

Centro de Investigación en Matemáticas, A.C.

CIMAT

**MODELO ESTADÍSTICO PARA DATOS
SOBRE AÑOS DE EXTINCIÓN,
EXPERIMENTANDO POR MEDIO DE
SIMULACIÓN**

T E S I S

que para obtener el grado de

Maestro en Ciencias con Especialidad en
Probabilidad y Estadística

p r e s e n t a

Norma Selomit Ramírez Uribe

Dr. Miguel Nakamura Savoy
Director de Tesis

Guanajuato, Gto.

Agosto de 2010

Agradecimientos

Quisiera agradecer a CONACYT por el apoyo que recibí para poder realizar mis estudios en esta maestría. Así mismo, agradezco a CIMAT, A.C. y a cada uno de los profesores que fueron parte de la formación brindada. De una manera muy especial, al Dr. Miguel Nakamura Savoy, por su paciencia y apoyo en cada instante de este camino, cada palabra, por que cada palabra fue una enseñanza.

A mi familia, mis tres ángeles: José Ángel, María de los Ángeles y Ángeles, que siempre han estado a mi lado, apoyando y respetando cada decisión tomada. Por supuesto a Mónica López, su impulso y palabras de aliento para buscar más.

Mis amigos, aquellos con los que compartí parte de esta travesía y en conjunto hicimos tanto: Pedro Salazar, Roxana Góngora, Boryana López, Gustavo Cano, Diego Rivera, Joel Iglesias y Carlos Campos.

De manera especial, muy especial, a Norma Leticia Escobar Zaragoza y Lina C. Vargas Serdio, por que no sólo fueron y son el apoyo, paciencia y enseñanza de dos grandes amigas, sino de mis hermanas.

Por supuesto no quiero dejar de lado a todas aquellas personas que me enseñaron que golpear un balón, rodar por el suelo, e intentar volar, puede ser el mejor de los escapes.

Resumen

Se ha observado a través de la historia la gran importancia que tiene el detectar la desaparición de especies, más aún en las últimas décadas, donde se ha hecho evidente el mayor interés por la conservación de especies. Un interesante e importante tema en el área de biología, es el de extinción de especies. Uno de los aspectos más importantes de este tema, en el enfoque del cuidado ambiental, es la evolución de las extinciones a través del tiempo. Ésto ya que si se tiene la posibilidad de detectar cambios bruscos (principalmente aumento) en la detección de extinciones, se pueden tomar medidas necesarias, a tiempo, para el cuidado de especies en peligro.

En esta tesis, los datos proporcionados son años de “extinción” de especies marinas. Cuando se publicó esta base de datos, sólo se complementó con análisis cualitativo, en donde afirmaron que las extinciones van en aumento a causa de la sobreexplotación. Esta compilación tuvo una fuerte crítica: la falta de calidad en el proceso de recopilación de la misma. Así pues, surgió el interés de otras personas para analizarlos cuantitativamente, para lo cual algunos biólogos propusieron la construcción de índices. Esto se hizo con la intención de mostrar la imprudencia de hacer conclusiones, basándose únicamente en el análisis cualitativo de dicha lista.

El propósito de esta tesina es entender de manera más profunda el proceso de registro de años de “extinción”, y por supuesto la variabilidad que dicho proceso pueda causar a los datos. Para ésto, primero se identifican factores que pudieran estar causando fuerte variabilidad en los datos. Después se propone un modelo estadístico que incorpora dichos factores, y que pretende representar de manera adecuada el proceso de recopilación de datos. Además, también se desea proponer alguna metodología razonable para identificar el estado de la extinción, con el objetivo de que las aseveraciones que se hagan con respecto a la evolución de las extinciones sean válidas y pertinentes.

Dada la propuesta de analizar los datos de extinción a través de índices, inicialmente se desea justificar si esta propuesta es sensata. La justificación se realiza a través de experimentos de simulación, los cuales se llevan a cabo implementando el modelo estadístico. Dichos experimentos consideran diferentes situaciones, para observar los distintos patrones que se pueden obtener en los resultados gráficos del índice propuesto. En algunas ocasiones, se obtienen patrones con los que sería razonable concluir, aún si estos son incorrectos: aumento o decremento en las extinciones. Así pues, al estudiar la evolución de las extinciones a través del tiempo, haciendo uso de índices,

se puede llegar a concluir erróneamente. Además, se observó la gran variabilidad que tienen los datos, a causa de la cantidad de factores que intervienen en la producción de los mismos, lo cual tiene como efecto enmascarar el posible efecto de la extinción en sí misma.

Después de justificar la inconveniencia de concluir a partir de un análisis con base en índices, se propone una metodología alternativa para abordar el problema de cuantificación del estado de extinción. A esta metodología se le conoce como verosimilitud simulada, la cual tiene su valor en mostrar e ilustrar la capacidad de poder realizar inferencia estadística formal. Para mostrar lo anterior, se realizan ejemplos de distintas situaciones imaginarias del fenómeno de extinción, donde se desea identificar el estado de extinción. Para ésto, se trabaja con el modelo estadístico y la implementación de la metodología de verosimilitud simulada. En algunos ejemplos, se logra identificar el estado de las extinciones de una forma aparentemente precisa, sin embargo, algunos ejemplos son más certeros que otros, con lo que se reconoce la gran dificultad que tiene el hacer inferencia sobre los datos de “extinciones”.

El trabajo realizado ha sido de gran importancia, ya que muestra de manera muy ilustrativa el proceso de recopilación de datos, así como la variabilidad que éste puede causar a los datos. Asimismo, se ha concientizado sobre la gran dificultad que tiene el realizar inferencia sobre los datos de extinción. En efecto, este trabajo no es definitivo y se puede mejorar en varios aspectos, como lo son: enriquecer el modelo estadístico, obtener una expresión analítica para las extinciones confirmadas y trabajar con metodología de verosimilitud convencional.

Índice general

Índice general

1. Introducción y contexto	1
1.1. Antecedentes	1
1.2. Planteamiento	4
1.3. Objetivos	7
2. Modelo estadístico	8
2.1. Componentes estocásticos	8
2.1.1. Extinción	9
2.1.2. Esfuerzo de muestreo	11
2.1.3. Último avistamiento	12
2.1.4. Confirmación	13
2.1.5. Rezago	14
2.2. Simulación	15
2.2.1. Resultados	15
2.2.2. Verosimilitud Simulada	20
3. Conclusiones	26
Bibliografía	29

Capítulo 1

Introducción y contexto

1.1. Antecedentes

En este capítulo se resumen algunos artículos centrales para la tesis en el área de biología marina. Ésto, con la finalidad de exponer diferentes opiniones acerca del aparente crecimiento de extinciones en el último siglo. Además, se presentarán conceptos importantes en el tema de extinción, los cuales son necesarios para el desarrollo ulterior de un modelo estadístico.

En 2003 se publicó un artículo por Dulvy *et al.* (2003) [9], en el cual se compila una colección de 133 casos de “extinciones” de especies marinas. Además, se abordan algunos conceptos relevantes para entender la variabilidad que existe en esos datos, tales como causas de extinción, métodos para la detección de extinción, rezago y recuperación.

En la compilación mencionada, las extinciones son consideradas en tres escalas: locales, regionales y globales, donde las escalas locales y regionales, funcionan como una advertencia para la conservación de especies. Estas escalas se definen por el espacio geográfico en el cual la población de cierta especie ha desaparecido [2]. En efecto, este concepto de escala ilustra un factor más de los que intervienen en este complejo proceso de extinción y recolección de datos.

Las causas atribuidas a los 133 casos de extinciones son explotación, pérdida de hábitat, especies invasivas, enfermedades, cambio de clima y contaminación. De las causas mencionadas, explotación y pérdida de hábitat son las que se consideran de mayor perjuicio para las especies marinas. En el caso de explotación, se toma en cuenta no solamente los excesos en la pesca de especies comerciales, sino los daños colaterales causados a especies no comerciales. Es complejo definir la causa de extinción de alguna especie, debido a que muchas de las causas se pueden confundir entre sí, incluso en

ocasiones, haber más de una causa. Evidentemente, para poder llegar a este análisis de causalidad, primero es necesario identificar las extinciones *per se*, lo cual puede llevarse a cabo de distintas maneras.

Los métodos para la detección de extinciones que se mencionan en Dulvy *et al.* (2003) son métodos directos e indirectos. Los primeros, mejor conocidos como monitores, los juzgan con grandes limitantes para detectar extinciones de una manera eficiente. Estas limitaciones se deben a la dificultad de muestrear hábitats marinos, así como, a la necesidad de muestras más grandes para especies poco comunes. Por otro lado se cuenta con métodos indirectos, divididos en comparaciones históricas y cuestionarios. La comparación histórica se basa en la comparación de dos tipos de listas. La primera, es la compilación de listas históricas de presencia o abundancia de especies y la segunda está formada por listas actuales de fauna. En ocasiones, se encuentran lugares en donde no se cuenta con dichas listas, ni la forma de compilarlas, pero sí se tiene presencia del hombre con un buen conocimiento ecológico. En estos casos, el conocimiento se plasma a través de cuestionarios cuidadosamente elaborados, con los que se puede determinar el estado de algunas especies. Para los 133 casos de extinciones, aproximadamente el 80% de éstas fueron compiladas a través de métodos indirectos, y sólo en un caso tenían datos de abundancia de población.

De la descripción de los métodos utilizados para la detección de extinciones, es evidente que no hay manera de saber el momento exacto de una extinción. Los datos llamados extinciones, en realidad son últimos avistamientos, entendiéndose como último avistamiento el registro del año en el que fue observada una especie por última vez. La razón por la que las extinciones son equiparadas con últimos avistamientos es que el año de último avistamiento se considera el más cercano al año de extinción. Cabe mencionar, que para poder considerar un último avistamiento como una extinción, es necesario un reporte de extinción, es decir, una confirmación que efectivamente dicha especie está extinta.

