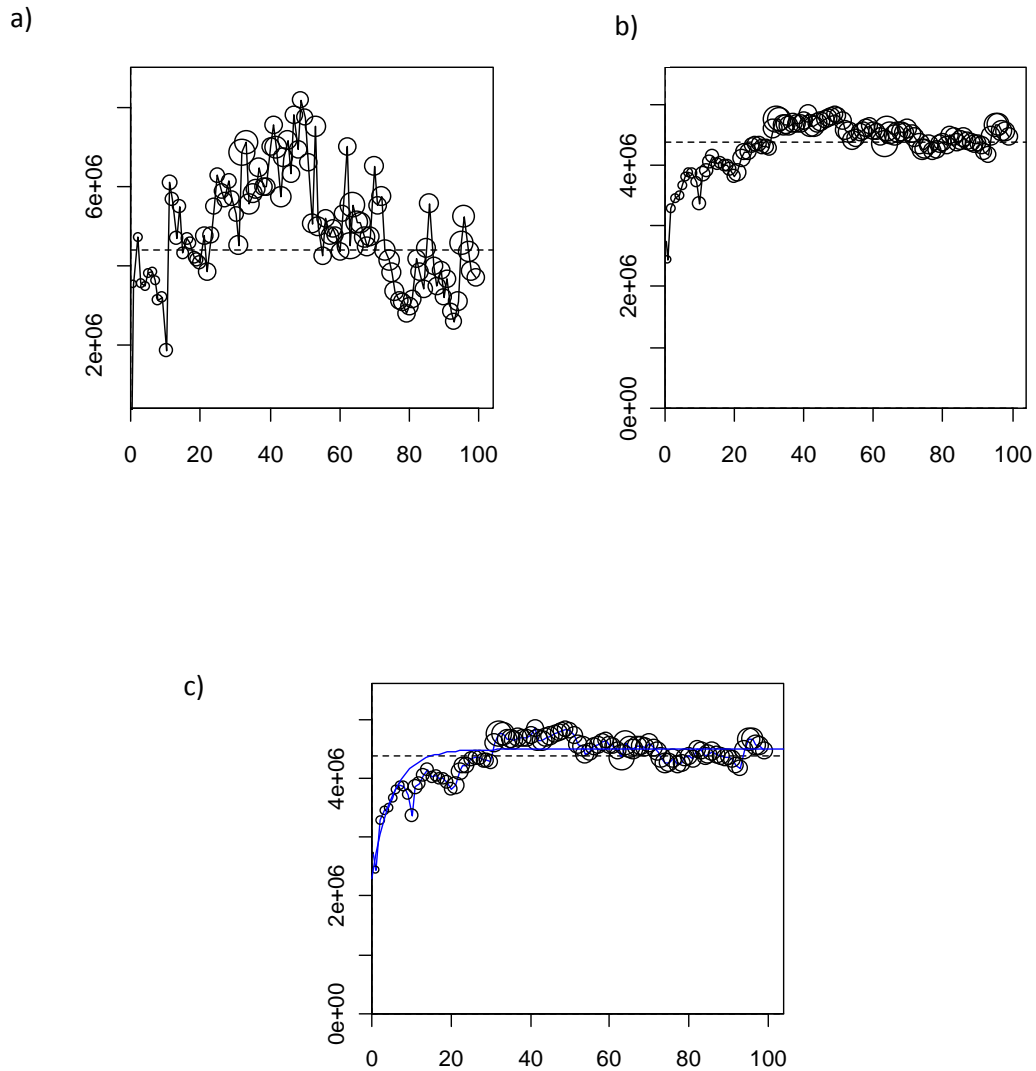
Primero se realizaron los variogramas bidireccionales con una inclinación de  $29^\circ$  en la dirección este-oeste y  $129^\circ$  en la dirección norte-sur con los datos originales, como no mostraron diferencias significativas (Anexo 2) se procedió a realizar un variograma experimental omnidireccional.



**Figura 37. Evaluación gráfica de la normalidad de la variable biomasa de peces epipelágicos**

El variograma experimental omnidireccional se realizó con los datos originales de la biomasa de peces epipelágicos (Fig. 38a) y también se realizó otro variograma con una transformación logarítmica (Fig. 38b).

Como el variograma experimental usando la transformación logarítmica de la variable es más robusto, se procedió a modelar este variograma experimental. Al variograma experimental se ajustó un modelo exponencial (Fig. 38c), los parámetros correspondientes se muestran en la Tabla 8. Se identificó que la primera estructura espacial se encontraba en un rango menor a 15km, lo cual demuestra que la variabilidad a pequeñas distancias es alta.

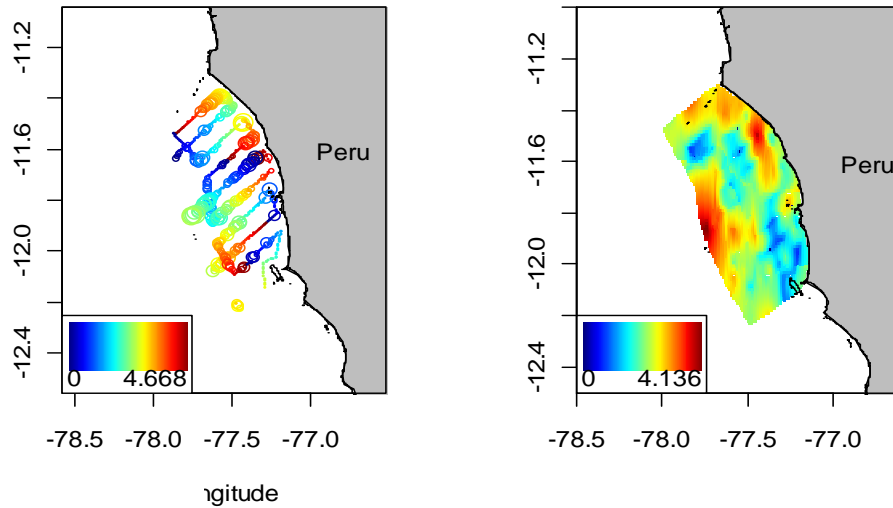


**Figura 38.** a) El variograma experimental de la variable biomasa de peces epipelágicos no transformada comparado con b) El variograma experimental usando la transformación logarítmica de la variable biomasa de peces epipelágicos. c) Modelación de b. Los círculos indican el número de puntos para cada punto en los variogramas. La escala horizontal es km.

*Tabla 8. Modelo ajustado de la variable biomasa de peces epipelágicos.*

MODELO	Rango (Km)	Meseta
Nugget(pepita)	-	2300000
Exponencial	15	2200000

Se obtuvo un mapa del logaritmo de la biomasa de peces epipelágicos  $\log(\text{NASC} + 1)$  en la zona Topineme (Fig. 39) utilizando el método de kriging ordinario con vecindad móvil. Se usó una grilla con celdas de 1.1km para realizar las interpolaciones y estimar la biomasa de peces epipelágicos en lugares no muestreados.



**Figura 39. a) Distribución espacial del  $\log_{10}(\text{NASC} + 1)$  b) Interpolación espacial  $\log_{10}(\text{NASC} + 1)$  mediante el método de kriging ordinario.**

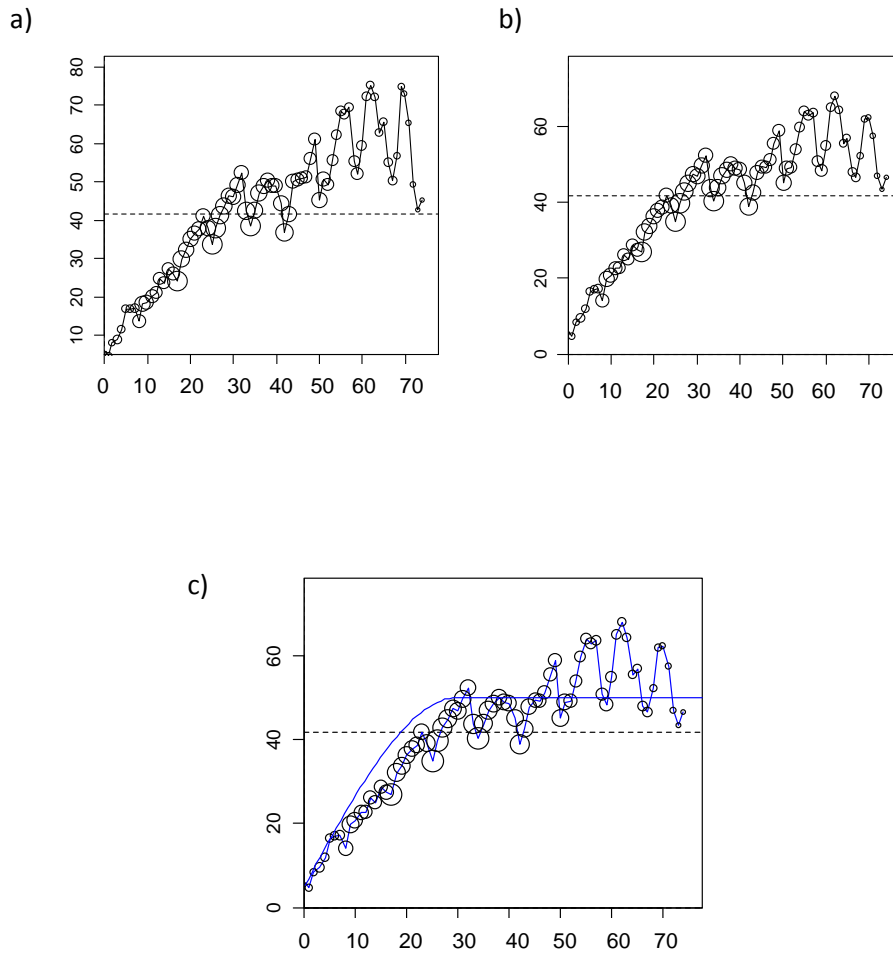
#### **4.2.2 Distribución espacial de la profundidad del límite superior de la zona de mínimo oxígeno**

Se calculó la prueba de normalidad Shapiro-Wilk y se rechazó la normalidad de la variable profundidad del límite superior de la ZMO ( $W=0.8328$  p-valor  $2.2 \times 10^{-16} < 0.05$ ) con el 95% de confianza. Se realizaron variogramas bidireccionales con una inclinación de  $29^\circ$  en la dirección este-oeste y  $129^\circ$  en la dirección norte-sur con los datos originales, como no mostraron diferencias significativas se procedió a realizar un variograma experimental omnidireccional (Anexo 3).

El variograma experimental omnidireccional se realizó con los datos originales (Fig. 40a) y también se realizó otro variograma con una transformación logarítmica (Fig. 40b) como se realizó con la biomasa de peces epipelágicos. Al variograma experimental usando la transformación logarítmica se ajustó un modelo esférico (Fig. 40c), los parámetros correspondientes se muestran en la Tabla 9. Se identificó que la mayor parte de la estructura espacial se encontraba en un rango de 30 km.

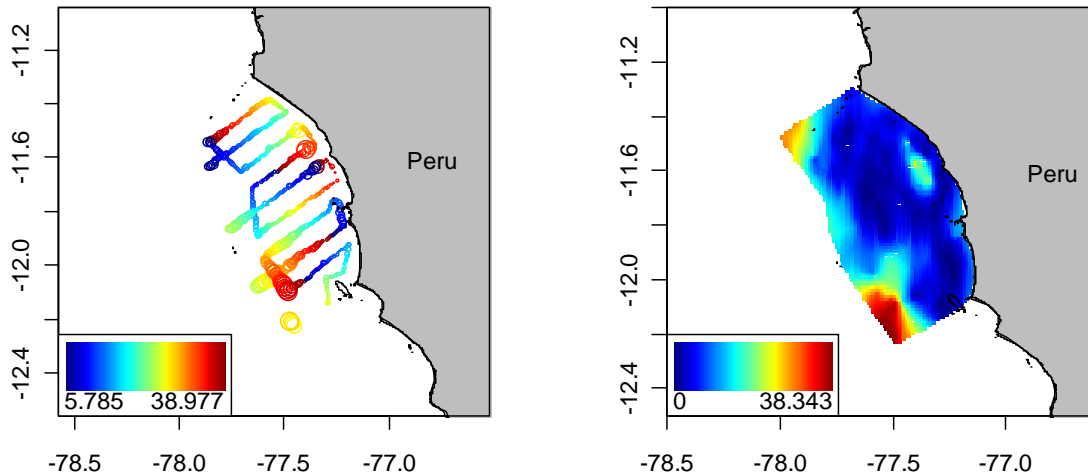
*Tabla 9. Modelo ajustado de la variable Profundidad del límite superior de la ZMO*

MODELO	Rango (Km)	Meseta
Nugget (pepita)	-	7
Esférico	30	43



**Figura 40. a) El variograma experimental de la variable original profundidad del límite superior de la ZMO no transformada comparado con b) El variograma experimental usando la transformación logarítmica de la profundidad del límite superior de la ZMO. c) Modelación de b. Los círculos indican el número de puntos para cada punto en los variogramas. La escala horizontal es km.**

Se obtuvo un mapa de la profundidad de la zona de mínimo oxígeno en la zona Topineme (Fig. 41) utilizando el método de kriging ordinario con vecindad móvil. Se usó una grilla con celdas de 1.1km para realizar las estimaciones de la profundidad del límite superior de la ZMO en lugares no muestreados.



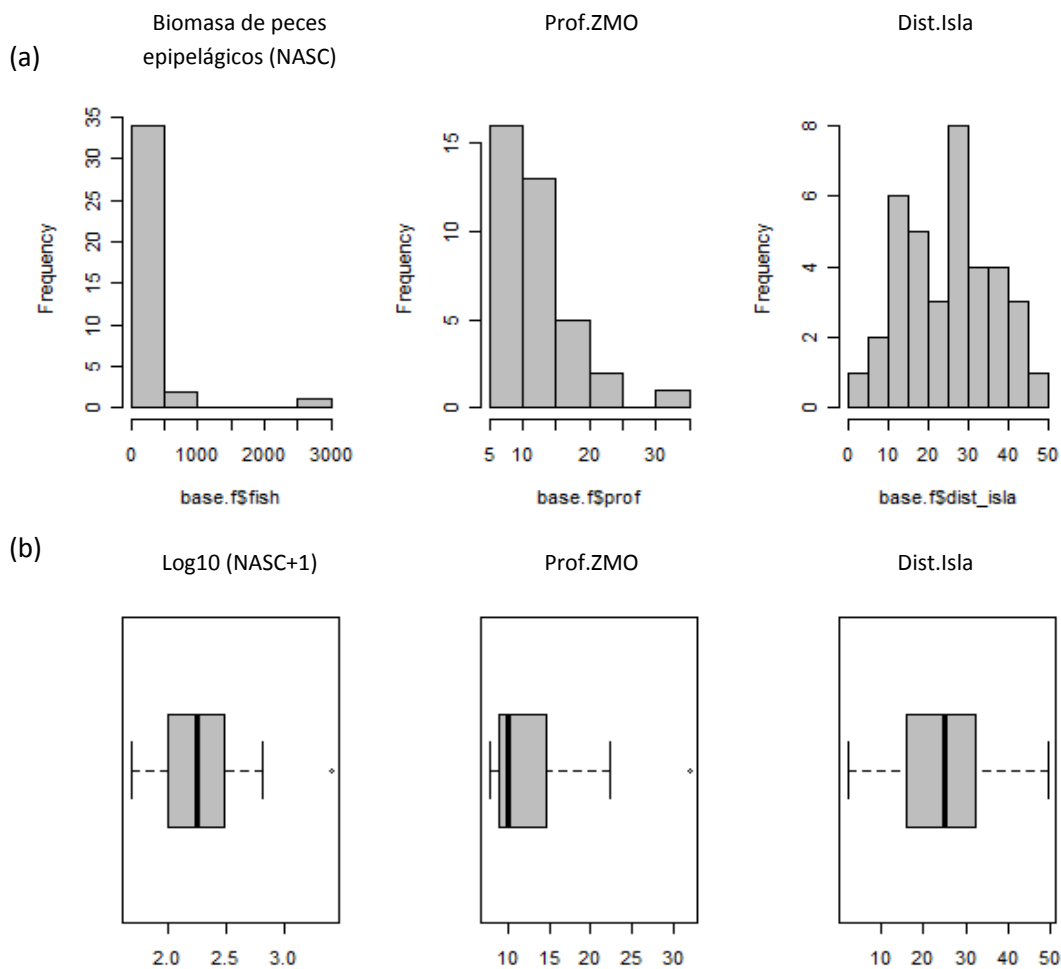
**Figura 41. a) Distribución espacial de la Profundidad del límite superior de la ZMO b) Interpolación espacial de la profundidad del límite superior de la ZMO mediante el método de kriging ordinario.**

### 4.3 Resultados de Modelos Lineales Generalizados

#### 4.3.1 Resultados MLG Presencia – Ausencia (Probabilidad de presencia de ARS)

Como ilustración del análisis exploratorio mediante histogramas de frecuencias de las covariables se presentan los datos provenientes de la escala 8.9km con tolerancia de tres días. En este caso particular los histogramas de frecuencias enseñan que las covariables *biomasa de peces epipelágicos* (NASC) y profundidad del límite superior de la ZMO no tienen una distribución normal (Fig. 42a), mientras que la distancia a la isla si se acerca a una distribución normal. Los p-valores que resultan de la prueba de Anderson – Darling para biomasa de peces epipelágicos (NASC), profundidad del límite superior de la ZMO (Prof.ZMO) y Distancia a la isla (Dist.Isla) son 4.177e-13, 0.0008185 y 0.9452, respectivamente. Debido a la de asimetría de la variable biomasa de peces epipelágicos (NASC), se transformó mediante  $\log_{10}(NASC+1)$  (Gutiérrez et al., 2008). Luego se realizaron los diagramas de caja, estos diagramas señalan la existencia de valores atípicos (círculos negros en la Fig. 42b), los cuales no fueron removidos debido a que no se encontró evidencia suficiente que nos lleve a pensar que no son observaciones reales.

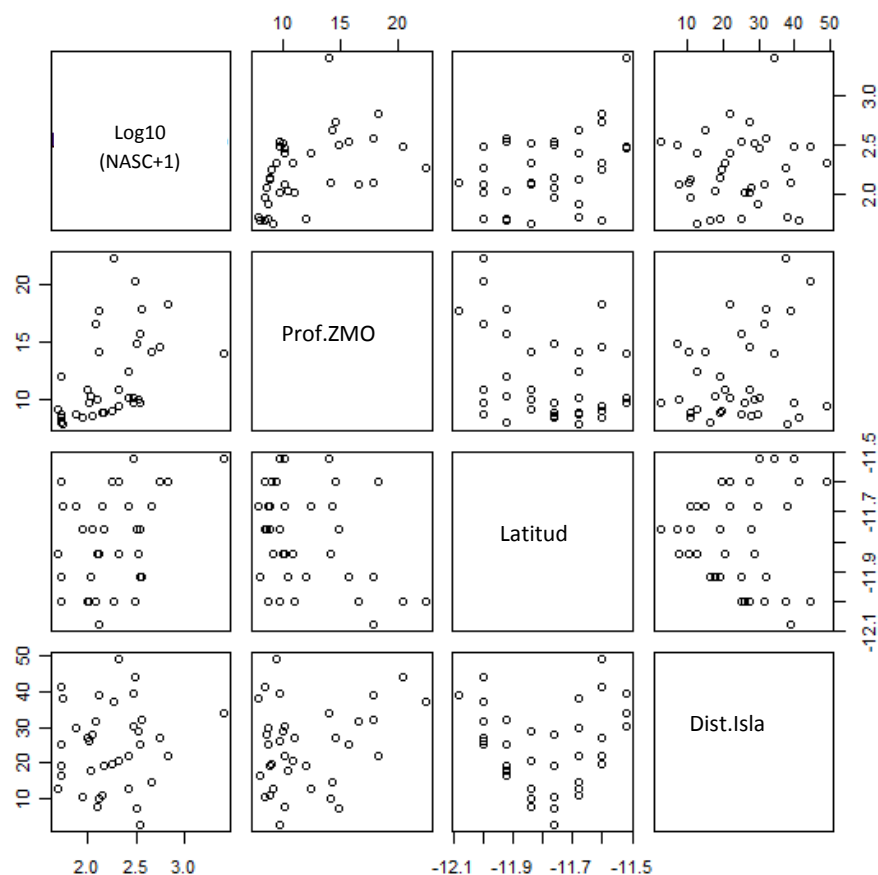
Las correlaciones entre las covariables biomasa de peces epipelágicos, profundidad del límite superior de la ZMO, latitud y Distancia a la isla fueron relativamente débiles ( $\rho < 0.8$ ), por lo tanto no hay evidencia de posible multicolinealidad entre dichas covariables (Tabla 10). En particular existe una ligera tendencia que a mayor profundidad del límite superior de la ZMO la biomasa de peces epipelágicos también sea mayor (Fig. 43). Estos procedimientos se realizaron para todas las escalas analizadas, los resultados se presentan en el Anexo 4, 5 y 6.



**Figura 42. (a) Histogramas de frecuencias y (b) Diagramas de caja de las covariables correspondientes a la escala 8.9km- más menos tres días.**

*Tabla 10. Correlación de Spearman ( $\rho$ ) entre las covariables correspondientes a la escala 8.9km - más menos tres días.*

	Log10(NASC+1)	Prof.ZMO	Latitud	Dist.Isla
Log10(NASC+1)	1	0.58	0.45	0.18
Prof.ZMO	0.58	1	-0.2	0.38
Latitud	0.45	-0.2	1	-0.26
Dist.Isla	0.18	0.38	-0.26	1



**Figura 43. Gráficos de relaciones entre covariables correspondientes a la escala 8.9km - más menos tres días.**

Para cada escala se calcularon posibles modelos ajustados a la presencia-ausencia de ARS, entre los cuales se comparo su bondad de ajuste y significancia global del modelo hasta escoger el “mejor” modelo (Anexo 7).

La aplicación de los MLG finales ajustados a la presencia-ausencia de ARS en relación a variables físicas y biológicas permitió obtener los siguientes resultados (Tabla 11):

- En general se observa que hay diferencia entre las escalas más grandes y las más pequeñas, las más notables en la profundidad del límite superior de la ZMO y  $\log_{10}(NASC+1)$ . Además respecto a la latitud mientras más pequeña es la escala, mayor es el efecto negativo y la probabilidad de encontrar ARS es menor a medida que se incrementa la latitud. Respecto a la distancia a la isla Pescadores es la covariable que permanece significativa en la mayoría de las escalas, ello induce a pensar que la probabilidad que ocurra un ARS se incrementa cuando la distancia a la isla es menor.
- En la escala más grande (8.9km) la profundidad del límite superior de la ZMO y distancia a la isla tienen un efecto significativo positivo y negativo respectivamente, es decir cuando se incrementa la profundidad del límite superior de la ZMO la probabilidad de encontrar un ARS también se incrementa, y cuando aumenta la distancia a la isla Pescadores la probabilidad de encontrar ARS disminuye.

- En la escala 6.6km la profundidad del límite superior de la ZMO y la distancia a la isla Pescadores también tienen un efecto significativo positivo y negativo respectivamente. Conforme se va reduciendo la escala, en 4.4km (mas menos dos días) la única variable significativa resulta ser la distancia a la isla Pescadores, para que luego en la escala 4.4km (mas menos un día) la covariable latitud también tenga un efecto significativo negativo junto con la variable distancia a la costa en la probabilidad de encontrar ARS.
- En las escalas más pequeñas (2.2km - 1.1km) mientras que la latitud y distancia a la isla Pescadores siguen siendo significativas negativamente, la covariable  $\log(NASC+1)$  ahora tiene un efecto significativo positivo, es decir cuando se incrementa  $\log(NASC+1)$  la probabilidad de encontrar ARS aumenta.

El modelo que explico mejor la probabilidad de encontrar ARS es el ajustado a la mayor escala (desviación explicada=39.28%). La bondad de ajuste de los modelos a la escala 4.4km no se acepto con la función de desvío sin embargo si se acepto con la prueba de Chi-cuadrado de Pearson. En el resto de modelos se acepto la bondad de ajuste con la función de desvío (p-valor >0.05). Todos los modelos ajustados resultaron ser significativos tanto con la prueba de Wald como de LRT (p-valor <0.05, Tabla 12). Para validar el GLM asumiendo una familia Binomial se utilizaron gráficos de gráficos con envolturas de los cuantiles de los residuos de desviación (debido a que los residuos no son normales y su varianzas no son iguales) y otros gráficos de diagnóstico (Anexo 8 y 9). Los “mejores” modelos ajustados según las funciones de enlace respectivas se muestran en la Tabla 13.



Tabla 11. Resultados de GLM (especificando la función de enlace) utilizados para estudiar la relación entre Ausencia-Presencia de ARS y las covariables acústicas (Logaritmo de la biomasa de peces epipelágicos (log (NASC+1)), profundidad del límite superior de la ZMO (Prof.ZMO), latitud y distancia a la Isla Pescadores (Dist.Isla). Para cada modelo el tamaño de muestra, coeficiente estimado y p-valor de las covariables son mostrados.

