

Centro de Investigación en Matemáticas A.C.

Modelación del Tamaño de Nidada
del Loro Corona Lila

tesis

que para obtener el grado de

Maestro en Ciencias
con especialidad en Probabilidad y Estadística

presenta:

Martín Alberto Basurto Arriaga

Dirigida por:

Dr. Miguel Nakamura Savoy

Guanajuato, Gto.

Índice general

Agradecimientos	1
1. Introducción	3
1.1. Antecedentes	4
1.2. El Loro Corona Lila y el Tamaño de Nidada	5
1.2.1. Teoría del Depredador	6
1.2.2. Teoría de la Latitud	7
1.2.3. Teoría de la Longevidad	9
1.2.4. Teoría de Costo-Beneficio	9
1.3. Objetivos	11
2. Modelación del Tamaño de Nidada del Loro Corona Lila	13
2.1. Premisas	13
2.2. Modelo General	15
2.2.1. Covariables	17
2.3. Modelo Adaptado al Loro Corona Lila en la región de Chamela	18
2.3.1. Densidad de Energía	19
2.3.2. Covariables Incorporadas	20
2.3.2.1. Índice de Lang	20
2.3.2.2. Temperatura Máxima	21
2.3.2.3. ENSO	22
3. Inferencia Estadística e Implementación	25
3.1. Estimación de Parámetros	26
3.2. Región de Incertidumbre	27
3.3. Implementación al Loro Corona Lila en la Región de Chamela	30
3.4. Resultados e Interpretación	33
3.4.1. Interpretación de los Parámetros	34

3.4.2. Estimación de la Media del Tamaño de Nidada	35
3.4.3. Región de Incertidumbre para la Media Estimada y los Parámetros .	36
3.4.4. Discusión de Años Atípicos	40
3.5. Resumen	41
3.6. Conclusiones	42
A. Apéndice 1	47
Bibliografía	49

Agradecimientos

Quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo otorgado para la realización de los estudios de la Maestría en Ciencias con especialidad en Probabilidad y Estadística que realice en el Centro de Investigación en Matemáticas A. C. (CIMAT). También debo agradecer a mis profesores por compartir sus conocimientos y contribuir en mi formación profesional. A mis compañeros y amigos que compartieron grandes momentos en estos años como estudiante en CIMAT. Oscar. Erasmo. Alberto. Cristina. Flor. Gerardo. Filo. Manuel.

Para la realización de esta tesis contamos con el apoyo del CONACYT que apporto los recursos necesarios para el proyecto “Análisis de la Vulnerabilidad del Socio-Ecosistema del Bosque Tropical Seco al Cambio Global en la Región de Chamela, Jalisco (CONACYT: 83441). En especial, agradezco a la Dra. Katherine Renton por la recolección de los datos utilizados en esta tesis y que amablemente compartió con nosotros. Del mismo modo, por el planteamiento del problema en el que se basa este trabajo y por la discusión sobre el tema, que en su carácter de especialista, apporto para el análisis del mismo.

Agradezco al Dr. Jorge Argáez y al Dr. José Montoya, por sus comentarios y observaciones a esta tesis; así como por su participación como miembros del jurado evaluador de la misma. Mención aparte merece mi agradecimiento al Dr. Miguel Nakamura por su dirección en este trabajo; sus ideas, recomendaciones y metodologías son parte esencial en esta tesis. Todavía más importante fueron su apoyo y paciencia. Debo mencionar que ha compartido conmigo conocimientos y consejos que van más allá de los que se tratan en la aulas; ha sido un placer compartir puntos de vista y opiniones a lo largo de los años recientes. Es ingenuo tratar de describir en un párrafo tantas razones para agradecer, por eso lo resumo diciendo: ¡Gracias Miguel!

Finalmente, tengo que agradecer a mi padre quien siempre a confiado en mí y me a apoyado en cada paso, sin ello nada sería igual. Agradezco a Liz por su compañía, amor y cariño durante el tiempo que compartimos. A mi familia por estar a mi lado, a mi abuela por estar ahí cada que necesito un hogar. Es difícil mencionar a todas y cada una de las personas que han sido parte importante en esta etapa de mi vida, pero agradezco profundamente su contribución a mi crecimiento personal.

Capítulo 1

Introducción

En este trabajo presentamos un modelo estocástico que permite estimar y explicar el tamaño de nidada del Loro Corona Lila (o simplemente el Loro). El modelo que proponemos, más que predecir el número de huevos por sí mismo, busca ser un modelo basado en argumentos biológicos que describa el mecanismo mediante el cuál el ave dispone el número de huevos puestos en cada nidada. La finalidad es determinar cómo influyen ciertos factores (covariables) en el tamaño de nidada del Loro.

El modelo propuesto será justificado paso a paso con sencillos argumentos que nos ayuden a entender su génesis. Del mismo modo, explicaremos cuáles son las covariables que incluimos como información adicional y cómo influyen en el comportamiento del Loro. Finalmente, mediante técnicas estadísticas, veremos cómo los datos observados del Loro se ajustan al modelo y obtendremos resultados e interpretaciones para esta especie en particular.

La importancia de estudiar el tamaño de nidada para los biólogos radica en que éste es normalmente el parámetro reproductivo más fácil de estimar y el mejor documentado en la bibliografía. El tamaño de nidada está intrínsecamente ligado al éxito reproductivo de una especie y de esto depende la continuidad de la especie. El número de nuevas crías está determinado por el número de huevos puestos y por el porcentaje de huevos que produzcan un polluelo. Por esta razón, el tamaño de nidada es un indicador inmediato del estado de la especie en relación a su evolución, ecología, ambiente y conducta. A modo de ejemplo, si el promedio de huevos puestos por nidada cae por debajo de cierto umbral (éste depende de cada especie) la continuidad de la especie estaría en riesgo en el corto plazo; esto se deduce de los modelos de poblaciones que estudian la supervivencia de las especies. Por otro lado, un aumento inusual en el tamaño promedio de nidada, podría ser una señal de alta depredación de la especie (generalmente, un aumento en la caza furtiva) o alguna otra causa

de la disminución de la población. El aumento en el número de huevos a su vez compromete la productividad futura al reducir la expectativa de vida de los padres, pues más huevos implica más esfuerzo. Todo esto pone de manifiesto la relevancia de medir el tamaño de nidada del Loro.

En este Capítulo 1 revisaremos los antecedentes de donde surge el interés por este problema, con lo cual buscamos poner en contexto al lector. Para ello haremos una breve revisión bibliográfica de algunas de las teorías sobre tamaño de nidada en aves. Terminaremos el capítulo describiendo los objetivos planteados en este trabajo.

1.1. Antecedentes

Chamela-Cuixmala es una reserva de la biosfera ubicada en la costa de Jalisco, entre Manzanillo y Puerto Vallarta. La reserva ha sido creada, esencialmente, para proteger y estudiar al bosque tropical seco o caducifolio de la región. Esta reserva se caracteriza por su variada y abundante fauna, además por mantener extensiones considerables de bosque tropical caducifolio con poca perturbación.¹

Desde 1971 la Estación de Biología Chamela de la UNAM, ha tenido como finalidad la protección de este sistema socio-ecológico, así como la generación de conocimientos científicos relacionados a éste. En tiempos recientes, a través de la “Red Mexicana de Investigación Ecológica de Largo Plazo (Red Mex-LTER)”, se ha creado un grupo de académicos investigadores denominado “Sitio Chamela de la Mex-LTER”. Este grupo interdisciplinario actualmente trabaja en el proyecto “Análisis de la Vulnerabilidad del Socio-Ecosistema de Bosque Tropical Seco al Cambio Global en la Región de Chamela, Jalisco”. El proyecto tiene por objetivo estudiar los efectos causados por los factores de impacto local como: el crecimiento poblacional, el cambio de uso de suelo, la contaminación y la alteración en los ciclos hidrológicos y biogeoquímicos. Y también por los provocados por los factores de impacto mundial como el calentamiento global y el adelgazamiento de la capa de ozono, sobre la estructura y funcionamiento del sistema socio-ecológico y su vulnerabilidad. El grupo de investigadores tiene por objetivo demostrar la importancia que tiene esta reserva de biosfera, entender los mecanismos que lo rigen y justificar su conservación.

¹Caducifolio: árbol o arbusto que pierde las hojas durante una época del año, conocida como época desfavorable.

La Dra. Katherine Renton, bióloga miembro del grupo, realiza estudios sobre la importancia de las aves en las comunidades tropicales y la aplicación de esquemas para su conservación. En particular, estudia la ecología y dinámica poblacional, la interacción planta-animal y la dinámica espacio-temporal de aves psitácidas dispersoras y depredadoras de frutos y semillas en el bosque tropical seco. Entre las aves ampliamente estudiadas por la Dra. Renton se encuentra el Loro Corona Lila. El trabajo aquí presentado surge de la colaboración entre los académicos de la Estación de Biología Chamela y el Dr. Miguel Nakamura del CIMAT y pretende sumarse a los esfuerzos realizados durante un largo tiempo para estudiar este Loro. Cabe mencionar que para este trabajo utilizamos la base de datos recabada durante 16 años por la Dra. Renton; estos datos corresponden a la observación directa de las nidadas del Loro en la región de Chamela.²

1.2. El Loro Corona Lila y el Tamaño de Nidada

El Loro Corona Lila (*Amazona finschi*) es un ave endémica mexicana, actualmente es una especie amenazada y su clasificación es: vulnerable. La especie ha desaparecido del 37% de las localidades donde se le registró históricamente y la población se calculaba entre 7,000 y 10,000 miembros en 2007 (Renton e Iñigo-Elías (2000)). Estos datos en sí mismos justifican la implementación de un esquema de conservación para el Loro. Adicionalmente, el Loro pertenece a un grupo de aves que desempeña un rol importante tanto en la depredación como en la diseminación de la vegetación, debido a sus hábitos alimenticios basados en frutos y granos.³ Esta última función está intrínsecamente relacionada con la *resistencia* y *resiliencia* del sistema socio-ecológico.⁴

Por tamaño de nidada nos referimos al número de huevos depositados en una sola temporada de puesta. Dentro de literatura de biología y ecología encontramos diversas teorías que tratan de explicar el tamaño de nidada de las aves. Si deseamos hacer modelación estadística y sustentar los resultados desde el punto de vista teórico es necesario conocer estas teorías, al menos *grosso modo*. Desde nuestra perspectiva hay dos cosas que serán importantes en

²Más información sobre el trabajo de la Dra. Renton puede ser consultada en http://www.ibiologia.unam.mx/directorio/r/renton%20_katherine.htm

³El Loro Corona Lila posee plumaje verde brillante, su frente y lores son rojos, además la corona y lados del cuello son de color lila (de ahí el nombre). Sus ojos son color ámbar, el anillo orbital y las patas son grises. Ambos sexos y los juveniles son similares.

⁴La resistencia de los sistemas ecológicos es su capacidad para absorber las perturbaciones y mantener sus interacciones y sus funciones, mientras que la resiliencia es la capacidad que tiene el sistema para regresar a su estado original — en términos de estructura y función ecológica — después de ser impactado por una perturbación.

el planteamiento de un modelo; en primer lugar está identificar un mecanismo o proceso de la puesta de huevos y en segundo lugar, identificar los factores (covariables) endógenos y exógenos que intervienen en él. En este caso, por tratarse de una sola especie de ave, nos enfocaremos a identificar las covariables exógenas (por ejemplo de tipo ambiental), ya que podemos considerar que las covariables endógenas (o propias de la especie) se mantienen constantes del mismo modo que lo hacen las covariables geográficas, debido a que estamos estudiando una sola especie en un solo lugar.⁵

A continuación revisaremos brevemente algunas de las teorías que intentan explicar el por qué varía el tamaño de nidada en las diversas especies de aves en el mundo. En general, existe una gran variedad de teorías pero entendemos que las aquí expuestas son las más aceptadas entre los biólogos; además, sintetizan las principales ideas sobre el tema.

1.2.1. Teoría del Depredador

La reducción del tamaño de nidada debido a un alto riesgo por depredación es una hipótesis bastante aceptada por dos razones. En primer lugar, porque la depredación de nidos se incrementa con el número de huevos puestos por nido, con lo que, pequeñas nidadas acortan el tiempo que las crías se quedan en su nido y se reduce el número de visitas de los padres al nido, las cuales atraen más depredadores. En segundo lugar, la supervivencia de los padres se reduce al aumentar el tamaño de la nidada, por lo que reducir el tamaño de la nidada mejora las perspectivas de supervivencia de los padres y la reproducción futura.

En un experimento realizado por Eggers et al. (2006), donde manipula la percepción del riesgo de depredación de nidos de Azulejos Siberianos (*Perisoreus infaustus*) mediante la reproducción de una mezcla de sonidos emitidos por los depredadores en las cercanías de los sitios de anidación, se observó que en respuesta a la exposición a esta señal acústica que simula un mayor riesgo de depredación de nidos, los Azulejos Siberianos eligen un nido que ofrezca más protección y reducen el tamaño de la nidada. Esta es la primera demostración experimental de ajuste de la puesta y la selección de tamaño de nido como resultado de la plasticidad fenotípica que refleja una respuesta a la percepción del riesgo de depredación de nidos; sin embargo, las investigaciones no establecen una relación causal entre depredación y tamaño de nidada.