A partir de la necesidad de contar con una confirmación para la obtención de datos de extinción, surge el concepto de *rezago*. Dicho concepto se define como los años que toma realizar un reporte de extinción después de un último avistamiento, el cual puede ayudar a reconocer cambios en la calidad de detección de extinciones a través del tiempo. En la gráfica de rezago contra últimos avistamientos de la Figura 1¹ se puede observar que en fechas más recientes, el rezago es menor. Ésto sugiere mejor calidad en la detección de extinciones en años más recientes, lo cual suena razonable dado el crecimiento de interés que últimamente el hombre ha mostrado a involucrarse

¹Tomada de Dulvy N. et al., (2003).

en actividades de conservación de especies.

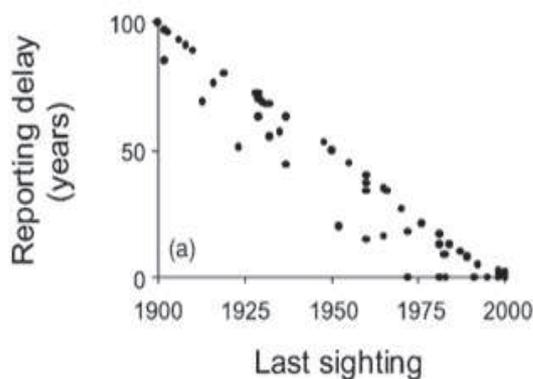


Figura 1. Gráfica de rezago.

Otro concepto relevante en el tema de extinción es el de recuperación. Este concepto se refiere a la capacidad que tiene una población para restablecerse una vez que ha sido suprimida la presión ejercida por el hombre. De lo anterior, uno podría preguntarse cuándo no es posible la recuperación de una especie, teniendo como consecuencia la extinción. En Dulvy *et al.* (2003), se argumenta que la respuesta depende de si la población se vuelve vulnerable, a cambios en el ecosistema o al efecto de la relación entre tasa de crecimiento *per capita* y el tamaño poblacional. Así pues, la recuperación puede depender de muchas variantes, de donde Dulvy opina que “aún teniendo gran abundancia de una especie, ésto no es suficiente para asegurar la recuperación de ésta”².

Una posición diferente a la recién planteada, está expresada en el artículo de Del Monte³ y Lluch (2004), en donde se afirma que existe evidencia basada en biología y genética de poblaciones, que apunta a que no hay razones evidentes para descartar que una población no pueda restablecerse. Basados en estos dos conceptos, se muestran ejemplos, donde especies tienen una muy buena recuperación, y otros donde no. En el caso donde no se tiene recuperación, no se puede asegurar que la razón principal, sea la sobreexplotación. Otro motivo plausible para la falta de recuperación, son señales térmicas a escala global que inciden en la temperatura del mar.

Cabe notar que no por desacreditar un proceso de extinción causado por el hombre, Del Monte y Lluch están justificando una negligencia hacia políticas de conservación. Por el contrario, escriben “resulta virtualmente imposible no reconocer el deterioro

²Dulvy *et al.* (2003)

³Dr. Pablo Del Monte Luna, Departamento de Pesquerías y Biología Marina, IPN, La Paz, B.C.S., 23096, México.

ambiental causado por el hombre, incluso desde tiempos geológicos, y sería éticamente incorrecto eximirlo de esa responsabilidad; pero también lo sería desconocer otras fuentes de influencia en la dinámica de las poblaciones explotadas, con tanta o mayor importancia que la humana, que pudieran servir para reducir la incertidumbre, o por lo menos conocerla, acerca de lo que sabemos sobre los sistemas naturales y planear el futuro de manera más eficiente”⁴. Así pues, la recuperación es otro factor importante que interviene en el proceso de extinción.

En otro artículo publicado por Del Monte *et al.* (2007), se realizó una revisión minuciosa de la base de datos de extinciones publicada por Dulvy *et al.* (2003). A partir de la revisión, dividen esta base de datos en dos conjuntos: sospechosos y verificados. Datos sospechosos, son aquellos últimos avistamientos de especies, que a pesar de contar con una confirmación de extinción, se poseen registros de dichas especies posteriormente a los años de confirmación. Así pues, basándose en la inclusión de datos como extinciones cuando no lo son, muestran la poca calidad de dicha compilación, concluyendo en la inconveniencia de hacer aseveraciones basándose en esta base de datos.

En resumen, existen diferentes opiniones sobre el comportamiento de las extinciones a través del tiempo, las cuales son basadas en biología marina y en registros de especies marinas. Estas opiniones dependen totalmente de la forma de interpretar dicha información, por lo que se podrían considerarlas subjetivas, ya que la misma información puede ser interpretada de formas opuestas. Así pues, sería conveniente poder identificar de manera objetiva el estado en que se encuentran las extinciones.

1.2. Planteamiento

El problema a tratar en esta tesis surge de una consulta realizada por el Dr. Del Monte. Como se ha comentado, su postura es muy clara. Su objetivo es establecer que no existe evidencia suficiente, para aseverar que las extinciones van en aumento y que su causa es debida a la sobreexplotación. Menos aún, considerando la base de datos analizada por Dulvy (2003). Esta posición se fundamenta en la revisión de dicha base de datos, donde se muestra la poca calidad de la misma.

Los únicos análisis realizados en la compilación de Dulvy (2003), con respecto a la evolución de extinciones de especies marinas han sido cualitativos. Por parte del Dr. Del Monte y su equipo, surgió el interés de analizarlos de una manera cuantitativa. En un manuscrito preliminar no publicado, ellos propusieron la creación de dos índices. El primero de ellos tiene la intención de representar una velocidad de extinciones como función del tiempo. Definen el *número promedio de extinciones por año*, denotada por NE , a través de

⁴Del Monte y Lluch (2004)

$$NE(t_i) = \frac{N(t_i) - N(t_{i-1})}{t_i - t_{i-1}}. \quad (1.1)$$

El segundo índice, denotado por RE , pretende representar la aceleración de dicha tasa de extinción. A este índice se le denomina *tasa anual de extinción* y se define por

$$RE(t_i) = \frac{\ln\left(\frac{N(t_i) - N(t_{i-1})}{N(t_{i-1}) - N(t_{i-2})}\right)}{t_i - t_{i-1}}. \quad (1.2)$$

En lo anterior, t_i es el año en que se tiene reporte de extinciones y $N(t)$ es el proceso que define el número de extinciones confirmadas hasta el tiempo t .

El índice (1.1) parece ser razonable, dado que

$$NE(t_i) = \frac{N(t_i) - N(t_{i-1})}{t_i - t_{i-1}} \approx \frac{d}{dt_i} [N(t_i)],$$

es decir, que NE es, en efecto una aproximación de la velocidad del proceso de extinciones. Por otro lado, no es razonable intentar representar la aceleración de NE a través del índice RE , dado que

$$RE(t_i) \not\approx \frac{d^2 NE(t_i)}{dt_i^2}.$$

Así pues, los análisis y conclusiones obtenidos por Del Monte y su equipo de trabajo a través de (1.2), no serán considerados.

Una vez especificados dichos índices, Del Monte *et al.* describen su relación con la variable tiempo. Lo hacen utilizando metodología de modelos lineales generalizados, sin tomar en cuenta que la estructura de datos es de series de tiempo. Se realizan ajustes de modelos haciendo comparaciones con el criterio de Akaike. Algunos de los modelos considerados, corresponden a ciertas hipótesis de trabajo que son de interés en el presente contexto. Por ejemplo, si la relación lineal tiene pendiente cero, el modelo corresponde a que no hay variabilidad en las tasas de extinción con respecto al tiempo.

La metodología basada en modelos lineales generalizados podría criticarse en varios sentidos. En primer lugar, las suposiciones usuales son cuestionables, puesto que los datos poseen una fuerte componente de temporalidad. También se pasa por alto conceptos relevantes que intervienen en la producción y obtención de datos. Como ejemplos, se equipara un dato de extinción con uno de último avistamiento, dando lugar a un concepto implícito de censura, y se ignora la posibilidad de que el registro de datos no haya sido homogéneo a lo largo de la historia. Como se verá, uno de los propósitos de esta tesis, es involucrar conceptos imperceptibles que pudieran tener influencia sobre el comportamiento temporal en los índices de extinción.

Con la finalidad de ilustrar el tipo de discusión que se suscita en el área de biología, se muestran aquí algunos resultados gráficos⁵, correspondientes a análisis realizados sobre tres conjuntos de datos: sospechosos, verificados y sospechosos más verificados.

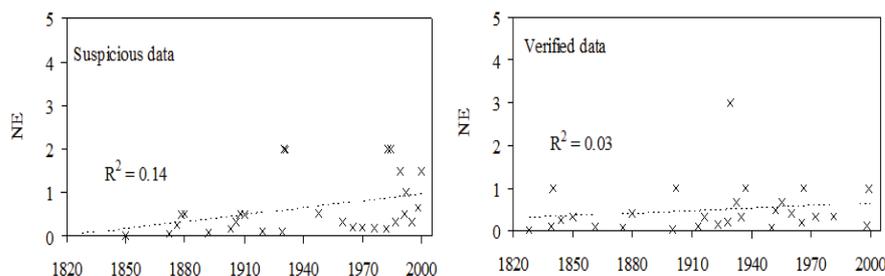


Figura 2. Gráficas de NE con datos sospechosos y verificados.

En el caso de restringir a datos sospechosos, se aprecia una tendencia creciente en el tiempo, lo cual sugiere incremento en las extinciones. Por otro lado, al usar datos verificados no se logra apreciar el mismo grado de tendencia. Lo anterior puede conducirnos a interpretar que la inclusión de datos sospechosos, da lugar a una evidencia espuria a favor de un incremento en la tasa de extinción. En efecto, al considerar la gráfica para el caso de los datos sospechosos más verificados de la Figura 3, se aprecia una tendencia creciente que pudiera ser también espuria debido al efecto de datos sospechosos. Ésta es la base del argumento de Del Monte *et al.* para establecer que la evidencia disponible no necesariamente debe interpretarse a favor de una hipótesis de incremento a tasa de extinción debido a sobreexplotación.