Función de enlace	Escala Espacial	Escala Temporal	Tamaño de muestra	Log(NASC+1)		Prof.ZMO		Latitud		Dist.Isla	
				Estimado	p-valor	Estimado	p-valor	Estimado	p-valor	Estimado	p-valor
logit	8.9km	más menos tres días	37	-	-	0.4122	0.03372(*)	-	-	-0.1570	0.01901(*)
cloglog	6.7km	más menos tres días	60	-	-	0.1788	0.00165(**)	-	-	-0.0759	0.00122(**)
cloglog	6.7km	más menos dos días	54	-	-	0.1645	0.00568(**)	-	-	-0.0938	0.00204(**)
cloglog	6.7km	más menos un día	53	-1.7145	0.01125(*)	0.1373	0.01588(*)	-	-	-0.0585	0.01893(*)
logit	4.4km	más menos tres días	123	-	-	0.0685	0.08691(.)	-	-	-0.0615	0.00181(**)
cloglog	4.4km	más menos dos días	109	-	-	-	-	-	-	-0.0229	0.08075(.)
cloglog	4.4km	más menos un día	105	-	-	-	-	-3.3584	0.00852(**)	-0.0292	0.06660(.)
cloglog	2.2km	más menos dos días	347	0.6002	0.06678(.)	-	-	-3.7065	0.00076(***)	-0.0643	0.00000(***)
cloglog	2.2km	más menos un día	320	0.6924	0.05930(.)	-	-	-5.8743	0.00003(***)	-0.0653	0.00000(***)
cloglog	1.1km	más menos un día	909	1.5575	0.00003(***)	-0.0871	0.03090(*)	-6.5620	0.00000(***)	-0.0596	0.00001(***)
logit	1.1km	día-día	500	1.5328	0.01744(*)	-	-	-19.1618	0.00279(**)	-0.1222	0.00971(**)

**Códigos de Significancia de las covariables:** <0, 0.001 > ‘\*\*\*’; <0.001, 0.01 > ‘\*\*’; <0.01, 0.05 > ‘\*’; < 0.05,0.1 > ‘.’

Tabla 12. Resultados de Bondad de ajuste y significancia de GLM utilizados para estudiar la relación entre Ausencia-Presencia de ARS y las covariables acústicas (log (peces+1), profundidad del límite superior de la ZMO, latitud y distancia a la Isla Pescadores).

Escala Espacial	Escala Temporal	Desviación Nula	Desviación Residual	Función de Desviación (p-valor)	Prueba de Pearson (p-valor)	AIC	BIC	Desviación Explicada%	Prueba Wald (p-valor)	Prueba LRT (p-valor)
8.9km	más menos tres días	38.634	23.457	0.913	0.0004	29.457	45.123	39.28%	0.024(*)	0.001(**)
6.7km	más menos tres días	73.304	51.821	0.669	0.458	57.821	76.387	29.31%	0.002(**)	2.16E-05(***)
6.7km	más menos dos días	65.631	50.533	0.492	0.54	56.533	74.467	23.00%	0.007(**)	0.001(**)
6.7km	más menos un día	67.923	54.817	0.264	0.377	62.817	86.579	19.30%	0.018(*)	0.004(**)
4.4km	más menos tres días	170.116	158.555	0.011	0.357	164.555	187.428	6.80%	0.008(**)	0.003(**)
4.4km	más menos dos días	150.877	147.935	0.005	0.423	151.935	166.701	1.95%	0.081(.)	0.086(.)
4.4km	más menos un día	142.796	133.547	0.02	0.367	139.547	161.471	6.48%	0.027(*)	0.01(*)
2.2km	más menos dos días	414.937	381.854	0.073	0.402	389.854	428.649	7.97%	9.74E-07(***)	3.09E-07(***)
2.2km	más menos un día	357.681	327.167	0.321	0.408	335.167	373.313	8.53%	2.42E-05(***)	1.08E-06(***)
1.1km	más menos un día	638.325	587.04	1	0.001	597.04	655.163	8.03%	3.67E-10(***)	1.95E-10(***)
1.1km	día-día	243.167	222.948	1	0.363	230.948	272.665	8.31%	0.015(*)	0.0002(***)

**Códigos de Significancia del modelo:** <0, 0.001 > ‘\*\*\*’; <0.001, 0.01 > ‘\*\*’; <0.01, 0.05 > ‘\*’; < 0.05, 0.1 > ‘.’

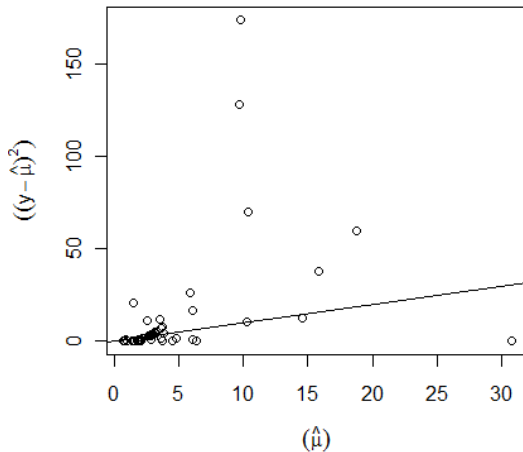
Tabla 13. MLG ajustados a la Ausencia-Presencia de ARS respecto de las covariables acústicas acústicas (Logaritmo de la biomasa de peces epipelágicos ( $\text{Log}(\text{NASC}+1)$ ), profundidad del límite superior de la ZMO ( $\text{Prof.ZMO}$ ), latitud ( $\text{Lat}$ ) y distancia a la Isla Pescadores ( $\text{Dist.Isla}$ ).  $\hat{y} = \hat{\pi}(x)$  probabilidad de presencia de ARS.

Escala Espacial	Escala Temporal	Función de enlace	Modelos estimados
8.9km	más menos tres días	Logit	$\hat{\pi}(x) = \frac{\exp(0.924 + 0.412 * \text{prof. ZMO} - 0.157 * \text{Dist. Isla})}{1 + \exp(0.924 + 0.412 * \text{prof. ZMO} - 0.157 * \text{Dist. Isla})}$
6.7km	más menos tres días	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(-0.065 + 0.1788 * \text{Prof. ZMO} - 0.0759 * \text{Dist. Isla})]$
6.7km	más menos dos días	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(0.322 + 0.1645 * \text{Prof. ZMO} - 0.093 * \text{Dist. Isla})]$
6.7km	más menos un día	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(3.546 - 1.715 * \text{Log}(\text{NASC} + 1) + 0.137 * \text{Prof. ZMO} - 0.058 * \text{Dist. Isla})]$
4.4km	más menos tres días	Logit	$\hat{\pi}(x) = \frac{\exp(0.733 + 0.0685 * \text{prof. ZMO} - 0.0615 * \text{Dist. Isla})}{1 + \exp(0.733 + 0.0685 * \text{prof. ZMO} - 0.0615 * \text{Dist. Isla})}$
4.4km	más menos dos días	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(0.201 - 0.023 * \text{Dist. Isla})]$
4.4km	más menos un día	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(-39.686 - 3.358 * \text{Lat} - 0.029 * \text{Dist. Isla})]$
2.2km	más menos dos días	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(-44.953 + 0.6002 * \text{Log}(\text{NASC} + 1) - 3.7065 * \text{Lat} - 0.064 * \text{Dist. Isla})]$
2.2km	más menos un día	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(-70.993 + 0.6924 * \text{Log}(\text{NASC} + 1) - 5.874 * \text{Lat} - 0.065 * \text{Dist. Isla})]$
1.1km	más menos un día	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(-81.033 + 1.558 * \text{Log}(\text{NASC} + 1) - 0.087 * \text{Prof. ZMO} - 6.562 * \text{Lat} - 0.059 * \text{Dist. Isla})]$
1.1km	día-día	Clog-log	$\hat{\pi}(x) = 1 - \exp[-\exp(-231.362 + 1.533 * \text{Log}(\text{NASC} + 1) - 19.16 * \text{Lat} - 0.122 * \text{Dist. Isla})]$

### 4.3.2 Resultados MLG conteos (Número de ARS - si está presente)

Previo al análisis de MLG conteos se llevan a cabo el análisis exploratorio para todas las escalas analizadas, los resultados se presentan en el Anexo 10, 11 y 12.

Se realizaron gráficos de la media respecto de la varianza estimada del número de ARS (Fig. 44) en todas las escalas (Anexo 13) para comprobar que la variable respuesta se distribuye como una Poisson, o en todo caso como una Binomial Negativa si la varianza crece en mayor proporción que la media.



**Figura 44. Relación entre la media y varianza estimada del número de ARS en el caso de sobredispersión, en este caso no se puede ajustar un modelo de Poisson. Escala 4.4 km - más menos dos días.**

Para cada escala se calcularon posibles modelos ajustados a la abundancia de ARS si está presente, entre los cuales se comparó su bondad de ajuste y significancia global del modelo hasta escoger el mejor “modelo” (ajustado) (Anexo 14).

La aplicación de los MLG finales ajustados al número de ARS si está presente en relación a variables físicas y biológicas permitió obtener los siguientes resultados (Tabla 14):

- En general se observa que la covariable  $\log_{10}(NASC+1)$  es significativa positivamente a partir de la escala 6.7km hasta las más pequeñas escalas, a mayor  $\log_{10}(NASC+1)$ , mayor número de ARS. Mientras más pequeña es la escala, menor es el efecto negativo de la latitud y el número de ARS se incrementa a medida que se incrementa la latitud. Respecto a la distancia a la isla Pescadores es la covariable que permanece significativa negativamente en la mayoría de las escalas, ello induce a pensar que el número de ARS se incrementa cuando la distancia a la isla es menor.
- En la escala 8.8km con tolerancia de más menos tres días, la profundidad del límite superior de la ZMO y distancia a la isla Pescadores tienen un efecto significativo positivo y negativo respectivamente, es decir cuando se incrementa la profundidad del límite superior de la ZMO el número de ARS también se incrementa, y si aumento la distancia a la isla el número de ARS disminuye.
- En las escalas 6.7km en general la distancia a la isla Pescadores también tienen un efecto significativo negativo. Además a partir de la escala de 4.4km la covariable  $\log_{10}(NASC+1)$  tiene un efecto significativo positivo, por lo cual cuando se incrementa  $\log_{10}(NASC+1)$  el número de ARS se incrementa.
- En las escalas (2.2km - más menos dos días) mientras que la latitud y distancia a la isla Pescadores siguen siendo significativas negativamente, la covariable  $\log_{10}(NASC+1)$  tiene un

efecto significativo positivo, por lo cual cuando se incrementa  $\log_{10}(NASC+1)$  el número de ARS se incrementa.

Los porcentajes de desviación explicada de los modelos encontrados son bastante buenos llegando a ser la mayor desviación explicada al 85% (Tabla 15). La bondad de ajuste de los modelos se aceptó en todas las escalas con la función de desvío y la Chi-cuadrado de Pearson (p-valor  $>0.05$ ). Asimismo todos los modelos ajustados resultaron ser significativos tanto con la prueba de Wald como de LRT (p-valor  $<0.05$ , Tabla 14). Para validar el GLM asumiendo una familia Poisson o Binomial Negativa se realizaron diversos gráficos de diagnóstico (Anexo 15). Los “mejores” modelos ajustados se muestran en la Tabla 16.

Tabla 14. Resultados de MLG (especificando distribuciones Poisson o Binomial Negativa y la función de enlace) utilizados para estudiar la relación entre número de ARS y las covariables acústicas (Logaritmo de la biomasa de peces epipelágicos ( $\log(\text{NASC}+1)$ )), profundidad del límite superior de la ZMO (prof.ZMO), latitud y distancia a la Isla Pescadores (Dist.Isla)). Para cada modelo el tamaño de muestra, coeficiente estimado y p-valor de las covariables son mostrados.

Distribución	Función de enlace	Escala Espacial	Escala Temporal	Tamaño de muestra	Log(NASC+1)		Prof.ZMO		Latitud		Dist.Isla	
					Estimado	p-valor	Estimado	p-valor	Estimado	p-valor	Estimado	p-valor
BN	log	8.9km	más menos tres días	29	-	-	0.124	2.11443E-05(***)	-	-	-0.133	1.28256E-21(***)
BN	log	6.7km	más menos tres días	42	-	-	-	-	-	-	-0.074	9.82745E-09(***)
BN	log	6.7km	más menos dos días	38	0.793	0.118	-	-	-2.191	0.143	-0.085	1.94994E-07(***)
BN	log	6.7km	más menos un día	35	1.022	0.060(.)	-	-	-3.676	0.027(*)	-0.081	6.29632E-06(***)
BN	log	4.4km	más menos tres días	65	1.086	0.000(***)	-	-	-3.309	0.0004 (***)	-0.092	5.55369E-21(***)
BN	log	4.4km	más menos dos días	57	1.047	0.008(**)	0.047	0.104	-3.069	0.009(**)	-0.103	3.8332E-18(***)
POISSON	log	4.4km	más menos un día	44	1.048	8.49877E-05(***)	-	-	-3.186	0.001(**)	-0.074	5.25713E-17(***)
POISSON	log	2.2km	más menos dos días	100	1.061	5.28253E-07(***)	-	-	-1.727	0.028(*)	-0.049	2.34112E-12(***)
POISSON	log	2.2km	más menos un día	80	0.845	0.0008(***)	-	-	-1.996	0.045(*)	-0.040	6.13654E-06(***)
POISSON	log	1.1km	más menos un día	107	0.437	0.0781(.)	0.004	0.8787	-1.254	0.202	-0.019	0.0854 (.)

**Códigos de Significancia de las covariables:**  $<0, 0.001 >$  ‘\*\*\*’;  $<0.001, 0.01 >$  ‘\*\*’;  $<0.01, 0.05 >$  ‘\*’;  $< 0.05, 0.1 >$  ‘.’

Tabla 15. Tabla 13. Resultados de Bondad de ajuste y significancia de MLG (especificando distribuciones Poisson o Binomial Negativa) utilizados para estudiar la relación entre número de ARS y las covariables acústicas (log (peces+1), profundidad del límite superior de la ZMO, latitud y distancia a la Isla Pescadores).

Distribución	Escala Espacial	Escala Temporal	Desviianza Nula	Desviianza Residual	Función de Desviianza (p-valor)	Prueba de Pearson (p-valor)	AIC	BIC	Desviianza Explicada%	Prueba de Wald (p-valor)	Prueba de LRT (p-valor)
BN	8.9km	más menos tres días	171.953	24.580	0.486	0.283	144.995	163.653	85.71%	2.67667E-22(***)	1.78E-11(***)
BN	6.7km	más menos tres días	86.087	39.684	0.484	0.413	224.044	240.470	53.90%	9.827e-09(***)	1.93E-08(***)
BN	6.7km	más menos dos días	77.392	35.642	0.391	0.527	207.326	233.702	53.95%	9.458e-07(***)	2.92E-06(***)
BN	6.7km	más menos un día	59.894	32.068	0.413	0.467	179.616	205.169	46.46%	9.318e-05(***)	0.0001(***)
BN	4.4km	más menos tres días	179.344	56.561	0.602	0.037	275.609	307.198	68.46%	1.55573E-19(***)	8.64E-14(***)
BN	4.4km	más menos dos días	156.025	46.928	0.673	0.518	246.913	283.429	69.92%	3.07174E-18(***)	3.59E-12(***)
POISSON	4.4km	más menos un día	133.866	42.512	0.283	0.231	167.055	188.956	68.24%	3.1741E-21(***)	1.12095E-19(***)
POISSON	2.2km	más menos dos días	180.588	102.232	0.288	0.061	355.054	383.815	43.39%	2.54989E-17(***)	6.91157E-17(***)
POISSON	2.2km	más menos un día	82.534	49.892	0.989	0.920	248.760	275.716	39.55%	1.82326E-07(***)	3.83345E-07(***)
POISSON	1.1km	más menos un día	48.109	39.648	1.000	1.000	289.423	326.151	17.59%	0.07384 (.)	0.07608 (.)

**Códigos de Significancia del modelo:** <0, 0.001 > ‘\*\*\*’; <0.001, 0.01 > ‘\*\*’; <0.01, 0.05 > ‘\*’; < 0.05,0.1 > ‘.’

Tabla 16. MLG ajustados al Número de ARS si está presente respecto de las covariables acústicas acústicas (Logaritmo de la biomasa de peces epipelágicos (Log (NASC+1)), profundidad del límite superior de la ZMO (Prof.ZMO), latitud (Lat) y distancia a la Isla Pescadores (Dist.Isla).

Escala Espacial	Escala Temporal	Función de enlace	Modelos estimados
8.9km	más menos tres días	Log	$\hat{y}(x) = \exp(3.1057 + 0.124 * \text{prof. ZMO} - 0.133 * \text{Dist. Isla})$
6.7km	más menos tres días	Log	$\hat{y}(x) = \exp(3.14 - 0.074 * \text{Dist. Isla})$
6.7km	más menos dos días	Log	$\hat{y}(x) = \exp(-24.358 + 0.793 * \text{Log(NASC + 1)} - 2.191 * \text{Lat} - 0.085 * \text{Dist. Isla})$
6.7km	más menos un día	Log	$\hat{y}(x) = \exp(-42.7268 + 1.022 * \text{Log(NASC + 1)} - 3.676 * \text{Lat} - 0.081 * \text{Dist. Isla})$
4.4km	más menos tres días	Log	$\hat{y}(x) = \exp(-38.472 + 1.086 * \text{Log(NASC + 1)} - 3.309 * \text{Lat} - 0.092 * \text{Dist. Isla})$
4.4km	más menos dos días	Log	$\hat{y}(x) = \exp(-35.9842 + 1.047 * \text{Log(NASC + 1)} + 0.047 * \text{prof. ZMO} - 3.069 * \text{Lat} - 0.103 * \text{Dist. Isla})$
4.4km	más menos un día	Log	$\hat{y}(x) = \exp(-37.413 + 1.048 * \text{Log(NASC + 1)} - 3.186 * \text{Lat} - 0.074 * \text{Dist. Isla})$
2.2km	más menos dos días	Log	$\hat{y}(x) = \exp(-21.2069 + 1.061 * \text{Log(NASC + 1)} - 1.727 * \text{Lat} - 0.049 * \text{Dist. Isla})$
2.2km	más menos un día	Log	$\hat{y}(x) = \exp(-24.1664 + 0.845 * \text{Log(NASC + 1)} - 1.996 * \text{Lat} - 0.04 * \text{Dist. Isla})$
1.1km	más menos un día	Log	$\hat{y}(x) = \exp(-15.1768 + 0.437 * \text{Log(NASC + 1)} + 0.004 * \text{Prof. ZMO} - 1.254 * \text{Lat} - 0.019 * \text{Dist. Isla})$



# **CAPÍTULO V**

## **DISCUSIÓN**

## **DE LA ECOLOGÍA**

La aplicación de dos familias de los métodos estadísticos (estadística espacial y modelos lineales generalizados) permitió un mejor entendimiento del comportamiento de forrajeo de las aves marinas en relación con sus presas y factores ambientales en el Sistema de la Corriente de Humboldt (SCH).

### ***Distribución de zonas de búsqueda restringida (ARS), peces epipelágicos y la profundidad del límite superior de la Zona de mínimo oxígeno (ZMO)***

La aplicación de métodos de procesos puntuales permitió enseñar que en todas las escalas estudiadas, incluso en las más pequeñas, las ARS tienen una clara tendencia a estar agregadas. Una primera hipótesis es que esto podría reflejar la existencia de cierto grado de autocorrelación entre las ARS ya que estas pueden provenir de un mismo viaje. También podría reflejar que el comportamiento de forrajeo de los predadores representa la distribución de sus presas así como disponibilidad de los mismos (Weimerskirch., 2007; Bertrand et al., 2005, 2008b; Fauchald., 2009) determinando la distribución de las ARS. Por otro lado la agregación de las ARS podría reflejar una característica conocida de los piqueros peruanos a buscar alimento en pequeños grupos a través de una red de forrajeo (Weimerskirch., et al., 2010). Sin embargo esta última hipótesis es la menos probable en nuestro caso debido a que la probabilidad que se hayan logrado equipar dos aves y que hayan salido en el mismo grupo de forrajeo fue muy baja.

La aplicación de la geoestadística permitió enseñar la existencia de estructuras espaciales de las variables continuas biomasa de peces epipelágicos y la profundidad del límite superior de la ZMO, modelar su autocorrelación espacial respectivamente para luego interpolar las distribuciones de dichas variables.

Las estructuras espaciales de la biomasa de peces epipelágicos tienen mayor variabilidad espacial que la profundidad del límite superior de la ZMO (Fig. 38a) debido al carácter gregario de peces como la anchoveta (Freón y Misund 1999; Bertrand et al., 2008a). Mientras que la profundidad del límite superior de la ZMO por ser una variable física oceanográfica reveló una leve variabilidad espacial con gradientes más suaves (Fig. 40a).

En el variograma de la distribución de los peces epipelágicos la mayor parte de la estructura se encontró en un rango <15 km lo cual demostró que la variabilidad de la biomasa de peces epipelágicos a pequeñas distancias es alta (Simmonds et al., 2009). En efecto esta estructura encierra a la vez los cardúmenes (metros a 10s m) y 'clusters' de cardúmenes (kms). También se observó una estructura con un rango de aproximadamente 30 km mostrando la existencia de agregaciones de mayor escala aún, es decir 'cluster de clusters' (10s km). Por otro lado en el variograma de la profundidad del límite superior de la ZMO la primera estructura resaltante (<30 km) se puede interpretar como la influencia de los procesos físicos de sub-mesoescala (100s m a km) como ondas internas y pequeños remolinos (Bertrand et al., 2008a).

Que se haya encontrado una diferencia en las estructuras de la distribución de la biomasa de peces epipelágicos y la profundidad del límite superior de la ZMO es ecológicamente lógico debido a la naturaleza de las variables. Sin embargo, desde un punto de vista estadístico es muy interesante, debido a que estas estructuras definen los "distintos modelos" utilizados para estimar dichas variables en un espacio determinado.

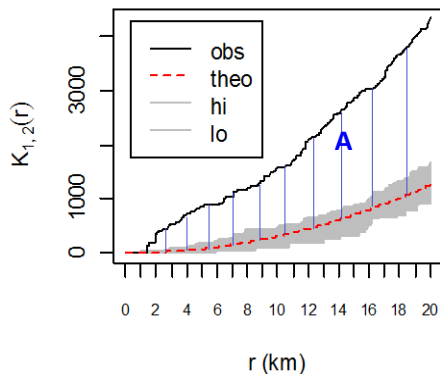
### ***Interacción de ARS con agregaciones de peces epipelágicos***

En la zona de estudio se encontraron importantes agregaciones de peces epipelágicos, como la anchoveta, (Fig. 34) disponibles en los alrededores de la Isla Pescadores. En general para escalas espaciales >1 km la agregación de las ARS de las aves con las agregaciones de peces sugiere que las aves pueden haberse alimentado eficientemente de agregaciones epipelágicas. En efecto los predadores tienden a agregarse en zonas donde la concentración de presas es mayor (Rose y Leggett, 1990). Además se sabe que el alimento primario de las aves marinas peruanas es casi exclusivamente la anchoveta (Jahncke., et al 2004; Zavalaga et al., 2010). Y teniendo en cuenta que el año 2008 se caracterizó por condiciones frías, las que están asociadas a una distribución más superficial de anchoveta (Bertrand et al., 2008a), entonces las agregaciones de anchoveta debieron ser accesibles, particularmente para los piqueros que sólo se alimentan de presas disponibles cerca de la superficie

(Weimerskirch et al., 2010). En efecto los piqueros son aves marinas adaptadas para volar eficientemente cuya estrategia de caza es conocida como “plunge-diving”, que consiste en lanzarse desde una altura para zambullirse rápidamente para atrapar su presa. Además esta estrategia no les permite bucear muy hondo (en promedio 2 m y máximo 6 m) en comparación con otras especies como los guanayes, que pueden bucear hasta profundidades entre los 10 y 50 m (Weimerskirch et al., 2010; Ludynia et al., 2010).

A escala muy fina (escala <1 km), las ARS y las agregaciones de peces se distribuyeron de forma independiente, es decir no se pudo detectar una correlación estrecha entre sus posiciones respectivas. Sin embargo no se puede afirmar que realmente no hubo interacción puesto que posiblemente existió un leve desfase temporal de horas (máx. 24h) entre los datos acústicos y de aves que nos limitan espacialmente para estudiar los procesos a una resolución <1 km. En efecto, se sabe que la anchoveta tiende a formar principalmente agregaciones, cuya distribución en el espacio y tiempo puede cambiar rápidamente (Gerlotto et al., 2006; Bertrand et al., 2008a). Por consiguiente puede ser que realmente sí haya existido una interacción de las ARS y las agregaciones de peces, sin embargo estamos por debajo de la resolución espacio-temporal de los datos para poder haberla detectado.

Además este desfase temporal también explicaría las diferencias significativas entre las funciones K de Ripley en las distintas escalas temporales estudiadas. Como resultado a medida que se incrementa la escala temporal disminuye el índice de asociación espacial (Fig. 35), que está determinado por la superficie limitada por la función doble K de Ripley empírica y teórica (Fig.45; Bertrand S., 2005; Bertrand S. et al., 2008b). Y aunque en el año 2008 se encontraron asociaciones espaciales positivas predador-presa, el ecosistema de surgencia de Humboldt se caracteriza por su variabilidad ambiental, lo cual influye en el comportamiento de muchas especies (Bertrand et al., 2008a; Espinoza y Bertrand., 2008; Swartzman et al., 2008). Por consiguiente sería interesante en el futuro estudiar cómo varía el índice de asociación espacial predador-presa de acuerdo a los distintos contextos ambientales y de presa entre años o temporadas.