⁵En la literatura en inglés el término usado para tamaño de nidada es *clutch size*.

1.2.2. Teoría de la Latitud

Las aves en las zonas tropicales generalmente ponen la mitad de huevos que los de las zonas del norte. Lack (1947) propuso que la variación latitudinal en el tamaño de las nidadas es resultado de las diferencias en la disponibilidad de alimentos. Esta limitación en los alimentos plantea la hipótesis de que el tamaño de la nidada se determina por el suministro de alimentos, con una selección natural a favor de nidadas con un número de crías viable para que el éxito de los padres pueda aumentar.

El tamaño de las aves que ocupan entornos no estacionales, es decir, donde hay poca variación estacional de la temperatura, es más pequeño que el de las aves que habitan en ambientes donde las temperaturas son más variables. Ambientes altamente estacionales pueden causar aumento de la mortalidad adulta (Ricklefs y Wikelski (2002)) ya que las aves deben sobrevivir a los períodos de temperaturas más bajas y a la disponibilidad de alimentos reducida o porque las aves deben asumir los riesgos asociados con la migración (Jetz et al. (2008)). Sin embargo, los entornos estacionales también tienen períodos en los que la disponibilidad de alimentos es muy alta. Esta combinación de abundancia de alimentos durante la temporada de reproducción y un aumento de la mortalidad durante el período no reproductivo ha favorecido la evolución de las grandes nidadas en los ambientes estacionales. Debido a la mayor disponibilidad de recursos, mayores nidadas pueden producirse y lograr más crías. Producir más crías aumenta la probabilidad de que alguna sobreviva al período de mayor mortalidad, cuando hay una reducción en la disponibilidad de alimentos. Cuanto mayor es el grado de estacionalidad o mayor es la fluctuación en la disponibilidad de recursos, mayor es la tendencia a grandes puestas.

La estacionalidad aumenta con el incremento de latitud al igual que los tamaños de nidada de las aves. Sin embargo, la estacionalidad también varía en algunos casos con la longitud. Por ejemplo, Europa occidental es menos estacional que Europa oriental y, como era de esperarse dado el efecto de la estacionalidad en la evolución del tamaño de nidada, las aves de Europa Oriental tienden a tener puestas más grandes que las aves en Europa occidental (Bell (1996)). La Figura 1.2.1 extraída de Jetz et al. (2008) muestra los patrones geográficos del tamaño de nidada de 5,290 especies en el mundo; este mapa ilustra el planteamiento que hay detrás la teoría de la latitud. La Figura 1.2.2 es un mapa global de lluvia media total; la similitud entre los patrones de lluvia y el tamaño de nidada se hace evidente.

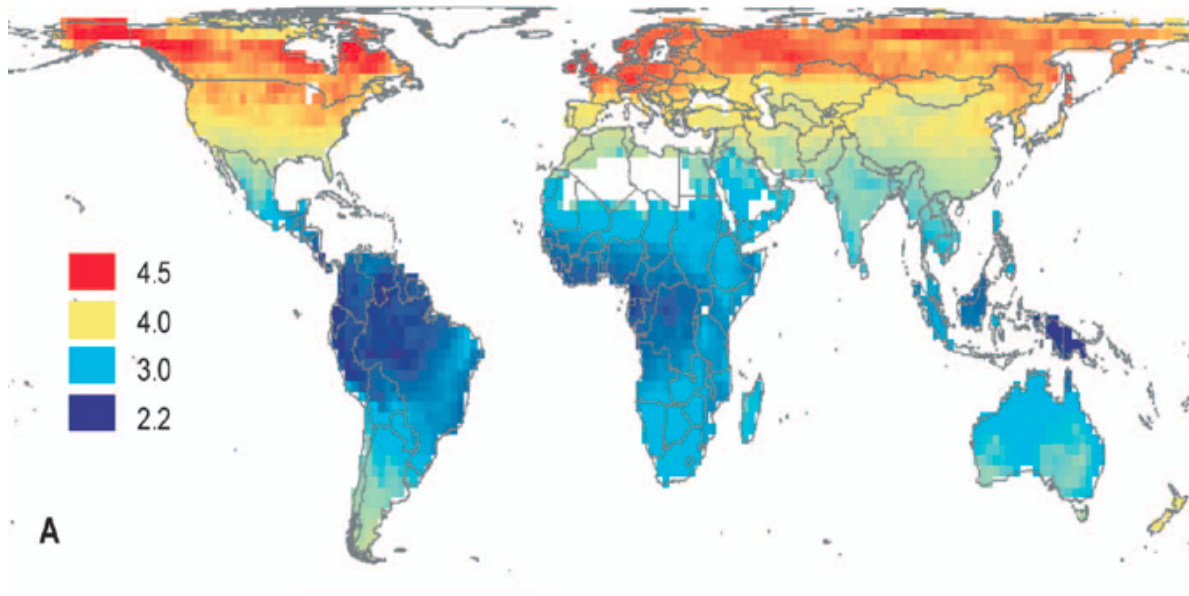


Figura 1.2.1: Mapa global de media del tamaño de nidada para 5,290 especies de aves (Jetz et al. 2008).

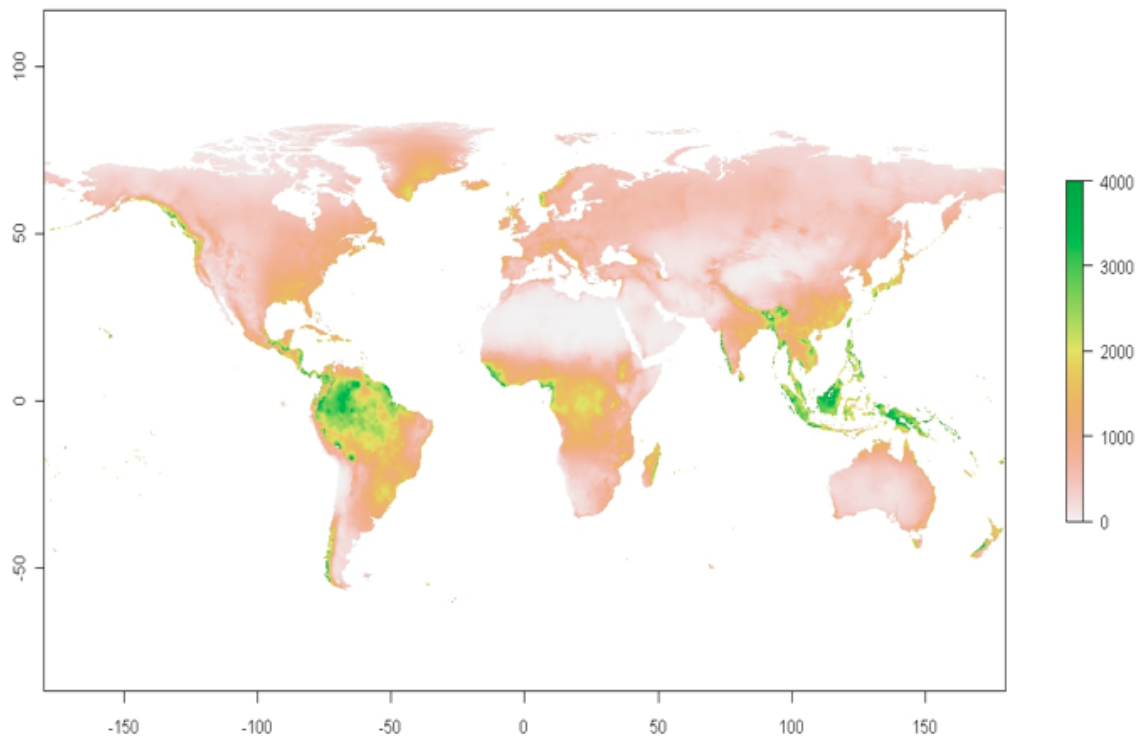


Figura 1.2.2: Mapa global de lluvia media total (medida en mm).

1.2.3. Teoría de la Longevidad

Esta teoría plantea que el éxito reproductivo en muchas especies depende de la edad. Tres hipótesis se han planteado para explicar este fenómeno: la hipótesis basada en la experiencia específica de la edad reproductiva, la hipótesis del esfuerzo reproductivo en función de la edad específica y la hipótesis basada en la selección natural, es decir, la desaparición progresiva de los fenotipos debido a la variación de la productividad individual y de la supervivencia.

Mauck et al. (2004) utilizaron datos de un estudio a largo plazo de Petreles de Leach (*Oceanodroma leucorhoa*) para presentar una simple prueba de que las predicciones son mutuamente excluyentes acerca de la relación entre el éxito de la reproducción temprana y la longevidad. No debería haber ninguna correlación entre el éxito reproductivo y la longevidad en las primeras puestas de un ave bajo la hipótesis de la experiencia. Pero sí habría una correlación negativa con la hipótesis de esfuerzo y una correlación positiva con la hipótesis de selección. Los autores encontraron una relación positiva significativa entre el éxito en los dos primeros intentos de puesta y la longevidad en esta población de aves marinas de larga vida. El estudio también sugiere que las crías de padres con baja productividad también fueron menos propensas a sobrevivir. Estos datos proporcionan uno de los más fuertes soportes hasta la fecha para la hipótesis de la selección natural.

1.2.4. Teoría de Costo-Beneficio

La hipótesis de costo-beneficio también se conoce como la hipótesis de asignación de recursos. El esfuerzo reproductivo se basa en intereses de reproducción a largo plazo. Un número grande de crías pueden requerir una gran inversión y un mayor riesgo en la supervivencia de los padres y por tanto, la reproducción en el futuro puede verse afectada. El incremento en la producción de huevos reduce la fortaleza física futura de las aves. Las mediciones de los costos de reproducción son esenciales para la comprensión de la evolución del esfuerzo reproductivo. Mientras que los efectos del aumento en la esfuerzo por la crianza de polluelos, la posterior supervivencia y la fecundidad han sido relativamente bien estudiado experimentalmente, los costos asociados con el aumento en el esfuerzo de la producción de huevos han recibido menos atención. En Nager et al. (2001), experimentalmente se incremento el esfuerzo en la producción de huevos de forma individual en gaviotas marcadas y se siguió su desempeño reproductivo en el año posterior. En la temporada siguiente a la producción de huevos aumentada, las hembras, pero no los machos, fueron menos propensas a ser vistas en la parcela de estudio y las aves que volvieron eran menos propensas a producir una nidada en comparación con las aves control. Además, las hembras del experimento invirtieron menos

en la producción de huevos en la temporada siguiente. Por lo tanto, este estudio proporciona evidencia de que hay una relación costo-beneficio entre el actual esfuerzo de la producción de huevos y de la condición física futura de las aves. Esta teoría propone que el fenómeno de puesta de huevos, busca optimizar el número de crías a lo largo de toda la vida reproductiva de la hembra y no sólo en una temporada de puesta.

Tiempo, energía y otros recursos son limitados y deben ser bien administrados. Las posibilidades de sobrevivir a los distintos riesgos asociados con la reproducción probablemente disminuirán a medida que aumenta la inversión parental, porque el tiempo y los recursos que podrían destinarse al mantenimiento de los padres y evitar la depredación se dedicó a la descendencia. Debido a que la selección natural favorece a los individuos que optimizan el éxito reproductivo (y no el éxito reproductivo durante un intento de reproducción simple o época de cría), la inversión óptima en un intento de reproducción particular, está influenciada por la duración de la vida. En otras palabras, los individuos de las especies con una menor esperanza de vida puede maximizar su tiempo de vida, invirtiendo más en la reproducción (por ejemplo, grandes puestas); individuos de especies con vidas más largas mejoran su condición física invirtiendo menos en la reproducción. Todo se resume en una negociación de costo-beneficio que optimice y garantice la continuidad de la especie y por lo tanto las aves se adaptan a su longevidad de vida, recursos y medio ambiente para lograr dicho objetivo.

Las teorías que acabamos de describir dan lugar a diversos modelos matemáticos que tratan de estimar el tamaño de nidada. No obstante, estos modelos, por lo general, tratan de explicar la variación del tamaño de nidada como efecto de la variación latitudinal y consideran más de una especie en el análisis (Jetz (2008)). O bien, consideran variables que no tenemos a nuestra disposición como tasa de metabolismo del ave, el peso del ave y el peso de los polluelos, la temperatura del aire entre otras, como en Royama (1969); además de consistir en el análisis de una sola especie en diferentes localidades. Por este motivo sigue siendo nuestro objetivo proponer un modelo para estudiar el tamaño de nidada del Loro Corona Lila en la región de Chamela. Debido a que estamos interesados en estudiar la variabilidad del número de huevos por nido de una sola especie y de una sola región, es decir, no hay variación latitudinal, ni de especie; más bien tenemos variación de individuos de la misma especie. El modelo específico que presentaremos en el siguiente capítulo estudia la variación del tamaño de nidada a través de los años en una misma especie de la misma región.

