

SCIENTIFIC ADVISORY BOARD

5TH MEETING

CANCUN (MÉXICO)

21 JUNE 2007

DOCUMENTO SAB-05-03a

ESTIMACIÓN DE MORTALIDAD NO REPORTADA

El personal de la CIAT viene realizando investigaciones encaminadas a detectar datos erróneos de mortalidad de delfines reportados por los observadores. Esto incluye la documentación de las percepciones del personal de las oficinas regionales acerca de los observadores, el examen de datos por los editores, y el uso de técnicas estadísticas para señalar informes atípicos que podrían indicar fabricación de datos. El presente documento describe procedimientos que podrían ser usados para estimar datos erróneos de mortalidad, y comenta brevemente sobre varios puntos relacionados con los requisitos de datos, actuales y futuros.

Procedimiento estadístico para estimar datos erróneos de mortalidad

A continuación se resume del procedimiento que se podría usar para elaborar una estimación preliminar de la mortalidad de delfines en lances sobre los mismos que no es reportada; se podría también considerar un método relacionado para estimar la mortalidad en lances sobre delfines que fueron reportados como lances no asociados. Este procedimiento consistiría en cinco pasos: i) identificación de los observadores cuyos datos son dudosos; ii) identificación de los lances específicos sobre delfines de esos observadores con mortalidades dudosas; iii) evaluación de la eficacia de los pasos (i)-(ii) para identificar datos dudosos bajo modificaciones conocidas de la mortalidad; iv) estimación de la mortalidad en los lances identificados en el paso (ii); y v) evaluación de la sensibilidad de la estimación del paso (iv) a los parámetros usados en los pasos (i)-(ii). Los pasos (i)-(ii) representan el método usado en la actualidad para revisar los datos de los observadores^{1,2}.

i. Identificación de los observadores cuyos datos serán considerados dudosos

Se usará una estimación de la probabilidad de que un observador esté asociado con tantos o más lances dudosos por casualidad para definir cuáles son los observadores que se considera producen datos dudosos (ver referencias citadas para mayor detalle). Se recomienda fijar el umbral para esta probabilidad inicialmente en 0,01; es decir, se someterán a análisis solamente los datos de los observadores con una probabilidad de menos de 1% de tener tantos o más lances dudosos. Este umbral es uno de dos parámetros usados en el algoritmo de filtración de datos que puede afectar la estimación de la mortalidad no reportada (ver punto (v)). En el punto (vi)a se comentan datos adicionales que podrían ser recolectados para ayudar a limitar la elección de estos parámetros.

ii. Identificación de lances con datos erróneos de mortalidad de delfines

Todos los lances sobre delfines reportados por los observadores identificados en (i) sin mortalidad de delfines reportada, pero en los que se hubiera predicho mortalidad, serán considerados como con datos de mortalidad erróneos. En términos del algoritmo de filtración de datos, esto incluiría todo lance reportado por los observadores identificados en (i) con un residual ≤ -0.5 .

¹ Lennert-Cody, C.E. and Berk, R.A. 2007. Statistical learning procedures for monitoring regulatory compliance: an application to fisheries data. Journal of the Royal Statistical Society Series A 170 Part 3: 1-19.

² Documento IRP-39-08b (<http://www.iattc.org/PDFFiles2/IRP-39-08b-Medicion-desempeno.pdf>)

iii. Simulaciones para evaluar el algoritmo de filtración de datos

Se realizarán simulaciones para evaluar la eficacia del algoritmo de filtración de datos bajo varios tipos de modificaciones de los datos de mortalidad. Se identificará un subconjunto apropiado de datos de observadores para las simulaciones, probablemente datos de aquellos observadores sin datos dudosos identificados en (i). La mortalidad reportada en los lances de estos observadores será modificada (por ejemplo, reemplazando lances con mortalidad de más de 10 animales con mortalidad 0) y luego revisada conforme a los pasos (i)-(ii). Se determinará el porcentaje de lances dudosos y de observadores con datos dudosos identificados correctamente. Se evaluará el algoritmo de filtración de datos con varios tipos de modificaciones de los datos (por ejemplo, sustitución de lances con mortalidad alta con mortalidad 0; reducción de la mortalidad por un porcentaje fijo).