**Figura 45. El Índice de asociación espacial está determinado por la superficie A limitada por la función K de Ripley empírica (obs) y teórica (theo).**

**Relación ARS – Abundancia de peces epipelágicos, profundidad del límite superior de la ZMO, distancia a la Isla Pescadores y latitud**

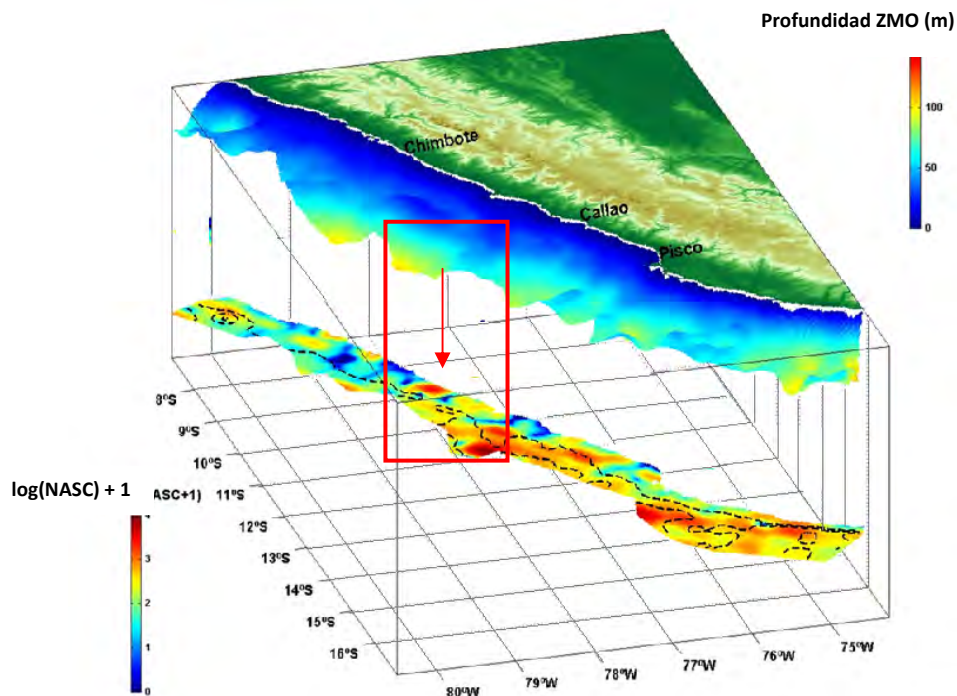
Los resultados de los modelos lineales generalizados mostraron que la biomasa de peces epipelágicos fue una covariable importante para determinar tanto la probabilidad de presencia de ARS como la abundancia de ARS en la zona de estudio. La relación entre la probabilidad de encontrar ARS y la biomasa de peces epipelágicos fue significativa y positiva a escalas espaciales comprendidas entre 1.1 y 2.2 km. Por otro lado la relación entre abundancia de ARS y biomasa de peces epipelágicos fue significativa positivamente hasta escalas espaciales de 6.7 km. La relación de la biomasa de peces epipelágicos con las ARS se explicó mejor con la abundancia de ARS que con la presencia de las ARS debido a la naturaleza de las variables respuesta respectivamente. Según los resultados cuando hay mayor biomasa de peces epipelágicos, hay presencia de un ‘cluster’ de cardúmenes y por consiguiente

hay mayor presencia y abundancia de ARS. Entonces, se puede concluir que las aves no tuvieron problema de accesibilidad de presa, el incremento en la abundancia de presa debió incrementar la habilidad de los predadores para rastrear el sistema y localizar exitosamente los peces epipelágicos (Fauchald et al., 2000; Weimerskirch, 2007; Fauchald, 2009).

La distancia a la Isla Pescadores es la única covariable que explica tanto la probabilidad de presencia de ARS como la abundancia de ARS en todas las escalas estudiadas. Cuando la distancia a la Isla Pescadores es menor hay mayor probabilidad de presencia de ARS y hay mayor número de ARS. Este resultado se debe probablemente al hecho de que, en presencia de presas accesibles cerca de la isla, las aves con crías forrajearon eficientemente cerca de la colonia, (en promedio entre 15 y 30 km, ver Anexo 10) ahorrando energía y por consiguiente no tuvieron la necesidad de buscar presas adicionales.

La relación con la latitud permitió enseñar que durante este experimento, la presencia y el número de ARS fueron mayores al sur oeste de la zona Topineme, y en el contexto de este año 2008 se debe posiblemente a que globalmente hubo una mayor concentración de anchoveta en esta zona.

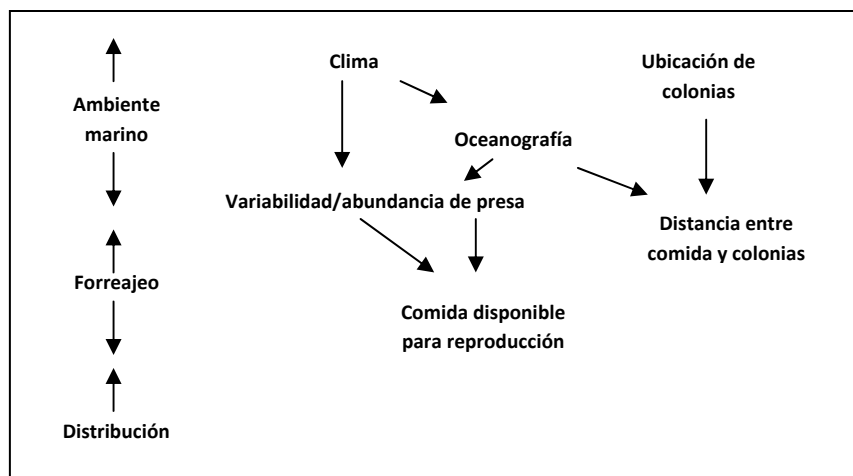
La relación entre probabilidad de presencia de ARS y la profundidad del límite superior de la ZMO fue significativa y positiva a grandes escalas. Las aves se alimentan principalmente de peces epipelágicos, y la abundancia de estos es mayor cuando, localmente, la profundidad del límite superior de la ZMO es mayor (Fig. 46, Bertrand et al., 2011). Lo cual se corroboró calculando correlaciones de Spearman entre la profundidad del límite superior de la ZMO y biomasa de peces epipelágicos (aprox.  $\rho = 0.5$ , p-valor= 0.0008). Según Fauchald (2009) las aves marinas muestran afinidad por áreas caracterizadas por propiedades físicas que mejoran o acumulan fuentes de alimento disponibles para la presa. La distribución espacial de estas fuentes de alimento puede concentrar la distribución de la presa y dar al predador cierta ventaja. En nuestro caso a sub-mesoescala procesos físicos 'convergentes' como ondas internas hacen bajar la profundidad de la oxiclina entonces se incrementa el rango vertical y las cualidades del hábitat de la anchoveta. En efecto estas ondas concentran nutrientes y plancton en parches por encima de la oxiclina, la presencia de estos parches durante el día es particularmente importante porque gran parte del zooplancton normalmente está distribuido por debajo de la oxiclina e inaccesible a la anchoveta (Bertrand et al., 2008a).



**Figura 46. Distribución espacial de la profundidad del límite superior de la ZMO y  $\log_{10}(\text{NASC} + 1)$ . A mayor profundidad del límite superior de la ZMO mayor abundancia de peces (Fuente: Bertrand et al., 2011)**

Para resumir las relaciones encontradas a lo largo de las escalas estudiadas: se observó que (i) efectivamente el comportamiento de forrajeo de las aves responde a la distribución de la anchoveta; (ii) las aves responden a la densidad de anchoveta y agregan su esfuerzo de forrajeo en los parches más cercanos y productivos (Fig. 48). (iii) Y globalmente la profundidad del límite superior de la ZMO no fue un factor limitante para que los piqueros se alimenten eficientemente.

En el 2008 las aves no sufrieron en su esfuerzo de forrajeo debido a las condiciones favorables de este contexto (biomasa alta y buena accesibilidad horizontal y vertical de la anchoveta). Sin embargo el SCH está expuesto a variabilidades ambientales y a la sobre-explotación humana (Gutiérrez et al., 2008). Razones suficientes para promover el desarrollo de estudios de esta índole en diversos contextos que brinden información suficiente para comprender mejor el ecosistema así como contribuir a la conservación de las aves marinas.



**Figura 48. Relación General entre variables oceanográficas, presa y aves marinas.**  
(Fuente: Diamond y Devlin., 2003)

## DE LAS METODOLOGÍAS ESTADÍSTICAS

Como lo hemos visto los métodos estudiados han permitido describir y entender mejor los procesos ecológicos. Sin embargo cada método tiene un campo de aplicación distinto y diversas ventajas y desventajas. La Tabla 17 muestra una comparación entre los métodos utilizados.

Tabla 17. Descripción de las metodologías utilizadas

Método estadístico	Función K de Ripley	Geoestadística	Modelos Lineales Generalizados (MLG)
<b>Objetivos</b>	Exploración de la distribución espacial de las ARS y la interacción espacial con agregaciones de peces.	Describir, modelar la estructura espacial de la biomasa de peces y profundidad del límite superior de la ZMO para realizar interpolaciones respectivamente.	Describir la relación espacial entre las ARS y covariables biológicas y físicas.
<b>Técnica</b>	Estadística espacial	Estadística espacial	Modelación paramétrica
<b>Basado en modelo</b>	Sí (Poisson)	Sí (Esférico, exponencial, etc.)	Sí (Binomial, Poisson, Binomial Negativa, etc.)
<b>Supuestos</b>	Estacionariedad	Estacionariedad	Independencia entre observaciones
<b>VARIABLES</b>	Continuas/ discretas	Continuas	Continuas/ discretas
<b>Permite pruebas de hipótesis</b>	Sí	No	Sí
<b>Información disponible a múltiples escalas</b>	Sí	Sí	No
<b>Autocorrelación espacial</b>	Sí	Sí	No
<b>Interacción entre dos variables</b>	Sí	Sí	Sí
<b>Interacción a mas de dos variables</b>	No	No	Sí
<b>Desventajas/limitaciones</b>	Extensiones con forma rectangular son más fáciles de calcular computacionalmente. El radio máximo debe ser 1/4 de la mayor distancia del área.	Cuando se tiene un gran número de datos esta técnica es computacionalmente lenta.	Necesidad de cruzar las diferentes bases de datos y usar escalas espacio-temporales para cada modelo.

La función K de Ripley y la geoestadística proveen valiosas herramientas estadísticas para el análisis espacial de datos. Ambas tienen en cuenta la autocorrelación espacial y permiten el estudio multi-escalas. Sin embargo la elección de una u otra prueba depende del tipo de datos con que se cuenta y del objetivo. Por un lado la geoestadística permitió identificar dependencias o estructuras espaciales de *variables continuas* mientras que la función K de Ripley permitió explorar la distribución espacial de *variables discretas*. Asimismo mediante la geoestadística se pudo modelar las estructuras espaciales para posteriormente realizar interpolaciones y estimar valores de la variable en las posiciones no muestreadas, mientras que con la función K de Ripley se pudo estudiar la interacción entre variables discretas.

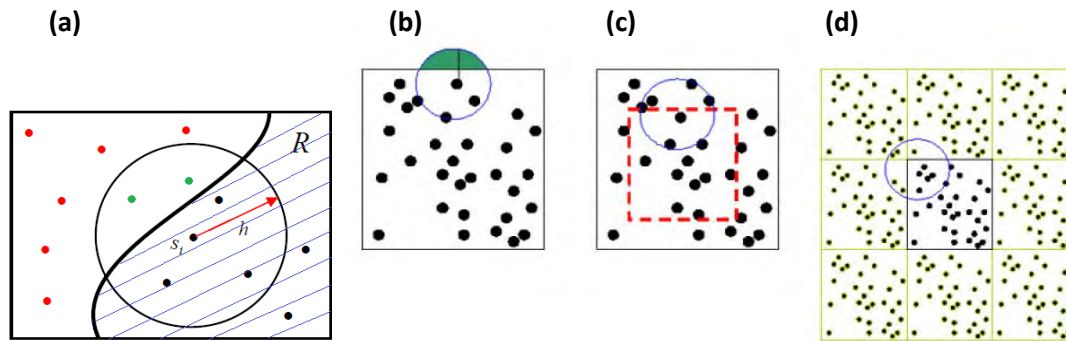
Cabe resaltar que también se tuvieron en cuenta otras metodologías estadísticas similares a la función K de Ripley como el *índice del vecino más próximo* o el *método de los cuadrantes*. Sin embargo no se usaron porque no permitían estudiar la distribución del proceso puntual en múltiples escalas en ambos casos (Marcoux et al., 2010). Y además el método de los cuadrantes que cuenta el número de eventos en subconjuntos (por lo general rectangulares) del área de estudio, es un método muy subjetivo pues el tamaño de los rectángulos se define de forma arbitraria. Mayor información sobre estas técnicas estadísticas es presentada en el Anexo 16.

Es importante mencionar que la estimación de la función K de Ripley puede estar “sesgada”, especialmente para valores grandes del radio (r), debido a efectos de borde (Ripley, 1981; Marcon y Puech, 2009; Perry et al., 2006; De la Cruz Rot., 2008). Los efectos de borde son aquellos puntos “no observados” que están ubicados cerca del límite del área de estudio. Entonces dado que el límite del área de estudio suele ser arbitrario el verdadero número de vecinos dentro de la distancia del radio (r) puede ser “*subestimado*” pues algunos de ellos pueden estar ubicados fuera del área de estudio (Fig. 49a; Goreaud y Pélissier, 1999). En nuestro caso debido a que usamos un área rectangular, se usó simplemente una corrección básica limitando los efectos de borde usando como radio máximo la cuarta parte de la longitud del menor lado del área (Baddley et al., 2010). Según los propósitos del estudio se puede “refinar” la estimación de la función K de Ripley mediante otras correcciones de los efectos de borde. Las correcciones del efecto de borde más conocidas son: (i) *el método de Ripley* que pondera los puntos próximos al borde según la porción del círculo que queda fuera del límite del área de estudio (Cressie., 1983; Fig.49b); (ii) *el método de tampón* (Guard area) que establece bandas dentro del área de estudio generando un área interna de menor tamaño que la original y un área tampón externa (buffer). Donde la función k de Ripley se calcula a partir de los puntos en el área interna pero al mismo tiempo se pueden contar los puntos que quedan en el área externa (Haase., 1995; De la Cruz Rot, 2008; Fig. 49c); y (iii) *El método de traslación toroidal* que replica el proceso puntual alrededor del área de estudio pues se asume que representa lo que no se ha observado, de tal manera que el área original está envuelto alrededor de un toroide (Ripley 1981; Haase., 1995; Fig. 49d).

También hay que tener en cuenta que cuando existen efectos de la intensidad local en el proceso puntual, es decir la intensidad varía a lo largo de la región de estudio, puede inducir a que cualquier dependencia espacial detectada pueda deberse más a efectos de la intensidad que a la interacción entre los eventos mismos rompiendo el supuesto de estacionariedad (Diggle et al., 2003). Muchos autores han desarrollado métodos para procesos no estacionarios (Ej. Marcon 2009; Moller y Waagepetersen, 2006; Gatrelletal et al., 1996), sin embargo agregando que el análisis se vuelve complicado no comprueban previamente la falta de estacionariedad. Muy pocos estudios han desarrollado métodos formales no subjetivos que prueben la estacionariedad en procesos puntuales espaciales y la literatura al respecto es bastante escasa. Guan (2009) recientemente ha desarrollado una versión de la prueba KPSS para evaluar la estacionariedad en procesos puntuales. En nuestro caso no fue realmente necesario evaluar previamente la estacionariedad debido a que no se utilizó la función k de Ripley para realizar inferencias de un modelo que caracterice el proceso puntual, por consiguiente la estacionariedad pudo ser asumida sin problemas (Fortín y Dale, 2005). Sin embargo sería interesante que en futuros trabajos se incluya este tipo de pruebas.

Con respecto a la geoestadística también se tuvieron en cuenta otras técnicas de interpolación como la *distancia inversa ponderada* ó *tendencia superficial*. Sin embargo no se usaron porque estas técnicas son útiles para realizar mapas de patrones espaciales (Fortín y Dale, 2005). A diferencia del kriging no brindan información precisa de la variable en lugares no muestreados ni una estimación de los errores muestreados. Además la tendencia superficial no debe usarse cuando hay muchas agregaciones pequeñas debido a que es un interpolador global y no local. También es interesante señalar que en vez de realizar variogramas y obtener directamente los parámetros para modelarlo se pudo declarar un modelo explícito a los datos, utilizando tendencias o covariables espaciales (Diggle y Ribeyro, 2007). Sin embargo en este caso los métodos de estimación usados se basan en el método de máxima verosimilitud y se hubiese tenido que cumplir con los supuestos de normalidad. En nuestro caso como se explicó previamente las variables utilizadas no se distribuyeron normalmente incluso haciendo uso de transformaciones de Box-Cox. Por otro lado también existen métodos geoestadísticos lineales

generalizados que tienen en cuenta distribuciones no gaussianas, sin embargo su aplicación es muy compleja y se ve limitada debido a su difícil implementación computacional.



**Figura 49. (a) Gráfica del efecto de borde representado por los círculos verdes que no forman parte de la región R. Correcciones del efecto de borde: (b) Método de Ripley (c) Método del tampón (d) Método de traslación toroidal (Fuente: De la Cruz Rot, 2008).**

Por otro lado los MLG han sido ampliamente aplicados en muchos campos de la investigación ecológica, como lo demuestra el creciente número de trabajos publicados (Ej. Laidre et al., 2010; Ainley et al., 2009; Burger et al., 2004). Los MLG a diferencia de la función  $k$  de Ripley o geoestadística no trabajan directamente a lo largo de escalas espaciales. Sin embargo permitieron trabajar con más de dos variables al mismo tiempo ya sean continuas, discretas o ambas.

En nuestro caso la variable respuesta tenía una alta proporción de ceros por consiguiente no se distribuía normalmente. Los MLG permitieron trabajar con otras distribuciones que representen mejor los datos ecológicos mediante dos etapas (Ainley et al., 2009), analizando la presencia-ausencia de ARS y la abundancia de ARS (solo presente). Sin embargo también existen otras metodologías similares a los MLG como el análisis de regresión lineal múltiple, los cuasi-MLG, los Modelos Aditivos Generalizados (GAM), los modelos Hurdle y modelos cero-inflado que no se usaron por diversas razones. El *análisis de regresión lineal múltiple* asume una distribución normal y los conteos de ARS no son normales, además no se pudo transformar la variable en todas las escalas estudiadas. Los *Modelos Aditivos Generalizados* son determinados por la naturaleza de la relación entre la variable respuesta y las variables predictoras y no por la asunción de alguna forma de relación paramétrica (Hastie y Tibshirani, 1990), es por ello que los GAM son “más flexibles”, permitiendo aditivamente una combinación de formas lineales y complejas en el mismo modelo. Sin embargo que sean más flexibles no es indicador que sean “mejores modelos”, se debe tener mucho cuidado de no sobre-ajustar los datos, es decir, aplicar un modelo excesivamente complejo con el fin de producir un buen ajuste. Los *Cuasi-Modelos Lineales Generalizados* se pueden usar en caso de sobredispersión tanto para distribuciones Binomial o Poisson, sin embargo el AIC no está definido, por lo tanto no se puede realizar un procedimiento de selección stepwise, backward (o forward). Los *Modelos Hurdle* y *modelos cero-inflado* son modelos más complejos y serían adecuados en caso la variable respuesta tuviera muchos ceros que causen la sobredispersión y no se logre resolver el problema con modelos más simples. Una breve descripción de estas técnicas estadísticas es presentada en el Anexo 17.

Para modelar la ausencia- presencia se utilizó la distribución Binomial porque la variable de respuesta estaba formada por ausencia (ceros) y presencia (unos). La función de enlace más común usada en los MLG Binomiales es la logística, por su facilidad interpretativa en función a probabilidades de presencia. Sin embargo existen otras funciones de enlace como las probit y cloglog también usadas en este trabajo y que pueden ajustarse mucho mejor a los datos, pero por falta de conocimiento e implementación computacional no se utilizan con frecuencia. La principal desventaja que se encontró fue la falta de gráficos de diagnóstico adecuados que tengan en cuenta que los residuos no son normales. Para lidiar con esto se usaron gráficos con envolturas de los cuantiles de los residuos de desviación.



En primera instancia se consideró trabajar con la distribución de Poisson, frecuentemente usada para modelar conteos. Sin embargo esta distribución asume que la media se incrementa proporcionalmente con la varianza, cuando en ciertas escalas la varianza se incrementó en mayor medida que la media, siendo esto evidencia de sobredispersión. Esta sobredispersión es causada posiblemente por diferencias entre los datos no representadas por el modelo. Y el mayor problema asociado con la sobredispersión es el riesgo de cometer el error de Tipo I. En estos casos, se debe especificar una distribución como la Binomial Negativa que incorpore en el modelo la varianza extra. En efecto White y Bennetts (2006) indican que los conteos de datos ecológicos exhiben una distribución Binomial Negativa más frecuentemente que una distribución Poisson o normal. La principal desventaja de la Binomial negativa es que hay muy pocos programas que la hayan implementado debido a que requiere de la estimación de otro parámetro adicional. Por otro lado no es recomendable usar la distribución Binomial Negativa cuando el tamaño de muestra es muy pequeño (Bourne et al., 2007).

Es importante tener en cuenta que modelar no es tan sencillo, la modelación en ciencias sigue siendo como un arte, debido a ello McCullagh (1979) propone algunos principios al "artista": (i) *"Todos los modelos son erróneos"*, aunque algunos son más útiles que otros y son estos los que debemos buscar; (ii) *"No debemos enamorarnos de un solo modelo"*, es importante reconocer que hay otros posibles modelos; y (iii) *"Verificar el ajuste del modelo a los datos"*, aunque estos procedimientos de diagnóstico no están completamente formalizados, algo de imaginación e introspección es requerida para determinar los aspectos del modelo que son más importantes.

# **CAPITULO VI CONCLUSIÓN**

En la introducción de esta tesis de tesis se planteó como objetivo principal aplicar metodologías estadísticas adecuadas para determinar en qué medida la presencia de aves marinas y su comportamiento espacial, reflejan las propiedades del ecosistema. La elección de las metodologías se realizó en base a las fortalezas y limitaciones de cada método tanto a nivel estadístico (eficacia frente a otros métodos de acuerdo al propósito, cumplimiento de supuestos estadísticos de los métodos, tipo de variables, datos autocorrelacionados, etc.) como de ecología marina (necesidad de análisis multi-escalas, sincronización espacio-temporal de los datos, resolución de los datos, etc.). Teniendo en cuenta esto se usaron las siguientes metodologías estadísticas: Función k de Ripley, geoestadística y modelos lineales generalizados.

Luego de aplicar las metodologías estadísticas propuestas se observó que la aplicación de la función K de Ripley demostró que las ARS tienen una clara tendencia a estar agregadas. La función doble K de Ripley enseñó que en general para escalas espaciales a partir de escalas >1 km las aves pudieron haberse alimentado eficientemente de agregaciones epipelágicas de peces. La geoestadística demostró que hay mayor variabilidad de la biomasa de peces epipelágicos a pequeñas distancias debido a la formación de cardúmenes (metros a 10s m) y 'clusters' de cardúmenes (kms). Los MLG demostraron que existe una relación positiva entre "presencia de ARS" y "biomasa de peces epipelágicos" en escalas <2.2 km. Asimismo también se demostró que existe una relación positiva entre la "abundancia de ARS" y "biomasa de peces epipelágicos" a escalas <6.7 km. Por lo tanto hay mayor presencia y abundancia de ARS cuando hay mayor concentración de peces epipelágicos. También los MLG enseñaron que existe mayor probabilidad que ocurra ARS cuando el ave se encuentra en una zona con mayor profundidad del límite superior de la ZMO en escalas >4.4 km, debido posiblemente a zonas de convergencia (Ej. Remolino anticiclónico u ondas internas) que concentran el zooplancton y atraen a los peces. Finalmente este método permitió enseñar que existe una relación negativa tanto entre la "presencia" y "abundancia" de ARS con la distancia a la Isla Pescadores en todas las escalas estudiadas, por lo cual se presume que las aves encontraron alimento disponible cerca de la colonia.

Las metodologías aplicadas en este trabajo respondieron eficientemente a nuestros objetivos. Sin embargo las técnicas de medición de los ecosistemas marinos van mejorando y a su vez los conocimientos en técnicas estadísticas se van renovando y actualizando. Por consiguiente las perspectivas futuras de estudios del complejo ecosistema marino incluyen estudiar la contribución de más factores biológicos y/o físicos, explorar las dinámicas en distintas escalas espacio-temporales en distintos contextos y sobre todo implica optar por adecuadas técnicas estadísticas que resulten en análisis más robustos que describan los datos y que cuantifiquen la relación entre predadores y variables medio ambientales eficientemente.