1.3. Objetivos

El objetivo en este trabajo es dar respuesta a las preguntas ¿Cómo determinan las aves, específicamente el Loro, el número de huevos que pone en cada nidada?, ¿Cuántos huevos pondrán? y ¿Qué factores intervienen en esa determinación? Para ello, hemos propuesto un modelo estocástico para el número de huevos por puesta del Loro. Como ya hemos dicho, buscamos un modelo que además de predecir el número de huevos, nos ayude a entender cómo se determina el número de huevos a poner y cuáles son los factores que participan en dicha decisión. En este sentido, el modelo es tanto cuantitativo como cualitativo. Dentro de las génesis usuales no hay una que de manera natural se adapte al mecanismo de puesta de huevos. En consecuencia, empleamos este modelo *ad hoc* con interpretación en términos biológicos. Puesto que queremos justificar el modelo desde el punto de vista de la biología, lo primero que tenemos que hacer es revisar la teoría que sobre el tema hay; esto ya lo hemos hecho previamente.

El siguiente paso será proponer un modelo basado en las teorías revisadas, para lo cual necesitamos traducir la ideas tras estas teorías a un lenguaje matemático y crear un modelo conceptual o general como el presentado en el Capítulo 2. Explicaremos con detalle la elección del modelo y las ideas detrás de éste, para así comprender su génesis. Después discutiremos sobre las covariables que aportan información al modelo y la forma de integrarlas como parámetros del modelo. Con esto, aprovecharemos los datos de conteos de huevos y las bases de datos de variables meteorológicas de varios años recolectados en la Estación de Biología Chamela.

El análisis de los datos disponibles y la interpretación de los resultados en el contexto interdisciplinario surgen como objetivos consecuentes. Para realizar el análisis de los datos usaremos herramientas estadísticas. Para la estimación de los parámetros del modelo usaremos el método de verosimilitud. Además, para respaldar el ajuste del modelo, hemos usado una *región de incertidumbre*; esta herramienta estadística será descrita en el Capítulo 3. Finalmente, interpretaremos los parámetros y obtendremos resultados y conclusiones del análisis realizado.

Capítulo 2

Modelación del Tamaño de Nidada del Loro Corona Lila

En este Capítulo 2, desarrollaremos y justificaremos un modelo que describe el mecanismo de puesta de huevos en aves. Lo haremos basándonos en las teorías revisadas en el capítulo anterior. Por esta razón estipularemos previamente las premisas en las que se sustenta el modelo. En primer lugar formularemos un modelo general, el cual describe el mecanismo de puesta de huevos y que nos permite estimar el número de huevos puestos por una hembra. Después discutiremos acerca de las covariables que pueden incorporarse al modelo y la pertinencia de hacerlo. Cerraremos el capítulo adaptando el modelo general al caso particular del Loro Corona Lila de la región Chamela. Esta formulación concreta, incorpora una función de densidad específica, a saber, la función de densidad LogNormal y daremos una justificación por medio de nociones de acumulación de energía. La expresión final incluye las covariables con cuyas mediciones contamos y que aportan información al modelo; estas covariables pertenecen, en su mayoría, al registro de la Estación de Biología Chamela.

2.1. Premisas

Después de revisar varias de las teorías sobre el tamaño de nidada tomamos como base para el desarrollo del modelo la teoría del costo-beneficio. Desde nuestro punto de vista, esta teoría puede englobar a las teorías restantes. Veamos por ejemplo que la teoría del depredador nos dice de que las aves ponen más o menos huevos según detecten o no la presencia de su depredador natural. En otras palabras, calculan el costo de poner una u otra cantidad de huevos, sabiendo que corre cierto nivel de riesgo de perder sus crías por un ataque del depredador al nido. Por otro lado cuando elige poner una cantidad menor de huevos, sabe que en caso de perder su primera puesta por un depredador, tendrá recursos suficientes para

hacer un nuevo intento en esa misma temporada de puesta.

La teoría de la latitud también puede pensarse dentro de la teoría de costo-beneficio. En este caso las aves optimizan sus posibilidades de crianza de acuerdo a la estacionalidad del hábitat en que se encuentran. Ante una inminente escasez de alimentos, bajan la producción de huevos para conservar parte de su energía y poder así, sobrevivir a la siguiente temporada de puesta. De esta forma la teoría del costo beneficio puede representar varias situaciones donde el objetivo es optimizar el número de huevos para dar continuidad a la especie. Como se mencionó, esta teoría plantea que el ave optimiza la producción de huevos a lo largo de su vida y no sólo durante una temporada específica. Ahora bien, el proceso de crianza de un polluelo tiene distintas etapas que abarcan la puesta del huevo, la incubación y la crianza. Después de estas etapas viene un tiempo intermedio entre cada temporada de puesta, y de esta forma se completa el ciclo de crianza.¹

Todo este proceso de crianza requiere la inversión de recursos, lo que puede verse como un consumo de energía. En este sentido, tomaremos como base la teoría del costo-beneficio y por optimización de los recursos entenderemos la optimización de la energía. Con esto estamos simplificando muchos fenómenos biológicos involucrados. Por ejemplo, la etapa de incubación requiere del consumo de energía para mantener caliente el huevo, a cierta temperatura; energía para mantener vivo a los padres, en este caso en forma de alimentos y energía para la locomoción de los mismos. No es necesario describir el proceso por el cual los padres convierten los recursos de la naturaleza en energía. Para nuestro propósito sólo nos basta saber que obtienen energía de sus alimentos y después por medio de algún proceso natural, esa energía es gastada en la producción de huevos, la crianza y la supervivencia posterior de los padres. El proceso por el cual las hembras convierten esta energía en huevos, entendido por huevo el cascarón y embrión, puede ser muy complejo, pero nosotros simplificaremos este proceso a la simple transformación de energía a huevos. En resumen, el ave obtiene energía de los alimentos, esta es invertida en el proceso de crianza, en particular en la producción de huevos e implícitamente el ave busca la optimización del gasto de energía a lo largo de todo su ciclo de vida.

¹El consumo de energía en la crianza de los polluelos es un fenómeno muy complejo que implica desde el gasto de energía en la construcción del nido hasta el costo de alimentación de cada cría hasta que esta abandona el nido. Para nuestro objetivo podemos ignorar este proceso.

2.2. Modelo General

Siguiendo con estas mismas ideas, daremos paso a la formulación de un modelo general que describe el proceso de puesta de huevos. Hemos mencionado el ciclo de crianza de las aves y vimos que entre cada ciclo hay una etapa intermedia, donde al inicio los padres se recuperan de la gran inversión de energía recién hecha y hacia al final los padres inician la acumulación de energía misma que requerirán para enfrentar la próxima temporada de puesta.

Un primer paso es postular la existencia de una variable de energía E que describe esta acumulación de recursos; cada hembra acumula una cantidad distinta de energía durante la etapa de acumulación de energía, es decir, que la cantidad de energía acumulada es una variable aleatoria. Esta variable de energía tiene una función de densidad continua que pertenece a una familia de densidades de probabilidad:

$$E \sim f(y; \omega), \quad \omega \in \Omega.$$

Cada etapa posterior al inicio de la temporada de puesta requiere un gasto de energía distinto, cierta proporción para la puesta, otra para la incubación, crianza y una proporción de energía para la supervivencia de los padres; sin esta reserva de energía, los padres estarían condenados a morir y no se estaría optimizando la reproducción a lo largo de su vida. Independientemente de cuál sea la proporción de energía invertida en cada etapa y cuál sea la energía de reserva que los padres requieran para su supervivencia, hay una cantidad de energía exclusiva para la producción de huevos. En este caso estamos hablando de un gasto de energía hecho particularmente por la hembra. Dependiendo de la especie de ave, el macho puede o no colaborar en la crianza de los polluelos, en este caso como es sabido en las especies de aves como los pericos y cotorras, las aves suelen ser monógamas y formar sociedades de parejas que colaboran en la crianza de sus polluelos. En general el macho contribuye en la elaboración del nido, aún cuando las aves ocupen huecos en árboles para anidar, como es el caso del Loro Corona Lila y en la posterior alimentación de la hembra o los polluelos. Este aporte por parte del macho ya está contemplado en las reservas de energía de la hembra que será en adelante quien tenga el desgaste energético.²

Retomando la construcción del modelo, esa cantidad específica de energía destinada a la puesta de huevos, está descrita por un modelo probabilístico caracterizado por una densidad

²Para más información sobre el comportamiento de anidación del Loro Corona Lila puede consultarse Renton y Salinas (1999).

de energía. Esta densidad ha sido denotada por $f(y; \omega)$. Tomaremos como un hecho que esta cantidad de energía reservada es la cantidad que optimiza el gasto de energía en todo el proceso. Con esto, no tenemos que pensar más en las etapas subsecuentes, ya que estamos aceptando que por medio de algún mecanismo la hembra ya ha optimizado y destinado la energía posible para producir huevos garantizando la energía suficiente para las etapas subsecuentes.

Ya que hemos hablado de una cantidad de energía reservada para la producción de huevos, podemos hablar del costo de producción. Poner cada huevo tendrá costo determinado para la hembra, digamos δ . Este valor δ es una cantidad de energía. De nueva cuenta no nos involucraremos en saber cómo la hembra convierte esta cantidad δ en un huevo; sólo postulamos que ese es el costo por huevo por simplificación del modelo. Otro supuesto que haremos es, que el hecho de que poner el primer huevo tiene el mismo costo que poner el segundo o uno posterior. No existe sustento biológico para pensar que poner el primer huevo cuesta más o menos energía que el último. El siguiente argumento nos servirá para entender qué es este costo δ : la hembra tiene un gasto total de energía Δ en poner n huevos entonces $\delta = \Delta/n$. En realidad más que importarnos la magnitud exacta de δ , nos importa la proporción que ocupa en relación a la energía reservada.³

Ahora, para hacer la formulación de un modelo tenemos que fusionar ambos conceptos: el de energía de reserva y el costo por huevo. Lo haremos de una forma simple. La energía aunque es una variable aleatoria, para el momento de la puesta de huevos ya ha sido fijada, ya que la acumulación de energía es previa al inicio del proceso. Entonces supongamos que una hembra acumuló energía y reserva una cantidad E para la producción de huevos, ya que el costo por poner un huevo es δ entonces esta hembra pondrá $\lfloor E/\delta \rfloor$ huevos.⁴ El excedente de esta cantidad no es suficiente para producir un huevo adicional, por el supuesto de costo homogéneo, y por lo tanto pasa al resto de energía que la hembra usa en las siguientes etapas.⁵

Como anunciamos desde el principio, estamos interesados en un modelo estocástico, por lo que este modelo está representado por una función de densidad, que en términos matemáticos tiene la siguiente expresión:

³Una forma alternativa de justificar un costo δ constante, es diciendo que aunque el costo por huevo podría sí variar de uno a otro, podemos pensar que el costo real es $\delta + \epsilon$ donde $\epsilon \sim (0, \sigma)$ y $\sigma \approx 0$.

⁴ $\lfloor E/\delta \rfloor$ representa un redondeo al menor entero inmediato del cociente E/δ .

⁵En este sentido estamos postulando que la hembra no tiene abortos, es decir, pone el número de huevos que le es posible.

$$P(x; \omega, \delta) = \int_{\delta x}^{\delta(x+1)} f(y; \omega) dy.$$

Esto se traduce como: la probabilidad de que la hembra ponga x huevos es la probabilidad de que la energía acumulada sea suficiente para poner x con un costo de δ por huevo.⁶

La siguiente paso es hacer una simplificación al modelo, lo cual podemos conseguir quitando el parámetro δ del modelo, esto haciendo $\delta = 1$. ¿Por qué es válido? En primer lugar no tenemos las herramientas suficientes para estimar o conocer el gasto de energía que un ave invierte en la producción de cada huevo. En segundo lugar no es de interés conocer o estimar esta cantidad, por lo que podemos decir que producir un huevo cuesta una unidad de energía y de esta forma estandarizamos la cantidad de energía y el costo por huevo, de modo que la densidad $P(x; \omega)$ se presupone medida en esa escala.⁷ Con esta simplificación el modelo se reduce a la forma:

$$P(x; \omega) = \int_x^{x+1} f(y; \omega) dy.$$

2.2.1. Covariables

Medir directamente la energía sería una tarea muy difícil de realizar; sin embargo, sabemos que la cantidad de energía depende directamente de otros factores o covariables que sí resultan fáciles de medir. Estas covariables deben ser incorporadas en el modelo, y de esta forma el parámetro de la función de energía será a su vez función de otras covariables. En términos generales tenemos que:

$$\omega = g(\vec{z}; \vec{\theta})$$

y por lo tanto,

$$P(x; \vec{z}, \vec{\theta}) = \int_x^{x+1} f(y; g(\vec{z}; \vec{\theta})) dy, \quad (2.2.1)$$

⁶El número de huevos está determinado por la cantidad de energía reservada y por el costo de cada huevo.

⁷Podemos suponer que el costo de un huevo es $\delta = 1$ y la cantidad de energía disponible es E , entonces es posible producir E/δ huevos. Esta misma cantidad se puede obtener si dividimos E por δ , esto es E/δ y hacemos $\delta = 1$.

donde \vec{z} es un vector de covariables y $\vec{\theta}$ es un vector de parámetros de la función g .