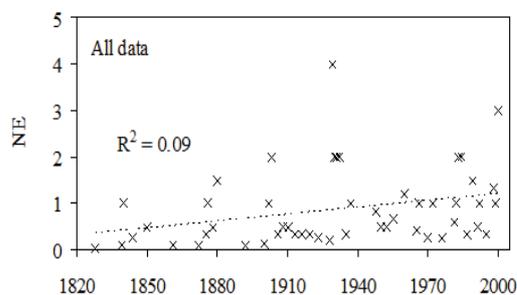


Figura3. Gráfica de NE con datos sospechosos más verificados.

⁵Figura 2 y 3, son gráficas tomadas de un manuscrito no publicado por Del Monte et al.

1.3. Objetivos

Del capítulo de introducción y contexto, se logran identificar algunas cuestiones importantes de entender, así como otras interesantes de desarrollar. En primera instancia, es necesario entender el proceso de compilación de datos, y más aún, la variabilidad que este proceso pueda tener y las posibles causas. Además, es de gran interés poder identificar el cambio del estado de las extinciones a través del tiempo de una forma adecuada. Ésto, con el fin de que las aseveraciones que se hagan en relación al estado de las extinciones, no sean basadas en opiniones subjetivas. Así pues, los objetivos de esta tesis se dividen en 5 puntos:

- Realizar una opinión y una crítica desde el punto de vista estadístico, sobre la metodología utilizada en el análisis de datos de extinciones abordada por el equipo de trabajo de Del Monte.
- Proponer un modelo estadístico, donde mediante el uso de la herramienta de simulación se pueda entender de mejor manera el fenómeno de extinción, particularmente la producción de datos.
- Justificar si el análisis de datos de extinciones a través del índice NE es sensato.
- Realizar una propuesta de metodología apropiada para analizar los datos que se tienen, donde los resultados que se obtengan reflejen de una manera certera el estado de las extinciones.
- Comunicar a los biólogos la argumentación obtenida.

Capítulo 2

Modelo estadístico

En este capítulo se presenta un modelo estadístico, que incorpora de manera encaenada componentes estocásticos importantes, para describir la producción de datos de extinciones de especies marinas. Dichos componentes serán descritos a través de diferentes modelos probabilísticos, con el fin de explorar y entender el efecto acumulado de todos ellos sobre la distribución de datos de extinción. A partir de dicho modelo integrado, considerando distintos escenarios para el fenómeno de extinción, se investigarán diferentes comportamientos de las extinciones a través del tiempo. Lo anterior se realizará mediante un proceso de experimentos de simulación. Como se verá, varios de los factores serán concebidos de manera hipotética, y no necesariamente existirán bases de datos para implementar un proceso formal de inferencia estadística. Así, el objetivo de este modelo estadístico, es simular de manera realista el complejo proceso de adquisición de datos de extinción, identificando sus componentes y sus posibles efectos.

2.1. Componentes estocásticos

A partir de los antecedentes bibliográficos —resumidos en el capítulo anterior— en torno al fenómeno de extinciones, los componentes más importantes que ya tienen un lugar preestablecido, tanto para la descripción del fenómeno como para la producción de datos, son: extinción, último avistamiento, confirmación y rezago. No se les confiere a estos elementos una naturaleza explícitamente probabilística. En efecto, una de las aportaciones importantes de este trabajo es reconocerlos como estocásticos de entrada y representarlos con modelos probabilísticos pertinentes.

Además de estos componentes ya mencionados, existe un concepto que no se había considerado previamente: la posible variación en el *esfuerzo de muestreo* de especies marinas como función del tiempo, y su influencia en la producción de datos. Por

esfuerzo de muestreo se entiende que la búsqueda histórica de especies marinas no necesariamente ha sido uniforme. Hay varias razones para establecer que en años recientes dicho esfuerzo de muestreo ha ido en aumento. Como ejemplo, considérese la Figura 4¹ donde se muestra el número de especies descritas como función del tiempo. En ella se aprecia que el ritmo de descubrimiento de nuevas especies no es uniforme a lo largo de la historia, sino que como pudiera esperarse, tiene mayor vigor en décadas recientes. Así pues, para describir el proceso de producción de datos de extinciones de especies marinas, es necesario considerar dos tipos de procesos: el proceso en el que sólo interviene la naturaleza y en el que interviene el humano. Estos dos tipos de procesos se superponen con el fin de considerar la influencia de todos aquellos factores trascendentes, lo cual es explicado a detalle más adelante.

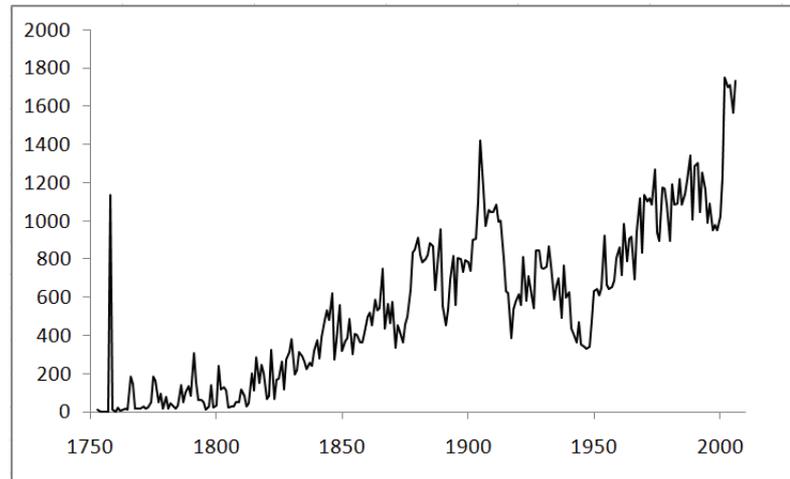


Figura 4. Gráfica de número de especies descritas por año.

2.1.1. Extinción

El tiempo de extinción de una especie marina se denotará como T , y se describirá técnicamente como un *tiempo de falla*, entendido en el contexto de análisis de supervivencia [7]. De este contexto se heredan términos técnicos que son útiles para estudiar la supervivencia. Como ejemplo, la distribución asignada a T será elegida con base en las características de la *función de riesgo*. Ésta última se define como

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)}, \quad t \geq 0,$$

donde $f(\cdot)$ y $F(\cdot)$ son, respectivamente, las funciones de densidad y distribución asociadas a $h(\cdot)$. La razón por la cual la especificación de F está basada en la función de

¹Proporcionada por Del Monte, de <http://www.marinespecies.org/aphia.php?p=stats>

riesgo, es por que ésta tiene la interpretación conveniente de ser la tasa instantánea de extinción al tiempo t , dado que la especie sobrevivió hasta el tiempo t ; es decir, refleja la vulnerabilidad de las especies como función del tiempo.

Esta manera de recurrir a conceptos de análisis de supervivencia para describir tiempos de extinción no es nueva. Dicha herramienta ha sido utilizada en [3], donde se dan argumentos ecológicos bajo ciertas suposiciones, tales como abundancia de poblaciones cerradas para mostrar que T tiene distribución aproximada inversa gaussiana. No obstante que este trabajo ilustra la pertinencia de conceptos de análisis de supervivencia para describir extinción, cabe mencionar que para los datos de especies marinas que se tienen, las suposiciones no necesariamente se cumplen. Primeramente, en esta lista no se cuenta con poblaciones cerradas. No obstante, aunque así fuera, para poder implementar lo propuesto en [3], es necesario contar con datos de series de tiempo de abundancia para cada una de las especies, con lo cual no se cuenta.

Para la especificación apropiada de una distribución de tiempo de falla, en la presente tesis se considerarán familias versátiles de modelos paramétricos. La pretensión es que puedan ser utilizadas para abarcar un rango muy amplio de situaciones interesantes en el análisis de datos de extinciones. La característica principal que interesa impartir a las grandes familias de funciones de riesgo, es que contengan como casos particulares funciones constantes, decrecientes y crecientes. Cuando la función de riesgo es constante, significa que las extinciones suceden de manera homogénea a través del tiempo. Si la función es creciente indica mayor vulnerabilidad; ésto se refleja en un incremento en el número de extinciones a través del tiempo. Por el contrario, funciones decrecientes indican menor vulnerabilidad y disminución en ritmos de extinción.

En esta tesis, por razones de flexibilidad y conveniencia, se adoptará la distribución Weibull de dos parámetros para los tiempos de extinciones de especies

$$F_T(t) = 1 - \exp \left[- (t/\eta)^\beta \right], \quad 0 \leq t < \infty, \quad \eta, \beta > 0,$$

donde η y β son respectivamente parámetros de escala y forma. Ésta distribución cumple con las características buscadas con economía de parámetros. Tiene la gran virtud de que al variar los parámetros de escala y forma, se obtiene una amplia gama de formas inducidas para la función de riesgo. En particular, se pueden obtener funciones de riesgo constantes, monótonas decrecientes y crecientes, donde cada una de estas funciones describe situaciones distintas en el contexto de extinción. La Figura 5 muestra tres ejemplos de funciones de riesgo asociadas a tiempos de falla distribuidos como Weibull.