Por otro lado hay que resaltar que un aporte agregado a este trabajo implicó no estudiar la ubicación de aves si no el comportamiento de aves mediante las ARS. En efecto, según Tremblay (et al., 2009) estudiar la asociación de ubicaciones de particulares eventos de comportamiento (ARS) con parámetros ambientales es una ruta prometedora que puede mejorar nuestro entendimiento de los mecanismos fundamentales del comportamiento de aves.

Finalmente mientras más exploraciones del ecosistema marino se realicen ayudaremos en mayor medida a mejorar el manejo del ecosistema y así garantizar su durabilidad. Sin embargo no es tarea fácil entender el comportamiento de los organismos así como la complejidad de la dinámica poblacional del ecosistema. Y por otro lado tampoco se cuenta con aparatos "perfectos" que midan las muchas variables implicadas en estudios de esta índole. Para hacer frente a dichas limitaciones se requiere ineludiblemente un enfoque multidisciplinario que involucre la colaboración de biólogos, ecólogos, científicos ambientales, estadísticos y matemáticos aplicados, e incluso los tomadores de decisiones del estado. Y es en particular aquí donde la estadística juega un rol importante para hacer frente a la naturaleza probabilística de los procesos naturales y sus respuestas a las perturbaciones provocadas por el hombre.

# **BIBLIOGRAFÍA**

- Agresti, A. (1996). Analysis of ordinal categorical data. doi: 10.1002/bimj.4710290113.
- Ainley, D. G., Dugger, K. D., Ford, R. G., Pierce, S. D., Reese, D. C., Brodeur, R. D., et al. (2009). Association of predators and prey at frontal features in the California Current: competition, facilitation, and co-occurrence. *Marine Ecology Progress Series*, 389, 271-294. doi: 10.3354/meps08153.
- Baddeley, A. (2008). Analysing spatial point patterns in R. CSIRO and University of Western Australia.
- Baddeley, A., & Turner, R. (2010). Package "spatstat." Retrieved from <http://www.spatstat.org>.
- Bahri, T. (2000). Spatial structure of coastal pelagic schools descriptors in the Mediterranean Sea. *Fisheries Research*, 48(2), 157-166. doi: 10.1016/S0165-7836(00)00176-4.
- Bailleul, C. C., & Guinet, C. F. (2010). Mesoscale eddies as foraging area of a deep-diving predator, the southern elephant seal. *Marine Ecology Progress Series*, 408, 251-264.
- Bakun, A., & Weeks, S. J. (2008). The marine ecosystem off Peru: What are the secrets of its fishery productivity and what might its future hold? *Progress In Oceanography*, 79(2-4), 290-299. Elsevier Ltd. doi: 10.1016/j.pocean.2008.10.027.
- Ballon, M. (2010). Acoustic study of macrozooplankton off Peru: biomass estimation, spatial patterns, impact of physical forcing and effect on forage fish distribution.
- Bertrand, A., Ballón, M., & Chaigneau, A. (2010a). Acoustic observation of living organisms reveals the upper limit of the oxygen minimum zone. *PloS one*, 5(4), e10330. doi: 10.1371/journal.pone.0010330.
- Bertrand, A., Ballón, M., Chaigneau, A., Grados, D., Quiroz, Z., Ayón, P., et al. (2011). High resolution macrozooplankton biomass distribution in relation to the depth of the upper oxygen minimum zone off Peru.
- Bertrand, A., Gerlotto, F., Bertrand, S., Gutiérrez, M., Alza, L., Chipollini, A., et al. (2008a). Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. *Progress in Oceanography*, 79, 264-277.
- Bertrand, A., Segura, M., Gutierrez, M., & Vasquez, L. (2004). From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations off Peru. *Fish and Fisheries*, 5(4), 296-316. doi: 10.1111/j.1467-2679.2004.00165.x.
- Bertrand, S., Díaz, E., & Lengaigne, M. (2008b). Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Progress in Oceanography*, 79, 379-389.
- Bertrand, S., Weimerskirch, Henri, Tremblay, Y., Castillo, R., Silva, J., Passuni, G., et al. (2010b). Seabirds and fishers competing for the same prey off Peru.
- Bez, N. (2009). Basics on geostatistics for fish data. IMARPE.
- Bez, N., & Rivoirard, J. (2001). Transitive geostatistics to characterise spatial aggregations with diffuse limits: an application on mackerel ichthyoplankton. *Fisheries Research*, 50(1-2), 41-58. doi: 10.1016/S0165-7836(00)00241-1.
- Bourne, C. M., Regular, P. M., Sun, B., Thompson, S. P., Trant, A. J., & A, J. (2007). Generalized Linear Model Analysis in Ecology. Memorial University of Newfoundland and Labrador.
- Breslow, N. E. (1996). Generalized Linear Models : Checking Assumptions and Strengthening Conclusions. *Statistica Applicata*, 8, 23-41. Retrieved from <http://faculty.washington.edu/norm/pubs/glm.pdf>.
- Burger, A. E., Hitchcock, C., & Davoren, G. K. (2004). Spatial aggregations of seabirds and their prey on the continental shelf off SW Vancouver Island. *Marine Ecology Progress Series*, 283, 279-292.
- Cameron, C. A., & K., T. P. (1998). Data, Regression analysis of count. Society.
- Cayuela, L. (2009). Modelos lineales generalizados (GLM). Universidad de Granada. Retrieved from [http://observatoriosierranevada.iecolab.es/images/e/ec/3\\_Modelos\\_lineales\\_generalizados.pdf](http://observatoriosierranevada.iecolab.es/images/e/ec/3_Modelos_lineales_generalizados.pdf).
- Chavez, Francisco, Bertrand, A., Guevaracarrasco, 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(2-4), 95-105. doi: 10.1016/j.pocean.2008.10.012.
- Cordeiro, G. M., & Demetrio, C. G. B. (2008). Modelos Lineares Generalizados e Extensosoes.
- Cox, D. R. (1980). Point Processes.
- Cressie. (1993). Spatial Patterns. *Statistics for spatial data* (pp. 577-723).
- Cueto, V. (2006). Escalas en ecología: Su importancia para el estudio de la seleccion de habitat en aves. *Hornero*, 21(1), 1-13.
- Cury, P., Shannon, L., & Shin, Y.-J. (2001). The Functioning of marine ecosystems. [osec.rutgers.edu](http://osec.rutgers.edu). Retrieved February 17, 2011, from [http://osec.rutgers.edu/courses/expl\\_oceans/07Cury.PDF](http://osec.rutgers.edu/courses/expl_oceans/07Cury.PDF).
- Davoren, M. W., Anderson, J. T., & Gail, K. (2003). Distributional patterns of a marine bird and its prey: habitat selection based on prey and conspecific behaviour. *Marine Ecology Progress Series*, 256, 229-242.

- De La Cruz, M. (2008). Métodos para analizar datos puntuales. *Introducción al Análisis Espacial de Datos en Ecología y Ciencias Ambientales: Métodos y Aplicaciones* (pp. 75-127).
- Dessard, H., Picard, N., Pelissier, R., & Collinet-Vautier, F. (2004). Ecology and management of a neotropical rainforest : lessons drawn from Paracou, a long-term experimental research site in French Guiana. *Ecology and Management of a Neotropical Rainforest* (pp. 177-190).
- Diamond, a W., & Devlin, C. M. (2003). Seabirds as indicators of changes in marine ecosystems: ecological monitoring on Machias Seal Island. *Environmental monitoring and assessment*, 88(1-3), 153-75. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14570414>.
- Diggle, P., & Ribeiro, P. (2007). *Model-bases Geostatistics*.
- Diggle, P.J., Eglén, S. J., & Troy, J. B. (2006). Modelling the bivariate spatial distribution of amacrine cells. *Case Studies in Spatial Point Process Modeling* (p. 215–233). Springer. Retrieved February 17, 2011, from <http://www.springerlink.com/index/v8j0n73756p00mh8.pdf>.
- Dixon, P. M. (2002). Ripley's K Function. *Encyclopedia of Environmetrics*, 3, 1796-1803.
- Dobson, A. J. (2002). *An introduction to Generalized Linear Models*. New York.
- Duffy, D. C. (1983a). Environmental uncertainty and commercial fishing: Effects on Peruvian guano birds. *Biological Conservation*, 26(3), 227-238. doi: 10.1016/0006-3207(83)90075-7.
- Duffy, D. C. (1983b). Competition for nesting space among peruvian guano birds. *The Auk*, 100, 680-688.
- Dunn, P. K., & Smyth, G. K. (1996). Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 5, 236-244. Retrieved from <http://www.statsci.org/smyth/pubs/residual.html>.
- Díaz, V. M. A. (2006). *Procesos puntuales. Simulación estocástica Modelos y Algoritmos*.
- Emery, X. (2008). *Apunte de Geoestadística*. Chile: Universidad de Chile.
- Espinoza, Pepe, & Bertrand, A. (2008). Revisiting Peruvian anchovy (*Engraulis ringens*) trophodynamics provides a new vision of the Humboldt Current system. *Progress In Oceanography*, 79(2-4), 215-227. Elsevier Ltd. doi: 10.1016/j.pocean.2008.10.022.
- Espinoza, Pepe, Bertrand, A., Lingen, C. D. van der, Garrido, S., & Rojas de Mendiola, B. (2009). Diet of sardine (*Sardinops sagax*) in the northern Humboldt Current system and comparison with the diets of clupeoids in this and other eastern boundary upwelling systems. *Progress In Oceanography*, 83(1-4), 242-250. doi: 10.1016/j.pocean.2009.07.045.
- Faraway, J. J. (2006). *Extending the Linear Model in R Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Methods.
- Fauchald, P., & Tveraa, T. (2003). Using First-Passage Time in the analysis of area-restricted search and habitat selection. *Ecology*, 84, 282-288.
- Fauchald, Per. (2009). Spatial interaction between seabirds and prey: review and synthesis. *Marine Ecology Progress Series*, 391, 139-151.
- Fortin, M.-J., & Dale, M. (2005). *Spatial Analysis*.
- Friedlaender, A. S., Halpin, P. N., Qian, S. S., Lawson, G. L., Wiebe, Peter H, Thiele, D., et al. (2006). Whale distribution in relation to prey abundance and oceanographic processes in shelf waters of the Western Antarctic Peninsula. *Marine Ecology Progress Series*, 317, 297-310.
- Fréon, P., Bouchon, M, Mullon, C., Garcia, C., & Niquen, M. (2008). Interdecadal variability of anchoveta abundance and overcapacity of the fishery in Peru. *Progress In Oceanography*, 79(2-4), 401-412. Elsevier Ltd. doi: 10.1016/j.pocean.2008.10.011.
- Fréon, P., Bouchon, Marilu, Domalain, G., Estrella, C., Iriarte, F., Lazard, J., et al. (2010). Impacts of the Peruvian anchoveta supply chains : from wild fish in the water to protein on the plate. *Globec International Newsletter*, (April), 27-31.
- 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), 386-462.
- Gallardo, A., & Maestre, F. T. (2008). *Metodos geoestadísticos para el análisis de datos ecológicos espacialmente explícitos. Introducción al Análisis Espacial de Datos en Ecología y Ciencias Ambientales: Métodos y Aplicaciones* (pp. 215-272). Madrid.
- Gatrell, A. C., Bailey, T. C., Diggle, Peter J., & Rowlingson, B. S. (1996). Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Transactions of the Institute of British Geographers*, 21, 256-274. doi: 10.2307/622936.

Gerlotto, Francois, Bertrand, S., Bez, N., & Gutierrez, M. (2006). Waves of agitation inside anchovy schools observed with multibeam sonar: a way to transmit information in response to predation. *ICES Journal of Marine Science*, 63(8), 1405-1417. doi: 10.1016/j.icesjms.2006.04.023.

Gerlotto, Francois, & Paramo, J. (2003). The three-dimensional morphology and internal structure of clupeid schools as observed using vertical scanning multibeam sonar. *Aquatic Living Resources*, 16(3), 113-122. doi: 10.1016/S0990-7440(03)00027-5.

Gerlotto, François. (1993). Identification and spatial stratification of tropical fish concentrations using acoustic populations. *Aquatic Living Resources*, 6(3), 243-254. doi: 10.1051/alr:1993025.

Goreaud, F., & Pelissier, R. (1999). On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science*, 10, 433-438. doi: 10.2307/3237072.

Guan, Y. (2008). A KPSS test for stationarity for spatial point processes. *Biometrics*, 64(3), 800-6. doi: 10.1111/j.1541-0420.2007.00977.x.

Gutierrez, M., Ramirez, A., Bertrand, S., Moron, O., & Bertrand, A. (2008). Ecological niches and areas of overlap of the squat lobster "munida" (*Pleuroncodes monodon*) and anchoveta (*Engraulis ringens*) off Peru. *Progress In Oceanography*, 79(2-4), 256-263. Elsevier Ltd. doi: 10.1016/j.pocean.2008.10.019.

Gutiérrez, Mariano, Swartzman, G., Bertrand, A., & Bertrand, S. (2007). Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from 1983?2003. *Fisheries Oceanography*, 16(2), 155-168. doi: 10.1111/j.1365-2419.2006.00422.x.

Haase, P. (1995). Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. *Journal of Vegetation Science*, 6(4), 575-582. doi: 10.2307/3236356.

Hastie, T. J., Tibshirani R.J. (1990). *Generalized Additive Models*.

Hauray, L. R., McGowan, J. A., & Wiebe, P.H. (1978). Patterns and processes in the time-space scales distributions. *Spatial pattern in plankton communities*, 3, 277-327.

Hsu, H. P. (1996). *Random processes. Schaum's Outline of Probability, Random Variables, and Random Processes*.

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*.

Jackman, S. (n.d.). *Generalized Linear Models: Generalized Linear Models*. Stanford University. Retrieved from <http://jackman.stanford.edu/papers/glm.pdf>.

Jahncke, J., & Checkley, D. M. (2004). Trends in carbon flux to seabirds in the Peruvian upwelling system : effects of wind and fisheries on population regulation. *Fisheries Oceanography*, 13(3), 208-223.

Johnson, P. E. (2006). *Residuals and analysis of fit Sample R output*. Retrieved from [http://pj.freefaculty.org/stat/GLM/GLM2\\_v2.pdf](http://pj.freefaculty.org/stat/GLM/GLM2_v2.pdf).

Keitt, T. H., Bjørnstad, O. N., Dixon, P. M., & Citron-Pousty, Steve. (2002). Accounting for spatial pattern when modeling organism environment interactions. *Ecography*, 25, 616625.

Laidre, K. L., Heide-Jørgensen, M. P., Heagerty, P., Cossio, A., Bergström, B., & Simon, M. (2010). Spatial associations between large baleen whales and their prey in West Greenland. *Marine Ecology Progress Series*, 402, 269-284.

Lancaster, J., & Downes, B. J. (2004). Spatial point pattern analysis of available and exploited resources. *Ecography*, 27, 94-102.

Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73(6), 1943-1964.

LIKOFF, L. (2007). *The Encyclopedia of Birds*. Retrieved from [http://books.google.com.pe/books?id=ILEMaU74oXgC&pg=PA372&dq=Likoff+guanay&hl=es&ei=\\_UNdTe72EYTGIQe9ubjqCg&sa=X&oi=book\\_result&ct=result&resnum=1&ved=0CCQQ6AEwAA#v=onepage&q&f=false](http://books.google.com.pe/books?id=ILEMaU74oXgC&pg=PA372&dq=Likoff+guanay&hl=es&ei=_UNdTe72EYTGIQe9ubjqCg&sa=X&oi=book_result&ct=result&resnum=1&ved=0CCQQ6AEwAA#v=onepage&q&f=false).

Lindsey, J. K. (1997). *Applying Generalized Linear Models. Technometrics (Vol. 40)*. doi: 10.2307/1270654.

Ludynia, K., Garthe, S., & Luna-Jorquera, G. (2009). Distribution and foraging behaviour of the Peruvian Booby (*Sula variegata*) off northern Chile. *Journal of Ornithology*, 151(1), 103-111. doi: 10.1007/s10336-009-0431-1.

MacLennan, D. (2002). A consistent approach to definitions and symbols in fisheries acoustics. *ICES Journal of Marine Science*, 59(2), 365-369. doi: 10.1006/jmsc.2001.1158.

Marcon, E., & Puech, F. (2009). Generalizing Ripley's k function to inhomogeneous populations.

Marcoux, M., Larocque, G., Auger-Méthé, M., Dutilleul, P., & Murray, M. H. (2010). Statistical analysis of animal observations and associated marks distributed in time using Ripley's functions. *Animal Behaviour*, 80, 329-337.

Mc Culloch, C. E., & Searle, S. R. (2001). *Generalized linear and mixed models*.

McCullagh, P. (1983). *Generalized linear models*. Department of Statistics, University of Chicago.

- Mitson, R. B., Simard, Y., & Goss, C. (1996). Use of a two-frequency algorithm to determine size and abundance of plankton in three widely spaced locations. *ICES Journal of Marine Science*, 53, 209-215. Retrieved from <http://dx.doi.org/10.1006/jmsc.1996.0024>.
- Mobley, J. A. (2009). *Birds of the world*.
- Moller, J., & Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*.
- Moller, J., & Waagepetersen, R. P. (2006). Modern statistics for spatial point processes. In A. U. Department of Mathematical Sciences (Ed.), *21st Nordic Conference on Mathematical Statistics*.
- O'Driscoll, R. L. (1998). Description of spatial pattern in seabird distributions along line transects using neighbour K statistics. *Marine Ecology Progress Series*, 165, 81-94.
- Peng, R. D. (2002). Multi-dimensional Point Process Models in R. (D. of S. Papers, Ed.). University of California. Retrieved from <http://www.escholarship.org/uc/item/3n6609wb#page-1>.
- Perry, G. L. W., Miller, B. P., & Enright, N. J. (2006). A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecology*, 187, 59-82.
- Perry, J. N., Liebhold, a M., Rosenberg, M. S., Dungan, J., Miriti, M., Jakomulska, A., et al. (2002). Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography*, 25(5), 578-600. doi: 10.1034/j.1600-0587.2002.250507.x.
- Pinaud, D., & Weimerskirch, Henri. (2007). At-sea distribution and scale-dependent foraging behaviour of petrels and albatrosses: a comparative study. *The Journal of animal ecology*, 76(1), 9-19. doi: 10.1111/j.1365-2656.2006.01186.x.
- PROFONANPE. (n.d.). Preparacion del programa nacional de Áreas Protegidas (PRONANP)- Propuesta de Reserva Nacional del Sistema de Islas y Puntas Guaneras (pp. 1-44).
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Applied Probability*, 13, 255-266.
- Ripley, B. D. (1977). Modelling Spatial Patterns. *Journal of the Royal Statistical Society*, 39, 172-212. Retrieved from <http://www.jstor.org/stable/2984796>.
- Ripley, B. D. (1979). Simulating Spatial Patterns: Dependent Samples from a Multivariate Densit. *Journal of the Royal Statistical Society*, 28, 109-112. Retrieved from <http://www.jstor.org/stable/2346831>.
- Ripley, B. D. (1981). Mapped Point Patterns. In J. W. & Sons (Ed.), *Spatial Statistics*.
- Ripley, B. D., & Kelly, F. P. (1977). Markov Point Processes. *London Math. Soc.*, 15, 188-192.
- Rivoirard, J., Simmonds, J, Foote, K. G., Fernandes, P., & Bez, N. (2000). Geostatistics for Estimating Fish Abundance.
- 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.
- Russell, R., Hunt, G., Coyle, K., & Cooney, T. (1992). Foraging in a fractal environment Spatial patterns in a marine predator-prey system. *Landscape Ecology*, 7(3), 195-209.
- Ryther, J. H. (1969). Photosynthesis and Fish Production in the Sea. *Science*, 166(3901), 72-76. doi: 10.1126/science.166.3901.72.
- Santora, J. A., Reiss, C. S., Loeb, V. J., & Veit, R. R. (2010). Spatial association between hotspots of baleen whales and demographic patterns of Antarctic krill *Euphausia superba* suggests size-dependent predation. *Marine Ecology Progress Series*, 405, 255–269.
- Schoenberg, F. (2000). *Point Processes*. Los Angeles.
- Schoenberg, F., Brillinger, D. R., & Guttorp, P. (2002). Point processes, spatial-temporal. *Encyclopedia of Environmetrics*, 3, 1573–1577.
- Seavy, N. E., Quader, S., Alexander, J. D., & Ralph, C. J. (2005). Generalized linear models and point count data: Statistical considerations for the design and analysis of monitoring studies. *Bird Conservation Implementation and Integration in the Americas*. USDA For. Ser. Gen. Tech. Rep. PSW-GTR-191, 744–753.
- Siegel, S., & Castellan, J. N. (2003). *Estadística no Paramétrica*.
- Simmonds, E. J., Gutierrez, M., Chipollini, Andres, Gerlotto, Francois, Woillez, M., & Bertrand, A. (2009). Optimizing the design of acoustic surveys of Peruvian anchoveta. *ICES Journal of Marine Science*, 66, 1341-1348. doi: 10.1093/icesjms/fsp118.
- Simmonds, John, & MacLennan, D. (2005). *Fisheries Acoustics*.
- Stoyan, D. (2006). *Fundamentals of Point Process Statistics. Case Studies in Spatial Point Process Modeling (Vol. 185, pp. 3-22)*. Springer New York.



Swartzman, G., Bertrand, A., Gutierrez, M., Bertrand, S., & Vasquez, L. (2008). The relationship of anchovy and sardine to water masses in the Peruvian Humboldt Current System from 1983 to 2005. *Progress In Oceanography*, 79(2-4), 228-237. Elsevier Ltd. doi: 10.1016/j.pocean.2008.10.021.

Swartzman, G., & Hunt, G. (2000). Spatial association between murre (Uria spp.), puffins (Fratercula spp.) and fish shoals near Pribilof Islands, Alaska. *Marine Ecology Progress Series*, 206, 297-309.

The Mendeley Support Team. (2011a). *Getting Started with Mendeley*. Mendeley Desktop. London: Mendeley Ltd. Retrieved from <http://www.mendeley.com>.

The Mendeley Support Team. (2011b). *Getting Started with Mendeley*. Mendeley Desktop. London: Mendeley Ltd. Retrieved from <http://www.mendeley.com>.

Tomczak, M., & Godfrey, J. S. (2005). *The Pacific Ocean. Regional Oceanography: an Introduction* (pp. 105-174).

Tovar Serpa, Augusto. (2000) *Diccionario ecológico, forestal, ambiental, Recursos naturales y conservación*. Lima, editorial CONCYTEC.

Tremblay, Y., Bertrand, S., Henry, R. W., Kappes, M. A., Costa, D. P., & Shaffer, S. A. (2009). Analytical approaches to investigating seabird–environment interactions: a review. *Marine Ecology Progress Series*, 391, 153-163. doi: 10.3354/meps08146.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. World.

Vlietstra, L. S. (2005). Spatial associations between seabirds and prey: effects of large-scale prey abundance on smallscale seabird distribution. *Marine Ecology Progress Series*, 291, 275-287.

Waagepetersen, R. P. (n.d.). Estimating functions for inhomogeneous cox processes.

Warren, J. D., Santora, J. A., & Demer, D. A. (2009). Submesoscale distribution of Antarctic krill and its avian and pinniped predators before and after a near gale. *Marine Biology*, 156, 479-491.

Weimerskirch, Henry. (2007). Are seabirds foraging for unpredictable resources? *Deep-Sea Research II*, 54, 211-223.

Weimerskirch, Henry, Bertrand, S., Silva, J., Marques, J. C., & Goya, E. (2010). Use of Social Information in Seabirds: Compass Rafts Indicate the Heading of Food Patches. *PloS one*, 5(3), e9928.

White, G., & Bennetts, R. (1996). Analysis of frequency count data using the negative binomial distribution. *Ecology*, 77(8), 2549-2557.

Wiens, J. A. (1989). Spatial scaling in ecology. *Functional Ecology*, 3, 385-397.