iv. Métodos para estimar la mortalidad de delfines en lances con datos dudosos

Se compararán varios métodos distintos para estimar la mortalidad de delfines en los lances identificados en (ii): mortalidad media por lance, mortalidad media por tonelada de atún aleta amarilla capturada, potenciación de gradientes estocástica (*stochastic gradient boosting*), y bosques aleatorios/bosques aleatorios por cuantiles. Se comparará el desempeño de estos cinco métodos mediante la construcción de un estimador en un conjunto apropiado de datos de aprendizaje, y luego estimando la mortalidad en un conjunto apropiado de datos de prueba con mortalidad conocida. El método seleccionado será aquél cuya estimación de mortalidad es más cercana a la mortalidad conocida del conjunto de datos de prueba.

Una alternativa sería considerar que todos los datos de los observadores identificados en (i) son dudosos, y estimar la mortalidad faltante para todos los lances de estos observadores, en lugar de limitar la estimación a aquéllos lances sin mortalidad reportada pero con mortalidad predicha. Esto es similar al procedimiento realizado para el cálculo del desempeño de los buques (ver referencias). Se compararán los resultados de los dos métodos.

v. Sensibilidad de la mortalidad estimada a los parámetros de filtración de datos

El procedimiento descrito en (i)-(iv) producirá una estimación de la cantidad de mortalidad de delfines no reportada que ocurre en los lances sobre delfines. A fin de determinar límites creíbles para esta estimación, serán necesarios datos independientes para establecer límites creíbles de los parámetros usados en el algoritmo de filtración de datos, y se recomienda que la toma de estos datos sea discutida en una reunión de los Programas de Observadores (ver (vi)). Por ahora, los parámetros usados en el algoritmo de filtración de datos serán variados, y las estimaciones de la mortalidad no reportada serán calculadas, a fin de determinar la sensibilidad de la mortalidad estimada a estos parámetros.

vi. Consideraciones de datos

Para poder estimar la cantidad de mortalidad de delfines no reportada para la flota internacional, es esencial que sean revisados los datos de todos los programas de observadores del APICD. Se recomienda que, como parte del proceso de revisión de datos, los programas intercambien información sobre buques y capitanes asociados con datos dudosos.

a. Datos adicionales para mejorar el algoritmo de filtración de datos

Se deberían obtener datos adicionales que pueden ser usados para mejorar el algoritmo de filtración de datos. Por ejemplo, información sobre las actividades de buzos durante y después del retroceso (tiempo en el agua, posición en la red, etcétera), y las características de las actividades de salabardeo, sería probablemente útil. Se deberían discutir en una reunión de los programas de observadores los tipos de datos adicionales que se podrían obtener.

b. Limitación de los parámetros usados en el algoritmo de filtración de datos

Para poder limitar la elección de parámetros usados en el algoritmo de filtración de datos, son necesarios datos independientes. Por ejemplo, una fuente que se debería considerar son datos sobre viajes de buques

que los observadores no acompañaron. Un análisis de los datos de patrones temporales en los observadores disponibles para viajes, y los buques en espera de observadores, por separado de los viajes realizados, podría brindar información sobre las preferencias de los buques con respecto a observadores particulares. Las fuentes de datos existentes que no se están usando en la actualidad para la filtración de datos incluyen la base de datos del PIR sobre infracciones, que podría ser usada para elaborar un índice de la probabilidad de que un capitán o buque intente influenciar a un observador para reportar datos erróneos. Los tipos de datos que se podrían recolectar (o que ya existen, pero no son usados) para brindar información independiente sobre los observadores que podrían estar reportando datos erróneos deberían ser discutidos en una reunión especial del Consejo Científico Asesor.