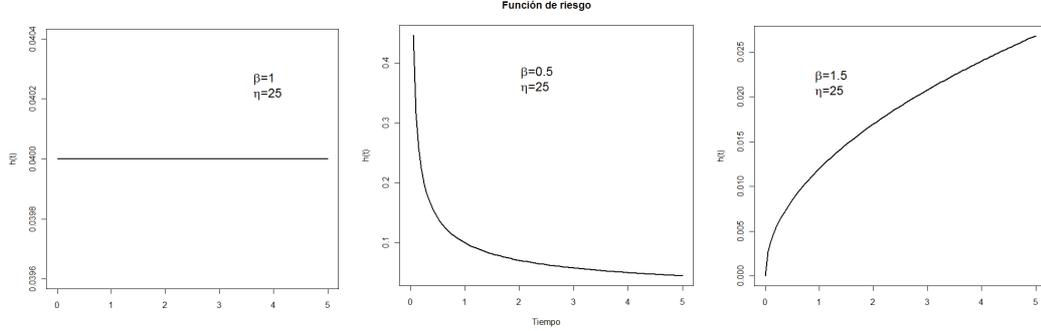


Figura 5. Ejemplos de funciones de riesgo, constante, decreciente y creciente.

2.1.2. Esfuerzo de muestreo

Como se mencionó, la razón por la cual se considera apropiado incluir este concepto, es la posibilidad de que la búsqueda de especies marinas no sea homogénea a través del tiempo. El no considerarlo, equivale a ignorar la posible confusión que esto pudiera provocar en la variabilidad de datos. Como ejemplo, el esfuerzo de muestreo pudiera tener influencia sobre los patrones observados anteriormente en las gráficas de NE (Figura 2 y 3).

El componente de esfuerzo de muestreo se modelará a través de un proceso de Poisson no homogéneo (PPNH), de nuevo por flexibilidad, conveniencia y valor ilustrativo. Ésto conlleva hacer la identificación del proceso de observación de especies con un proceso de conteo de cierta intensidad. El parámetro de este proceso de observación es la intensidad misma, y es razonable concebir que esta intensidad está relacionada con el esfuerzo de muestreo. En esta tesis no se explorará a fondo la naturaleza de dicha relación, debido a que no es muy claro como cuantificar esfuerzo de muestreo. Lo que si se realizará, es considerar una familia flexible de funciones de intensidad, para fines de explorar y entender la influencia que podría tener el esfuerzo de muestreo.

La familia de funciones de intensidad λ se representa a través de un polinomio de cuatro entradas, donde dicha intensidad al tiempo t está dada por

$$\lambda(t) = at^b + ct^d + et^f + gt^h,$$

y donde $a, b, c, d, e, f, g, h \in \mathbb{R}^+ \cup \{0\}$. Dicho polinomio se postula como una función creciente, con base a lo mencionado previamente sobre el crecimiento del esfuerzo de muestreo en años recientes.

Dos métodos para simular un PPNH, llamados de integración y rechazo, son descritos en [6]. El método que se utiliza para generar las búsquedas de especies marinas,

es el de rechazo. A continuación se presenta una pequeña descripción de este método. Supóngase que existe una constante $\bar{\lambda}$ tal que se cumple que $\lambda(t) \leq \bar{\lambda}$ para todo t . Sean $T_1^*, T_2^*, T_3^*, \dots$ tiempos sucesivos de llegada de un proceso de Poisson homogéneo con intensidad $\bar{\lambda}$. Si se acepta el i -ésimo tiempo de llegada con probabilidad $\lambda(T_i^*)/\bar{\lambda}$, independientemente de los otros tiempos de llegada, entonces la secuencia T_1, T_2, \dots de tiempos de llegada aceptados, forman una secuencia de tiempos de llegada de un PPNH con función de intensidad $\lambda(t)$.

2.1.3. Último avistamiento

Retomando lo visto en la parte de antecedentes, los datos que tradicionalmente se analizan como extinciones, son últimos avistamientos confirmados. El proceso para generar últimos avistamientos, considerando la posibilidad de que la búsqueda de especies no sea homogénea, se realiza integrando 2 componentes, extinción y esfuerzo de muestreo, el primero correspondiente a un proceso de la naturaleza y el segundo a un proceso del hombre. Cada uno de estos componentes es considerado bajo una amplia gama de situaciones, de tal forma que permite observar diferentes patrones en la producción de últimos avistamientos. A continuación se explica dicho proceso, utilizando apoyo visual de diagramas que ejemplifiquen los pasos del mismo.

1. Primero se genera un PPNH entre un límite inferior y superior. Las cruces representan el proceso humano de búsqueda de especies, es decir, los elementos generados por el PPNH.



Figura 6. Diagrama para el PPNH.

2. Después se genera un tiempo de extinción, el cual se superpone en el proceso de búsqueda de especies descrito anteriormente. Para que este tiempo de extinción sea registrado, es necesario que cumpla algunas condiciones. El tiempo de extinción puede caer en tres zonas: antes de la primera búsqueda, después del límite superior donde se genera el PPNH y entre estos dos puntos. Si el tiempo de extinción es generado antes de la primera búsqueda, es equivalente a la extinción de una especie que nunca fue descubierta, por lo que esta extinción no es tomada en cuenta. Cuando se genera después del límite superior tampoco se toma en cuenta, ya que es equivalente a una extinción que todavía no sucede. Así pues, el tiempo de extinción se registrará, si éste se encuentra entre la primera

búsqueda y el límite superior del PPNH; contextualmente representa que existe la posibilidad de “observar” dicha extinción.

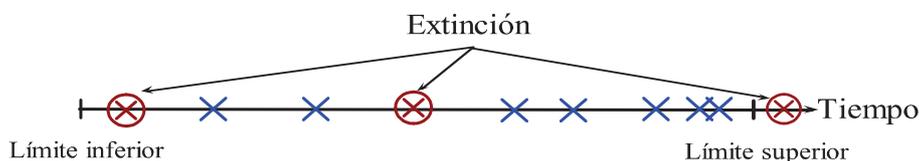


Figura 7. Diagrama que ilustra la ocurrencia de extinciones.

Cabe mencionar que aquellos elementos del PPNH que se encuentren después de la extinción generada, representan búsquedas fallidas.

- Finalmente, un último avistamiento es aquel elemento del PPNH generado, que cumpla la condición, de ser el mayor avistamiento menor que el tiempo de extinción generado.

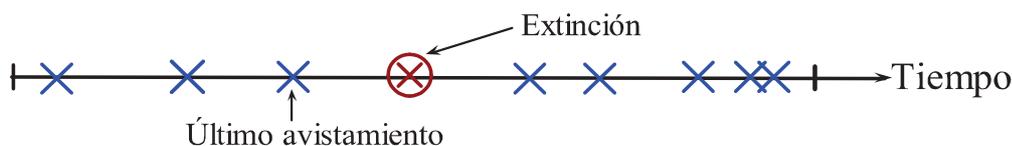


Figura 8. Diagrama del proceso para generar últimos avistamientos.

Dado un último avistamiento, es necesario identificar si éste será considerado como dato de extinción, es decir si se le asignará una confirmación de extinción. Es importante recalcar, que el componente de extinción representa al fenómeno *per se*, y datos de extinción, son aquellos últimos avistamientos que cuentan con confirmación y que se analizan con etiqueta de extinción.

2.1.4. Confirmación

La confirmación juega un rol primordial en la generación de datos, dado que ésta es necesaria para que un último avistamiento sea rotulado o no como dato de extinción. En la literatura consultada, el concepto de confirmación parece ser más cualitativo que cuantitativo. Conlleva implícitamente la noción de que en el momento de detectar un avistamiento, no es posible determinar en ese entonces si habrá sido el último o equivalentemente, saber si ese momento estará relacionado con una extinción. Como en estudios biológicos en general, la detección de un espécimen es prueba confirmatoria de su presencia, pero la no detección no implica la ausencia de la especie. En este trabajo asumiremos la postura de que el concepto de confirmación está basado en la siguiente

idea: que si no se ha colectado una especie por un periodo de tiempo largo a pesar de sí haberse hecho exploración continuada, ésto es un indicio de su extinción confirmada. La naturaleza inherentemente estocástica de este concepto de confirmación, se debe a que el proceso de exploración y detección son a su vez aleatorios.

Para implementar esta premisa de trabajo, se considerará en esta tesis el siguiente mecanismo aleatorio para generar una confirmación. Aún después de que una especie se haya extinguido, suponemos que el proceso de exploración por parte de humanos sigue realizándose bajo el mismo proceso PPNH. Así sucederán instancias de exploración en las cuales la especie evidentemente no será detectada, y si el número de estas detecciones fallidas alcanza cierto umbral, será razón para declarar una extinción confirmada. Sea k un número entero fijo. Se considerará una confirmación de una extinción al momento en que se cumplan k detecciones fallidas. Este mecanismo recoge la noción de que la extinción confirmada, tiene que ver con dejar de detectar la especie. El proceso de confirmación así propuesto, depende de la intensidad de muestreo (es decir del PPNH) y de un parámetro k que representa el grado de rigor con el que se asegura una extinción.

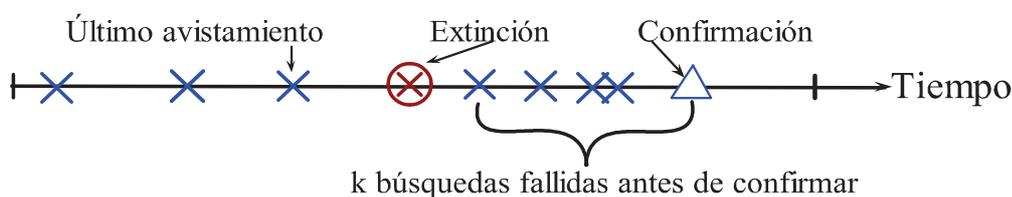


Figura 9. Diagrama para el proceso de confirmación.

2.1.5. Rezago

Recordando, el concepto de rezago se define como los años que hay entre un reporte confirmado de extinción y la fecha del último avistamiento. El mecanismo anteriormente descrito para generar confirmaciones, proporciona de manera automática el tiempo de rezago. La distribución implícita de estos tiempos de rezagos depende, como se ha descrito, de los parámetros del PPNH y de k .