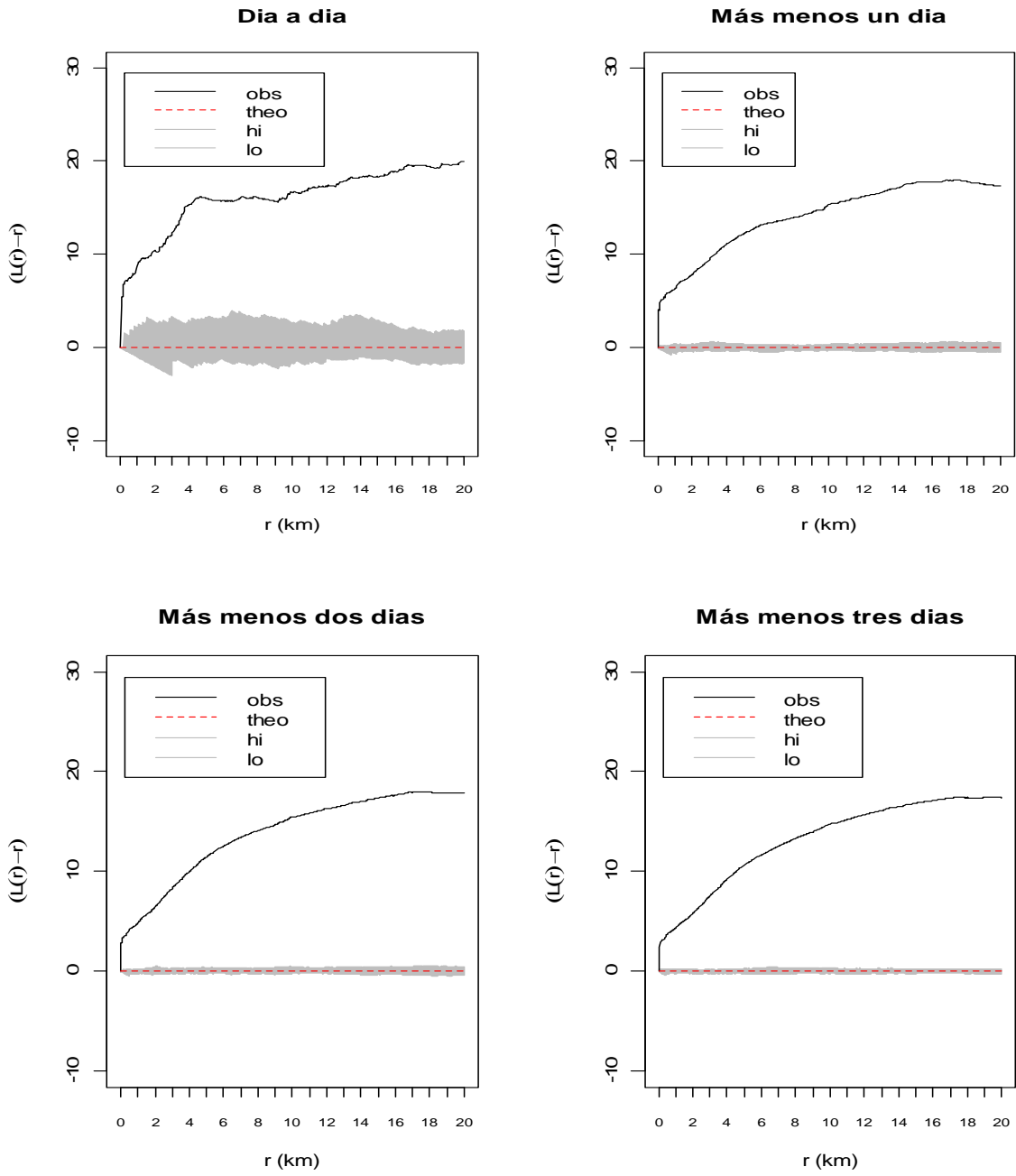
Zavalaga, C. B., Halls, J., Mori, G., Taylor, S. A., & Dell’Omo, G. (2010). At-sea movement patterns and diving behavior of Peruvian boobies *Sula variegata* in northern Peru. *Marine Ecology Progress Series*, 404, 259-274.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Journal Of Statistical Software*, 27(8), 1-25. Retrieved from <http://www.jstatsoft.org/v27/i08>.

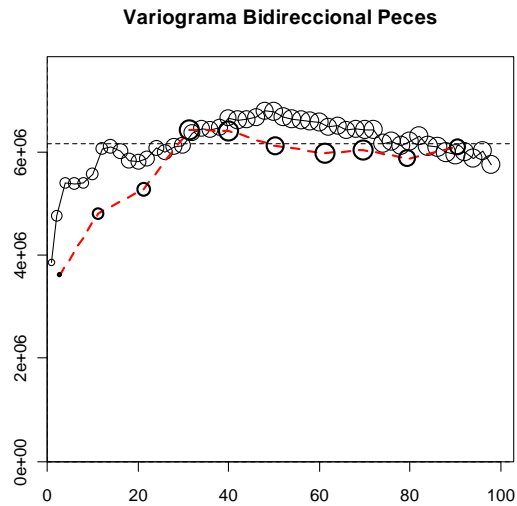
Zuur, A. F., Ieno, E. N., J.Walker, N., Saveliev, A. A., & Smith, G. M. (2008). *Mixed Effects Models and Extensions in Ecology with R*.

# **ANEXOS**

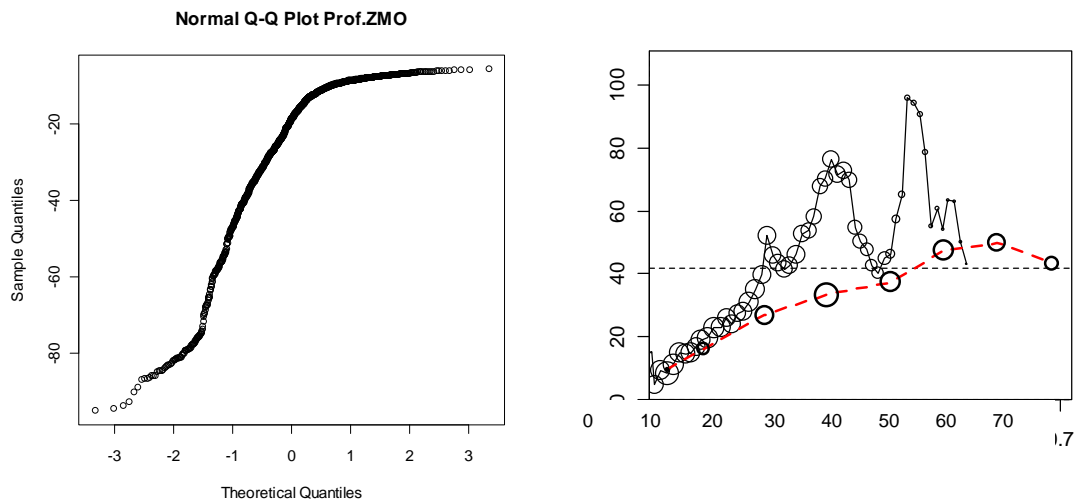
ANEXO 1: Función  $L(r) - r$  de ARS para cada escala temporal. Función teórica de  $L(r) - r = 0$  de Proceso Homogéneo de Poisson (theo). Envoltura de la distribución teórica  $L(r) - r$  ( $<hi$  y  $>lo$ ). Radio en Km.



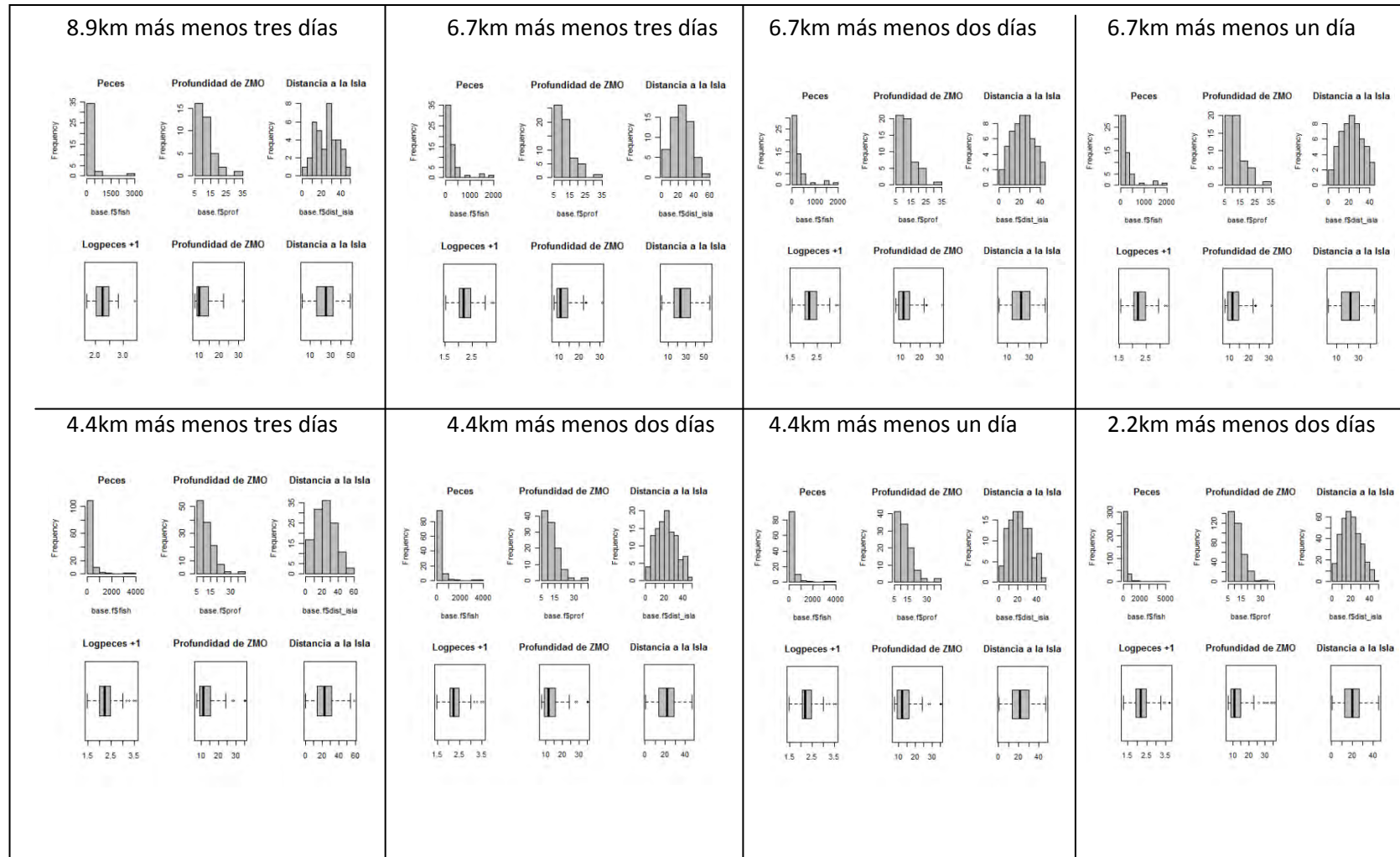
ANEXO 2: Variograma bidireccional de la abundancia de peces. Los círculos indican el número de puntos para cada punto en los variogramas. La escala horizontal es km.



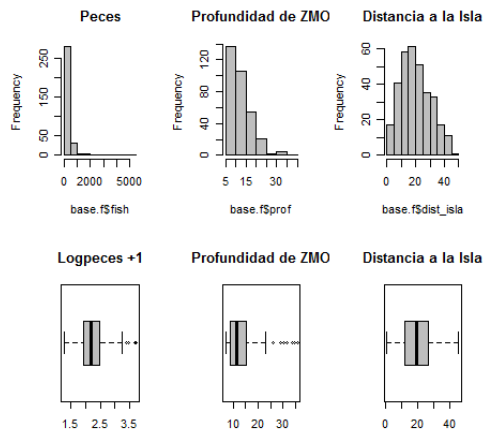
ANEXO 3: Grafico exploratorio de la normalidad y variograma bidireccional de la profundidad del límite superior de la ZMO. Los círculos indican el número de puntos para cada punto en los variogramas. La escala horizontal es km.



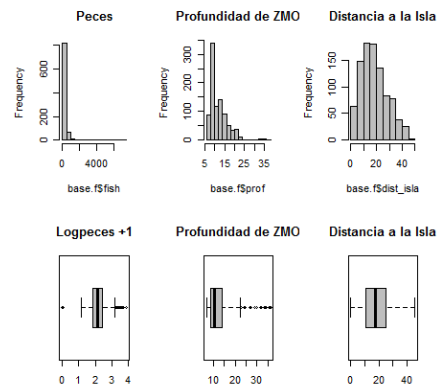
ANEXO 4: Histogramas de frecuencia y diagramas de caja de covariables de MLG ausencia-presencia de ARS.



## 2.2km más menos un día



## 1.1km más menos un día



ANEXO 5: Resultados de coeficientes de correlación de Spearman de las covariables y prueba de Anderson-Darling (A\_D) de la variable ARS del MLG ausencia-presencia de ARS.

ESCALA ESPACIAL	ESCALA TEMPORAL	Covariables	Log10(NASC+1)	Prof.ZMO	Latitud	Dist.Isla	A-D
0.08	Más menos tres días	Log10(NASC+1)	1	0.62	0.34	0.07	2.88E-14
		Prof.ZMO	0.62	1	-0.27	0.12	4.55E-05
		Latitud	0.34	-0.27	1	0.04	0.0545
		Dist.Isla	0.07	0.12	0.04	1	0.9257
0.06	Más menos tres días	Log10(NASC+1)	1	0.58	0.34	0.12	2.20E-16
		Prof.ZMO	0.58	1	-0.32	0.22	5.09E-07
		Latitud	0.34	-0.32	1	-0.01	0.03364
		Dist.Isla	0.12	0.22	-0.01	1	0.6369
0.06	Más menos dos días	Log10(NASC+1)	1	0.54	0.46	0.25	2.20E-16
		Prof.ZMO	0.54	1	-0.26	0.47	1.96E-05
		Latitud	0.46	-0.26	1	-0.22	0.06215
		Dist.Isla	0.25	0.47	-0.22	1	0.863
0.06	Más menos un día	Log10(NASC+1)	1	0.53	0.48	0.26	2.20E-16
		Prof.ZMO	0.53	1	-0.26	0.48	4.41E-05
		Latitud	0.48	-0.26	1	-0.25	0.06179
		Dist.Isla	0.26	0.48	-0.25	1	0.7662
0.04	Más menos tres días	Log10(NASC+1)	1	0.59	0.4	0.15	2.20E-16
		Prof.ZMO	0.59	1	-0.21	0.29	9.48E-13
		Latitud	0.4	-0.21	1	-0.05	0.003058
		Dist.Isla	0.15	0.29	-0.05	1	0.5977
0.04	Más menos dos días	Log10(NASC+1)	1	0.54	0.49	0.23	2.20E-16
		Prof.ZMO	0.54	1	-0.19	0.5	5.69E-10
		Latitud	0.49	-0.19	1	-0.25	0.006779
		Dist.Isla	0.23	0.5	-0.25	1	0.734
0.04	Más menos un día	Log10(NASC+1)	1	0.54	0.45	0.24	2.20E-16
		Prof.ZMO	0.54	1	-0.23	0.53	1.17E-09
		Latitud	0.45	-0.23	1	-0.3	0.01371
		Dist.Isla	0.24	0.53	-0.3	1	0.4362
0.02	Más menos dos días	Log10(NASC+1)	1	0.53	0.54	0.22	2.20E-16

		Prof.ZMO	0.53	1	-0.15	0.48	2.20E-16
		Latitud	0.54	-0.15	1	-0.14	1.73E-06
		Dist.Isla	0.22	0.48	-0.14	1	0.009805

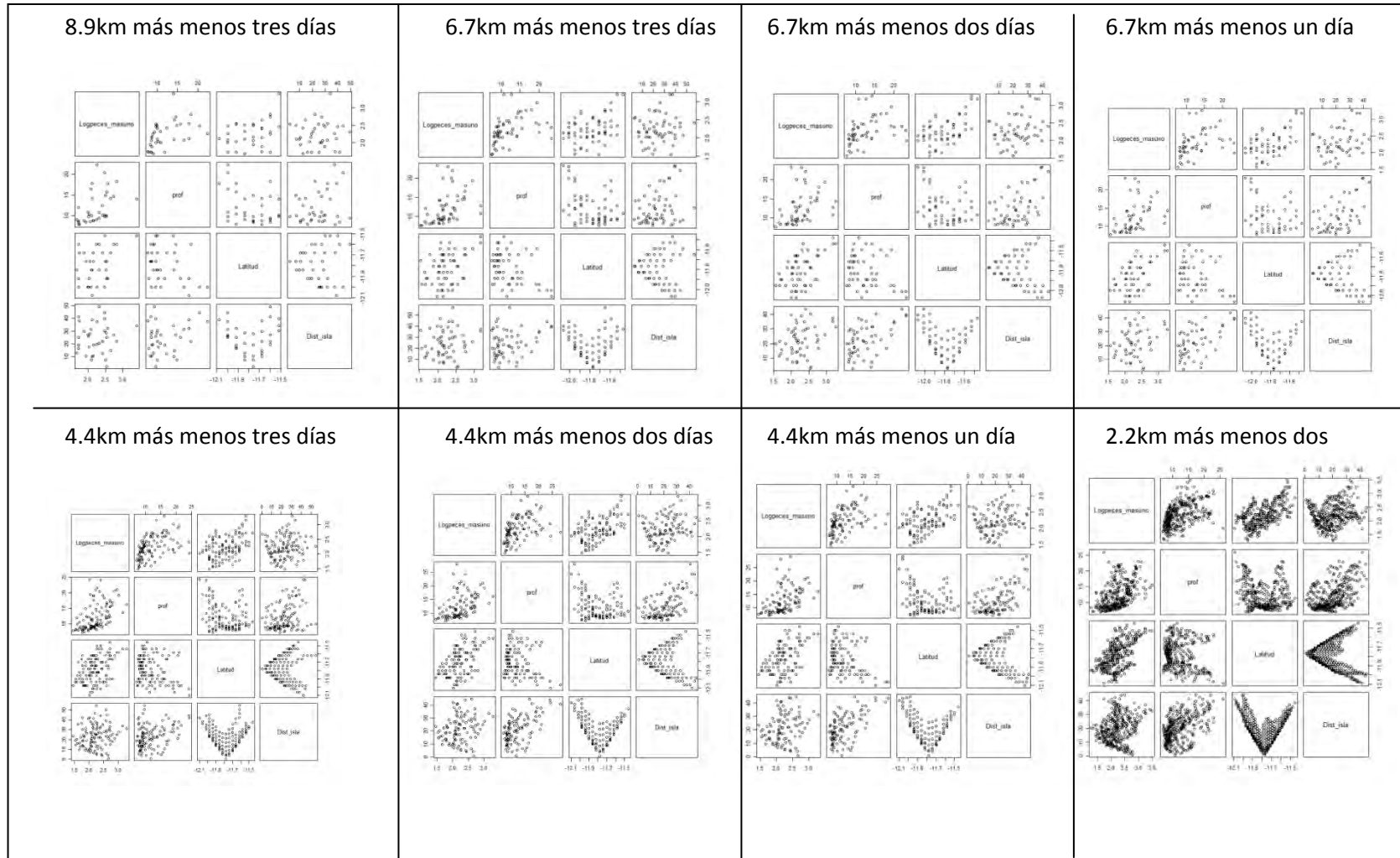
0.02	Más menos un día	Log10(NASC+1)	1	0.54	0.51	0.21	2.20E-16
		Prof.ZMO	0.54	1	-0.17	0.5	2.20E-16
		Latitud	0.51	-0.17	1	-0.23	2.29E-05
		Dist.Isla	0.21	0.5	-0.23	1	0.0007523

0.01	Más menos un día	Log10(NASC+1)	1	0.5	0.5	0.12	2.20E-16
		Prof.ZMO	0.5	1	-0.18	0.46	2.20E-16
		Latitud	0.5	-0.18	1	-0.25	2.42E-08
		Dist.Isla	0.12	0.46	-0.25	1	6.46E-13

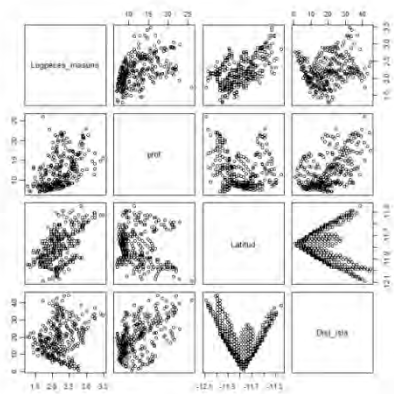
0.01	Día-día	Log10(NASC+1)	1	0.42	0.35	0.05	2.20E-16
		Prof.ZMO	0.42	1	-0.44	0.65	2.20E-16
		Latitud	0.35	-0.44	1	-0.64	4.94E-15
		Dist.Isla	0.05	0.65	-0.64	1	1.21E-10



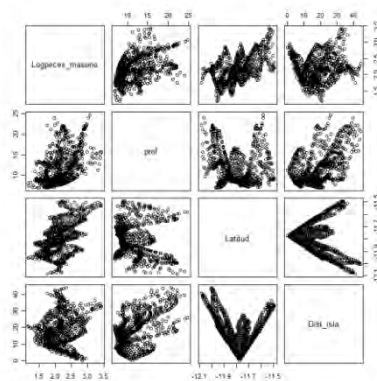
ANEXO 6: Graficas de correlación de covariables usadas en el MLG ausencia-presencia de ARS.



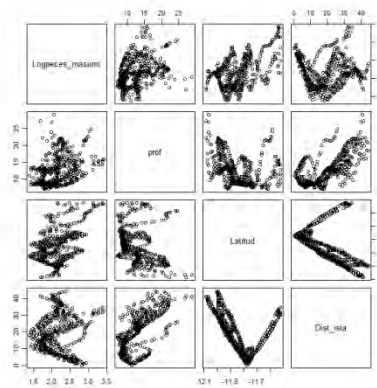
### 2.2km más menos un día



### 1.1km más menos un día



### 1.1km día-día



ANEXO 7: Resultados GLM ausencia-presencia de ARS

Escala Espacial	Escala Temporal	Modelo	Tamaño de muestra	Log(peces+1)		Prof.ZMO		Latitud		Dist.Isla	
				Estimado	p-valor	Estimado	p-valor	Estimado	p-valor	Estimado	p-valor
8.9km	Más menos tres días	Logit	37	-1.336	0.538	0.519	0.139	-0.518	0.911	-0.150	0.037
		Logit Modelo stepAIC	37			0.412	0.034			-0.157	0.019
		Probit	37	-0.518	0.664	0.242	0.137	-1.173	0.647	-0.074	0.027
		Probit Modelo stepAIC	37			0.218	0.021			-0.078	0.012
		Cloglog	37	-0.468	0.693	0.191	0.135	-1.582	0.500	-0.059	0.038
		Cloglog Modelo stepAIC	37			0.177	0.025			-0.057	0.036
		Logit sin outliers	36	-4.454	0.178	0.664	0.071	0.535	0.914	-0.193	0.033
		Logit l sin outliers Modelo stepAIC	36	-4.244	0.108	0.641	0.030			-0.189	0.020
		Probit sin outliers	36	-4.488	0.208	1.302	0.156	6.193	0.203	-0.293	0.059
6.7km	Más menos tres días	Logit	60	-1.314	0.356	0.384	0.034	0.586	0.849	-0.116	0.002
		Logit Modelo stepAIC	60			0.301	0.006			-0.120	0.002
		Probit	60	-0.777	0.333	0.224	0.021	0.355	0.843	-0.064	0.002
		Probit Modelo stepAIC	60			0.176	0.003			-0.066	0.001
		Cloglog	60	-0.880	0.301	0.222	0.010	0.328	0.862	-0.074	0.001
		Cloglog Modelo stepAIC	60			0.179	0.002			-0.076	0.001
6.7km	Más menos dos días	Logit	53	-0.544	0.726	0.280	0.033	-1.394	0.683	-0.154	0.004
		Logit Modelo stepAIC	53			0.288	0.007			-0.160	0.002
		Probit Modelo stepAIC	53	-0.294	0.741	0.162	0.026	-0.799	0.690	-0.091	0.002
		Probit Modelo stepAIC	53			0.169	0.005			-0.094	0.001
		Cloglog	53	-0.591	0.518	0.167	0.018	-0.507	0.812	-0.091	0.003
		Cloglog Modelo stepAIC	53			0.165	0.006			-0.094	0.002
6.7km	Más menos un día	Logit	54	-1.564	0.299	0.172	0.127	-1.988	0.536	-0.091	0.034
		Logit Modelo stepAIC	54	-2.216	0.042	0.207	0.035			-0.085	0.038

		Probit	54	-0.890	0.303	0.098	0.125	-1.127	0.560	-0.053	0.031
		Probit Modelo stepAIC	54	-1.266	0.035	0.119	0.031			-0.050	0.035
		Cloglog	54	-1.329	0.150	0.116	0.073	-1.217	0.570	-0.061	0.017
		Cloglog Modelo stepAIC	54	-1.715	0.011	0.137	0.016			-0.058	0.019

4.4km	Más menos tres días	Logit	123	0.099	0.887	0.056	0.303	-0.741	0.673	-0.061	0.002
		Logit Modelo stepAIC	123			0.068	0.087			-0.062	0.002
		Probit	123	0.042	0.923	0.035	0.305	-0.416	0.700	-0.036	0.002
		Probit Modelo stepAIC	123			0.041	0.091			-0.037	0.001
		Cloglog	123	0.047	0.924	0.039	0.303	-0.509	0.689	-0.042	0.002
		Cloglog Modelo stepAIC	123			0.045	0.104			-0.043	0.001

4.4km	Más menos dos días	Logit	109	0.164	0.834	-0.002	0.977	-2.036	0.299	-0.040	0.105
		Logit Modelo stepAIC	109					-1.766	0.162	-0.037	0.054
		Probit	109	0.117	0.809	-0.002	0.965	-1.299	0.285	-0.025	0.098
		Probit Modelo stepAIC	109					-1.097	0.158	-0.023	0.051
		Cloglog	109	0.060	0.914	-0.002	0.957	-1.345	0.343	-0.027	0.102
		Cloglog Modelo stepAIC	109							-0.023	0.081