Discutamos un poco sobre qué factores pueden intervenir en la cantidad de energía que acumula un ave. Ya que la principal fuente de energía del ave son sus alimentos, entonces podríamos cambiar la pregunta a ¿qué factores determinan la abundancia de alimentos? Pronto vienen a la mente los factores ambientales como: lluvia, temperatura, insolación, humedad de la tierra, etc. Otro factor es el tamaño de la población pues a mayor número de individuos menor la ración de alimentos que cada uno recibe. La estacionalidad del clima también influye puesto que la abundancia de frutos depende de que tan corta o larga haya sido la época de lluvias. Así podemos listar una serie de factores que influyen en la abundancia de alimentos en cada ciclo de crianza. Estas covariables son las que, en caso de contar con su medición, determinan la forma de la función de energía y han sido consideradas en la función de densidad (2.2.1).

2.3. Modelo Adaptado al Loro Corona Lila en la región de Chamela

Hemos presentado un modelo general que captura las ideas de la teoría del costo-beneficio. En el modelo general la función de energía es una función de distribución genérica y las covariables son un vector \vec{z} de variables genéricas. Pero hay una gran variedad de familias de distribución y debemos elegir alguna que sea sensata para modelar un comportamiento de acumulación de energía. Del mismo modo, hay un sinnúmero de covariables que participan en la disposición de energía para las aves.

A partir de este punto construiremos un modelo particular para modelar el tamaño de nidada del Loro Corona Lila de la región Chamela. Haremos una adaptación para incluir las covariables que tenemos a disposición y que aporten información adicional. Entonces tenemos que advertir que la formulación aquí planteada es simplemente una propuesta que a nuestro juicio es muy razonable. Además, la propuesta planteada obedece a la disponibilidad de datos de la Estación de Biología Chamela. La virtud de la formulación general es que se basa directamente en teorías bien establecidas en biología; nosotros proponemos una adaptación *ad hoc* a nuestras necesidades y limitaciones. Quien posea más datos o información puede adaptar fácilmente el modelo a sus circunstancias. El trabajo aquí presentado es así, el producto de una combinación entre modelación empírica y modelación teórica.

2.3.1. Densidad de Energía

Imagine el lector que un ave tiene una cantidad de energía inicial W y que en un siguiente momento $t = 1$ esta cantidad aumenta o disminuye; entonces la cantidad de energía en el momento 1 puede escribirse como $E = W \cdot S_1$ donde S_1 tiene función de distribución F y $\mathbb{E}(S_1) = 1$. Si la cantidad de energía aumenta $S_1 > 1$ y si disminuye $S_1 < 1$. Ahora en un segundo periodo $t = 2$, $E = W \cdot S_1 \cdot S_2$. Si se continua con este proceso, después de n periodos tendremos la cantidad de energía acumulada es $E = W \cdot S_1 \cdots S_{n-1} \cdot S_n$ donde cada S_i ($i = 1, 2, \dots, n$) tiene función de distribución F . Observemos ahora que si tomamos logaritmos de cada lado de la igualdad tenemos que

$$\log(E) = \log(W) + \log(S_1) + \cdots + \log(S_n) = \log(W) + \sum_{i=1}^n \log(S_i). \quad (2.3.1)$$

Para n suficientemente grande, aplicamos el Teorema del Límite Central y obtenemos que el lado derecho de la igualdad (2.3.1) tiene distribución aproximadamente Normal. Por lo tanto $E \sim \text{LogNormal}$. Esta construcción multiplicativa constituye una de las génesis comunes del modelo LogNormal (Limpert et al. (2001)).

El proceso que acabamos de describir constituye una aproximación razonable al mecanismo de acumulación de energía en las aves. En consecuencia, la elección de la distribución LogNormal como una primera densidad de energía para este modelo es razonablemente sensata.

La función de densidad LogNormal tiene la forma:⁸

$$f(y; \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(\log(y) - \mu)^2}{2\sigma^2} \right\}, \quad \text{para } y > 0.$$

Recordemos algunas interpretaciones que tienen los parámetros de la densidad LogNormal, para fines de entender de mejor manera su rol en los análisis que seguirán. La mediana de la distribución LogNormal esta dada por e^μ , con lo que μ es el logaritmo de la mediana. El cociente entre la media y la mediana es igual a $e^{\frac{\sigma^2}{2}}$. Esto significa que μ por sí solo tiene información sobre la mediana, y σ describe la tasa entre media y mediana, es decir, es una medida de asimetría de la distribución. Esto también puede interpretarse de acuerdo a la

⁸Esta distribución cuenta con dos parámetros μ y σ , siendo el primero un parámetro de localización y el segundo de escala.

magnitud de σ , si $\sigma > 1$ la distribución tiene cola pesada a la derecha; mientras que si $\sigma \leq 1$ la distribución se vuelve más simétrica y concentrada alrededor de la media. La varianza tiene la forma $(e^{2\sigma^2} - e^{\sigma^2})e^{2\mu}$ con lo que para σ constante, la varianza es proporcional a e^μ .

2.3.2. Covariables Incorporadas

Mencionamos de que son muchos los posibles factores que afectan las disponibilidad de alimentos, que es la fuente de energía de las aves. Lamentablemente, no es posible medir todas las variables; sin embargo, en la Estación de Biología Chamela se cuenta con la medición de ciertas variables ambientales como cantidad de lluvia, temperatura máxima, temperatura promedio, etc. También contamos con los datos internacionales del fenómeno ENSO. Algo que debemos notar es que las variables disponibles son variables ambientales, es decir extrínsecas; las variables intrínsecas como peso del ave, edad, tamaño, etc. podrían influir en la captación de energía del ave, pero su medición no está disponible. También hay otras clases de variables ambientales que tampoco están disponibles, como tipo de suelo, topografía, capacidad hidrológica, etc.

Tanto las variables como la forma de incorporarlas al modelo son mera propuesta nuestra. Las variables serán incorporadas al modelo ya que después de un análisis empírico hemos identificado su aporte informativo para la estimación del número de huevos. Sin embargo, es totalmente plausible que cualquier otra estructura o cualesquiera otras variables también aporten información valiosa; pero como ya mencionamos, nos adaptamos a nuestras limitaciones y el modelo queda abierto a cualquier modificación factible.

Las covariables incorporadas en el modelo son: Índice de Lang, Temperatura Máxima y ENSO. A continuación explicaremos por qué elegimos cada uno de estas covariables. Un punto importante antes de continuar, es que debemos aclarar que las variables son medidas en años hidrológicos y no en años calendario, debido a que los ciclos de las aves se rigen por los ciclos hidrológicos.⁹Las mediciones de estas covariables pueden encontrarse en el Apéndice 1.

2.3.2.1. Índice de Lang

La precipitación es el factor detonante de la floración de las plantas y por eso natural su incorporación en el modelo. Al realizar distintos análisis encontramos que no sólo la cantidad de agua que cae sobre una región importa para la cantidad disponible de alimentos en ese año,

⁹Un año hidrológico es periodo de tiempo que abarca de abril a marzo del año solar siguiente.

sino que también es importante la temperatura que se presenta durante esa época. Porque un año que presenta lluvias abundantes y altas temperaturas es muy distinto a un año con lluvias abundantes y bajas temperaturas, como sería intuitivo pensarlo.

El Índice de Lang es un indicador termopluviométrico que sirve a los meteorólogos para clasificar el clima de una región o una temporada. El Índice de Lang se calcula simplemente dividiendo la cantidad de lluvia en el año hidrológico medida en milímetros (*mm*) entre la temperatura promedio medida en grados Celsius($^{\circ}C$) del mismo periodo,

$$I = \frac{p}{t}.$$

Se considera un año muy seco cuando el Índice de Lang está por debajo de 20, año seco si está entre 20 y 40, y año húmedo si es mayor a 40. En la zona de Chamela se presentan años en los tres niveles debido a la estacionalidad del ecosistema. Chamela presenta un clima tropical cálido subhúmedo con dos estaciones bien definidas: la temporada de sequía que se presenta de noviembre a junio y algunas veces se ve interrumpida por lluvias ligeras o fuertes en diciembre o enero, y la temporada de lluvias de julio a octubre y que puede prolongarse hasta enero.

2.3.2.2. Temperatura Máxima

La temperatura promedio de la zona suele tener poca variación; sin embargo, las temperaturas extremas tienen mayor variación y ésta provoca alteraciones en la vegetación. La temperatura máxima es el promedio de las temperaturas máximas anuales. La temperatura máxima depende de varios factores, entre ellos la insolación, la mayor o menor humedad, la continentalidad, los vientos y de otros. Así que esta variable puede servir como un indicador de las otras variables involucradas, de las cuales no tenemos mediciones.

La temperatura es un factor biometeorológico ligado directamente con el crecimiento y desarrollo de la vegetación. De esta manera, existen umbrales de temperatura mínima y máxima en los cuales la temperatura óptima debe situarse, de lo contrario se producen daños permanentes o la muerte de la vegetación. Ante temperaturas muy bajas la temporada de floración tiende a retrasarse o bien, ante temperaturas muy altas los frutos tienden a madurarse antes de que las aves puedan aprovecharlos y de ahí la importancia de la temperatura en la producción de alimentos para el Loro.

Esta noción nos indica que la temperatura máxima debe ser considerada dentro del modelo. Aún más, intuitivamente las temperatura extremas, en este caso altas, afectan negativamente la acumulación de energía; y en efecto, un comportamiento de este tipo será evidenciado en el Capítulo 3.

2.3.2.3. ENSO

El Niño Southern Oscillation (ENSO) es el más importante fenómeno de asociación entre atmósfera y océano que causa la variabilidad del clima global en escalas de tiempo interanuales, comúnmente conocido simplemente como fenómeno de El Niño. En términos generales podemos decir que es un fenómeno que ocurre cuando hay un proceso de calentamiento o enfriamiento (El Niño o La Niña, respectivamente) del Océano Pacífico lo cual provoca un cambio atípico en los ciclos hidrológicos prácticamente en todo el planeta.

En la región de Chamela se tienen bien identificados los efectos producidos por el fenómeno ENSO. Cuando un año exhibe el fenómeno de El Niño, se presentan lluvias abundantes y atípicas en los meses de diciembre y enero con lo que la temporada de lluvias se prolonga más de lo habitual, incluso hasta el mes de febrero.¹⁰ Un año que exhibe el fenómeno de La Niña presenta lluvias muy altas entre los meses de julio y septiembre y las lluvias en el resto de los meses son considerablemente más bajas de lo normal.¹¹

La razones para tomar en cuenta la variable ENSO es que esta describe de cierto modo la forma en que se distribuye la lluvia a lo largo del año hidrológico. Así si tenemos un año Niño, habrá mayores precipitaciones en los meses de diciembre y enero; éste es el periodo previo a las puesta del Loro Corona Lila. Por otro lado, un año Niña presenta mayores precipitaciones en la época que es cuando el Loro ya ha puesto sus huevos. Entonces la variable ENSO aporta información a la forma en cómo se distribuyen las lluvias en el año hidrológico.

El índice ENSO, es medido en una escala estandarizada entre -3 y 3 . Los valores extremos indican un año fuertemente marcado por el fenómeno ENSO, correspondiendo los valores positivos a El Niño y los valores negativos a La Niña. Valores cercanos al 0 son considerados años regulares.¹²

¹⁰Este es el motivo por el que se denomina fenómeno del Niño, en alusión a las lluvias atípicas en torno a la Navidad.

¹¹Para más detalles sobre los efectos del fenómeno ENSO en la región de Chamela puede consultarse Martínez (2011).

¹²Para más información sobre el ENSO puede consultarse la página oficial de la División de Ciencias Físicas del Laboratorio de Investigaciones del Sistema Terrestre de Estados Unidos, <http://www.esrl.noaa.gov/psd/>

La formulación del modelo termina con la incorporación de la función de distribución LogNormal y de estas variables elegidas. Esto se logra haciendo μ y σ función de las covariables, esto es,

$$(\mu, \sigma) = g(\vec{z}; \vec{\theta}).$$

En este caso $z_1 = \text{Índice de Lang}$, $z_2 = \text{Temperatura Máxima}$, $z_3 = \text{Índice ENSO}$ y $\vec{\theta} \in \mathbb{R}^5$. Nuestra propuesta de función g es una función exponencial con argumento lineal, por simplificación del modelo, con lo que:¹³

$$\begin{aligned} \mu &= \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\}, \text{ y} \\ \sigma &= \exp\{\theta_4\}. \end{aligned}$$

Con esta forma funcional de g y recordando la interpretación de los parámetros de la distribución LogNormal, estamos postulando que es la mediana de la densidad de energía quien depende directamente de las covariables elegidas. Por otra parte, al tomar $\sigma = e^{\theta_4} = \text{constante}$ tenemos que la varianza de la variable de energía es proporcional $e^{2\mu}$. Además, también se cumple que la relación entre media y mediana se mantiene constante. Así, se simplifica el análisis del modelo.

El modelo tiene finalmente la forma:

$$P(x; \vec{\theta}) = \int_x^{x+1} \frac{1}{y \exp\{\theta_4\} \sqrt{2\pi}} \exp\left\{ \frac{-(\log(y) - \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\})^2}{2(\exp\{\theta_4\})^2} \right\} dy; \quad (2.3.2)$$

$$y > 0, \quad x \in \mathbb{N} \cup \{0\}.$$

Notar no obstante la economía de la notación, que en realidad lo que se está planteando es una distribución condicional del tamaño de nidada dadas las covariables, y que dicha probabilidad condicional depende de $\vec{\theta}$. Usamos la notación $P(x; \vec{\theta})$ en lugar de $P_{\vec{\theta}}(x|\vec{z})$.