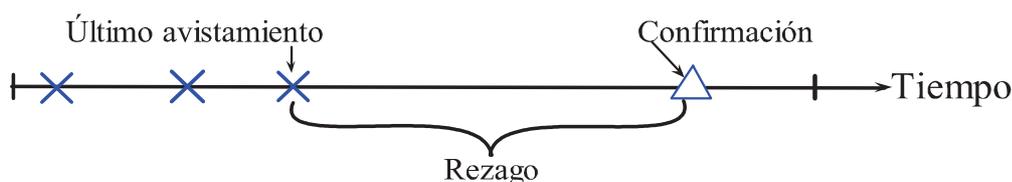


Figura 10. Diagrama del rezago.

2.2. Simulación

2.2.1. Resultados

A partir del modelo estadístico descrito en el capítulo anterior, se experimenta por medio de simulaciones, considerando distintos escenarios del fenómeno extinción. Estos escenarios son elegidos en base a la función de riesgo y el esfuerzo de muestreo, con el fin de observar de una forma clara, la influencia que tiene el esfuerzo de muestreo en los patrones de la gráfica del índice NE . Es verdad que la diversidad de escenarios posibles bajo este modelo, no es tan rico comparado con la realidad. Es decir, se tiene conciencia que el modelo propuesto seguramente no está considerando otros factores, que influyen en el fenómeno de extinción y producción de datos. A pesar de esto, se encuentran patrones de las gráficas de rezago y NE muy similares a los obtenidos con los datos originales, donde opuesto a lo que podría pensarse, la razón del patrón en la gráfica de NE no es el incremento de extinciones, sino el aumento de esfuerzo de muestreo. De modo que, se muestra que el esfuerzo de muestreo puede ser un factor de confusión en la fuente de variabilidad de los datos. Más aún lo sería si se incluyeran otros factores, que también pudieran tener influencia en la variación de la producción de datos.

A continuación se presentan algunos resultados de las simulaciones, donde la función de riesgo se supone constante, por lo que la distribución asociada es exponencial. Ésto es un caso particular de Weibull con parámetro de forma igual a 1. Para el caso del esfuerzo de muestreo se considerarán funciones de intensidad crecientes, decrecientes y constantes. Los resultados de las simulaciones se dividen en dos partes, entradas y salidas. Las entradas muestran las características bajo las cuales se realizó la simulación y las salidas las gráficas de rezago e índice NE .

1. En este primer ejemplo, se lleva a cabo la simulación con una función de intensidad creciente. La gráfica obtenida de NE en la Figura 12 tiene una tendencia creciente a través del tiempo, patrón similar al visto en las gráficas de las Figuras 2 y 3, correspondientes a los datos de extinciones. Lo anterior a pesar de no ser real un incremento, sugiere que las extinciones van en aumento. Además, el factor clave por el que se encuentra dicho patrón, es el esfuerzo de muestreo. Por otro lado, la gráfica de rezago al igual que en el caso del índice NE , tiene tendencia parecida a la obtenida con los datos de extinción (Figura 1).

ENTRADAS

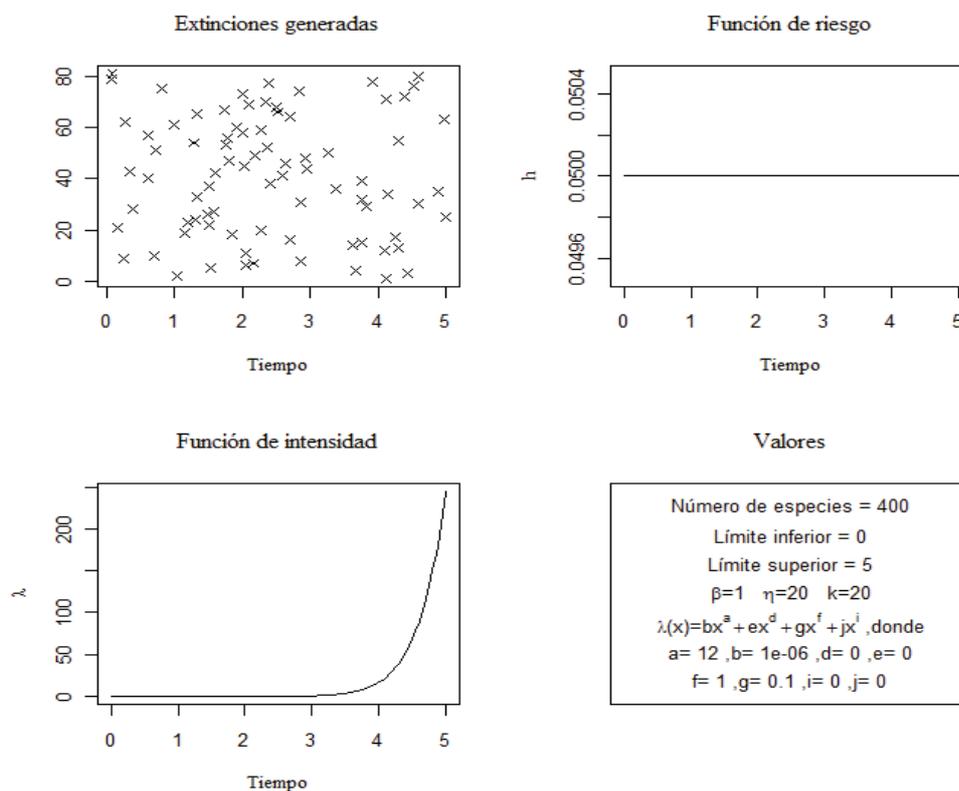


Figura 11. Entradas para la simulación con intensidad creciente

SALIDAS

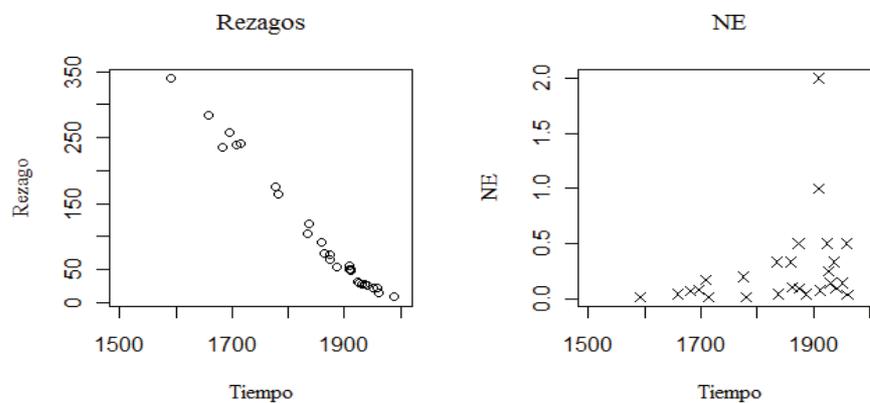


Figura 12. Gráficas de rezago y NE, bajo las condiciones de la Figura 11.

2. Al considerar en la simulación una función de intensidad decreciente, se observa de la Figura 14 en la gráfica de NE una tendencia decreciente en el tiempo. De la misma manera que el ejemplo anterior, dicha gráfica podría ser interpretada de manera incorrecta, ya que parece razonable asumir que las extinciones están disminuyendo, a pesar de no ser así. Por otro lado, en la gráfica de rezago en la Figura 14 se observa un comportamiento creciente a través del tiempo. Ésto es bastante sensato, debido a la combinación del criterio de confirmación dado y la función de intensidad decreciente. Como la intensidad de muestreo es mayor en años cercanos a 1500, se alcanza de manera más rápida el umbral para la confirmación en los años mencionados.

ENTRADAS

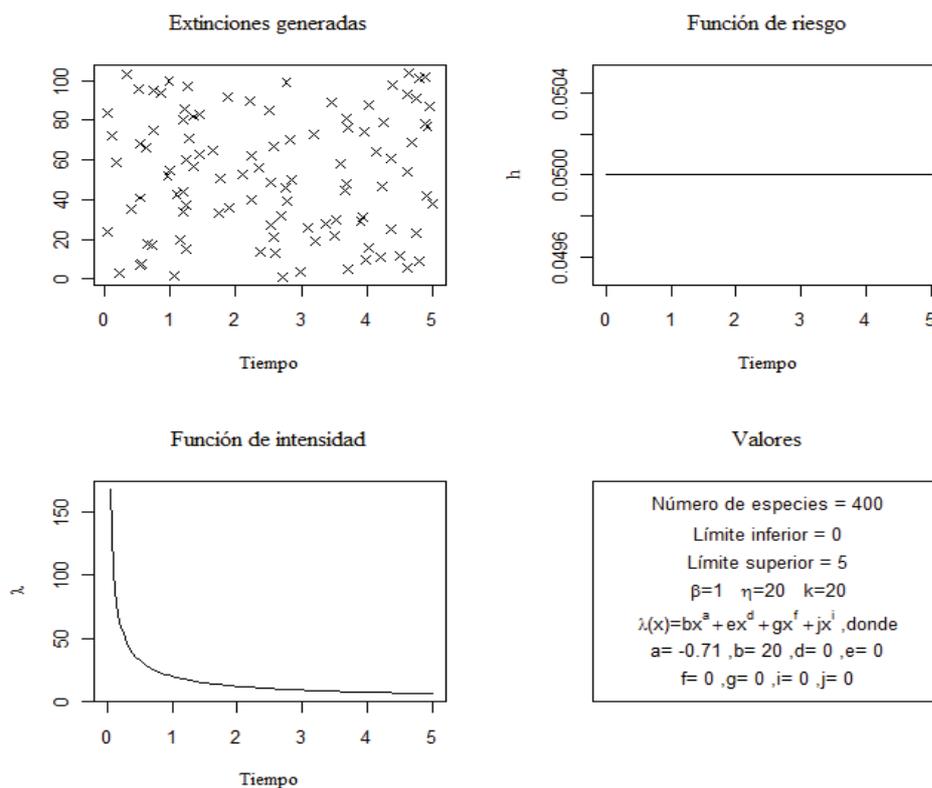


Figura 13. Entradas para la simulación con intensidad decreciente.