4.4km	Más menos un día	Logit	105	-0.188	0.822	-0.003	0.958	-3.808	0.082	-0.031	0.226
		Logit Modelo stepAIC	105					-4.088	0.008	-0.035	0.088
		Probit	105	-0.061	0.904	-0.005	0.891	-2.362	0.072	-0.018	0.246
		Probit Modelo stepAIC	105					-2.429	0.006	-0.021	0.097
		Cloglog	105	-0.150	0.811	0.002	0.957	-3.087	0.077	-0.027	0.173
		Cloglog Modelo stepAIC	105					-3.358	0.009	-0.029	0.067

2.2km	Más menos dos días	Logit	347	0.726	0.148	0.001	0.988	-4.352	0.002	-0.074	0.000
		Logit Modelo stepAIC	347	0.731	0.075			-4.362	0.001	-0.074	0.000
		Probit	347	0.450	0.128	0.002	0.945	-2.563	0.002	-0.044	0.000
		Probit Modelo stepAIC	347	0.462	0.057			-2.590	0.000	-0.043	0.000

		Cloglog	347	0.670	0.098	-0.010	0.777	-3.847	0.001	-0.063	0.000
		Cloglog Modelo stepAIC	347	0.600	0.067			-3.707	0.001	-0.064	0.000

2.2km	Más menos un día	Logit	320	0.817	0.125	-0.004	0.931	-6.600	0.000	-0.072	0.000
		Logit Modelo stepAIC	320	0.792	0.077			-6.550	0.000	-0.073	0.000
		Probit	320	0.497	0.114	-0.002	0.947	-3.708	0.000	-0.041	0.000
		Probit Modelo stepAIC	320	0.486	0.065			-3.683	0.000	-0.041	0.000
		Cloglog	320	0.761	0.084	-0.010	0.781	-6.011	0.000	-0.064	0.000
		Cloglog Modelo stepAIC	320	0.692	0.059			-5.874	0.000	-0.065	0.000

1.1km	Más menos un día	Logit	910	1.230	0.003	-0.055	0.176	-6.155	0.000	-0.068	0.000
		Logit Modelo stepAIC	910	0.912	0.006			-5.552	0.000	-0.078	0.000
		Probit	910	0.546	0.010	-0.015	0.441	-2.954	0.000	-0.036	0.000
		Probit Modelo stepAIC	910	0.470	0.008			-2.791	0.000	-0.039	0.000
		Cloglog	910	1.216	0.001	-0.062	0.113	-5.925	0.000	-0.064	0.000
		Cloglog Modelo stepAIC	910	1.629	0.000	-0.084	0.048	-6.905	0.000	-0.064	0.000
		Logit sin outliers	909								
		Logit sin outliers Modelo stepAIC	909								
		Cloglog sin outliers	909	1.558	0.000	-0.087	0.031	-6.562	0.000	-0.060	0.000

1.1km	Día-día	Logit	501	1.085	0.087	-0.001	0.986	-11.650	0.012	-0.071	0.071
		Logit Modelo stepAIC	501	1.082	0.070			-11.651	0.012	-0.071	0.044
		Probit	501	0.440	0.147	-0.002	0.939	-4.553	0.012	-0.026	0.111
		Probit Modelo stepAIC	501	0.010	0.125			-4.545	0.013	-0.027	0.064
		Cloglog	501	1.071	0.078	0.000	0.993	-11.504	0.011	-0.070	0.065
		Cloglog Modelo stepAIC	501	1.069	0.062			-11.505	0.011	-0.071	0.040
		Logit sin outliers	500	1.492	0.027	0.010	0.837	-19.294	0.003	-0.127	0.015
		Logit sin outliers Modelo stepAIC	500	1.533	0.017			-19.162	0.003	-0.122	0.010
		Probit sin outliers	500	0.655	0.046	0.004	0.876	-8.563	0.004	-0.056	0.021
		Probit sin outliers Modelo	500	0.670	0.031			-8.544	0.004	-0.054	0.014

stepAIC										
	Cloglog sin outliers	500	1.448	0.024	0.010	0.824	-18.686	0.003	-0.123	0.014
	Cloglog sin outliers Modelo stepAIC	500	1.490	0.015			-18.525	0.002	-0.118	0.009

Escala Espacial	Escala Temporal	Modelo	DESVIANZA NULA	DEVIANZA RESIDUAL	Funcion Desviianza (p-valor)	Prueba de Pearson (p-valor)	AIC	BIC	DESVIANZA EXPLICADA %	Prueba de Wald (p-valor)	Prueba LRT (p-valor)
8.9km	Más menos tres días	Logit	38.634	22.495	0.894	0.004	32.495	58.604	41.773	0.104	0.003
		<b>Logit Modelo stepAIC</b>	<b>38.634</b>	<b>23.457</b>	<b>0.913</b>	<b>0.000</b>	<b>29.457</b>	<b>45.123</b>	<b>39.283</b>	<b>0.024</b>	<b>0.001</b>
		Probit	38.634	23.036	0.877	0.086	33.036	59.145	40.373	0.043	0.004
		Probit Modelo stepAIC	38.634	24.309	0.890	0.002	30.309	45.975	37.078	0.012	0.001
		Cloglog	38.634	24.221	0.836	0.367	34.221	60.330	37.306	0.087	0.006
		Cloglog Modelo stepAIC	38.634	26.112	0.831	0.027	32.112	47.778	32.411	0.040	0.002
		Logit sin outliers	38.139	19.568	0.945	0.002	29.568	55.404	48.692	0.114	0.001
		Logit I sin outliers Modelo stepAIC	38.139	19.580	0.958	0.008	27.580	48.248	48.661	0.056	0.000
		Probit sin outliers	35.467	9.471	1.000	1.000	19.471	45.306	73.297	0.380	0.000

6.7km	Más menos tres días	Logit	73.304	51.336	0.615	0.322	61.336	92.279	29.968	0.012	0.000
		Logit Modelo stepAIC	73.304	52.622	0.640	0.337	58.622	77.188	28.213	0.002	0.000
		Probit	73.304	51.424	0.612	0.410	61.424	92.367	29.848	0.005	0.000
		Probit Modelo stepAIC	73.304	52.731	0.636	0.434	58.731	77.297	28.065	0.001	0.000
		Cloglog	73.304	50.272	0.656	0.393	60.272	91.216	31.419	0.007	0.000
		<b>Cloglog Modelo stepAIC</b>	<b>73.304</b>	<b>51.821</b>	<b>0.669</b>	<b>0.458</b>	<b>57.821</b>	<b>76.387</b>	<b>29.306</b>	<b>0.002</b>	<b>0.000</b>

6.7km	Más menos dos días	Logit	65.631	49.608	0.449	0.385	59.608	89.498	24.414	0.028	0.003
		Logit Modelo stepAIC	65.631	50.732	0.484	0.521	56.732	74.666	22.701	0.006	0.001
		Probit Modelo stepAIC	65.631	49.562	0.451	0.402	59.562	89.452	24.483	0.016	0.003
		Probit Modelo stepAIC	65.631	50.625	0.488	0.529	56.625	74.559	22.864	0.003	0.001
		Cloglog	65.631	49.206	0.465	0.404	59.206	89.096	25.026	0.028	0.002

			Cloglog Modelo stepAIC	65.631	50.533	0.492	0.540	56.533	74.467	23.004	0.007	0.001
--	--	--	------------------------	--------	--------	-------	-------	--------	--------	--------	-------	-------

6.7km	Más menos un día	Logit		67.923	55.992	0.200	0.252	65.992	95.695	17.566	0.083	0.018
		Logit Modelo stepAIC		67.923	56.381	0.218	0.311	64.381	88.143	16.993	0.047	0.009
		Probit		67.923	56.049	0.199	0.284	66.049	95.752	17.482	0.057	0.018
		Probit Modelo stepAIC		67.923	56.389	0.218	0.340	64.389	88.151	16.981	0.030	0.009
		Cloglog		67.923	54.479	0.242	0.313	64.479	94.182	19.793	0.037	0.009
		Cloglog Modelo stepAIC		67.923	54.817	0.264	0.377	62.817	86.579	19.296	0.018	0.004

4.4km	Más menos tres días	Logit		170.116	158.334	0.008	0.302	168.334	206.456	6.925	0.043	0.019
		Logit Modelo stepAIC		170.116	158.555	0.011	0.357	164.555	187.428	6.796	0.008	0.003
		Probit		170.116	158.504	0.008	0.301	168.504	206.626	6.826	0.035	0.020
		Probit Modelo stepAIC		170.116	158.709	0.010	0.358	164.709	187.582	6.706	0.006	0.003
		Cloglog		170.116	158.392	0.008	0.342	168.392	206.513	6.892	0.037	0.020
		Cloglog Modelo stepAIC		170.116	158.623	0.010	0.392	164.623	187.496	6.756	0.006	0.003

4.4km	Más menos dos días	Logit		150.877	146.073	0.004	0.353	156.073	192.987	3.184	0.339	0.308
		Logit Modelo stepAIC		150.877	146.126	0.006	0.401	152.126	174.274	3.148	0.106	0.093
		Probit		150.877	146.058	0.004	0.353	156.058	192.972	3.194	0.317	0.306
		Probit Modelo stepAIC		150.877	146.126	0.006	0.401	152.126	174.275	3.148	0.099	0.093
		Cloglog		150.877	146.114	0.004	0.352	156.114	193.027	3.157	0.348	0.313
		Cloglog Modelo stepAIC		150.877	147.935	0.005	0.423	151.935	166.701	1.949	0.081	0.086

4.4km	Más menos un día	Logit		142.796	133.643	0.014	0.295	143.643	180.182	6.410	0.112	0.057
		Logit Modelo stepAIC		142.796	133.734	0.019	0.360	139.734	161.658	6.346	0.025	0.011
		Probit		142.796	133.776	0.014	0.304	143.776	180.315	6.317	0.091	0.061
		Probit Modelo stepAIC		142.796	133.845	0.019	0.362	139.845	161.769	6.269	0.019	0.011
		Cloglog		142.796	133.481	0.014	0.305	143.481	180.021	6.523	0.119	0.054
		Cloglog Modelo stepAIC		142.796	133.547	0.020	0.367	139.547	161.471	6.477	0.027	0.010

2.2km	Más menos dos días	Logit	414.937	382.811	0.063	0.398	392.811	441.304	7.742	0.000	0.000
		Logit Modelo stepAIC	414.937	382.811	0.068	0.412	390.811	429.606	7.742	0.000	0.000
		Probit	414.937	383.023	0.062	0.405	393.023	441.516	7.691	0.000	0.000
		Probit Modelo stepAIC	414.937	383.028	0.067	0.420	391.028	429.823	7.690	0.000	0.000
		Cloglog	414.937	381.779	0.068	0.371	391.779	440.272	7.991	0.000	0.000
		Cloglog Modelo stepAIC	414.937	381.854	0.073	0.402	389.854	428.649	7.973	0.000	0.000

2.2km	Más menos un día	Logit	357.681	327.801	0.298	0.401	337.801	385.484	8.354	0.000	0.000
		Logit Modelo stepAIC	357.681	327.809	0.312	0.414	335.809	373.955	8.352	0.000	0.000
		Probit	357.681	328.083	0.294	0.420	338.083	385.766	8.275	0.000	0.000
		Probit Modelo stepAIC	357.681	328.088	0.308	0.432	336.088	374.235	8.273	0.000	0.000
		Cloglog	357.681	327.095	0.308	0.389	337.095	384.778	8.551	0.000	0.000
		Cloglog Modelo stepAIC	357.681	327.167	0.321	0.408	335.167	373.313	8.531	0.000	0.000

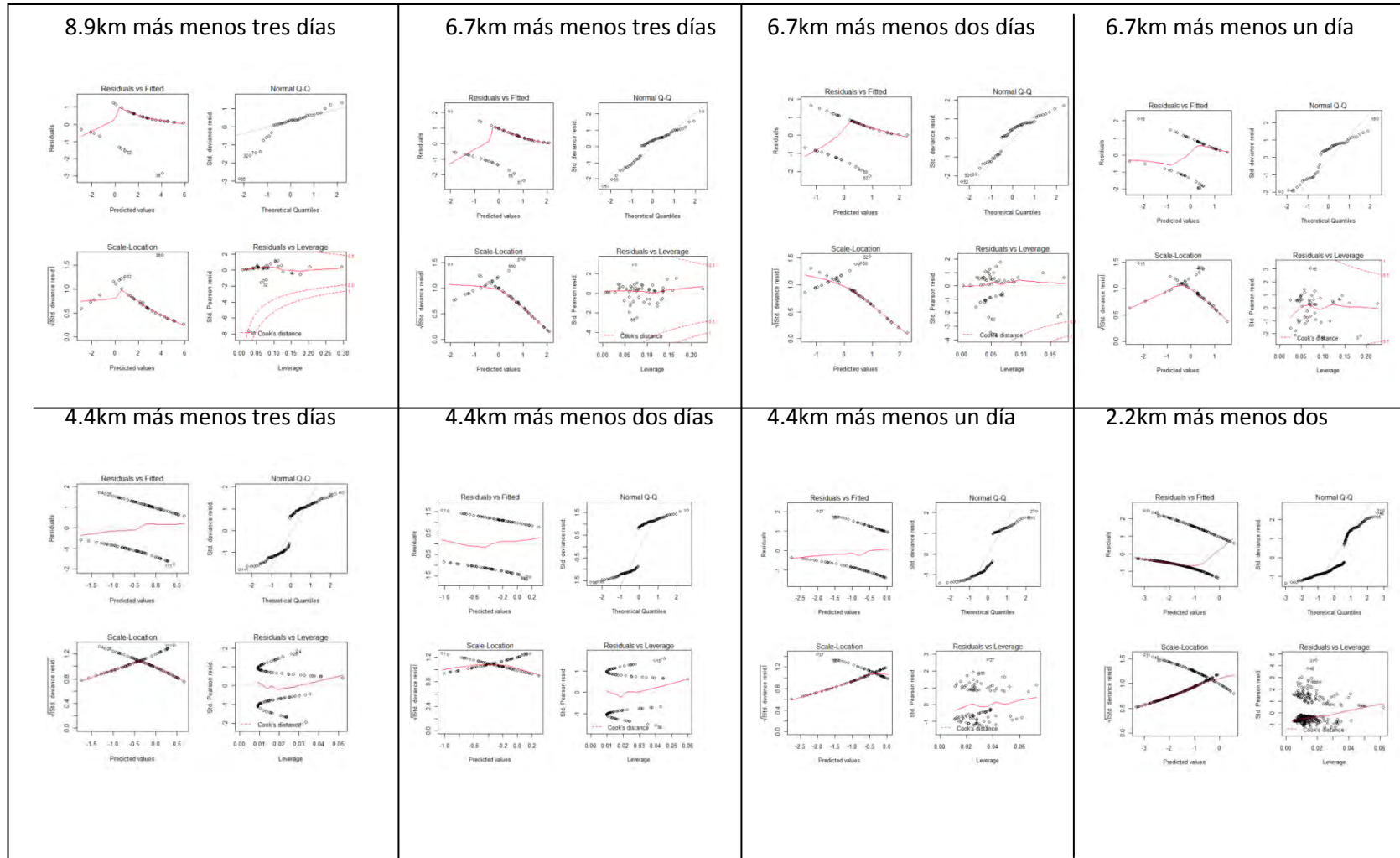
1.1km	Más menos un día	Logit	642.691	597.619	1.000	0.010	607.619	665.753	7.013	0.000	0.000
		Logit Modelo stepAIC	642.691	599.571	1.000	0.279	607.571	654.079	6.709	0.000	0.000
		Probit	642.691	600.008	1.000	0.109	610.008	668.143	6.641	0.000	0.000
		Probit Modelo stepAIC	642.691	600.730	1.000	0.273	608.730	655.238	6.529	0.000	0.000
		Cloglog	642.691	596.175	1.000	0.003	606.175	664.310	7.238	0.000	0.000
		Cloglog Modelo stepAIC	642.691	596.175	1.000	0.003	606.175	664.310	7.238	0.000	0.000
		Logit sin outliers	638.325	588.568	1.000	0.002	598.568	656.691	7.795	0.000	0.000
		Logit sin outliers Modelo stepAIC	638.325	588.568	1.000	0.002	598.568	656.691	7.795	0.000	0.000
		Cloglog sin outliers	638.325	587.040	1.000	0.001	597.040	655.163	8.034	0.000	0.000

1.1km	Día-día	Logit	248.575	235.439	1.000	0.000	245.439	297.605	5.285	0.123	0.011
		Logit Modelo stepAIC	248.575	235.439	1.000	0.000	243.439	285.172	5.285	0.065	0.004
		Probit	248.575	236.476	1.000	0.001	246.476	298.642	4.867	0.146	0.017
		Probit Modelo stepAIC	248.575	236.482	1.000	0.001	244.482	286.215	4.865	0.079	0.007

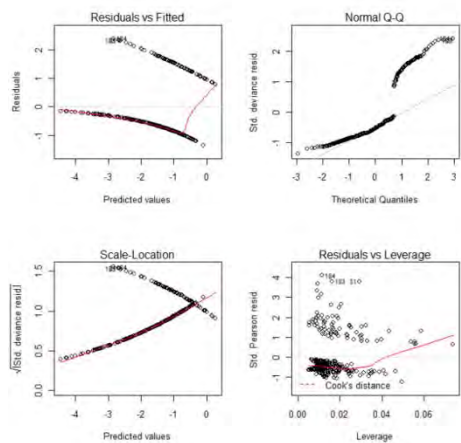


Cloglog	248.575	235.263	1.000	0.000	245.263	297.430	5.355	0.112	0.010
Cloglog Modelo stepAIC	248.575	235.264	1.000	0.000	243.264	284.996	5.355	0.058	0.004
Logit sin outliers	243.167	222.906	1.000	0.337	232.906	285.052	8.332	0.035	0.000
Logit sin outliers Modelo stepAIC	243.167	222.948	1.000	0.363	230.948	272.665	8.315	0.015	0.000
Probit sin outliers	243.167	223.522	1.000	0.247	233.522	285.668	8.079	0.053	0.001
Probit sin outliers Modelo stepAIC	243.167	223.547							
Cloglog sin outliers	243.167	222.806	1.000	0.378	232.806	284.952	8.373	0.030	0.000
Cloglog sin outliers Modelo stepAIC	243.167	222.855							

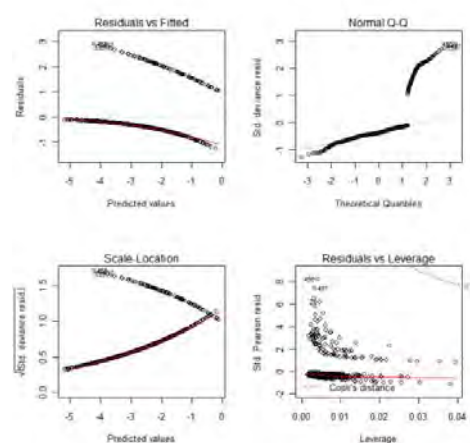
ANEXO 8: Validación de MLG finales Ausencia-presencia de ARS.



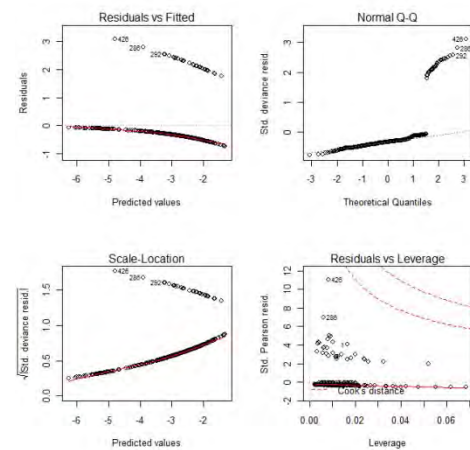
### 2.2km más menos un día



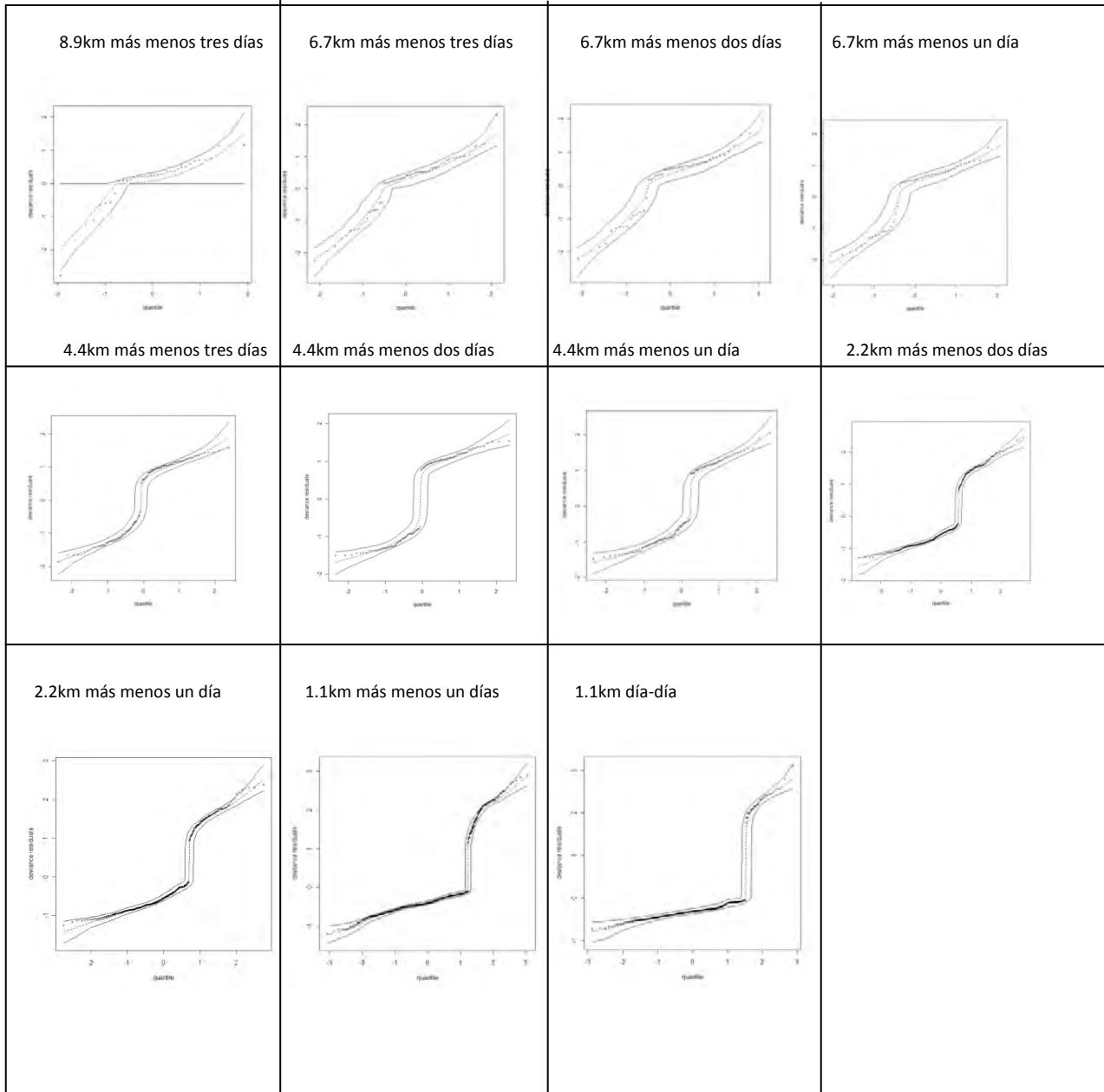
### 1.1km más menos un día



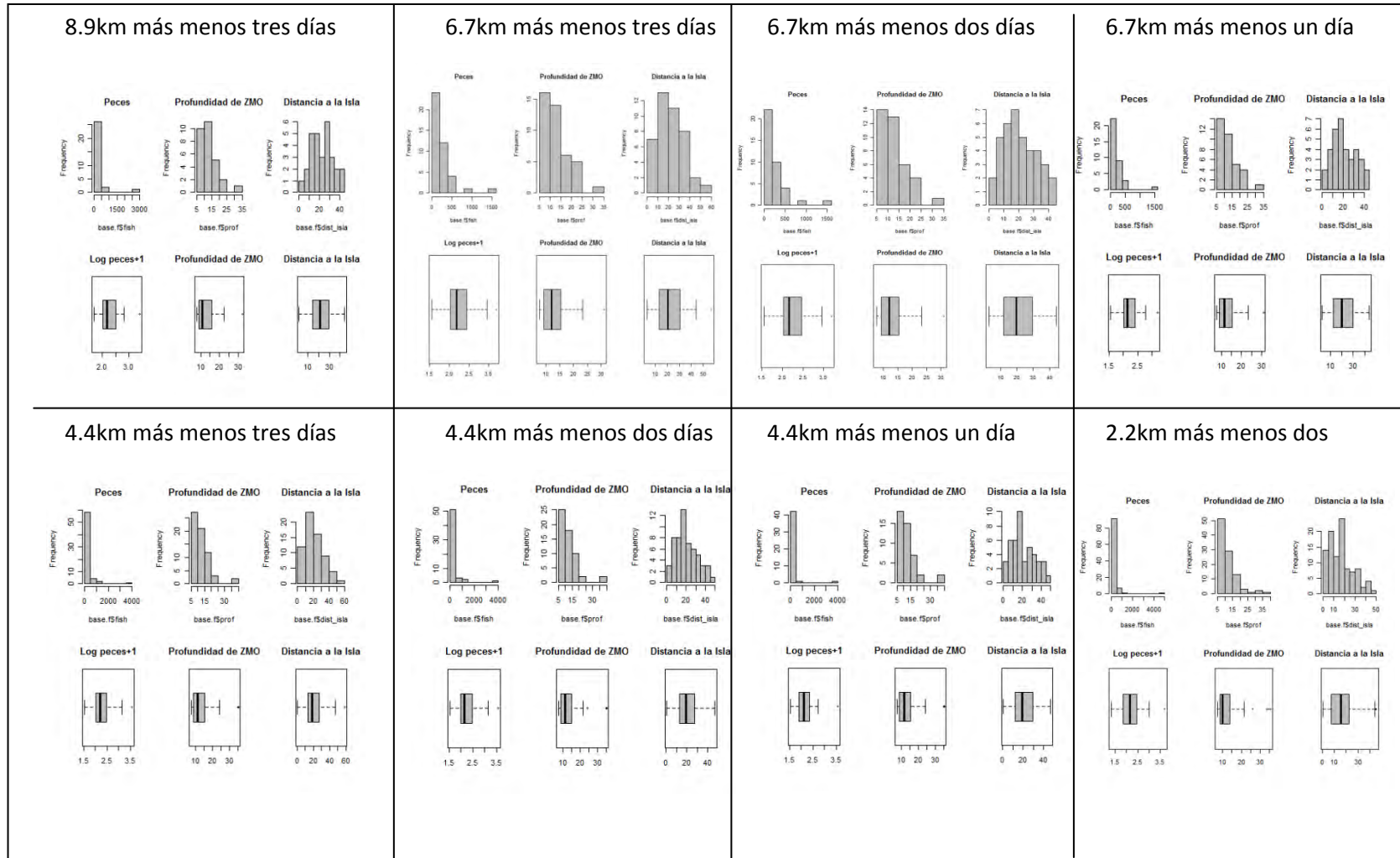
### 1.1km día-día



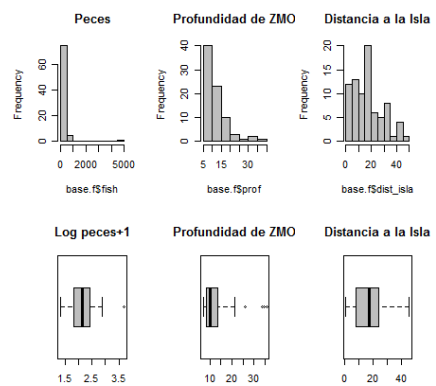
ANEXO 9: Validación de MLG finales Ausencia-presencia de ARS. Gráficos de envolturas de los cuantiles de los residuos de desvianza.