¹³La función exponencial nos asegura tener valores positivos de los parámetros como lo requiere la distribución LogNormal.

Capítulo 3

Inferencia Estadística e Implementación

En este Capítulo 3 implementaremos el modelo sobre una base de datos de observaciones de nidos de Loro Corona Lila en la región de Chamela. Para ello, haremos la estimación estadística de los parámetros; usaremos el método de verosimilitud y posteriormente, haremos una interpretación contextual de los parámetros involucrados. Además, describiremos y utilizaremos la herramienta estadística que llamaremos *Región de Incertidumbre*. Presentaremos una aplicación de este instrumento estadístico, con lo cual podremos esbozar la incertidumbre de las estimaciones del modelo. El capítulo termina con una serie de conclusiones y comentarios que surgen de este trabajo, tanto de la formulación del modelo como del análisis realizado.

Los datos aquí usados corresponden a la observación directa de 135 nidos a lo largo de 16 años en la Estación de Biología Chamela y sus alrededores. La base de datos fue recolectada por la Dra. Renton de la Estación de Biología Chamela durante el periodo comprendido de 1996 a 2011. Debemos mencionar que la recopilación de datos, debido a su aparente escasez, ha requerido de un gran esfuerzo. Sólo para darnos una idea, el proceso de recolección de los datos, implica localizar y seguir las parvadas de Loros por varios días, a través del bosque (entre cerros y cañadas) hasta localizar el lugar de anidada de la temporada. Después hay que ubicar el nido, que generalmente se encuentra en un hueco en la parte más alta de los árboles. Luego se tiene que escalar el árbol para colocar una cámara de video y esperar en el lugar (acampando a la intemperie) otro días más hasta que la puesta se haya completado. La dificultad y el largo tiempo invertido en el proceso trae como consecuencia que en algunos años tengamos muy pocos conteos; esto ofrece un reto adicional en términos de la estimación de parámetros, como veremos más adelante.

Recapitulando, en el capítulo anterior hemos formulado un modelo general para modelar

el tamaño de nidada de aves. En particular, hicimos la adaptación de este modelo al caso del Loro que habita en la región de Chamela, lo cual incluyó la justificación de la elección de las covariables y su incorporación al modelo. Finalmente, obtuvimos la función de densidad

$$P(x; \vec{\theta}) = \int_x^{x+1} \frac{1}{y \exp\{\theta_4\} \sqrt{2\pi}} \exp \left\{ \frac{-(\log(y) - \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\})^2}{2(\exp\{\theta_4\})^2} \right\} dy \quad (3.0.1)$$

para $y > 0$, $x \in \mathbb{N} \cup 0$; donde $z_1 =$ Índice de Lang, $z_2 =$ Temperatura Máxima, $z_3 =$ Índice ENSO y $\vec{\theta} \in \mathbb{R}^5$.

Por definición, un nido es considerado como tal sólo cuando se observa al menos un huevo en su interior, por lo tanto en los 135 conteos no observamos ceros (no hay nidos con 0 huevos). Por esta razón debemos hacer una adaptación al modelo; esto lo conseguimos con una simple truncación.¹ Así la función de densidad (3.0.1) tiene su versión truncada en la forma:

$$P_0(x; \vec{\theta}) = \frac{P(x; \vec{\theta})}{1 - P(0; \vec{\theta})} \quad x \in \mathbb{N}. \quad (3.0.2)$$

donde x es tamaño de nidada, es decir, el número de huevos en el nido.

3.1. Estimación de Parámetros

Para la estimación de los parámetros utilizaremos el método de verosimilitud. La función de verosimilitud del modelo es,

$$\begin{aligned} L(\vec{\theta}; \vec{x}) &= \prod_{i=0}^n P_0(x_i; \vec{\theta}) \\ &= \prod_{i=0}^n \frac{P(x_i; \vec{\theta})}{1 - P(0; \vec{\theta})}, \end{aligned}$$

donde \vec{x} es el vector de las observaciones de los nidos y la función de log-verosimilitud es,

¹Esto se consigue dividiendo entre $1 - P(0; \vec{\alpha})$.

$$\begin{aligned}
l(\vec{\theta}; \vec{x}) &= \log L(\vec{\theta}; \vec{x}) \\
&= \sum_{i=0}^n \log(P(x_i; \vec{\theta})) - \log(1 - P(0; \vec{\theta})) \\
&= \sum_{i=0}^n \log \left\{ \int_{x_i}^{x_i+1} \frac{1}{y \exp\{\theta_4\} \sqrt{2\pi}} \exp \left\{ \frac{-(\log(y) - \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\})^2}{2(\exp\{\theta_4\})^2} \right\} dy \right\} \\
&\quad - \log \left\{ 1 - \int_0^1 \frac{1}{y \exp\{\theta_4\} \sqrt{2\pi}} \exp \left\{ \frac{-(\log(y) - \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\})^2}{2(\exp\{\theta_4\})^2} \right\} dy \right\}.
\end{aligned}$$

El estimador de máxima verosimilitud es,

$$\hat{\vec{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4) = \underset{\vec{\theta}}{\operatorname{argmax}} \left\{ l(\vec{\theta}; \vec{x}) \right\}.$$

3.2. Región de Incertidumbre

Como estadísticos sabemos que la naturaleza estocástica del modelo conlleva cierto grado de incertidumbre en las estimaciones, tanto de parámetros como de funciones de ellos que son de nuestro interés. La *Región de Incertidumbre (RI)* es una herramienta útil para darnos una idea de la naturaleza de esta incertidumbre. La idea detrás de la *RI* se basa en el siguiente teorema:²

Teorema. Sea X_1, X_2, \dots, X_n variables aleatorias independientes con función de distribución $f(x; \vec{\theta})$. Sea $\vec{\theta}_0 \in \Theta \subseteq \mathbb{R}^n$, tal que $X_i \sim f(x; \vec{\theta}_0)$, $i = 1, 2, \dots, n$. Entonces bajo condiciones de regularidad, se cumple que $-2\log R(\vec{\theta}_0; \vec{x})$ converge a una distribución $\chi_{(k)}^2$ (k grados de libertad), donde k es la dimensión del vector $\vec{\theta}$ y $R(\vec{\theta}_0; \vec{x}) = L(\vec{\theta}_0; \vec{x})/L(\hat{\vec{\theta}}; \vec{x})$ es la función de verosimilitud relativa.

Este teorema nos dice que si q_α es el cuantil tal que $P[\chi_{(k)}^2 \leq q_\alpha] = 1 - \alpha$, $0 \leq \alpha \leq 1$, entonces,

$$P[-2\log R(\vec{\theta}; \vec{x}) \leq q_\alpha] \approx P[\chi_{(k)}^2 \leq q_\alpha] = 1 - \alpha.$$

²Una demostración del Teorema puede consultarse en Serfling (1980).

Esto es equivalente a,

$$P \left[\log L(\vec{\theta}; \vec{x}) \geq \log L(\hat{\vec{\theta}}; \vec{x}) - \frac{q_\alpha}{2} \right] \approx 1 - \alpha. \quad (3.2.1)$$

Así todos los valores de $\vec{\theta}$ que satisfacen la desigualdad $\log L(\vec{\theta}; \vec{x}) \geq \log L(\hat{\vec{\theta}}; \vec{x}) - \frac{q_\alpha}{2}$ de la ecuación (3.2.1) pertenecen a un conjunto llamado región de confianza del $(1 - \alpha)\%$, lo cual denotamos como $RC_{\vec{\theta}}(\alpha)$. La justificación para el nombre es que

$$P(\vec{\theta} \in RC_{\vec{\theta}}(\alpha)) \approx 1 - \alpha, \quad \forall \vec{\theta},$$

es decir, la probabilidad de $RC_{\vec{\theta}}(\alpha)$ contenga a $\vec{\theta}$ es aproximadamente $1 - \alpha$.

Definamos $RI_{H(\vec{\theta})}(\alpha) := H(RC_{\vec{\theta}}(\alpha)) = \left\{ w \in \mathbb{R}^j \mid w = H(\vec{\theta}) \text{ para algún } \vec{\theta} \in RC_{\vec{\theta}}(\alpha) \right\}$.

Proposición. Sea $\vec{\theta} \in \mathbb{R}^k$ y $H : \mathbb{R}^k \rightarrow \mathbb{R}^j$ una función. Como $\vec{\theta} \in RC_{\vec{\theta}}(\alpha)$ implica $H(\vec{\theta}) \in RI_{H(\vec{\theta})}(\alpha)$, entonces

$$(1 - \alpha) \approx P \left[\vec{\theta} \in RC_{\vec{\theta}}(\alpha) \right] \leq P \left[H(\vec{\theta}) \in RI_{H(\vec{\theta})}(\alpha) \right].$$

La demostración es una implicación directa del teorema y de la propiedad de la medida de probabilidad que nos dice que, si $A \subseteq B$ entonces $P(A) \leq P(B)$.

Nota: El conjunto $RI_{H(\vec{\theta})}(\alpha)$ no merece ser llamado “región de confianza para $H(\vec{\theta})$ ” porque no tenemos control sobre la probabilidad de cobertura; sin embargo, es una región “conservadora”. Por ello la hemos llamado *Región de Incertidumbre*, ya que sí aporta información sobre la incertidumbre para $H(\vec{\theta})$. En el caso en que $H : \mathbb{R}^k \rightarrow \mathbb{R}^j$ es una función uno-a-uno, entonces $RI_{H(\vec{\theta})}(\alpha)$ sí es una región de confianza para $H(\vec{\theta})$ en el sentido de que $P \left[H(\vec{\theta}) \in RI_{H(\vec{\theta})}(\alpha) \right] \approx 1 - \alpha$. Un problema interesante que plantea esta situación es el problema de calibrar una *RI* para que sí tenga probabilidad de cobertura predeterminada. Por ejemplo, ¿qué $\alpha' < \alpha$ debe utilizarse para construir $RI_{H(\vec{\theta})}(\alpha')$ de tal manera que $RI_{H(\vec{\theta})}(\alpha')$ sí posea cobertura $1 - \alpha'$? En la literatura no parece haber una definición de manera general para este concepto.

Por ejemplo, si tomamos $H(\vec{\theta}) = \mathbb{E}(X) = \sum_i i * P(x = i; \vec{\theta})$, la media de la variable aleatoria discreta X y si aplicamos esta función para cada $\theta \in RC_{\vec{\theta}}(\alpha)$, α fijo, tenemos que $H(\vec{\theta}) \in RI_{\mathbb{E}(X)}(\alpha)$. Al variar $\vec{\theta} \in RC_{\vec{\theta}}(\alpha)$ se genera un conjunto $RI_{\mathbb{E}(X)}(\alpha)$ que tiene una probabilidad mayor o igual a $(1 - \alpha)$ de cubrir el valor desconocido de $\mathbb{E}(X)$.

En general, esta herramienta nos sirve para describir la naturaleza de la incertidumbre en la estimación de $H(\theta)$. Además, su valía radica en que con ella estamos preparados para resolver preguntas de interés para los biólogos que surjan en el futuro, y que estén formuladas directamente en términos de los parámetros. En lo que respecta a este trabajo, hemos usado la *RI* para analizar la variabilidad del estimador de máxima verosimilitud y el tamaño medio de nidada, a manera de ilustración.

Entre las posibles funciones o medición de interés están: la media, la varianza, los cuantiles, intervalos para los parámetros y la función de distribución. Por citar un ejemplo más de la utilidad de la herramienta, imaginemos que por razones de supervivencia de la especie los biólogos se interesan en la proporción de nidos con sólo un huevo; ahora podemos simplemente implementar el método para estimar $P(1; \theta)$ y dar un intervalo que la cubre con probabilidad mayor o igual a $1 - \alpha$.

3.3. Implementación al Loro Corona Lila en la Región de Chamela

Los datos del Loro dentro de la región de Chamela, cuyo proceso de recolección describimos al inicio de este capítulo, consta de 135 nidos registrados durante 16 años. La Tabla 1 resume la información de los conteos.

Tabla 1. Descripción de los datos observados.

Número de				
Año	Observaciones	Media	Varianza	Rango
1996	6	2.50	0.30	2-3
1997	8	2.50	0.28	2-3
1998	6	2.50	0.30	2-3
1999	10	2.70	0.23	2-3
2000	6	3.33	0.26	3-4
2001	7	2.14	0.14	2-3
2002	10	2.50	0.27	2-3
2003	8	2.75	0.21	2-3
2004	6	2.66	0.26	2-3
2005	7	1.71	0.23	1-2
2006	11	1.81	0.16	1-2
2007	9	2.66	0.25	2-3
2008	8	2.37	0.26	2-3
2009	8	2.50	0.28	2-3
2010	12	2.40	0.44	1-3
2011	13	2.61	0.25	2-3
Total	135	2.48	0.35	1-4

Lo primero que debemos hacer notar sobre los datos, es que en más de la mitad los años contamos con sólo 8 o menos observaciones. Estos tamaños de muestra, pequeños, son típicos en el contexto, y son una realidad a la que habrá que sujetarse. Por lo demás, la media empírica es aproximadamente 2.5, pero hay años con una media por abajo de 2 y otro por arriba de 3, es decir, se presenta variabilidad temporal en las observaciones; algunos años se observaron 1 o 4 huevos, lo cual podría considerarse un evento atípico. Es precisamente el objetivo de este trabajo es tratar de entender, qué causa esta variación en la media.