SALIDAS

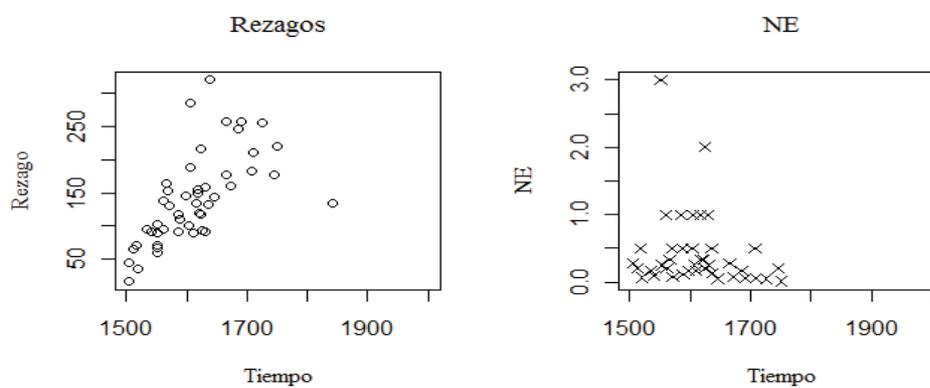


Figura 14. Gráficas de rezago y NE bajo las condiciones de la figura 10.

3. Por último, en el caso de la simulación con una función de intensidad constante, no se observa tendencia alguna en la gráfica de NE (Figura 16). Así pues, de los tres ejemplos, éste es el único caso en el cual se logra reflejar el estado con el que fueron generadas las extinciones. Asimismo, en el caso del rezago no se aprecia ninguna tendencia (Figura 16), es decir, no hay cambios significativos en los rezago a través del tiempo.

ENTRADAS

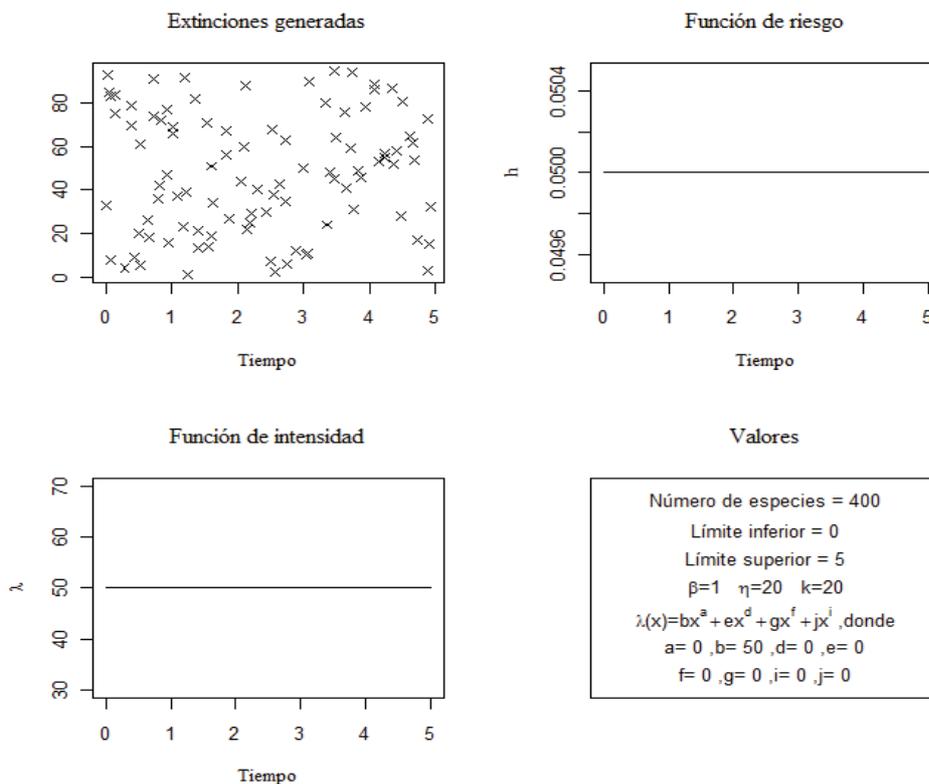


Figura 15. Entradas para la simulación con intensidad constante.

SALIDAS

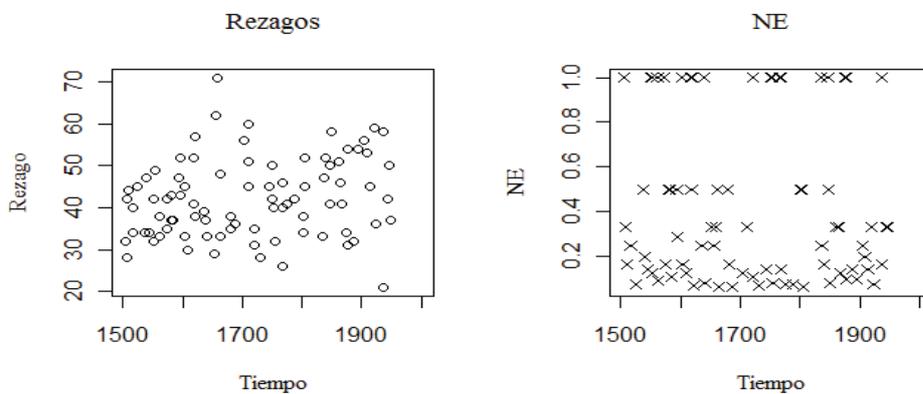


Figura 16. Gráficas de rezago y NE bajo las condiciones de la figura 14.

2.2.2. Verosimilitud Simulada

Hasta este punto, se ha logrado mostrar que analizar la base de datos de extinciones por medio del índice NE , con el fin de conocer el estado de las extinciones no es concluyente. La razón es que dicho índice puede mostrar comportamientos similares ante contextos totalmente diferentes. Una pregunta natural es ¿qué constituiría evidencia real de que ha habido una variación en tasas de extinción atribuidas a sobreexplotación? En la terminología de los elementos que se han discernido en esta tesis, esto significa realizar inferencia sobre las formas que tienen la intensidad en el proceso de exploraciones, y la función de riesgo que describe al proceso de extinción. Si se asume una parametrización de estas dos funciones como se ha descrito en el proceso de simulación anterior, esto se parafrasea como hacer inferencia sobre ciertos parámetros específicos. Así, el problema técnico estadístico se reduce a estimación y pruebas para los parámetros $a, b, c, d, e, f, g, h, \eta$ y β , con base en datos (la base de datos de extinciones). Siendo el interés primordial realizar inferencia sobre los parámetros η y β , esto plantea que los parámetros a, b, c, d, e, f, g y h son los que se conocen como parámetros de estorbo.

Este problema de inferencia podría abordarse por medio de una función de verosimilitud. Como se habrá notado los ingredientes estocásticos que hay en el proceso de simulación son muchos y muy complejos, y han sido introducidos de manera implícita. Esto es, no se conoce una forma explícita para la densidad de los datos de extinción como función de los parámetros $a, b, c, d, e, f, g, h, \eta$ y β . Sin embargo, si uno está dispuesto a suponer las formas genéricas para las funciones intensidad y riesgo sí se cuenta con un proceso de simular cualquier número de observaciones con cualquier configuración de parámetros. Lo anterior sugiere una metodología basada en el concepto de *verosimilitud simulada* para realizar inferencia. Este procedimiento permitiría realizar inferencia formal bajo un juego de suposiciones. Estas suposiciones estarán y deberán estar naturalmente sujetas a crítica; lo que se hace a continuación indica una forma de realizar inferencia suponiendo que las grandes familias de funciones hayan sido juzgadas por biólogos como sensatas y pertinentes.

La idea intuitiva de la verosimilitud simulada es aprovechar la posibilidad de obtener datos simulados arbitrariamente de una distribución para estimar con ellos una función de densidad, es decir, la verosimilitud. Esta idea de un procedimiento para aproximar la función de verosimilitud usando simulaciones se detalla en Lerman y Manski (1981). En Arias y Cox (1999) discuten la importancia de desarrollar métodos para la simulación de probabilidades en el contexto de máxima verosimilitud simulada, al igual que el rol del ruido de la simulación. En Franco y Zuluaga (2006) se muestra una aplicación de esta metodología. Se presenta una solución propuesta por Keane

y Sauer (2005), basada en verosimilitud simulada, para al problema de condiciones iniciales en la estimación de procesos estocásticos discretos.

En esta tesis la estimación de verosimilitud con base en simulaciones se realizará con estimadores de densidad kernel. Es necesario repasar ciertos elementos básicos de estimación con un kernel. Sea X_1, \dots, X_n los datos observados de una distribución F . Un kernel es definido como una función suave K tal que $K(x) \geq 0$, $\int K(x) dx = 1$, $\int xK(x) = 0$ y $\sigma_K^2 = \int x^2 K(x) dx > 0$. Dado un kernel K y un número positivo h , llamado ancho de banda, el estimador de densidad kernel está definido como

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right).$$

El ancho de banda es un parámetro de suavización, que indica cuanto contribuye cada punto muestral al estimado en el punto x . Si h es muy pequeño, el estimador de densidad kernel degenera en una colección de muchos picos. Si h es demasiado grande, entonces el estimado se sobresuaviza y se obtiene casi una distribución uniforme.

Más adelante se describe el proceso para obtener la verosimilitud aproximada bajo ciertos supuestos para los datos de extinciones confirmados. Los parámetros de la función de intensidad a, b, c, d, e, f, g y h se consideran conocidos, mientras que los correspondientes a la función de riesgo, β y η desconocidos. Tales supuestos se adoptan por practicidad y conveniencia, ya que el costo computacional crece de manera muy rápida con el aumento de número de parámetros desconocidos. Además, esencialmente el interés es hacer inferencia sobre los parámetros de la función de riesgo.

Sea $\underline{x} = (x_1, x_2, \dots, x_n)$ los datos de extinciones confirmadas, los cuales siguen una distribución conjunta implícita $G(\underline{x} : \beta, \eta)$ (desconocida). Bajo el supuesto que el modelo estadístico propuesto refleja de una manera sensata la producción de datos de extinción, las observaciones obtenidas por medio de simulación \underline{y} siguen una distribución conjunta implícita $G'(\underline{y} : \beta, \eta)$, la cual sería aproximada a $G(\underline{x} : \beta, \eta)$. Como se mencionó, es de interés encontrar una aproximación de la verosimilitud $L(\beta, \eta : \underline{x})$, denotada por $\tilde{L}(\beta, \eta : \underline{x})$. A partir de $\tilde{L}(\beta, \eta : \underline{x})$ se obtiene una aproximación a la verosimilitud perfil de β , $\tilde{L}_p(\beta; \underline{x})$.

El proceso para obtener $\tilde{L}(\beta, \eta : \underline{x})$, a partir del método de verosimilitud simulada es:

1. Sean $Q, m_1, m_2 \in \mathbb{N}$ y $\underline{y}_{(i,j)k}$ la k -ésima simulación tomando valores para β y η de β_i y η_j respectivamente, donde $i = \{1, 2, \dots, m_1\}$ y $j \in \{1, 2, \dots, m_2\}$. Dado el modelo propuesto, se realizan Q simulaciones de observaciones $\underline{y}_{(i,j)}$ bajo una distribución $G'(\bullet; \beta_i, \eta_j)$.

Simulaciones

$$\begin{aligned}
(\beta_1, \eta_1) &\longrightarrow \underline{y_{(1,1)1}}, \dots, \underline{y_{(1,1)Q}} \sim G'(\bullet; \beta_1, \eta_1), \\
&\vdots \\
(\beta_{m_1}, \eta_1) &\longrightarrow \underline{y_{(m_1,1)1}}, \dots, \underline{y_{(m_1,1)Q}} \sim G'(\bullet; \beta_{m_1}, \eta_1), \\
&\vdots \\
(\beta_1, \eta_{m_2}) &\longrightarrow \underline{y_{(1,m_2)1}}, \dots, \underline{y_{(1,m_2)Q}} \sim G'(\bullet; \beta_1, \eta_{m_2}), \\
&\vdots \\
(\beta_{m_1}, \eta_{m_2}) &\longrightarrow \underline{y_{(m_1,m_2)1}}, \dots, \underline{y_{(m_1,m_2)Q}} \sim G'(\bullet; \beta_{m_1}, \eta_{m_2}).
\end{aligned}$$

2. Dado cada uno de los conjuntos simulados bajo las mismas condiciones, se lleva a cabo estimación de densidad en cada una de las observaciones de los datos de extinción \underline{x} .

Estimación de densidad

$$\begin{aligned}
\left(\underline{y_{(1,1)1}}, \underline{y_{(1,1)2}}, \dots, \underline{y_{(1,1)Q}} \right) &\longrightarrow \hat{g}(x_1; \beta_1, \eta_1), \dots, \hat{g}(x_n; \beta_1, \eta_1), \\
&\vdots \\
\left(\underline{y_{(m_1,1)1}}, \underline{y_{(m_1,1)2}}, \dots, \underline{y_{(m_1,1)Q}} \right) &\longrightarrow \hat{g}(x_1; \beta_{m_1}, \eta_1), \dots, \hat{g}(x_n; \beta_{m_1}, \eta_1), \\
&\vdots \\
\left(\underline{y_{(1,m_2)1}}, \underline{y_{(1,m_2)2}}, \dots, \underline{y_{(1,m_2)Q}} \right) &\longrightarrow \hat{g}(x_1; \beta_1, \eta_{m_2}), \dots, \hat{g}(x_n; \beta_1, \eta_{m_2}), \\
&\vdots \\
\left(\underline{y_{(m_1,m_2)1}}, \underline{y_{(m_1,m_2)2}}, \dots, \underline{y_{(m_1,m_2)Q}} \right) &\longrightarrow \hat{g}(x_1; \beta_{m_1}, \eta_{m_2}), \dots, \hat{g}(x_n; \beta_{m_1}, \eta_{m_2}).
\end{aligned}$$

3. A partir de las densidades estimadas, se calculan los valores de la verosimilitud aproximada para cada par de valores de los parámetros β y η .

Estimación de verosimilitud

$$\begin{aligned}
 \tilde{L}(\beta_1, \eta_1) &= \prod_{i=1}^n \hat{g}(x_i; \beta_1, \eta_1), \\
 &\vdots \\
 \tilde{L}(\beta_{m_1}, \eta_1) &= \prod_{i=1}^n \hat{g}(x_i; \beta_{m_1}, \eta_1), \\
 &\vdots \\
 \tilde{L}(\beta_1, \eta_{m_2}) &= \prod_{i=1}^n \hat{g}(x_i; \beta_1, \eta_{m_2}), \\
 &\vdots \\
 \tilde{L}(\beta_{m_1}, \eta_{m_2}) &= \prod_{i=1}^n \hat{g}(x_i; \beta_{m_1}, \eta_{m_2}).
 \end{aligned}$$

Finalmente, dada la estimación de la verosimilitud, se obtiene la verosimilitud perfil de β

$$\tilde{L}_p(\beta; \underline{x}) = \max_{\eta} \tilde{L}(\beta, \eta; \underline{x})$$

con el fin de estimar β . El interés de estimar β , surge a partir que éste refleja el cambio del estado de las extinciones a través del tiempo. Si el parámetro es menor que uno, se tiene función de riesgo decreciente, si es igual a uno constante, y si es mayor que 1 creciente.

A continuación se muestra un par de ejemplos de la implementación de la metodología de verosimilitud simulada descrita. Se utilizan datos generados de un proceso de extinción cuyo comportamiento es conocido, con el objetivo de evaluar el potencial de realizar inferencia estadística. Se generan datos de extinciones confirmadas, donde los parámetros de la función de riesgo son $\beta = 1$ y $\eta = 20$. Inferir acerca del proceso de extinción se equipara a identificar razonablemente el valor verdadero del parámetro β .

En la Figura 18 se muestra la gráfica basada en el índice NE descrito en la Sección de Planteamiento. Se utiliza una rejilla con 21 valores entre 0.001 y 3 para β , otra rejilla con 21 valores entre 15 y 25 para η , $\lambda(t) = 1^{-6}t^{12} + 0.1t$ y 50 simulaciones. En ella se aprecia un comportamiento creciente a pesar de que la tasa de extinción es constante por especificación de diseño. En la Figura 17 se muestra la gráfica de la verosimilitud perfil simulada correspondiente. En esta gráfica, se observa que la forma de $\tilde{L}_p(\beta; \vec{x})$ denota una moda que se concentra alrededor de $\beta = 1$. La relevancia de esta ilustración, es que sí es posible inferir un valor desconocido de β . Ésto, con base en datos de extinción confirmada, que han sido confundidos por las nociones estocásticas

de intensidad de muestreo, rezago en el reporte y posible variación de la intensidad de extinción.

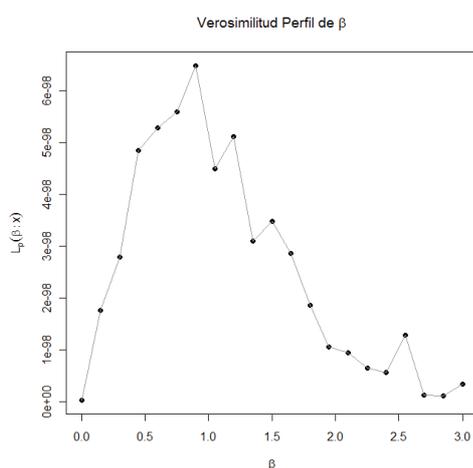


Figura 17. Verosimilitud perfil de β cuando el valor verdadero de β es 1.

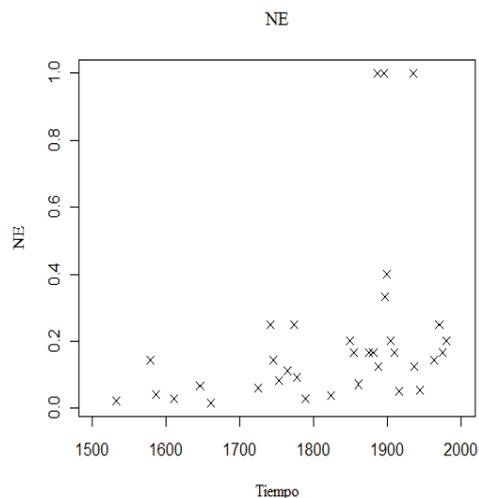


Figura 18. Gráfica de NE cuando β es 1.

Sin embargo, no todas las situaciones experimentales son tan virtuosas como el ejemplo anterior, en el que el valor $\beta = 1$ es inferido con gran precisión. En efecto, y a manera ilustrativa se presenta el caso $\beta = 2.7$, que corresponde a una situación en la cual la tasa de extinción va en franco incremento. Se considera una rejilla con 29 valores entre 0.001 y 4 para β , otra rejilla con 21 valores entre 15 y 25 para η , $\lambda(t) = 1 - 6t^{12} + 0.1t$ y 25 simulaciones. Se obtiene la gráfica de NE presentada en la Figura 20 y la verosimilitud perfil presentada en la Figura 19. A diferencia del ejemplo anterior, en este experimento hay mayor incertidumbre acerca del valor desconocido de β , representado por una verosimilitud perfil que se encuentra menos concentrada y mal posicionada respecto al valor $\beta = 2.7$. Lo anterior denota que realmente es muy complejo inferir acerca de la tasa de extinción en esta situación ficticia. Pero la importancia de la metodología propuesta, es que por lo menos sabemos acerca de esta dificultad, cosa que con la mera observación de la Figura 20 es imposible. Si una aplicación real resultara ser de esta segunda clase de ejemplo, la conclusión es que sería

inadecuado aseverar sin lugar a duda que la tasa de extinción se está incrementando.

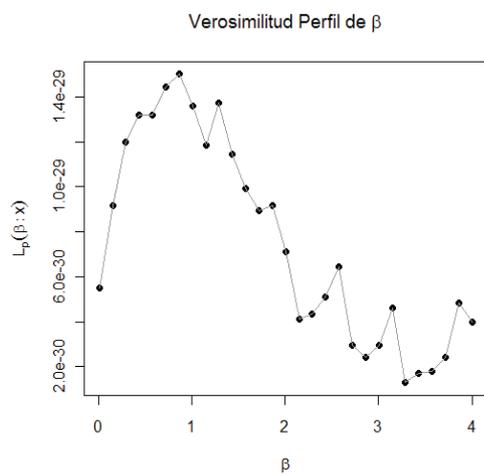


Figura 19. Verosimilitud perfil de β cuando el valor verdadero de β es 2.7.

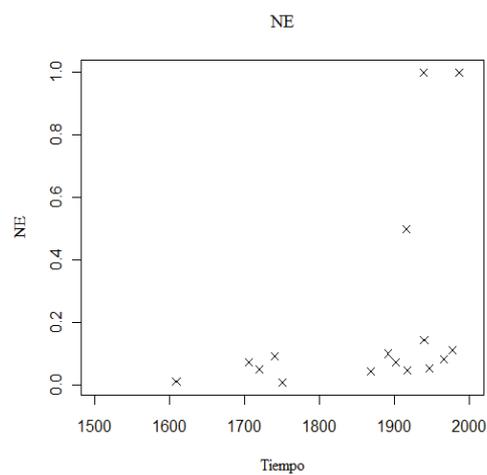


Figura 20. Gráfica de NE cuando β es 2.7

Capítulo 3

Conclusiones

Al revisar algunos artículos del gremio de biología relacionados con el tema de extinción, se observó que el proceso para la obtención de datos es sumamente complejo. Existen muchos factores que forman parte de dicho proceso—y posiblemente algunos otros—que no han sido considerados explícitamente para el análisis de datos de extinciones. En Dulvy (2003) se hacen aseveraciones del *status quo* de las extinciones basado en análisis cuantitativo, concluyendo que el aumento de extinciones a través del tiempo a causa de sobreexplotación. De lo anterior, surgió la inquietud de Del Monte y su equipo por argumentar con base en un análisis cuantitativo, que no necesariamente existe evidencia suficiente para aseverar que el crecimiento de extinciones va en aumento causado por la sobreexplotación. La forma en que abordan estos datos, está basada en metodología de modelos lineales generalizados para estudiar la relación entre una variable de respuesta y el tiempo. Esta metodología no se considera apropiada, ya que podría criticarse en varios sentidos. Primeramente, las suposiciones usuales son cuestionables, puesto que los datos sí poseen una fuerte componente de temporalidad. Además, se pasa por alto conceptos relevantes que intervienen en la producción y obtención de datos. Como ejemplos se tiene: censura, diferente calidad de información y la posibilidad de que el registro de datos no haya sido homogéneo a lo largo de la historia.

La premisa principal con la que se ha abordado este trabajo, radica en pensar el problema entre manos con nociones de razonamiento estadístico. Ésto implica dos preceptos: concebir la existencia de aleatoriedad en los mecanismos que dan lugar a los datos, e incertidumbre en la conclusión obtenida a partir del análisis de ellos. Para describir la aleatoriedad de datos, se utilizan modelos estadísticos individuales. Éstos son integrados de manera encadenada a un gran modelo estadístico el cual es utilizado para la obtención de datos. Para abordar la incertidumbre en las inferencias, se utilizaron técnicas basadas en simulación y verosimilitud.

Para poder describir el proceso de compilación de datos a través de un modelo estadístico, primero hubo la necesidad de identificar algunos componentes críticos. Como ejemplo está el esfuerzo de muestreo, donde se detectó que el no incluirlo en el análisis, podría causar confusión entre éste y la evolución real de las extinciones. Lo anterior tendría como consecuencia una dudosa interpretación de los resultados. A partir de este componente, se demuestra la necesidad de variables exógenas, para una mejor descripción de la producción y obtención de datos. Es decir, la necesidad de incluir factores que no están directamente ligadas con la extinción *per se*, pero que permitirían extraer inferencia válida acerca de ésta.

A partir del modelo estadístico propuesto, se implementó un gran proceso de simulación, con el que se puede entender de una mejor manera los componentes implícitos en el fenómeno de extinción, así como en la producción de datos. A través de tal implementación, se efectuaron diversos estudios, donde se logró obtener algunas conclusiones incluyendo una validación informal. Por esto último se entiende que a través de la regulación de parámetros y conceptos se logra reproducir un comportamiento cualitativamente similar a lo que se observó en la literatura base; no significa en este momento estimación formal ni pruebas de bondad de ajuste, debido a que no existen los datos necesarios. Posteriormente se mostró la impertinencia de la gráfica *NE*. Se encontraron casos de simulaciones con riesgo constante, en los que la gráfica *NE* muestra tendencias para las cuales parecería ser razonable concluir una tasa no constante de extinciones. Más aún, no sólo se trata de que el índice *NE* sea inadecuado, sino que en general el uso de índices para la cuantificación del estado de las extinciones no parece ser razonable. Debido a que son tantos los factores que influyen en la extinción *per se* y en la producción de datos, se considera que es muy complicado que un solo índice pueda reflejar de manera apropiada el estado de las extinciones.

La solución pertinente parecería radicar en investigar sobre cierta componente que se identifique plenamente con el proceso de extinción, en lugar de recurrir a un índice sumario. Así en esta tesis, se ha descrito la extinción con una función de riesgo, y para realizar inferencia sobre esta función, se propuso representarla paramétricamente y recurrir a metodología de verosimilitud simulada. El uso de verosimilitud simulada, ilustró que es posible realizar inferencia estadística formal y correcta respecto a la tasa de extinción. En efecto, en un ambiente de datos simulados se mostró lo que significa haber descubierto el comportamiento del fenómeno de extinción, es decir, que habrá circunstancias en las que será posible realizar inferencia precisa del parámetro de forma de la función de riesgo, mientras que en otras ocasiones la incertidumbre inductiva será de magnitud considerable. Cabe mencionar que el uso de verosimilitud simulada no es del todo práctico, en razón del gran requerimiento de computación que ello implica.

No obstante que el trabajo realizado ha resultado ser ilustrativo, se reconoce que el

modelo estadístico propuesto puede ser cuestionado en varios sentidos, al igual que el uso de verosimilitud simulada. Muchas de las formulaciones sugeridas y suposiciones, como las formas específicas de la función de riesgo y la de la intensidad de muestreo, obedecieron a practicidad y conveniencia en lugar de fidelidad. Como ejemplos, se supuso homogeneidad por especies en la función de riesgo, en el PPNH y en el criterio de confirmación. Así pues, a partir de estas debilidades surgen algunas de las consideraciones para trabajo a futuro:

- a) Enriquecer el modelo estadístico, lo cual incluye retroalimentación con biólogos y agregar otros posibles componentes de confusión.
- b) Obtener analíticamente una expresión para la distribución de fechas de últimos avistamientos. Este punto se vislumbra viable utilizando propiedades analíticas de procesos de Poisson, y haciendo uso de suposiciones simplificadoras, tales como independencia entre el proceso de extinción y el de muestreo por parte de humanos.
- c) Proceder con metodología basada en verosimilitud convencional.
- c) Lograr mayor eficiencia en el cómputo.

Bibliografía

- [1] Arias, C., and Cox, T.L. (1999), “Maximum Simulated Likelihood: A Brief Introduction for Practitioners”, *Agricultural & Applied economics*, 421.
- [2] Carlton, J.T., Geller, J.B, Reaka-Kudla, M.L. and Norse, E.A. (1999), “Historical Extinctions in the Sea”, *Annual Review of Ecology and Systematics*, 30, 515-538.
- [3] Dennis, B., Munholland, P.L. and Scott, J.M. (1991), “Estimation of growth and extinction parameters for endangered species”, *Ecological Monographs*, 6(12), 115-143.
- [4] Franco, L.C. y Zuluaga, F.I. (2006), “El problema de condiciones iniciales en la estimación de procesos estocásticos discretos: algunos elementos teóricos”, *Revista de Ingenierías Universidad de Medellín*, 5(008), 35-41.
- [5] Geyer, C.J., y Thompson, E.A. (1992), “Constrained Monte Carlo Maximum Likelihood for Dependent Data”, *Journal of the Royal Statistical Society*, 54, 3, 657-699.
- [6] Burnecki, K., Härdle, W. and Weron, R., “An Introduction to Simulation of Risk Processes”, *Hugo Steinhaus Center for Stochastic Methods, Wroclaw*.
- [7] Lawless, J.F. (1982), *Statistical Model and Methods for Lifetime Data*, John Wiley & Sons, New York.
- [8] Lerman, S.R. and Manski, C.F. (1981), “On the Use of Simulated Frequencies to Approximate Choice Probabilities”. In: *Structural Analysis of Discrete Data*, edited by Manski C.F. and McFadden, D., Cambridge: The MIT press, 305-319.
- [9] Dulvy, N. *et al.* (2003), “Extinction vulnerability in marine populations”, *Fish and Fisheries*, 4, 25-64.
- [10] Del Monte, P. *et al.* (2007), “Marine extinctions revisited”, *Fish and Fisheries*, 8, 107-122.
- [11] Del Monte, P. y Lluch, D. (2004), “Reflexiones sobre la explotación, recuperación y extinción de poblaciones naturales”, *Interciencia*, 29, 003, 158-162.
- [12] <http://www.marinespecies.org/aphia.php?p=stats>