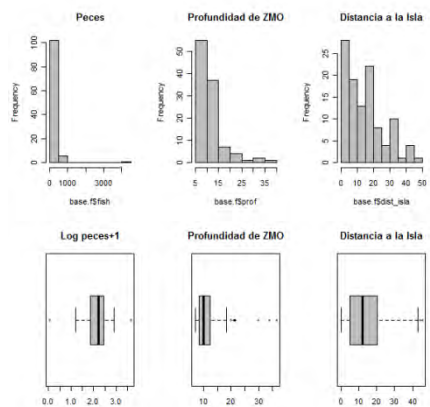
ANEXO 10: Histogramas de frecuencia y diagramas de caja de covariables de MLG conteos de ARS.



## 2.2km más menos un día



## 1.1km más menos un día



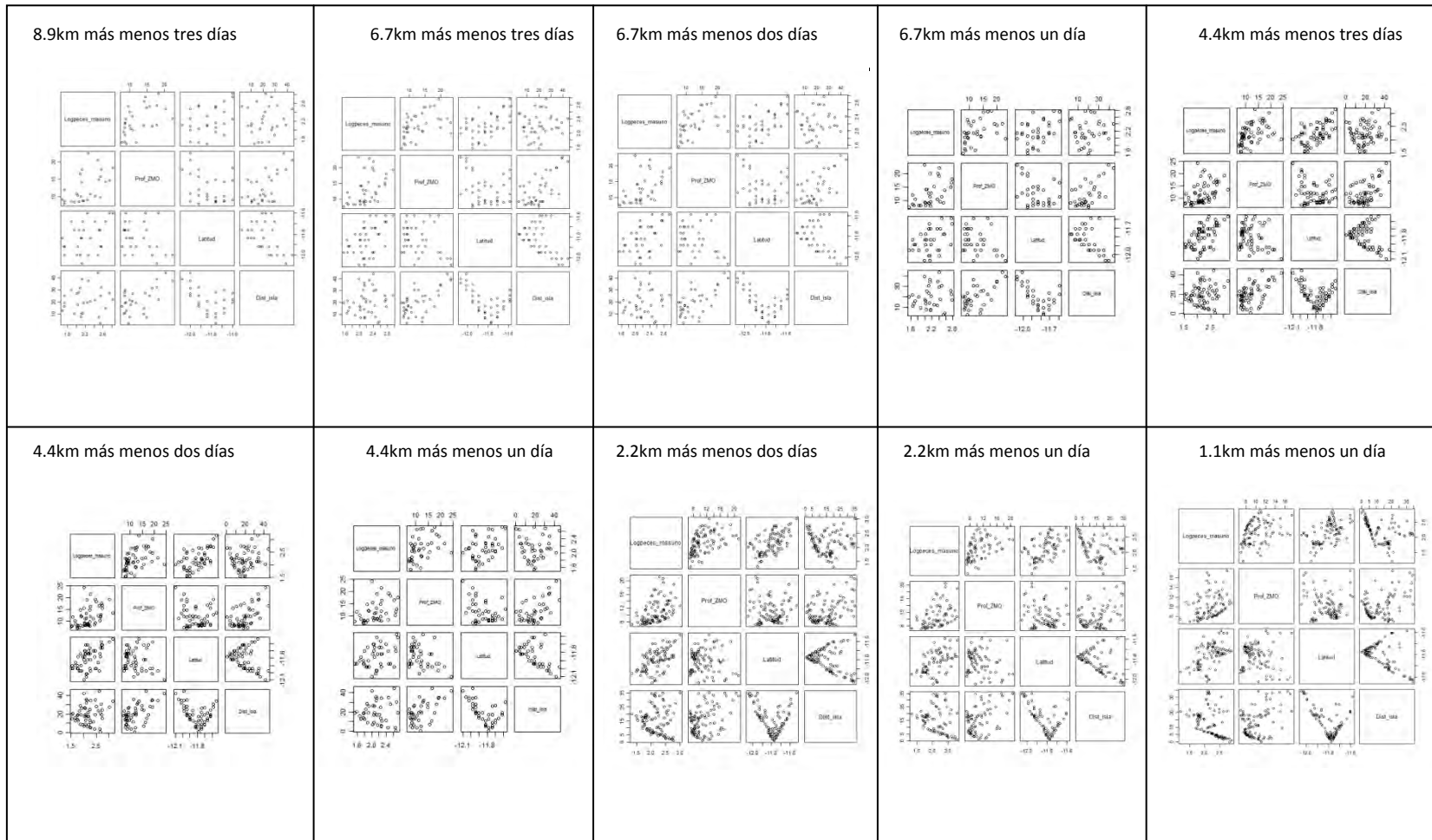
ANEXO 11: Resultados del análisis de correlación de Spearman de las covariables y prueba de Anderson-Darling (A\_D) de la variable ARS del MLG conteos de ARS.

ESCALA ESPACIAL	ESCALA TEMPORAL	Covariables	Log10(NASC+1)	Prof.ZMO	Latitud	Dist.Isla	A-D
0.08	Más menos tres días	Log10(NASC+1)	1	0.61	0.36	0.11	0.001008
		Prof.ZMO	0.61	1	-0.31	0.44	0.002383
		Latitud	0.36	-0.31	1	-0.4	
		Dist.Isla	0.11	0.44	-0.4	1	0.9567
0.06	Más menos tres días	Log10(NASC+1)	1	0.51	0.48	0.13	1.21E-06
		Prof.ZMO	0.51	1	-0.28	0.57	0.0002777
		Latitud	0.48	-0.28	1	-0.43	
		Dist.Isla	0.13	0.57	-0.43	1	0.4913
0.06	Más menos dos días	Log10(NASC+1)	1	0.54	0.51	0.1	3.31E-06
		Prof.ZMO	0.54	1	-0.2	0.55	0.002612
		Latitud	0.51	-0.2	1	-0.44	
		Dist.Isla	0.1	0.55	-0.44	1	0.5211
0.06	Más menos un día	Log10(NASC+1)	1	0.51	0.38	0.09	3.15E-05
		Prof.ZMO	0.51	1	-0.34	0.54	0.001231
		Latitud	0.38	-0.34	1	-0.56	
		Dist.Isla	0.09	0.54	-0.56	1	0.4131
0.04	Más menos tres días	Log10(NASC+1)	1	0.55	0.51	0.11	4.96E-12
		Prof.ZMO	0.55	1	-0.13	0.49	2.61E-06
		Latitud	0.51	-0.13	1	-0.27	
		Dist.Isla	0.11	0.49	-0.27	1	0.5724
0.04	Más menos dos días	Log10(NASC+1)	1	0.55	0.49	0.12	1.48E-12
		Prof.ZMO	0.55	1	-0.12	0.54	5.01E-06
		Latitud	0.49	-0.12	1	-0.34	
		Dist.Isla	0.12	0.54	-0.34	1	0.534
0.04	Más menos un día	Log10(NASC+1)	1	0.47	0.33	-0.03	6.56E-07
		Prof.ZMO	0.47	1	-0.35	0.5	0.0001757
		Latitud	0.33	-0.35	1	-0.59	
		Dist.Isla	-0.03	0.5	-0.59	1	0.4424

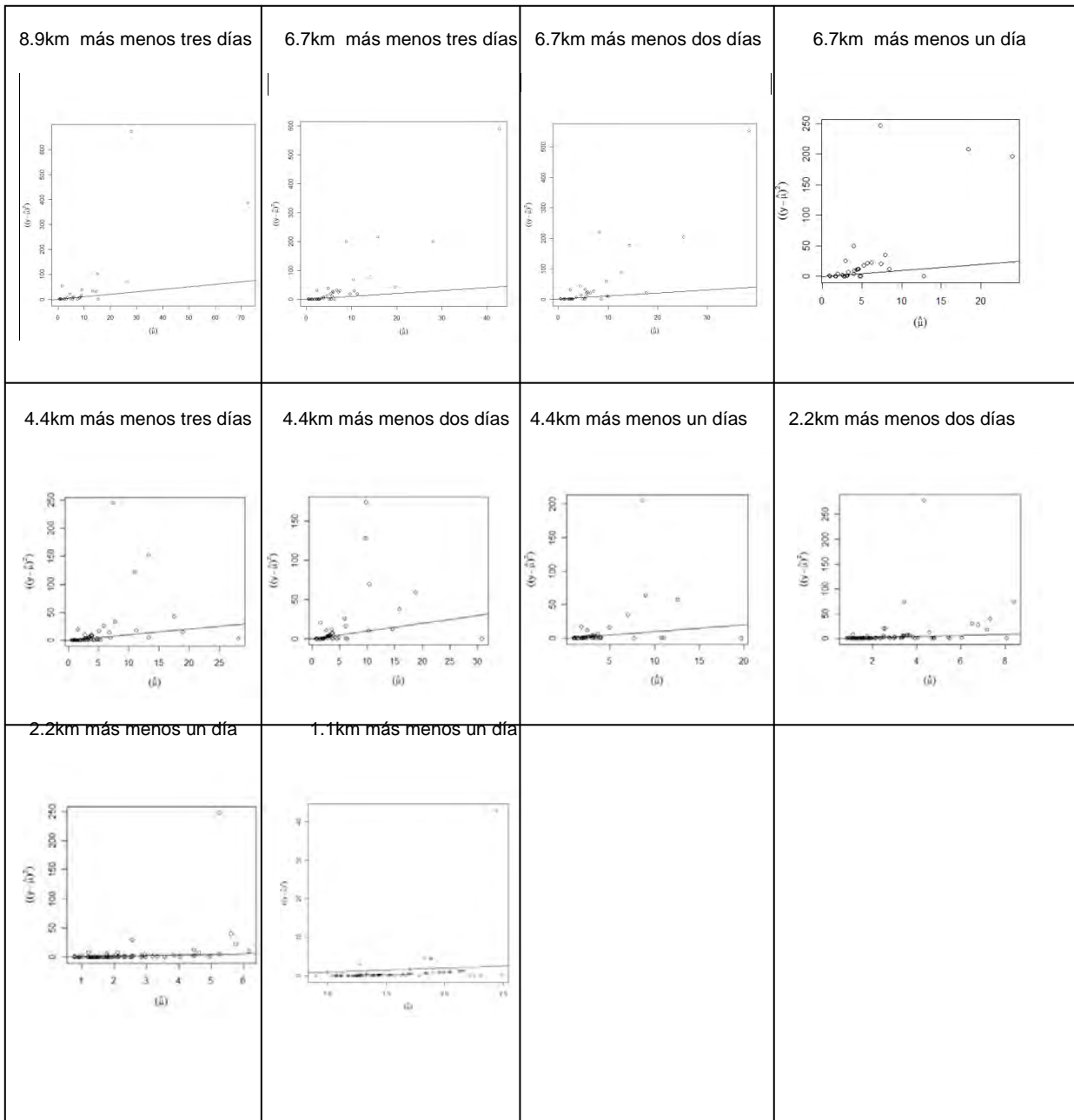
0.02	Más menos dos días	Log10(NASC+1)	1	0.56	0.51	-0.16	5.73E-12
		Prof.ZMO	0.56	1	-0.02	0.28	1.13E-08
		Latitud	0.51	-0.02	1	-0.31	
		Dist.Isla	-0.16	0.28	-0.31	1	0.003685
0.02	Más menos un día	Log10(NASC+1)	1	0.53	0.45	-0.3	4.28E-08
		Prof.ZMO	0.53	1	-0.09	0.28	8.63E-07
		Latitud	0.45	-0.09	1	-0.52	
		Dist.Isla	-0.3	0.28	-0.52	1	0.02799
0.01	Más menos un día	Log10(NASC+1)	1	0.36	0.47	-0.55	5.19E-09
		Prof.ZMO	0.36	1	-0.13	0.22	5.71E-06
		Latitud	0.47	-0.13	1	-0.42	
		Dist.Isla	-0.55	0.22	-0.42	1	9.21E-06
0.01	Día-día	Log10(NASC+1)	1	0.18	0.23	-0.28	0.00179
		Prof.ZMO	0.18	1	-0.71	0.71	0.002303
		Latitud	0.23	-0.71	1	-0.9	
		Dist.Isla	-0.28	0.71	-0.9	1	0.009263



ANEXO 12: Graficas de correlación de covariables usadas en el MLG conteos de ARS.



ANEXO 13: Graficas de relación media-varianza en MLG conteos de ARS.



ANEXO 14: Resultados GLM conteos de ARS con función de enlace log

Escala Espacial	Escala Temporal	Modelo	Tamaño de muestra	Log(peces+1)		Prof.ZMO		Latitud		Dist.Isla	
				Estimado	p-valor	Estimado	p-valor	Estimado	p-valor	Estimado	p-valor
8.9km	Más menos tres días	Poisson	29	1.1834	0.0001	-0.0286	0.2985	-2.3808	0.0050	-0.1116	0.0000
		Poisson Poisson Modelo stepAIC	29	0.9874	0.0000			-2.0439	0.0096	-0.1156	0.0000
		Poisson sin outliers		-0.2863	0.5090	0.1129	0.0001	0.5483	0.5943	-0.1114	0.0000
		Poisson sin outliers Modelo stepAIC				0.1003	0.0000			-0.1131	0.0000
		Binomial Negativo	29	0.2291	0.7193	0.0875	0.0581	-0.8405	0.5982	-0.1182	0.0000
		Binomial Negativo Modelo stepAIC	29			0.1015	0.0034			-0.1175	0.0000
		Binomial Negativo sin outliers		0.3964	0.3949	0.1020	0.0050	-1.0336	0.3838	-0.1331	0.0000
		<b>Binomial Negativo sin outliers Modelo stepAIC</b>				<b>0.1238</b>	<b>0.0000</b>			<b>-0.1333</b>	<b>0.0000</b>
6.7km	Más menos tres días	Poisson	42	1.3758	0.0000	-0.0571	0.0249	-2.8580	0.0005	-0.0941	0.0000
		Poisson Poisson Modelo stepAIC	42	1.3758	0.0000	-0.0571	0.0249	-2.8580	0.0005	-0.0941	0.0000
		Poisson sin outliers		-0.2198	0.6076	0.0471	0.1058	0.3361	0.7513	-0.0782	0.0000
		Poisson sin outliers Modelo stepAIC				0.0367	0.0811			-0.0790	0.0000
		Binomial Negativo	42	0.6178	0.3328	0.0069	0.8823	-1.6737	0.2919	-0.0859	0.0000
				<b>Binomial Negativo Modelo stepAIC</b>	42						<b>-0.0743</b>
6.7km	Más menos dos días	Poisson	38	1.4261	0.0000	-0.0539	0.0428	-3.1009	0.0004	-0.0915	0.0000
		Poisson Poisson Modelo stepAIC	38	1.4261	0.0000	-0.0539	0.0428	-3.1009	0.0004	-0.0915	0.0000
		Poisson sin outliers		-0.2082	0.6503	0.0133	0.6862	0.6419	0.5802	-0.0568	0.0000
		Poisson sin outliers Modelo stepAIC								-0.0581	0.0000
		Binomial Negativo	38	0.6830	0.2938	0.0133	0.7822	-1.9777	0.2419	-0.0880	0.0000
				<b>Binomial Negativo Modelo stepAIC</b>	38	0.7934	0.1175			<b>-2.1905</b>	<b>0.1433</b>
6.7km	Más menos un día	Poisson	35	1.3430	0.0001	-0.0244	0.3929	-4.4370	0.0000	-0.0864	0.0000

		Poisson Poisson Modelo stepAIC	35	1.1680	0.0000			-4.1245	0.0001	-0.0902	0.0000
		Poisson sin outliers		0.5491	0.2714	-0.0045	0.8957	-2.4790	0.0632	-0.0621	0.0001
		Poisson sin outliers Modelo stepAIC						-1.3863	0.1186	-0.0510	0.0000
		Binomial Negativo	35	0.9044	0.1815	0.0141	0.7745	-3.4299	0.0668	-0.0839	0.0000
		Binomial Negativo Modelo stepAIC	35	1.0215	0.0597			-3.6763	0.0270	-0.0813	0.0000

4.4km	Más menos tres días	Poisson	65	1.0359	0.0001	-0.0010	0.9626	-3.0900	0.0001	-0.0920	0.0000
		Poisson Poisson Modelo stepAIC	65	1.0285	0.0000			-3.0741	0.0000	-0.0922	0.0000
		Poisson sin outliers		1.1339	0.0000	-0.0104	0.6844	-3.1409	0.0001	-0.0931	0.0000
		Poisson sin outliers Modelo stepAIC		1.0651	0.0000			-3.0026	0.0000	-0.0948	0.0000
		Binomial Negativo	65	0.7714	0.0619	0.0252	0.4044	-2.5775	0.0250	-0.0913	0.0000
		Binomial Negativo Modelo stepAIC	65	0.9807	0.0026			-3.0475	0.0020	-0.0869	0.0000
		Binomial Negativo sin outliers		0.7938	0.0357	0.0361	0.2047	-2.6674	0.0125	-0.0991	0.0000
		Binomial Negativo sin outliers Modelo stepAIC		1.0859	0.0003			-3.3092	0.0004	-0.0924	0.0000

4.4km	Más menos dos días	Poisson	57	1.0524	0.0001	0.0433	0.0462	-2.9361	0.0006	-0.1019	0.0000
		Poisson Poisson Modelo stepAIC	57	1.0524	0.0001	0.0433	0.0462	-2.9361	0.0006	-0.1019	0.0000
		Poisson sin outliers		0.6571	0.0248	0.0803	0.0008	-2.3212	0.0082	-0.1132	0.0000
		Poisson sin outliers Modelo stepAIC		0.6571	0.0248	0.0803	0.0008	-2.3212	0.0082	-0.1132	0.0000
		Binomial Negativo	57	1.0471	0.0081	0.0468	0.1042	-3.0687	0.0087	-0.1031	0.0000

4.4km	Más menos un día	Poisson	44	0.8560	0.0053	0.0447	0.0513	-2.5931	0.0144	-0.0837	0.0000
		Poisson Poisson Modelo stepAIC	44	0.8560	0.0053	0.0447	0.0513	-2.5931	0.0144	-0.0837	0.0000
		Poisson sin outliers		0.8011	0.0236	0.0320	0.2754	-2.6349	0.0180	-0.0811	0.0000
		Poisson sin outliers Modelo stepAIC		1.0477	0.0001			-3.1864	0.0012	-0.0743	0.0000
		Binomial Negativo	44	0.9953	0.0232	0.0378	0.2101	-2.9681	0.0329	-0.0854	0.0000
		Binomial Negativo Modelo stepAIC	44	1.3206	0.0004			-3.7654	0.0027	-0.0789	0.0000

2.2km	Más menos dos días	Poisson	100	0.9209	0.0001	0.0298	0.1400	-1.6279	0.0464	-0.0583	0.0000
-------	--------------------	---------	-----	--------	--------	--------	--------	---------	--------	---------	--------

	Poisson Modelo stepAIC	100	0.9209	0.0001	0.0298	0.1400	-1.6279	0.0464	-0.0583	0.0000
	Poisson sin outliers		1.1569	0.0000	-0.0164	0.4892	-1.9211	0.0208	-0.0458	0.0000
	<b>Poisson sin outliers Modelo stepAIC</b>		<b>1.0610</b>	<b>0.0000</b>			<b>-1.7273</b>	<b>0.0277</b>	<b>-0.0490</b>	<b>0.0000</b>
	Binomial Negativo	100	0.8781	0.0050	0.0287	0.2470	-1.5789	0.1197	-0.0567	0.0000
	Binomial Negativo Modelo stepAIC	100	1.0688	0.0001			-2.0045	0.0336	-0.0505	0.0000
	Binomial Negativo sin outliers		1.0703	0.0002	-0.0098	0.7038	-1.7967	0.0557	-0.0451	0.0000
	Binomial Negativo sin outliers Modelo stepAIC		1.0099	0.0000			-1.6740	0.0579	-0.0472	0.0000

2.2km	Más menos un día	Poisson	80	0.5845	0.0436	0.0471	0.0286	-1.6822	0.0980	-0.0570	0.0000
		Poisson Modelo stepAIC	80	0.5845	0.0436	0.0471	0.0286	-1.6822	0.0980	-0.0570	0.0000
		Poisson sin outliers		0.8928	0.0030	-0.0073	0.7723	-2.0829	0.0445	-0.0378	0.0003
		<b>Poisson sin outliers Modelo stepAIC</b>		<b>0.8449</b>	<b>0.0008</b>			<b>-1.9957</b>	<b>0.0446</b>	<b>-0.0396</b>	<b>0.0000</b>
		Binomial Negativo	80	0.6136	0.0739	0.0409	0.1050	-1.7023	0.1502	-0.0544	0.0000
		Binomial Negativo Modelo stepAIC	80	0.6136	0.0739	0.0409	0.1050	-1.7023	0.1502	-0.0544	0.0000
		Binomial Negativo sin outliers		0.8918	0.0031	-0.0073	0.7740	-2.0808	0.0451	-0.0378	0.0003
	Binomial Negativo sin outliers Modelo stepAIC		0.8438	0.0009			-1.9933	0.0454	-0.0395	0.0000	

Escala Espacial	Escala Temporal	Modelo	DESVIANZA NULA	DEVIANZA RESIDUAL	Funcion Desviianza (p-valor)	Prueba de Pearson (p-valor)	AIC	BIC	DESVIANZA EXPLICADA%	Prueba de Wald (p-valor)	Prueba LRT (p-valor)
8.9km	Más menos tres días	Poisson	441.516	113.933	0.000	0.000	222.753	246.426	74.20%	1.06E-63	1.21E-69
		Poisson Poisson Modelo stepAIC	441.516	115.032	0.000	0.000	221.852	240.790	73.95%	1.70E-63	1.84E-70
		Poisson sin outliers	157.927	33.620	0.054	0.028	133.464	156.423	78.71%	6.67E-23	6.42E-26
		Poisson sin outliers Modelo stepAIC	157.927	34.057	0.084	0.045	129.901	143.676	78.44%	9.17E-25	1.26E-27
		Binomial Negativo	95.800	25.275	0.391	0.681	166.422	194.829	73.62%	3.82E-11	
		Binomial Negativo Modelo stepAIC	94.532	25.222							