Implementaremos el modelo en estos datos con dos finalidades. La primera es mostrar que en términos generales el modelo propuesto es sensato para la modelación del tamaño de nidada. La segunda razón, es obtener resultados relevantes para el ámbito de biología.

Lo primero que haremos será la estimación de los parámetros para el modelo cuya expresión está dada en (3.0.1). La función de log-verosimilitud correspondiente tiene la forma particular,

$$l(\vec{\theta}; \vec{x}) = \sum_{i=0}^{135} \log \left\{ \int_{x_i}^{x_{i+1}} \frac{1}{y \exp\{\theta_4\} \sqrt{2\pi}} \exp \left\{ \frac{-(\log(y) - \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\})^2}{2(\exp\{\theta_4\})^2} \right\} dy \right\} \\ - \log \left\{ 1 - \int_0^1 \frac{1}{y \exp\{\theta_4\} \sqrt{2\pi}} \exp \left\{ \frac{-(\log(y) - \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\})^2}{2(\exp\{\theta_4\})^2} \right\} dy \right\}$$

donde cada x_i , $i = 1, 2, \dots, 135$, es una de las observaciones y $\vec{\hat{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4) \in \mathbb{R}^5$.

Mediante métodos numéricos computacionales calculamos $\hat{\theta}$, el estimador de máxima verosimilitud. Los valores obtenidos fueron:

$$\begin{aligned} \hat{\theta}_0 &= 1.29, \\ \hat{\theta}_1 &= 0.0078, \\ \hat{\theta}_2 &= -0.0428, \\ \hat{\theta}_3 &= 0.0224, \\ \hat{\theta}_4 &= -1.8575. \end{aligned}$$

La interpretación de estos parámetros se hará en la siguiente sección.

Antes de continuar, podemos mencionar que para la estimación usamos la rutina *optim* del software estadístico R. Esta rutina usa por defecto el método de optimización Nelder-Mead que es generalmente robusto; sin embargo, hay que hacer una buena elección de los puntos de arranque. Como observamos los valores obtenidos tienen magnitudes muy distintas, por lo que un punto de arranque inadecuado causa la no convergencia del método de optimización, o bien, arroja falsos estimadores. La elección de los puntos de arranque debe hacerse progresivamente, es decir, incorporando un parámetro a la vez y probando por ensayo y error, hasta encontrar un punto arranque útil. En parte, esta dificultad se remedia con la forma funcional de g , propuesta en el capítulo anterior $(\mu, \sigma) = g(\vec{z}; \vec{\theta})$. Recordemos que la estructura propuesta es:

$$\begin{aligned}\mu &= \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\}, \text{ y} \\ \sigma &= \exp\{\theta_4\}.\end{aligned}$$

Ahora, sabemos que la estimación puntual de los parámetros no es suficiente para una inferencia adecuada acerca de sus valores desconocidos. En términos estrictos, para la cuantificación de la incertidumbre en la estimación puntual de cada componente de $\vec{\theta}$ deberíamos utilizar la función de verosimilitud perfil. En el análisis que realizamos no implementamos la verosimilitud perfil por dos razones. Primero, nos percatamos de que podría tornarse complejo el cómputo pero en segundo lugar, y razón de mayor peso, es que consideramos más importante contar con una herramienta general que pudiera implementarse para cualquier función de los parámetros. Por un lado, las covariables asociadas a los parámetros y la forma estructural de incorporarlas son una mera propuesta, la cual debe ser sometida al juicio del biólogo. Por otro lado, no podemos anticipar en este momento la función de parámetros que el biólogo considerará como directamente pertinente para sus intereses. Por todo ello, haremos uso de la *RI* para cuantificar calidad de la inferencia.

A continuación describiremos el algoritmo general para la construcción de la *Región de Incertidumbre*. El proceso es el siguiente:

- Definamos $H : \mathbb{R}^k \rightarrow \mathbb{R}^j$ una función de $\vec{\theta}$ que sea de interés para el análisis. Fijamos α el nivel de confianza para $RC_{\vec{\theta}}(\alpha)$ siguiendo la idea descrita en la sección anterior; suponer $\hat{\vec{\theta}}$ ya calculado, usando los datos \vec{x} .
- Observemos que el parámetro $\vec{\theta} \in \mathbb{R}^5$. Luego elegimos una vecindad $V(\hat{\vec{\theta}})$ alrededor de $\hat{\vec{\theta}}$, es decir, $V(\hat{\vec{\theta}}) \subseteq \mathbb{R}^5$.³
- Seleccionamos aleatoriamente m vectores, $\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_m \in V(\hat{\vec{\theta}})$.
- Para un cuantil q_α verificamos si $\vec{\theta}_i, i = 1, 2, \dots, m$, satisface la desigualdad $\log L(\vec{\theta}; \vec{x}) \geq \log L(\hat{\vec{\theta}}; \vec{x}) - \frac{q_\alpha}{2}$; en caso de que $\vec{\theta}_i$ satisfaga estar en la $RC_{\vec{\theta}}(\alpha)$ calculamos y guardamos $H(\vec{\theta}_i)$. En otro caso desecharmos la elección.
- Finalmente, la colección de valores $H(\vec{\theta}_i)$, conforman una región valores con probabilidad de al menos $(1-\alpha)\%$ de cubrir el valor desconocido $H(\vec{\theta})$, lo que hemos denominado $RI_{H(\vec{\theta})}(\alpha)$.

³La elección de $V(\hat{\vec{\theta}})$ es sólo para acotar el rango de θ .

Cabe mencionar algunos comentarios sobre cómo elegir $V(\hat{\vec{\theta}})$ y los $\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_m \in V(\hat{\vec{\theta}})$. Debido a la naturaleza del estimador de máxima verosimilitud $\hat{\vec{\theta}}$ se debe tener cierto cuidado con la elección de $V(\hat{\vec{\theta}})$ para evitar problemas computacionales o esfuerzos innecesarios. La elección de $V(\hat{\vec{\theta}})$ va de la mano con la elección de los $\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_m \in V(\hat{\vec{\theta}})$. En nuestro caso, la elección se hizo eligiendo valores centrados en $\hat{\vec{\theta}}$. Recordando que $\hat{\vec{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)$ podemos generar valores $\epsilon_i \sim N(0, \sigma_i)$, $i = 1, 2, 3, 4, 5$, donde σ_i depende directamente de la magnitud de $\hat{\theta}_i$ cuidando siempre no excluir valores válidos. Así, cada $\vec{\theta} \in V(\hat{\vec{\theta}})$ es de la forma $\vec{\theta} = (\theta_1 + \epsilon_1, \theta_2 + \epsilon_2, \theta_3 + \epsilon_3, \theta_4 + \epsilon_4, \theta_5 + \epsilon_5)$. El número m de vectores depende de la proporción de ellos que satisfagan la desigualdad mencionada. En nuestro análisis, por ilustrar un caso, elegimos $m = 50000$ y $\alpha = .05$ obtuvimos que aproximadamente sólo entre el 1% y el 2% de los vectores $\vec{\theta}$ caen en la $RC_{\vec{\theta}}(.05)$. Tomamos $\epsilon_1 \sim N(0, .1)$, $\epsilon_2 \sim N(0, .002)$, $\epsilon_3 \sim N(0, .02)$, $\epsilon_4 \sim N(0, .05)$ y $\epsilon_5 \sim N(0, .1)$. En nuestra experiencia intentar con m más grande vuelve el proceso más tardado y en nuestra perspectiva no aporta información adicional.

La implementación de la *RI* la hicimos con el fin de ilustrar la variabilidad de los parámetros y de la media estimada de huevos por año. En el modelo propuesto por vía de una distribución LogNormal, la medida de centralidad dada por la mediana, como vimos, es más sencilla. Pero ilustraremos el caso con la media, en virtud de que es un concepto más ampliamente favorecido por biólogos, y que tenemos metodología general para cualquier función de los parámetros. En breve mostraremos los resultados obtenidos.

3.4. Resultados e Interpretación

Primero veamos qué tan válido o sensato es utilizar las covariables postuladas en el modelo. Planteamos la prueba de hipótesis

$$H_0 : \vec{\theta} = \vec{0},$$

que es lo mismo que

$$H_0 : \vec{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (0, 0, 0, 0, 0).$$

Si calculamos el p-valor, definido como

$$\text{p-valor} = P[-2\log R(\vec{0}, \vec{x}) \leq \chi_{(5)}]$$

obtenemos que p-valor ≈ 0 . Esto quiere decir que existe evidencia muy sustancial de que el vector de parámetros no es $(0,0,0,0,0)$.

La conclusión inmediata es que esta propuesta de covariables sí aporta información a la distribución del tamaño de nidada. Recordemos que las covariables involucradas son: $z_1 =$ Índice de Lang, $z_2 =$ Temperatura Máxima, $z_3 =$ Índice ENSO y $\vec{\theta} \in \mathbb{R}^5$. El estimador de máxima verosimilitud es $\hat{\vec{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)$ con

$$\begin{aligned}\hat{\theta}_0 &= 1.29, \\ \hat{\theta}_1 &= 0.0078, \\ \hat{\theta}_2 &= -0.0428, \\ \hat{\theta}_3 &= 0.0224, \\ \hat{\theta}_4 &= -1.8575.\end{aligned}$$

3.4.1. Interpretación de los Parámetros

El parámetro θ_0 no está vinculado a ninguna covariable, es decir, que su aporte al valor de μ no depende del valor que tomen las covariables. Recordemos que $\mu = \exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_3\}$. Por lo anterior θ_0 se traduce en un cantidad basal de energía. Visto de otro modo, es al aporte mínimo de energía que le otorga al Loro el ecosistema de Chamela.

El parámetro θ_1 está asociado al Índice de Lang y al tener signo positivo implica que a mayor nivel del índice mayor será μ . Traducido de otra forma, esto significa que cuánto más húmedo sea un año, mayor será el valor μ y esto implica en sentido contrario que en un año seco menor será el valor de μ . Entonces concluimos que la cantidad de lluvias en el año provoca un aumento en el tamaño de nidada.

El parámetro θ_2 está asociado con la temperatura máxima. El signo negativo, nos dice que afecta de forma negativa al tamaño de nidada. Esto es, cuanto mayor sea la temperatura máxima del año, menor será el valor de μ . Lo anterior hace mucho sentido con el planteamiento descrito en la elección de esta covariable, donde dijimos que temperaturas extremas causa alteraciones el floración y fructificación de la vegetación, que son la fuente directa de energía del Loro.

El parámetro θ_3 asociado con el índice ENSO, describe cómo afecta la forma en que la lluvia se distribuye durante el año hidrológico. Recordemos que el índice ENSO tiene una escala que va de -3 a 3 por lo cual a veces toma valores cercanos a cero; en estos años el índice ENSO aporta poco al valor de μ ; sin embargo, sí podemos decir que cuando el índice toma valores cercanos a los extremos, ya sea 3 o -3 , entonces μ aumenta o disminuye según sea el caso. En otras palabras, cuando un año presenta el fenómeno de El Niño (lluvias atípicas en diciembre y enero, cercanas a la temporada de puesta) el valor de μ aumenta. Por otro lado, un año que presenta el fenómeno de La Niña (disminución de lluvias todo el año excepto en la época de lluvias) el valor de μ disminuye.

En el caso de estos cuatro parámetros tengamos en cuenta que las escalas de los valores de las covariables son distintas y que su efecto de cada covariable es multiplicativo y exponencial con respecto a μ por lo cual el efecto real de cada parámetro no es inmediato. Esto sería pertinente investigarlo con mucho mayor detalle una vez que las covariables y el modelo hayan sido ratificados por el biólogo.

Por último, recordemos que definimos $\sigma = \exp(\theta_4)$. Dado el valor estimado de θ_4 , tenemos que $\sigma = \exp(-1.8575)$. Evocando las propiedades de los parámetros de la distribución LogNormal descritas en el Capítulo 2, tenemos que para este caso el cociente entre la media y la mediana de la variable de energía es igual a $e^{\frac{\sigma^2}{2}} = 1.012 \approx 1$ lo cual nos dice que para este modelo la mediana es aproximadamente igual a la media de la variable de energía (no se debe confundir con la media del modelo completo); además la varianza de la variable de energía es proporcional a $e^{2\mu}$ y la cantidad proporcional es $e^{2\sigma^2} - e^{\sigma^2} = .025$ estos dos hechos nos dicen que la forma de la densidad de energía es simétrica y concentrada alrededor de la media.

3.4.2. Estimación de la Media del Tamaño de Nidada

Ya que hemos estimado $\hat{\theta}$ y que tenemos una interpretación para cada parámetro, continuaremos con la estimación del tamaño promedio de nidada. Lo haremos calculando la media estimada como función de $\vec{\theta}$. Entonces tomemos

$$H(\vec{\theta}) = \mathbb{E}(X) = \sum_{i=1}^{\infty} i * P_0 \left[x = i; \vec{\theta} \right].$$

Esta media estimada será comparada con la media o promedio observado. La media observada es una estimación no paramétrica del número promedio de huevos de los nidos observados en un año en particular, es decir, la media muestral por año.