		Binomial Negativo sin outliers	186.760	25.460	0.327	0.217	148.212	176.199	86.37%	1.37E-22
		Binomial Negativo sin outliers Modelo stepAIC	171.953	24.580	0.486	0.283	144.995	163.653	85.71%	2.68E-22

6.7km	Más menos tres días	Poisson	416.132	159.658	0.000	0.000	300.315	327.691	61.63%	1.64E-48	2.62E-54
		Poisson Poisson Modelo stepAIC	416.132	159.658	0.000	0.000	300.315	327.691	61.63%	1.64E-48	2.62E-54
		Poisson sin outliers	203.801	115.770	0.000	0.000	245.893	272.782	43.19%	7.85E-15	3.45E-18
		Poisson sin outliers Modelo stepAIC	203.801	116.064	0.000	0.000	242.187	258.320	43.05%	2.40E-16	8.87E-20
		Binomial Negativo	92.351	39.885	0.343	0.439	227.660	260.512	56.81%	2.65E-07	
		Binomial Negativo Modelo stepAIC	86.087	39.684							

6.7km	Más menos dos días	Poisson	351.856	148.227	0.000	0.000	278.336	304.711	57.87%	2.07E-40	6.23E-43
		Poisson Poisson Modelo stepAIC	351.856	148.227	0.000	0.000	278.336	304.711	57.87%	2.07E-40	6.23E-43
		Poisson sin outliers	131.887	86.094	0.000	0.000	201.003	226.557	34.72%	7.23E-08	2.72E-09
		Poisson sin outliers Modelo stepAIC	131.887	86.435	0.000	0.000	195.345	205.566	34.46%	4.79E-10	1.56E-11
		Binomial Negativo	77.302	35.520	0.350	0.467	209.244	240.895	54.05%	2.74E-06	
		Binomial Negativo Modelo stepAIC	77.392	35.642							

6.7km	Más menos un día	Poisson	228.062	117.192	0.000	0.000	230.390	255.943	48.61%	4.34E-23	4.75E-23
		Poisson Poisson Modelo stepAIC	228.062	117.931	0.000	0.000	229.129	249.571	48.29%	9.37E-24	1.03E-23
		Poisson sin outliers	90.160	66.295	0.000	0.000	165.767	190.424	26.47%	0.0002132	8.50E-05
		Poisson sin outliers Modelo stepAIC	90.160	67.797	0.000	0.000	163.268	178.063	24.80%	4.60E-05	1.39E-05
		Binomial Negativo	59.899	31.987	0.368	0.425	181.532	212.196	46.60%	0.0002463	
		Binomial Negativo Modelo stepAIC	59.894	32.068							

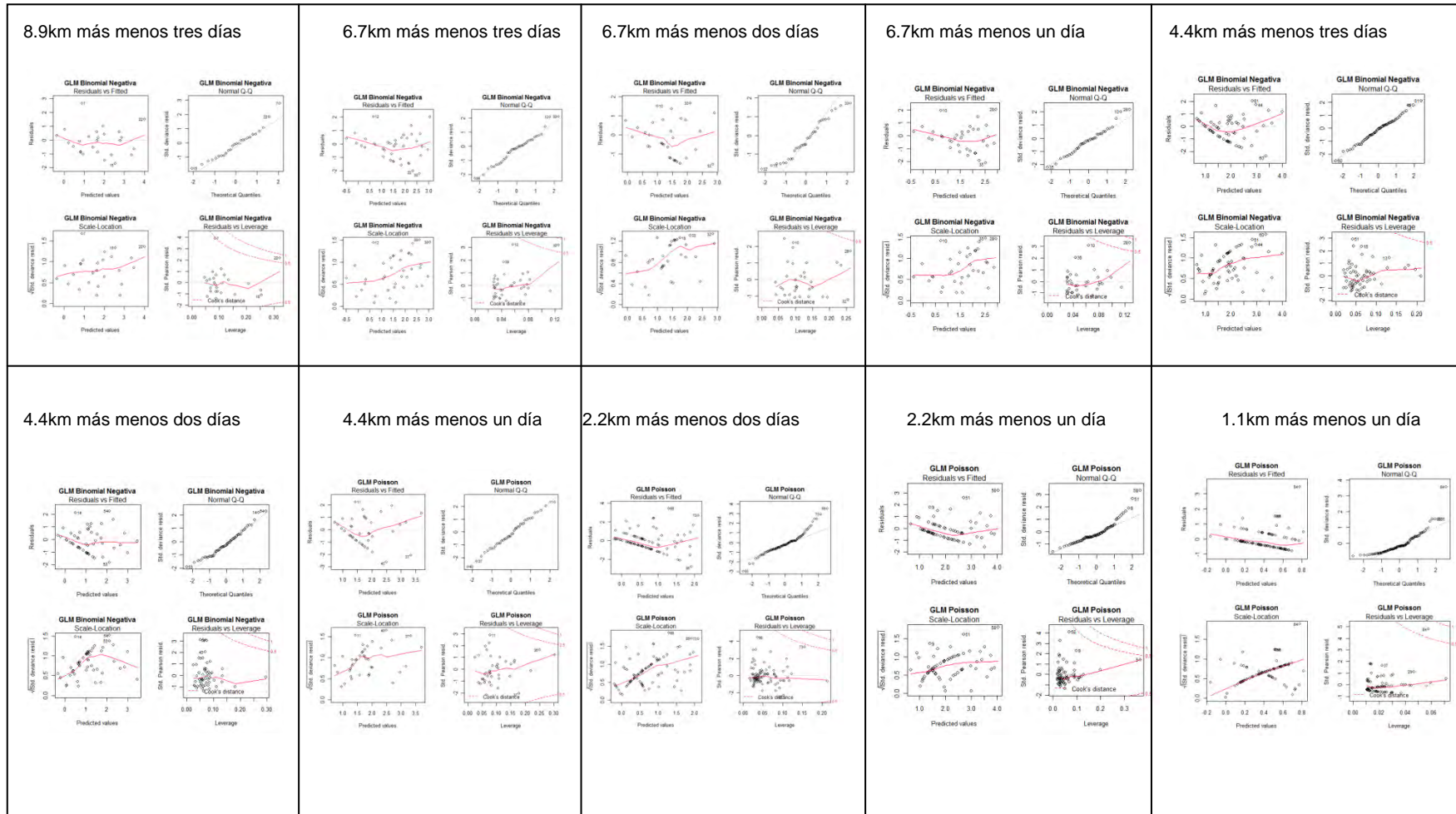
4.4km	Más menos tres días	Poisson	352.751	124.185	0.000	0.000	316.081	347.824	64.80%	7.01E-47	2.69E-48
		Poisson Poisson Modelo stepAIC	352.751	124.187	0.000	0.000	314.083	339.478	64.79%	7.65E-48	2.83E-49
		Poisson sin outliers	308.381	85.103	0.012	0.000	270.018	301.450	72.40%	2.62E-47	3.69E-47
		Poisson sin outliers Modelo stepAIC	308.381	85.270	0.014	0.000	268.186	293.331	72.35%	3.04E-48	4.27E-48
		Binomial Negativo	152.783	55.782	0.630	0.166	287.515	325.607	63.49%	5.87E-15	

			Binomial Negativo Modelo stepAIC	153.054	56.564							
			Binomial Negativo sin outliers	181.760	55.612	0.601	0.025	276.028	313.935	69.40%	1.70E-19	
			Binomial Negativo sin outliers Modelo stepAIC	179.344	56.561	0.602	0.037	275.609	307.198	68.46%	1.56E-19	
4.4km	Más menos dos días	Poisson		312.377	91.563	0.001	0.000	262.193	292.623	70.69%	8.09E-49	1.25E-46
		Poisson Poisson Modelo stepAIC		312.377	91.563	0.001	0.000	262.193	292.623	70.69%	8.09E-49	1.25E-46
		Poisson sin outliers		310.723	77.301	0.010	0.001	245.316	275.570	75.12%	1.86E-50	2.42E-49
		Poisson sin outliers Modelo stepAIC		310.723	77.301	0.010	0.001	245.316	275.570	75.12%	1.86E-50	2.42E-49
		Binomial Negativo		156.025	46.928	0.673	0.518	246.913	283.429	0.699	3.072E-18	
4.4km	Más menos un día	Poisson		181.333	70.283	0.002	0.000	203.806	231.648	61.24%	4.51E-24	4.35E-23
		Poisson Poisson Modelo stepAIC		181.333	70.283	0.002	0.000	203.806	231.648	61.24%	4.51E-24	4.35E-23
		Poisson sin outliers		133.866	41.358	0.286	0.183	167.900	195.277	69.11%	1.22E-20	3.86E-19
		Poisson sin outliers Modelo stepAIC		133.866	42.512	0.283	0.231	167.055	188.956	68.24%	3.17E-21	1.12E-19
		Binomial Negativo		95.000	35.635	0.624	0.477	192.371	225.781	62.49%	7.51E-11	
		Binomial Negativo Modelo stepAIC		91.462	35.833							
2.2km	Más menos dos días	Poisson		234.238	140.308	0.002	0.000	400.020	436.072	40.10%	9.75E-20	1.92E-19
		Poisson Modelo stepAIC		234.238	140.308	0.002	0.000	400.020	436.072	40.10%	9.75E-20	1.92E-19
		Poisson sin outliers		180.588	101.741	0.275	0.057	356.563	392.514	43.66%	1.29E-16	3.06E-16
		Poisson sin outliers Modelo stepAIC		180.588	102.232	0.288	0.061	355.054	383.815	43.39%	2.55E-17	6.91E-17
		Binomial Negativo		127.426	72.326	0.960	0.355	374.501	417.763	43.24%	4.79E-10	
		Binomial Negativo Modelo stepAIC		126.118	72.885							
		Binomial Negativo sin outliers		127.116	70.657	0.965	0.753	350.922	394.063	44.41%	2.11E-10	
		Binomial Negativo sin outliers Modelo stepAIC		126.462	70.431	0.972	0.780	349.064	385.016	44.31%	4.91E-11	
2.2km	Más menos un día	Poisson		141.571	91.957	0.089	0.000	297.716	331.536	35.05%	2.87E-10	4.35E-10
		Poisson Modelo stepAIC		141.571	91.957	0.089	0.000	297.716	331.536	35.05%	2.87E-10	4.35E-10
		Poisson sin outliers		82.534	49.807	0.986	0.910	250.675	284.370	39.65%	7.01E-07	1.36E-06

Poisson sin outliers Modelo stepAIC	82.534	49.892	0.989	0.920	248.760	275.716	39.55%	1.82E-07	3.83E-07
Binomial Negativo	86.817	53.180	0.974	0.208	285.170	325.754	38.74%	2.48E-06	
Binomial Negativo Modelo stepAIC	86.817	53.180							
Binomial Negativo sin outliers	82.149	49.560	0.987	0.913	252.674	293.108	39.67%	7.72E-07	
Binomial Negativo sin outliers Modelo stepAIC	81.918	49.494	0.990	0.925	250.757	284.452	39.58%	2.13E-07	



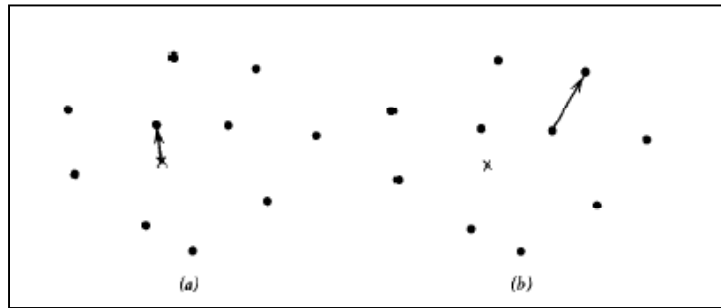
ANEXO 15: Validación de MLG finales conteos de ARS.



Anexo 16

**Índice del vecino más próximo (Ripley, 1981)**

Suponiendo que el proceso puntual este formado por arboles el índice del vecino más próximo relaciona la distancia entre un árbol elegido al azar y su vecino más próximo (árbol-árbol) con la distancia entre un punto situado al azar y el árbol más cercano (punto-árbol).



**Figura 45. Tipos de medidas de distancia. • Denota un árbol, X un punto muestreado. (a) Punto al árbol más cercano (b) Árbol al árbol más cercano.**

El índice del vecino más próximo (NNI) está definido entonces como:

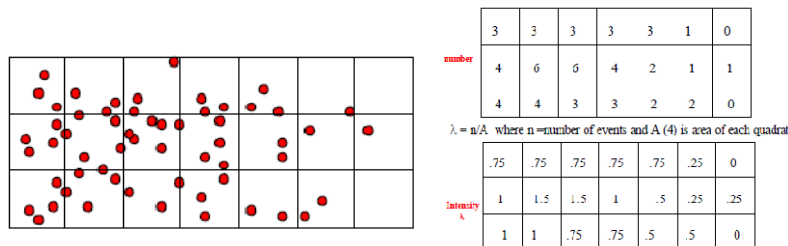
$$NNI = \frac{d(NN)}{d(\text{aleatorio})}$$

donde:  $d(NN) = \sum_{i=1}^N \frac{\min(d_{ij})}{N}$ ;  $d(\text{aleatorio}) = 0.5 \sqrt{\frac{\text{area}}{N}}$

Si el índice es menor a uno el proceso exhibe agregación, si el índice es mayor a uno el proceso es regular y si es igual a uno es completamente aleatorio.

**Método de los cuadrantes para probar aleatoriedad espacial completa (Baddley, 2008)**

Una prueba clásica para probar la hipótesis nula de aleatoriedad espacial completa (CSR) es la  $\chi^2$  basada en el conteo de cuadrantes (Fig. xx). Una alternativa a los contrastes basados en distancias consiste en dividir A en m subregiones o cuadrados de igual area y utilizar el contador de numero de eventos en cada cuadrado para la prueba de CSR. Las subregiones se eligen de forma arbitraria. Supongamos que A es el cuadrado unidad y está dividido en una red k xk de subregiones cuadradas, por tanto  $m = k^2$ . Sean  $n_i, i = 1, \dots, m$  los contadores de eventos en los cuadrados resultantes de esta partición de A y denotamos  $\bar{n} = n/m$  a la media muestral de los  $n_i$ .



**Figura 46. (a) Resumen de puntos en una región R. (b) Arriba: Numero de puntos en cuadrantes de la región R. Abajo: Intensidad por cuadrante de la región R.**

Un estadístico obvio para contrastar CSR en  $A$  es el criterio  $\chi^2$  de Pearson.

$$\chi^2 = \sum_{i=1}^m (n_i - \bar{n})^2 / \bar{n}$$

Donde  $m$  es el número de regiones y  $\bar{n} = \sum_{i=1}^m n_i / m$ , es el número esperado de eventos por región si la intensidad es homogénea. Bajo la hipótesis nula, este estadístico tiene distribución  $\chi_{m-1}^2$ , asumiendo que  $\bar{n}$  no es muy pequeño, en la práctica se pide  $\bar{n} > 5$ .

La elección del tamaño del cuadrante es una asunción implícita, queda abierta la duda sobre “cuál debe ser el tamaño de cuadrante”. El poder de esta prueba de cuadrantes depende del tamaño de los cuadrantes. Se aproxima a cero cuando los cuadrantes son muy grandes porque no cuenta con datos o se genera alta variación. Y cuando se toma en cuenta cuadrantes muy pequeños para mejor detalle espacial se genera alta variación.

### **Modelos Aditivos Generalizados (Zuur, 2008)**

Los Modelos Aditivos generalizados (GAM) son modelos no paramétricos en donde la variable respuesta puede tener una distribución no normal como en los MLG y está definida como:

$$Y_i = \alpha + f(X_i) + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma^2) \quad y \quad f(X_i) = \sum_{j=1}^p \beta_j b_j(X_i)$$

donde  $Y_i$  son independientes,  $\beta$  son los parámetros y los  $b_j(X_i)$  llamadas funciones básicas son los suavizadores. Por lo tanto la función  $f(X_i)$  es una curva de suavización que relaciona la variable respuesta con las variables predictoras mediante los suavizadores. Así los GAM permiten a los datos determinar la forma de las curvas de respuesta. Como resultado características como la bimodalidad y asimetría pronunciada en los datos pueden ser fácilmente detectadas.

Existen muchas clases de suavizadores para implementar los GAM, los más conocidos son: (i) Suavizador LOESS (en librería *gam* del R por defecto ajusta un modelo polinómico de orden 2) usa un método simple pero robusto llamado “algoritmo back-fitting” que estima cada suavizador por vez. La cantidad de suavización está determinada por el tamaño de la ventana definida por los nodos y solo en este caso se puede determinar la cantidad óptima de nodos en base a comparaciones visuales de los suavizadores. Se recomienda usar 3 nodos si se cuenta con menos de 30 observaciones y 5 nodos si se cuenta con más de 100 observaciones; (ii) Los splines cúbicos, B-splines, splines naturales, thin-splines y splines de suavización (o splines de suavización). En este caso la curva es dividida en segmentos usando cierto número de nodos y un modelo lineal, cuadrático o cúbico es ajustado a los datos en cada segmento. Los splines de suavización son introducidos como la mejor opción porque minimizan el criterio de los mínimos cuadrados penalizados. Algunos paquetes (librería *mgcv* en R) utilizan un término llamado grados de libertad efectivos (*edf*). Este es un valor entre 0 e infinito, un alto número de grados de libertad efectivos (*edf*) significa que es menos lineal la función de suavizado utilizada.

### **Cuasi Modelos Lineales Generalizados (McCulloch y Searle, 2001; McCullagh y Nelder, 1983)**

El término función cuasi-verosimilitud fue presentado por Robert Wedderburn en 1974 para describir una función de propiedades similares a las que tiene la función de log-verosimilitud. La estimación de cuasi-verosimilitud es una manera de permitir la sobredispersión, es decir, mayor variabilidad en los datos que la esperada en los modelos estadísticos utilizados. Los modelos de cuasi-verosimilitud puedan ajustarse usando una extensión directa del algoritmo usado para modelos lineales generalizados. Y a menudo se utiliza en modelos para datos de recuento o de datos binarios agrupados, es decir, datos que de otro modo serían modelados usando la distribución Binomial o Poisson. Los métodos cuasi-verosimilitud tienen la ventaja de la relativa simplicidad de cálculo, velocidad y robustez, ya que pueden hacer uso de los algoritmos más sencillos desarrollados para ajustar modelos lineales generalizados.

#### **Construcción de la función de cuasi-verosimilitud:**

Se define las siguientes propiedades del  $\log f_Y(y_i)$  derivando respecto de  $\mu_i$ :

$$E\left(\frac{d \log f_Y(y_i)}{d \mu_i}\right) = 0 \text{ y } \text{var}\left(\frac{d \log f_Y(y_i)}{d \mu_i}\right) = \frac{1}{a(\phi)b''(t)}$$

Se busca una cantidad en lugar del  $\log f_Y(y_i)$  que tenga las propiedades anteriores. Así se define:

$$q_i = \frac{y_i - \mu_i}{a(\phi)b''(\mu_i)}$$

que satisface las mismas condiciones, y además asume que  $\text{var}(y_i) \propto b''(\mu_i)$ .

Entonces se define el logaritmo de cuasi-verosimilitud como:

$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{a(\phi)b''(t)} dt$  el cual por definición, tiene derivada con respecto de  $\mu_i$  igual a  $q_i$ . Además  $Q_i$  es construida usando solamente información sobre como cambia la varianza con la media. Algunos ejemplos de esta función de cuasi-verosimilitud para un numero de varianzas comunes se muestran en la Tabla xx.

Finalmente para encontrar el estimador de  $\beta$  por máxima cuasi-verosimilitud (MQL) resolvemos la ecuación de máxima cuasi-verosimilitud:

$$\frac{\partial}{\partial \beta} \sum Q_i = 0$$

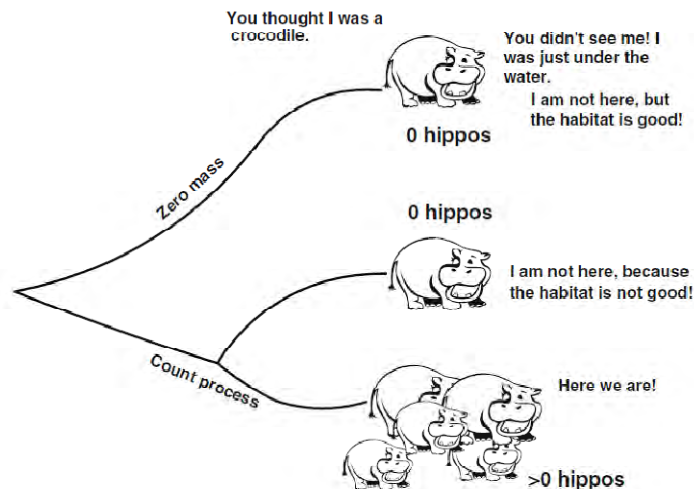
### **Modelos cero-inflado (Zeileis et al., 2008)**

Los modelos cero-inflado son otros modelos capaces de lidiar con exceso de ceros contados. Son modelos que resultan de dos componentes combinando un punto masivo en cero con una distribución de conteos como la Poisson (ZIP) o Binomial Negativa (ZINB). Hay dos fuentes de ceros (Fig. 47) ya que los ceros pueden provenir de las dos componentes. Para modelar el estado no observable se usa un modelo binario.

Formalmente, la densidad del modelo cero inflado es una combinación del punto masivo en cero  $I_{\{0\}}(y)$  y una distribución de conteo  $f_{\text{conteo}}(y; x, \beta)$ . La probabilidad de observar un conteo cero es inflada con la probabilidad  $\pi = f_{\text{cero}}(0; z, \gamma)$  modelada por un MLG Binomial:

$$f_{\text{cero inflado}}(y; x, z, \beta, \gamma) = f_{\text{cero}}(0; z, \gamma) \cdot I_{\{0\}}(y) + (1 - f_{\text{cero}}(0; z, \gamma)) \cdot f_{\text{conteo}}(y; x, \beta)$$

donde  $I(\cdot)$  es la función indicador .



**Figura 47. Croquis del principio fundamental de los modelos mixtos (ZIP y ZINB). En el conteo de los hipopótamos en las ubicaciones, uno puede medir un cero porque el hábitat no es bueno (a los hipopótamos no les gusta las covariables), o debido a un pobre diseño experimental y observadores sin experiencia (o experimentados observadores, pero dificultad para observar las especies).**

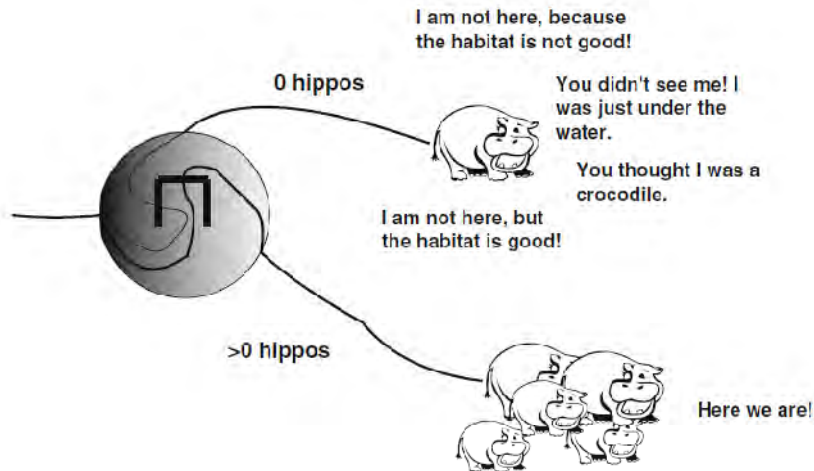
**Modelos Hurdle (Zeileis et al., 2008)**

Los modelos Hurdle están formados por dos componentes: Un componente de conteo truncado que es empleado para conteos positivos y un componente Hurdle que modela los conteos cero. A diferencia de los modelos cero inflado, no hay dos fuentes de ceros (Fig. 48): el modelo de recuento sólo se utiliza si el "Hurdle" (ocurrencia de conteos positivos) es excedido por la ocurrencia de ceros. El modelo de conteo es típicamente una regresión de Poisson o Binomial negativa truncada, es decir sin ceros. Para modelar el "Hurdle" puede ser empleado un modelo Binomial o una distribución de conteo (Poisson o Binomial Negativa).

Formalmente el modelo Hurdle combina el modelo de conteo  $f_{\text{conteo}}(y; x, \beta)$  y el modelo Hurdle cero  $f_{\text{cero}}(0; z, y)$  :

$$f_{\text{Hurdle}}(y; x, z, \beta, \gamma) = \begin{cases} f_{\text{cero}}(0; z, y) & \text{si } y = 0 \\ (1 - f_{\text{cero}}(0; z, y)) \cdot f_{\text{conteo}}(y; x, \beta) / (1 - f_{\text{conteo}}(0; x, \beta)) & \text{si } y > 0 \end{cases}$$

donde los parámetros son estimados por el método de máxima verosimilitud.



**Figura 48. Boceto del modelo Hurdle. Hay dos procesos: uno es la causa ceros frente a los no-ceros, el otro proceso está explicando los conteos de no-ceros. Esto se expresa con el Hurdle (obstáculo) en el círculo, hay que atravesarlo para llegar a conteos de no-ceros. El modelo no hace ninguna distinción entre los diferentes tipos de ceros.**