La Tabla 2 resume el tamaño de nidada promedio observado por año así como el promedio estimado y la diferencia entre ambas cantidades. En general, no se observa una discrepancia notoria salvo en los casos, correspondientes a los años 2003 y 2005. Aún así, esta diferencia es menor a 0.5 en valor absoluto.

Tabla 2. Promedio observado vs promedio estimado.

Año	Promedio		Diferencia
	Observado	Estimado	
1996	2.50	2.56	-0.06
1997	2.50	2.58	-0.08
1998	2.50	2.35	0.15
1999	2.70	2.75	-0.05
2000	3.33	3.04	0.29
2001	2.14	2.38	-0.24
2002	2.50	2.22	0.28
2003	2.75	2.27	0.48
2004	2.66	2.64	0.02
2005	1.71	2.17	-0.46
2006	1.81	1.99	-0.18
2007	2.66	2.79	-0.13
2008	2.37	2.41	-0.04
2009	2.50	2.59	-0.09
2010	2.41	2.43	-0.02
2011	2.61	2.45	0.16

La Figura 3.4.1 contrasta las medias estimadas contra las observadas y a partir de ésta podemos señalar que el comportamiento estimado corresponde al comportamiento observado en el tamaño de nidada. Se observa una clara monotonía en el comportamiento de ambas trayectorias. Debemos recordar que para los parámetros usamos estimación paramétrica involucrando los 135 datos observados, mientras que para las medias estimadas usamos las covariables de cada año. Las medias observadas en muchos casos corresponde a la media muestral de a lo más 8 datos. Esto podría explicar en cierto modo, el no ajuste del modelo en ciertos años.

3.4.3. Región de Incertidumbre para la Media Estimada y los Parámetros

Hemos visto que la media observada y la media estimada es en esencia equivalente para la mayoría de los años; sin embargo, como parte de un análisis más elaborado debemos decir

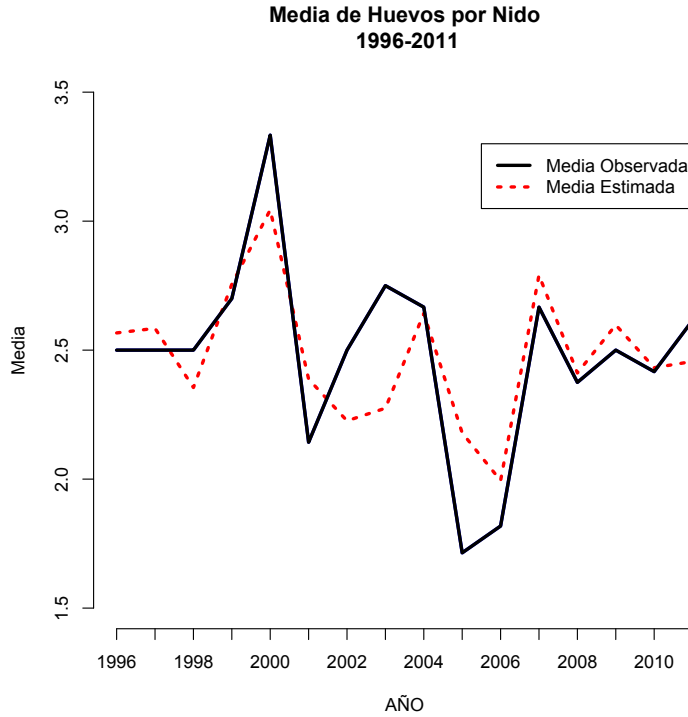


Figura 3.4.1: Media observada vs. media estimada.

algo sobre de la incertidumbre que hay en la estimación. Para ello haremos uso de la RI , la cual nos ayudará a darnos una idea de la magnitud de la incertidumbre involucrada en las estimaciones puntuales.

Para ello aplicaremos el algoritmo descrito antes y tomaremos $\vec{\theta} \in \mathbb{R}^5$ y $H(\vec{\theta}) = \mathbb{E}(X) = \sum_{i=1}^{\infty} i * P_0(x = i; \vec{\theta})$ como antes. Elegimos $\alpha = .05$ y en consecuencia $q_{.05} = 11.07$ es el cuantil correspondiente. Así el conjunto de todos los vectores $\vec{\theta} \in \mathbb{R}^5$ que satisfagan la desigualdad

$$\log L(\vec{\theta}; \vec{x}) \geq \log L(\hat{\vec{\theta}}; \vec{x}) - \frac{11.07}{2}$$

pertenecen a la $RC_{\vec{\theta}}(.05)$ como ya explicamos. Luego, el resultado de aplicar el algoritmo completo es una $RI_{\mathbb{E}(X)}(.05)$.

La Figura 2 muestra algunas trayectorias individuales de la $RI_{\mathbb{E}(X)}(.05)$, mientras que la Figura 3 muestra la $RI_{\mathbb{E}(X)}(.05)$ completa y sobre ella sobre puestos los promedios observados y estimados.

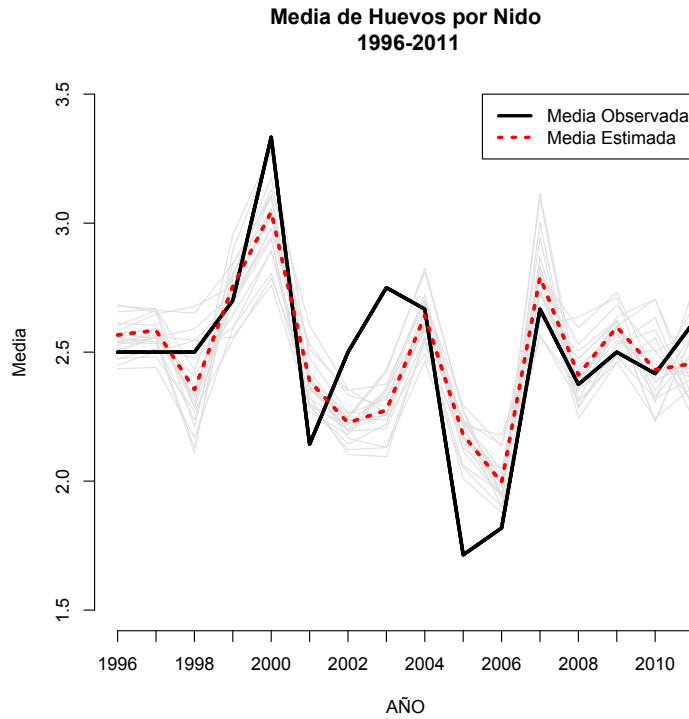


Figura 3.4.2: Trayectorias particulares de $RI_{\mathbb{E}(X)}(.05)$.

A partir de la observación de la $RI_{\mathbb{E}(X)}(.05)$ todos los promedios observados caen dentro, excepto tal vez en cuatro años. Estos años son: 2001, 2002, 2003 y 2005. En el resto de los años podemos decir que la estimación es razonable. La Tabla 3 muestra las $RI_{\mathbb{E}(X)}(.05)$ para cada año y podemos observar que, salvo el los años ya mencionados, el valor estimado se encuentra dentro de la RI . No debemos olvidar que estas RI son regiones conservadoras por lo que suelen ser más amplias que las regiones de confianza, pero no podemos establecer la relación exacta en este momento.

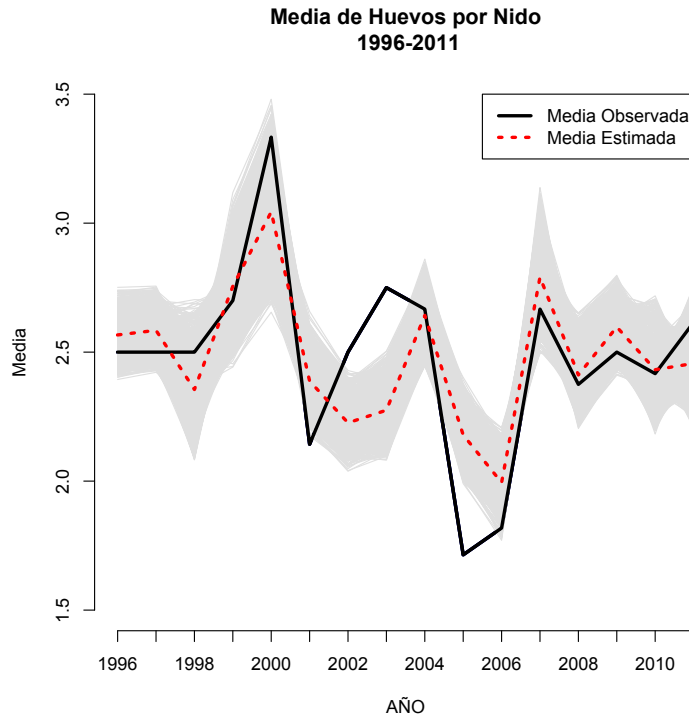


Figura 3.4.3: $RI_{E(X)}(.05)$. completa.

Tabla 3. Promedio observado vs promedio estimado.

Año	Promedio		RI(.05)
	Observado	Estimado	
1996	2.50	2.56	(2.39, 2.75)
1997	2.50	2.58	(2.42, 2.75)
1998	2.50	2.35	(2.08, 2.70)
1999	2.70	2.75	(2.44, 3.11)
2000	3.33	3.04	(2.65, 3.48)
2001	2.14	2.38	(2.17, 2.65)
2002	2.50	2.22	(2.03, 2.42)
2003	2.75	2.27	(2.08, 2.50)
2004	2.66	2.64	(2.44, 2.86)
2005	1.71	2.17	(1.99, 2.40)
2006	1.81	1.99	(1.77, 2.20)
2007	2.66	2.79	(2.49, 3.13)
2008	2.37	2.41	(2.20, 2.65)
2009	2.50	2.59	(2.41, 2.79)
2010	2.41	2.43	(2.18, 2.71)
2011	2.61	2.45	(2.20, 2.75)

También hicimos uso de la RI para cuantificar la incertidumbre alrededor de la estimación puntual de los parámetros. El procedimiento es similar, basta con elegir la función $H(\vec{\theta}) = \theta_i$, $i = 1, 2, \dots, 5$, según sea el caso. El resultado es una $RI_{\theta_i}(.05)$, pues mantenemos nuestra elección de $\alpha = .05$. La Tabla 4 muestra las RI para cada parámetro.

Tabla 4. Región de Incertidumbre Parámetros

Parámetro	Estimación	
	puntual	$RI(.05)$
θ_0	1.2900	(.9419, 1.596)
θ_1	0.0078	(.0033, .0123)
θ_2	-0.0428	(-.0526, -.0307)
θ_3	0.0224	(-.0141, .0573)
θ_4	-1.8575	(-2.082, -1.614)

A partir de la Tabla 4 podemos ilustrar cómo es la variabilidad de cada uno de los parámetros involucrados. Aquí se refuerza la interpretación de los parámetros y su covariable asociada. Podemos concluir que en efecto las covariables tienen cierta significancia en el modelo y las RI son consistentes con la interpretación que dimos para cada parámetros. No olvidemos que las RI son regiones “conservadoras” lo cual amplía el rango de valores posibles y esa puede ser parte de la explicación de que la $RI_{\theta_3}(.05)$ asociado al índice ENSO contenga el valor cero, además de que desconocemos la forma de la función de verosimilitud perfil para poder descartar ese caso con entera certeza.

3.4.4. Discusión de Años Atípicos

Los años 2001, 2002, 2003 y 2005 en nuestro análisis fueron considerados como *outliers*. Para determinar las posibles causas de esta diferencia se requiere del estudio de puntual de cada año y de la experiencia de los biólogos de la estación.

Desde un punto de vista estadístico sí podemos decir es que las estimaciones en estos años están hechas con 10, 8 y hasta 7 datos como puede consultarse en la Tabla 1. Pocos datos causa mucha incertidumbre en la estimación. Otro punto que debemos observar es que los 4 años tienen 4 de los 5 índices de Lang más bajos, lo cual puede sugerir un cambio en la estructura del modelo, por ejemplo cambiando la forma funcional de μ .

Cuando discutimos la importancia del tamaño medio de nidada dijimos que este era un indicador del estado de la especie. Estas discrepancias podrían sugerir un comportamiento anormal en el Loro derivado de alguna alteración en su medio. Por ejemplo, cuando el promedio observado es mayor que el promedio esperado consistentemente como ocurre en los años 2002 y 2003, esto puede deberse a una respuesta a la depredación. Generalmente este perico es usado como mascota y extraído de su hábitat natural.

El año 2005 particularmente es un año atípico ya que se observaron nidos con sólo un huevo. Este hecho, aunado a que únicamente tenemos 7 conteos provoca que tengamos este año el promedio de huevos observado más bajo de todos. También recordemos que teóricamente, años con un número bajo huevos es el producto de un inusual y mayor esfuerzo de los padres por aumentar el tamaño de nidada o por sobrevivir en la temporada inmediatamente anterior.

Todos los argumentos aquí dados son un ejemplo del tipo de análisis que deberá realizarse conjuntamente entre biólogos y estadísticos del proyecto. Lo atípico o no de estas observaciones debe ser evaluado por expertos en el Loro, en este caso, la Dra. Renton. Como reiteramos, el modelo es susceptible a las modificaciones respectivas que resulten de este análisis biológico. Pero sí da una clara idea de las conclusiones que pueden obtenerse de su implementación respecto a la aplicación de medidas de conservación del Loro.

3.5. Resumen

El tamaño de nidada resulta ser de interés especial para el estudio de las aves. En particular estamos interesados en el tamaño de nidada o número de huevos puesto por nido del Loro Corona Lila, lo cual resulta ser un fenómeno aleatorio que puede modelarse con una función de densidad. Aquí hemos presentado un modelo adaptado que incorpora covariables ambientales para la modelación del tamaño de nidada del Loro.

Este modelo está basado en la teoría de Costo-Beneficio, la cual es una de las cuatro principales teorías para el tamaño de nidada. El modelo incorpora ideas de las otras teorías como la posibilidad de medir el efecto de depredación como covariable o algunas otras características particulares de la especie. De hecho está en armonía con la teoría de la latitud en el sentido de que al cambiar la ubicación geográfica del ave cambia su nidada promedio, esto como consecuencia de un cambio en los niveles de las covariables ambientales. Pero es

la teoría del Costo-Beneficio la que incorpora toda información para crear un modelo que explica cuál es el mecanismo llevado a cabo para la determinación del número de huevos puesto por el Loro.

El modelo contiene la noción de una variable de energía, la cual tiene una función de densidad que hemos propuesto sea LogNormal, en virtud de la génesis presentada aquí. Posteriormente propusimos una estructura funcional para la incorporación de las covariables, lo cual tiene cierto grado de sensatez debido al contexto, y porque a partir de la prueba de hipótesis correspondiente podemos concluir que las covariables elegidas aportan información al modelo. Estas covariables son el índice de Lang, que es el cociente entre la lluvia total y la temperatura media; la temperatura máxima, tomada como el promedio de las temperaturas máximas del año y el índice ENSO.

Con el modelo ya establecido hicimos la estimación de las medias por año usando para ello los datos de las observaciones directa de nidos y las covariables ambientales mencionadas. El resultado es que el modelo tiene en lo general un buen ajuste. La Figura 3.5.1 muestra los resultados año por año; en negro se muestra la densidad observada, mientras que las marcas (cruces) muestran la densidad estimada. La línea negra muestra la media observada y la línea punteada muestra la media estimada.

Como parte de la colaboración entre los miembros del grupo interdisciplinario queda pendiente obtener retroalimentación acerca del modelo. Ello incluye analizar la pertinencia de las covariables en el modelo y validar la elección de la teoría de Costo-Beneficio como soporte del modelo. También revisar las posibles causas de los datos atípicos, y finalmente plantear más preguntas de interés que puedan resolverse por implementación del modelo.

3.6. Conclusiones

Acerca de la modelación realizada en este trabajo podemos comentar que es el resultado de una interacción entre modelación de expertos, donde se consultó literatura de ecología y biología en particular sobre el tema, y por otra lado de modelación empírica donde a partir de los datos inferíamos cierta información que fue combinada con la información aportada por las fuentes expertas. De ahí que la formulación del modelo general mantenga todos sus fundamentos meramente en la teoría, mientras que el modelo adaptado sí incorpora trabajo empírico.

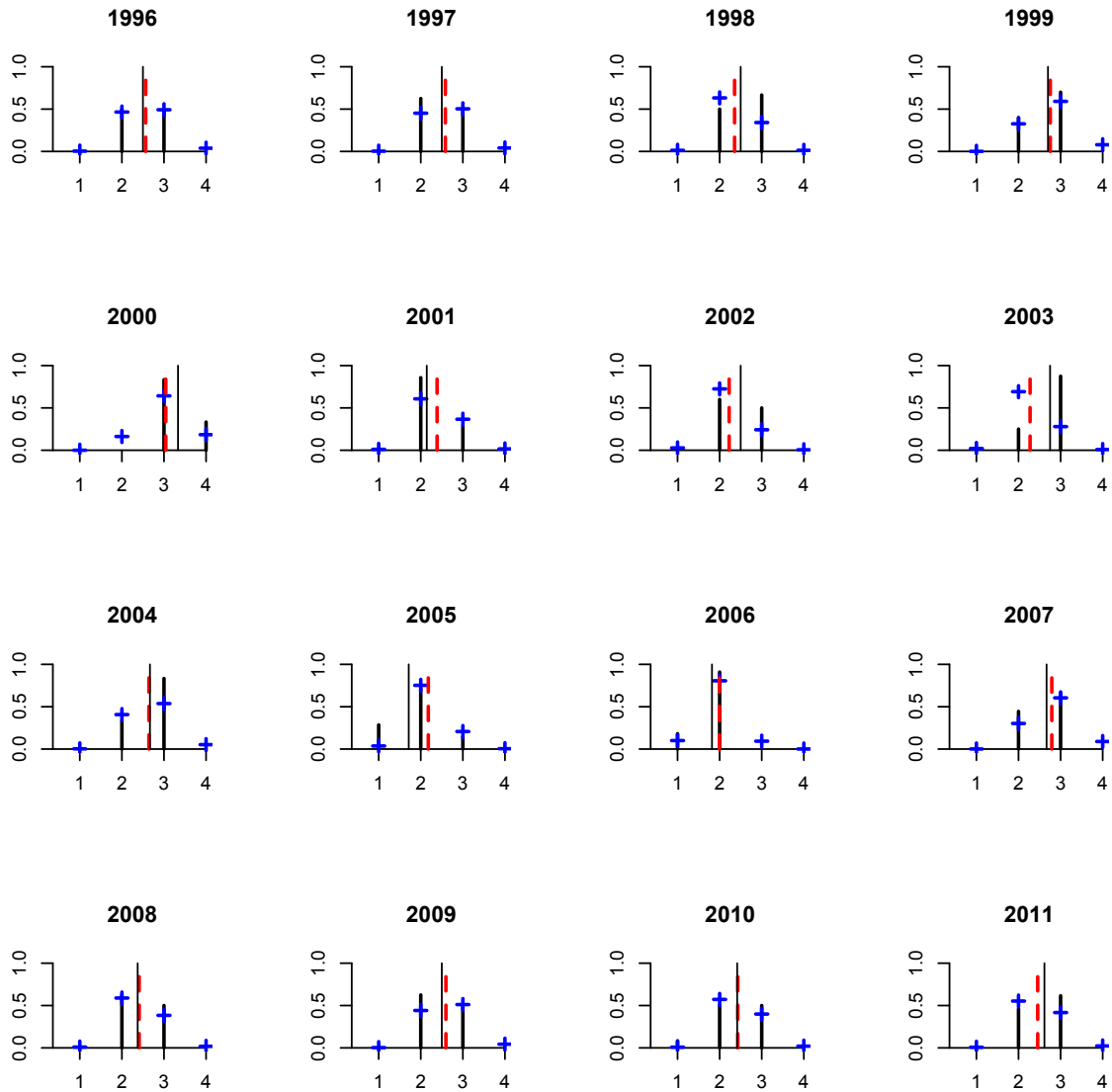


Figura 3.5.1: Ajuste año por año.

Una primera conclusión sobre el modelo general es que su planteamiento no es un caso particular de ninguna especie y es perfectamente adaptable a cualquier ave, por lo cual una posible generalización queda como un trabajo posible. Tanto la densidad de la variable energía como la estructura funcional de los covariables pueden adaptarse a las necesidades del problema planteado.

Dentro del trabajo empírico realizado está en primer lugar, probar distintas funciones de densidad como lo son: Gamma, Beta y Normal. La elección de la LogNormal como ya dijimos, responde a una mera propuesta pero que tiene interpretación y un sentido de congruencia con nuestro objeto de modelación. Otro punto, que es directamente un resultado de modelación empírica es la elección de las covariables y la forma funcional de incorporarlas al modelo. La elección trató de mantener los principios de parsimonia de la modelación además de estar limitada por la disponibilidad de datos. Esta parte implicó probar con una gran variedad de variables y otra tanto de formas funcionales; al final elegimos una forma funcional estética, simplificadora, útil y pertinente, en el sentido de que las covariables sí poseen sentido e información en el marco biológico en que trabajamos.

Sobre el no ajuste del modelo en ciertos años, ya mencionamos que estuvimos restringidos por las pocas observaciones con que se cuenta en cada año (no debemos olvidar el esfuerzo que requiere la recolección de datos). Pero también es posible que este no ajuste se derive de la falta, ya sea por omisión o por falta de ella, de alguna variable que contenga información relevante para el modelo. Ejemplo de éstas son: el tamaño del ave, el tamaño de los polluelos, el tamaño de las comunidades del Loro, el efecto de la depredación, la disponibilidad de árboles para anidación o el efecto de los huracanes cercanos a la región, sólo por mencionar algunas posibles.

Entre las principales extensiones de este trabajo están la generalización del modelo como ya mencionamos, pero también está una posible reformulación del modelo o las covariables después de la confrontación del modelo con biólogos expertos en la materia. Es posible que este modelo tenga aplicaciones en áreas muy diversas; por ejemplo, si en lugar de interesarnos el tamaño de nidada nos interesamos en el tamaño de familia en una sociedad, ¿será este modelo sensato en esa situación? Por otra parte es posible que surjan nuevas interrogantes a partir de las conclusiones obtenidas y que puedan resolverse con el modelo y las herramientas aquí presentadas. Tengamos en cuenta que los datos no ajustados podrían no ser atípicos en esencia si no corresponden a una verdadera señal obtenida del modelo como explicamos a su debido tiempo.

En la parte estadística puramente, queda como posibilidad la implementación de las verosimilitudes perfiles en donde corresponde realizarlo. Sin embargo, yo apuntaría un mayor esfuerzo hacia la explotación total de la herramienta que aquí hemos llamado *Región de Incertidumbre*. Como mencionamos anteriormente su generalidad la convierte en una herramienta estadística potencial, y entonces se abre una amplia gama de incógnitas a partir de ella. Por ejemplo, ¿cómo calibrar la incertidumbre que hay alrededor de las estimaciones?, ¿qué tipo de funciones se favorecen y cuáles no?, ¿qué relaciones se pueden establecer (además de la descrita aquí) entre las *RI* y las regiones de confianza en ámbitos más generales? En nuestra perspectiva este es un tema que tiene mucho por ofrecer.

Apéndice A

Apéndice 1

Los datos contenidos en este apéndice son públicos y puede encontrarse página oficial de la Estación de Biología Chamela, <http://www.ibiologia.unam.mx/ebchamela/> y en la página oficial del Laboratorio de Investigaciones del Sistema Terrestre de Estados Unidos, <http://www.esrl.noaa.gov/psd//enso/>.

Covariables Ambientales

Año	Lang	T. Máxima	ENSO
1996	32.34	33.20	-0.67
1997	31.87	33.23	-0.22
1998	24.96	35.26	2.93
1999	48.72	34.68	-1.31
2000	47.01	32.56	-1.67
2001	20.98	32.35	-0.75
2002	18.11	33.70	0.24
2003	22.56	34.80	1.62
2004	34.96	33.84	0.57
2005	18.31	34.78	1.37
2006	12.61	34.90	-0.25
2007	39.76	34.20	1.30
2008	36.06	34.35	-1.85
2009	36.02	33.55	-0.90
2010	31.05	35.35	2.10
2011	36.74	33.83	-2.45

Bibliografía

- [1] Bell, C. 1996. *The relationship between geographic variation in clutch size and migration pattern in the Yellow Wagtail*. Bird Study 43: 333-341.
- [2] Eggers, S., Griesser, M., Nystrand, M. y Ekman, J. 2006. *Predation risk induces changes in nest-site selection and clutch size in the Siberian Jay*. Proceedings of the Royal Society B 273(1587): 701-706.
- [3] Jetz, W., Sekercioglu, C. y Böhring-Gaese, K. 2008. *The worldwide variation in avian clutch size across species and space*. PLoS Biology 6(12): e303.
- [4] Lack, D. 1947. The significance of clutch size. Ibis 89: 302-352.
- [5] Limpert, E., Stahel, W. y Abbt, M. 2001. *Log-normal distributions across the Sciences. Keys and Clues*. BioScience 51(5): 341:352.
- [6] Mauck, R., Huntington, C. y Grubb, T. 2004. *Age-specific reproductive success: evidence for the selection hypothesis*. Evolution 58: 880-885.
- [7] Martinez, F. 2011. *Clustering para datos funcionales*. Universidad de Guanajuato. Tesis de licenciatura.
- [8] Nager, R., Monaghan, P. y Houston, D. 2001. *The cost of egg production: increased egg production reduces future fitness in gulls*. Journal of Avian Biology 32: 159-166.
- [9] Renton, K. y Iñigo-Eliás, E. 2003. *Evaluación del estado de conservación de las poblaciones de loro corona lila (Amazona finschi) en México*. Universidad Nacional Autónoma de México. Instituto de Biología. Informe final SNIBCONABIO proyecto No. AS001
- [10] Renton, K. y Salinas, A. 1999. *Nesting behaviour of the Lilac-crowned Parrot*. Wilson Bulletin 111: 488-493.
- [11] Ricklefs, R. y Wikelski, M. 2002. *The physiology/life-history nexus*. Trends in Ecology and Evolution 17: 462-468.

- [12] Royama, T. 1969. *A model for the global variation of clutch size in birds*. *Oikos* 20: 562-567.
- [13] Serfling, R. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley.