

FISHGLOB_data: un conjunto integrado de datos relacionados a la biodiversidad de peces a partir de muestreos científicos de arrastre de fondo.

Aurore A. Maureaud, Juliano Palacios-Abrantes, Zoë Kitchel, Laura Mannocci, Malin L. Pinsky, Alexa Fredston, Esther Beukhof, Daniel L. Forrest, Romain Frelat, Maria L. D. Palomares, Laurene Pecuchet, James T. Thorson, P. Daniël van Denderen & Bastien Mérigot

06/2024

- **NOTA: esta es una traducción al español del artículo:** *FISHGLOB_data: an integrated dataset of fish biodiversity sampled with scientific bottom-trawl surveys*, escrito por Aurore A. Maureaud y colaboradores y que se puede encontrar en la siguiente liga. No cuenta con la sección de Referencias.

– Traducido por Juliano Palacios. Revisión técnica de Guillermo Palacios.

- **Forma de Citar:** Maureaud, A. A., Palacios-Abrantes, J., Kitchel, Z., Mannocci, L., Pinsky, M. L., Fredston, A., Beukhof, E., Forrest, D. L., Frelat, R., Palomares, M. L. D., Pecuchet, L., Thorson, J. T., Denderen, P. D. van and Mérigot, B., 2024. FISHGLOB_data: an integrated dataset of fish biodiversity sampled with scientific bottom-trawl surveys. *Scientific Data*, 11 (1), 24.

Resumen

Los muestreos científicos a lo largo de las plataformas continentales y pendientes marinas que utilizan red de arrastre de fondo son programas de monitoreo ecológico destinados a estudiar las comunidades marinas asociadas con el fondo marino. Estos muestreos informan sobre la ocurrencia de taxones, su abundancia y/o peso en el espacio y el tiempo, y contribuyen al manejo pesquero, así como a la investigación de poblaciones ecológicas y de la biodiversidad. Los muestreos de red de arrastre de fondo se llevan a cabo en todo el mundo y representan una oportunidad única para comprender la biogeografía oceánica, la macroecología y el cambio global. Sin embargo, combinar estos datos con el fin de realizar análisis entre ecosistemas significa un desafío. Aquí se presenta una base de datos integrados provenientes de 29 muestreos con red de arrastre de fondo en las aguas nacionales de 18 países, que cubren un total de 2,170 taxones de peces y 216,548 lances, realizados entre 1963 y 2021. Los datos, estandarizados y pre-procesados, se encuentran disponibles para consulta pública. En el presente ensayo se describen el procedimiento empleado para crear la base de datos, las alertas y los métodos de estandarización que se desarrollaron con el objetivo de apoyar a los usuarios a realizar análisis espacio-temporales, con huellas de muestreos regionales estables. El objetivo principal de esta base de datos es apoyar la investigación, la gestión y la conservación marina en el contexto del cambio global.

Introducción y Resumen

Los programas espacio-temporales de monitoreo de biodiversidad que informan sobre la ocurrencia de taxones, abundancia y biomasa, entre otros, son importantes para el estudio de la biodiversidad, la

distribución de las especies y sus respuestas al cambio global, así como para orientar la toma de decisiones en materia de conservación y gestión de recursos. Dichos programas varían ampliamente en sus estrategias de muestreo, extensiones temporales y espaciales y alcance taxonómico, de acuerdo a las diferencias en prioridades taxonómicas, financiamiento, accesibilidad y posibilidades de muestreo de los países. Esto ha dado lugar a un panorama heterogéneo de programas de monitoreamiento de la biodiversidad, con algunos grupos taxonómicos y ecosistemas muestreados de manera más sistemática que otros. Si bien los programas de larga duración desempeñan un papel clave en la creación de información científica básica, la formulación de políticas públicas y el estudio de la biodiversidad y de los ecosistemas ante el cambio climático, mantener dichos programas a largo plazo requiere recursos sustanciales que generalmente son difíciles de obtener. Es por eso que pocos programas de monitoreo muestrean regularmente conjuntos completos de especies en grandes áreas y durante largos períodos de tiempo, razón por la cual los datos existentes representan una importante oportunidad para la investigación y la gestión.

Desde su inicio en la década de los 60s, los muestreos científicos con red de arrastre de fondo (SBTS - *scientific bottom-trawl surveys*, por sus siglas en inglés) han representado una fuente de datos única entre las estrategias de monitoreo de biodiversidad espacio-temporales, en términos de longitud temporal y extensión espacial. Dichos muestreos fueron inicialmente diseñados para proporcionar estimaciones de abundancia y peso a las evaluaciones de las poblaciones de peces (*stock assesment*) y poder determinar niveles sostenibles de explotación pesquera. Los SBTS son independientes de los datos de captura pesquera comercial y recreativa (es decir, son datos independientes de la pesquería) y su objetivo es muestrear las comunidades marinas que habitan las profundidades con un equipo (red) de arrastre de fondo. A diferencia de la pesca de arrastre, los SBTS siguen protocolos previamente establecidos que incluyen un diseño de muestreo específico y una duración del arrastre baja y estandarizada, garantizando así un muestreo homogéneo de la región y de las comunidades de animales marinos [12]. Una síntesis global, cuyo objetivo es muestrear comunidades marinas demersales (i.e., que viven cerca del fondo marino) que habitan plataformas y taludes continentales en distintas partes del océano, identificó cerca de 100 SBTS activos [13]. Los SBTS son notables por tener series temporales largas, un muestreo regular (normalmente una vez al año o más), una amplia cobertura espacial y alta diversidad de taxones identificados. El fuerte vínculo entre los SBTS y la gestión y actividad pesquera ha asegurado un financiamiento relativamente estable con algunos programas muestreando regiones específicas de manera continua por más de sesenta años (por ejemplo, el Mar del Norte y el noreste de los Estados Unidos de América). Además de recolectar datos biológicos cruciales para la evaluación de poblaciones de peces, los SBTS también han demostrado ser valiosos para el estudio de la ocurrencia de especies y las redes tróficas dentro del área muestreada [14-15], así como para identificar alteraciones a largo plazo en la biodiversidad y las especies de comunidades marinas, debido al cambio climático [17-20].

Los SBTS regionales se llevan a cabo en varias partes del mundo, por lo que representan una oportunidad única para estudiar la biogeografía y la macro ecología marina. Particularmente en lo que respecta a comprender la dinámica de distribución de las especies en el espacio y el tiempo [13][21], así como para efectuar comparaciones y síntesis entre ecosistemas [22]. Estos esfuerzos requieren integrar múltiples SBTS, lo que crea nuevas oportunidades de investigación. Iniciativas pasadas que combinaron distintos SBTS, han ampliado nuestro conocimiento sobre los impactos de la pesca en la abundancia de especies sensibles [23], cambios en estructuras tróficas [24][25], patrones fisiológicos macro ecológicos de larga escala de las especies, cambios en la distribución de éstas [29], dinámicas de biodiversidad y abundancia de comunidades naturales, [30] así como las respuestas a los cambios de la temperatura en la composición de las mismas [32][33].

Sin embargo, combinar conjuntos de datos de muestreo es un desafío central en la síntesis ecológica. Es ampliamente reconocido que las bases de datos ecológicos (siendo cada una de ellas una representación irremplazable de un ecosistema en un lugar y tiempo únicos) rara vez se recopilan, comparten y mantienen en un formato ideal para uso y re-uso público [34]. Esto sucede incluso cuando se requiere que los datos se compartan públicamente,[36] en parte debido a los desafíos prácticos de crear repositorios estables de datos y hacer que los flujos de trabajo de codificación sean completamente reproducibles [37]. Entre los conjuntos de datos ecológicos, las SBTS son difíciles de combinar debido a los diferentes niveles de experiencia de las agencias que colectan las informaciones, la capturabilidad de las especies, la falta de metadatos, los diferentes formatos en que se encuentran los datos, las diversas unidades de las variables registradas, los diferentes procesamientos de datos crudos disponibles para los usuarios y de un esfuerzo de muestreo desigual en tiempo y espacio. Existen ensayos previos que han disponibilizado datos de muestreo (<https://datras.ices.dk/>,

<https://oceanadapt.rutgers.edu/>, <https://apps-st.fisheries.noaa.gov/dismap/DisMAP.html>), pero las SBTS aún no se han reconciliado formalmente en todos los continentes, no se han documentado en un formato fácil de usar ni puesto a disposición para la comunidad científica en general.

Las soluciones para avanzar hacia conjuntos de datos ecológicos mejores y más accesibles incluyen prácticas de ciencia abierta que fomenten la reproducibilidad computacional como FAIR (*Findable, Accessible, Interoperable, and Reproducible*), permitan trabajar en equipos grandes y diversos donde los miembros aporten conocimientos y perspectivas únicas y lleven a publicar conjuntos de datos diseñados para su re-utilización y vinculados a repositorios en línea. Aplicar estas soluciones para sintetizar conjuntos de datos de SBTS proporcionaría un servicio invaluable a la comunidad científica, ya que esta colección dispar de muestreos independientes podría convertirse en una de las fuentes de datos más grandes referentes a la ecología marina.

En este artículo se presenta FISHGLOB_data, un conjunto de datos pre-procesados de comunidades de peces que compila 29 SBTS regionales de plataformas continentales y taludes de América del Norte y Europa. Se detallan los pasos metodológicos diseñados para desarrollar este conjunto de datos espacio-temporales con el objetivo de lograr una mayor transparencia y reproducibilidad Fig. 1. A partir de la experiencia previa de los autores con datos de SBTS, se produce un conjunto de datos que pueden ser utilizados en diversas aplicaciones relacionadas con poblaciones y diversidad comunitaria, o la presencia-ausencia, abundancia y biomasa de taxones. También se presenta la extensión de la base de datos, así como información relevante para usuarios que estén interesados en utilizarla para fines de investigación científica.

En comparación con los recursos disponibles ya mencionados, FISHGLOB_data proporciona la primera compilación mundial de datos públicos de SBTS que cruza continentes en un formato fácil de usar. FISHGLOB_data facilita el uso de SBTS para la comunidad mediante la aplicación de métodos novedosos que estandarizan la huella espacio-temporal de cada región de estudio y localizan cambios en la identificación de especies, cuestiones que han sido preocupaciones y desafíos persistentes en el análisis de SBTS y más allá.

Al hacer uso de un flujo de trabajo y códigos abiertos, los usuarios pueden actualizar la base de datos con SBTS más recientes, modificar, ajustar y/o reproducir los métodos de procesamiento, así como proporcionar retroalimentación según corresponda. Esta base de datos complementa el inventario previo de metadatos existentes de SBTS al proporcionar una versión amigable para el usuario de los SBTS disponibles públicamente. Muchos otros SBTS siguen siendo de índole particular. Mientras que el inventario de metadatos de SBTS publicado anteriormente mejoró la localización y accesibilidad de dichos muestreos, la base de datos presentada en este artículo mejora su uso general, la interoperabilidad y la reproducibilidad, avanzando así hacia los principios FAIR [42]. De manera más general, este artículo contribuye a las prácticas de ciencia abierta, especialmente en torno a métodos y recursos[34][43] para la comunidad científica ecológica y pesquera que utilizan este tipo de datos en su investigación.

Los métodos e infraestructura presentados aquí son lo suficientemente generales como para integrar otros SBTS que existen pero que pueden no estar disponibles públicamente [13]. Por ejemplo, el programa de muestreo de arrastre de fondo MEDITS (que agrupa SBTS del norte del Mediterráneo realizados anualmente desde 1994 [44]) y el muestreo de arrastre de fondo de Islandia realizado desde 2005 [45] no están disponibles públicamente. En esfuerzos separados a este artículo, estos y otros estudios han sido integrados siguiendo la metodología descrita en este trabajo, después de obtener permiso de los proveedores de datos y expertos regionales e instituciones. Maureaud et al. [13] proporciona metadatos y contactos para varios estudios que no están disponibles públicamente. El establecimiento de métodos estandarizados de procesamiento de datos facilita las colaboraciones incluso cuando los datos brutos de los estudios no están disponibles públicamente, ya que los datos pueden ser accesibles para proyectos de investigación específicos. Una metodología estandarizada y una red de usuarios también ayudan a mejorar la visibilidad de los esfuerzos de estudio existentes, generar nuevas oportunidades científicas y colaboraciones potenciales, y facilitar la transferencia de conocimientos entre comunidades científicas y países.

Metodos

En esta sección se explican los pasos metodológicos desarrollados y empleados para producir los datos de FISHGLOB_data. Primero, se estandarizó y armonizó cada SBTS en un formato único (Fig. 1, (I)). Segundo, se desarrolló una metodología de etiquetado por SBTS para informar sobre posibles armonizaciones taxonómicas adicionales y asegurar una huella espacio-temporal consistente entre ellos (Fig. 1, (II)). Tercero, se produjeron resúmenes de los muestreos y entre muestreos para la validación técnica del procesamiento de datos (Fig. 1, (III)). Finalmente, se presentan instrucciones para realizar investigaciones utilizando la base de datos (Fig. 1, (IV)).

Compilación de datos de encuestas públicas

Los SBTS son programas regionales que muestrean comunidades demersales marinas que habitan en las plataformas continentales y taludes. Hoy en día muchos de estos muestreos son de acceso abierto: los europeos están disponibles a través del Consejo Internacional para la Exploración del Mar (*ICES*, por sus siglas en inglés) y el Instituto de Investigación Marina (*IMR*, por sus siglas en inglés), y los muestreos de Estados Unidos y Canadá están disponibles a través de la Administración Nacional Oceánica y Atmosférica estadounidense (*NOAA*, por sus siglas en inglés) y Pesca y Océanos Canadá (*DFO*, por sus siglas en inglés). Con base en esfuerzos regionales previos que integraron SBTS (<https://oceanadapt.rutgers.edu/>, <https://apps-st.fisheries.noaa.gov/dismap/DisMAP.html>, <https://datras.ices.dk/>) [26][28][29][33][46][47], se cumplieron 29 SBTS de acceso abierto desde 1963 hasta 2020 (Tabla Suplementaria S1, Fig. 2) que muestrean comunidades demersales desde las plataformas continentales subtropicales hasta los taludes polares marinos de América del Norte (Estados Unidos y Canadá) y Europa.

Estandarización de datos

Aunque los SBTS incluyen gremios ecológicos similares (peces y macroinvertebrados localizados en el fondo marino), con equipos de muestreo (es decir, redes de arrastre de fondo), también existen muchas diferencias en los formatos de datos, las variables reportadas, los métodos de muestreo, el tipo preciso de equipo, estación y ubicación de muestreo, así como en otros aspectos. Aquí se desarrolló un proceso de estandarización aprovechando el conocimiento y los métodos de expertos de SBTS para facilitar la investigación ecológica comparativa entre regiones de estudio. Investigadores que utilicen estos datos deben tener en cuenta que aun existen diferencias entre los muestreos que no pueden abordarse de manera fácil y/o sistemática. Se estandarizó cada conjunto de datos de manera individual a nivel de lance. Un “lance” es un evento único de pesca que constituye la unidad de muestreo; y una “estación” es el lugar donde se realiza el lance. Cada SBTS registra las capturas de arrastre de fondo por lance. La estandarización de muestreos se hizo de la siguiente manera: (i) se eliminaron datos que no pasaron los filtros de calidad, o que no estaban relacionados con las variables focales; (ii) cuando posible, se estimaron las variables focales faltantes a partir de la información disponible; (iii) se armonizó la taxonomía de acuerdo a una nomenclatura consistente; y (iv) se convirtieron los valores a unidades estandarizadas. A continuación se explica cada paso de manera detallada (Fig. 1).

Filtros de calidad

Se eliminaron todos los lances que no fueron considerados válidos según las “banderas” de control de calidad de cada muestreo (por ejemplo, lances en los que se indicaba que la red había fallado). Para ser incluido en la base de datos, un lance necesitaba tener geolocalización precisa, fecha, identificación taxonómica, abundancia y/o peso de taxones, duración del lance y/o área barrida, y equipo utilizado. La duración del lance y el área barrida son *proxies* del esfuerzo de muestreo en términos de tiempo de muestreo y área cubierta por el arrastre de fondo. Solo se mantuvieron los lances muestreados con equipos de arrastre de fondo tipo *otter-trawl* mientras que los muestreos realizados con redes de arrastre de tipo viga fueron retirados ya que muestrean comunidades demersales con una capturabilidad y detectabilidad diferentes [13][48] (Tabla

Suplementaria S1). Los muestreos con redes de arrastre de viga se centran en partes más someras de la plataforma continental y se realizan a una escala menor que la mayoría de los muestreos con redes de arrastre de fondo tipo *otter-trawl*. Compilar una base de datos de redes de arrastre de viga sería un esfuerzo futuro interesante que requeriría un inventario significativo y la creación de consorcios entre regiones. Más adelante se proporcionan métodos adicionales de control de calidad en la sección de *Validación Técnica*.

Variables focales

El esfuerzo se centró en asegurar que la información esencial para la investigación ecológica de poblaciones y comunidades para todos los lances estuviera disponible. Esta información incluyó variables de geocalización del muestreo (latitud, longitud, profundidad, número de estación), fecha de muestreo (año, mes, día), área barrida (área superficial muestreada), duración del lance (tiempo total de muestreo), y capturas para cada taxón de peces muestreado (abundancia y/o peso). Se seleccionaron muestreos que reportaran todos los individuos de peces, excluyendo aquellos que se centraron únicamente en las especies comerciales más importantes. La definición y visión general de estas y otras variables se detallan en la Tabla Suplementaria S2.

Algunos muestreos sólo registraron capturas específicas de taxones en abundancia (es decir, número de individuos), mientras que otras lo hicieron sólo en peso. Se hizo un esfuerzo para calcular las unidades de abundancia y biomasa faltantes para cada taxón y encuesta a partir de la información obtenida. Para los muestreos de *ICES*, se estimó el peso por taxón por lance utilizando datos de longitud de los datos brutos del muestreo y los coeficientes de relación longitud-peso ‘*a*’ y ‘*b*’ obtenidos de *FishBase*[49], utilizando la ecuación $w = al^b$, donde *w* es el peso, y *l* la longitud. La información de abundancia y longitud para varios muestreos de la *NOAA* no está (públicamente) disponible, por lo tanto, la base de datos sólo contiene peso de taxones para estos muestreos.

Taxonomía

La información taxonómica en cada muestreo fue curada utilizando el *World Register of Marine Species* (*WoRMS*), una clasificación y catálogo taxonómico marino ampliamente utilizado y reconocido [50]. Si bien *WoRMS* integra *FishBase* como una columna vertebral taxonómica de referencia, se le utilizó porque algunas STBS ya usan sus códigos para reportar taxones. Sólo se conservaron los taxones de peces muestreados por los muestreos seleccionados porque los macroinvertebrados no fueron muestreados de manera tan exhaustiva o consistente en todos los SBTS. Esto se hizo seleccionando taxones incluidos en las siguientes clases taxonómicas: Elasmobranchii, Holocephali, Chondrostei, Holostei y Teleostei. En los muestreos de *ICES*, se proporciona el ‘*AphiaID*’ para cada taxón, por lo que se determinó cada nombre científico correspondiente y su clasificación utilizando el paquete “*worms*” de *R* [51]. Todos los demás muestreos solo proporcionaron el nombre científico. En estos casos, cada nombre se verificó con la columna vertebral taxonómica de *WoRMS* de noviembre de 2023, y fueron todos actualizados conforme a su nombre válido y clasificación más reciente. Esto aseguró que los nombres dados en el pasado y que ya no se consideran válidos se armonizaran con los nombres recientes. Para algunos muestreos específicos, se corrigieron inconsistencias en la identificación taxonómica que los autores conocían. También se asoció cada nombre científico con su ‘*AphiaID*’ válido y rango de clasificación. Los nombres y códigos de los datos brutos de la encuesta se conservaron en la base de datos para facilitar la homogeneización futura (los nombres de las columnas precedidos por ‘*verbatim*’) [52]. Cuando un nombre no coincidía directamente con la columna vertebral taxonómica de *WoRMS*, se utilizó la herramienta de coincidencia difusa de *WoRMS* y se seleccionaron las coincidencias de manera manual. Los nombres que no se pudieron emparejar se eliminaron de la base de datos. Adicionalmente, se incluyó el ‘*SpecCode*’ de *FishBase*[49] cuando lo había, utilizando su versión actualizada, disponible a través del paquete “*rfishbase*” de *R* [53]. Este enlace se incluyó para facilitar la interoperabilidad con *FishBase* ya que este conjunto de datos se utiliza ampliamente en estudios ecológicos. *WoRMS* sigue la columna vertebral taxonómica de *FishBase*, por lo que la coincidencia de nombres científicos aceptados puede ser suficiente para permitir la interoperabilidad entre SBTS y *FishBase*.

Unidades estandarizadas

Los datos crudos de cada muestreo no incluyen las mismas métricas de abundancia y no todos registran el área barrida por cada lance. La duración del lance y el área barrida son necesarias para convertir las abundancias crudas en métricas estandarizadas comparables entre muestreos, como la captura por unidad de esfuerzo (CPUE) y la captura por unidad de área (CPUA) [54]. Para calcular CPUE y CPUA, se dividió el peso y/o la abundancia por la duración del lance y el área barrida, respectivamente. Para varios muestreos de NOAA la variable de área barrida estaba directamente disponible o ya se utilizaba para estandarizar la abundancia; sin embargo ICES, DFO e IMR no siempre contaban con dicha información. Para calcular el área barrida cuando esta variable no se reportaba directamente, se utilizó la distancia de muestreo y la apertura del equipo (cuando disponibles), o, como alternativa, la duración del lance, la velocidad del buque y la apertura del equipo [54–56]. Para algunos lances, el área barrida no se pudo calcular debido a la falta de datos. En estos casos se ajustó un modelo lineal con el área barrida como variable de respuesta y datos del lance como posibles variables predictoras, utilizando combinaciones de duración del lance, profundidad de muestreo, longitud del barrido, país y buque de muestreo. Los valores faltantes de área barrida fueron proyectados a partir del modelo lineal específico para cada muestreo. Todos estos modelos se detallan en el script ‘*get_dstras.R*’ para los muestreos de ICES. Cuando no se pudo obtener la apertura del equipo a nivel de lance para calcular áreas barridas, se utilizó una apertura estandarizada comunicada por expertos en SBTS obtenidas del DFO. En la Tabla Suplementaria S3 se indican más detalles específicos de la fuente de unidades estandarizadas.

En el conjunto final de datos llamados *FISHGLOB_data*, se proporcionan la(s) estimación(es) de abundancia y/o peso, la abundancia y/o peso estandarizado por la duración del lance (CPUE), y la abundancia y/o peso estandarizado por el área barrida (CPUA). Como se mencionó anteriormente en *Variables focales*, no todas las variables están disponibles para cada muestreo y algunas son estimaciones que utilizan conversiones de longitud-peso en lugar de mediciones a bordo. Se proporcionan las variables focales agregadas por taxón por lance en la base de datos final, que es el nivel de agregación que permite una mejor integración entre todas las regiones de muestreo.

Métodos de estandarización opcionales

Se incluyeron procedimientos adicionales de estandarización de datos que los usuarios pueden implementar según sus necesidades de investigación. En particular, se implementaron dos procedimientos frecuentemente utilizados para la identificación de anomalías en los datos relacionados con (i) taxones, que pueden no estar consistentemente identificados a lo largo del tiempo por el muestreo; y (ii) para ubicaciones y años que caen fuera de una huella espacial que ha sido consistentemente muestreada a lo largo del tiempo para cada combinación de encuesta y temporada disponible.

Identificación de anomalías taxonómicas temporales. Si bien la estandarización taxonómica resuelve muchas de las discrepancias en las bases de datos, aun pueden persistir inconsistencias. Por ejemplo, algunos muestreos han cambiado el nivel taxonómico en el que registran ciertos taxones a partir de guías de campo mejoradas que permiten la identificación de un taxón a nivel de especie, en lugar de género, como en años anteriores. La clasificación taxonómica fue actualizada para casos donde las anomalías era conocidas por los expertos en el SBTS, con el fin de mantener la consistencia a lo largo del tiempo (p.ej., mover especies de un grupo particular al rango de género). Sin embargo, aún pueden aparecer inconsistencias en los datos finales, razón por la cual se identificó la presencia de taxones en cada SBTS y año que pudieran requerir atención adicional. Para que fuera identificado, un taxón debería de estar presente en menos del 95% de los años muestreados o haber tenido un cambio de presencia/ausencia (o viceversa) en menos de cuatro veces a lo largo de la serie temporal. Este método se realizó en el ‘*nombre_aceptado*’, que ya incorporaba la armonización taxonómica en ‘*nombre_verbatim*’ (Tabla Suplementaria S2). Estos criterios fueron diseñados para detectar especies que pudieran haber cambiado de convención de nomenclatura pero que no fueron identificadas durante la armonización taxonómica. Dependiendo del uso previsto, los usuarios pueden tener que verificar estos registros con expertos familiarizados con el SBTS específico.

Banderas de muestreo espacio-temporales. Aunque los SBTS proporcionan una amplia cobertura de datos espacio-temporales, la huella espacial de cada muestreo a menudo varía a lo largo del tiempo debido a restricciones logísticas o falta de oportunidades en algún año en particular. Se estableció un sistema de alerta para identificar alguna huella espacial que haya sido muestreada consistentemente a lo largo del tiempo. Se siguieron dos métodos principales para asegurar una cobertura espacial y temporal consistentes: (i) un método de cobertura temporal y espacial en cuadrícula; y (ii) un método de estandarización desarrollado previamente para la base de datos BioTime [57–59] (Fig. 3a–c). Para (i), se dividieron los lances por separado en una cuadrícula hexagonal de área igual para cada unidad temporal en cada muestreo (es decir, una unidad temporal de muestreo es un código definido por la combinación de ‘muestreo’ y ‘temporada’ y/o ‘trimestre’ de la Tabla Suplementaria S2, como NS-IBTS trimestre 1). Se utilizó el paquete “dggridR” de R [60] con resoluciones espaciales de 7 y 8, que corresponden a celdas hexagonales de aproximadamente 23,320 km² y 7,770 km², respectivamente. Cada combinación de una celda en la cuadrícula y el año en que se muestreó se denominó “celda de cuadrícula-año”. Luego se identificó la base más grande de celdas y años consecutivos de manera que todas las celdas de la cuadrícula seleccionadas fueron muestreadas en todos los años (umbral de 0% de ausencia). También se identificó la base más grande de celdas y años de manera que tuvieran un máximo de 2% de todas las celdas de la cuadrícula en cada año (umbral de 2% de ausencia). Se identificaron los lances de muestreos que no estaban en los conjuntos básicos de 0% de resolución 7, básico de 0% resolución 8, 2% resolución 7, y 2% resolución 8 (ver diagramas conceptualizados en Fig. 3a,b). Se debe tener en cuenta que para estos dos procesos de maximización (0% y 2%), tanto las celdas como los años podrían estar señalados. Para (ii), se dividieron los lances en una cuadrícula pre-diseñada para cada unidad temporal por muestreo [59,61]. La resolución de la cuadrícula es específica para cada muestreo y se estableció como 1/5 del rango latitudinal y longitudinal. Cada celda incluía las ubicaciones de los lances (combinación de latitud y longitud). Primero, se identificó un lance si cualquier celda de la cuadrícula correspondiente no fuera muestreada al menos 4 veces en un año individual [61]. Una celda que presenta un mínimo de 4 eventos de muestreo garantiza que ésta no esté sesgada por ruido de muestreo generado por haber pocos eventos. Segundo, se eliminaron las celdas (que incluyen al menos 4 ubicaciones de muestreo por año) que incluían menos de 10 años de muestreo Fig. 3c. Se consideran 10 años de muestreo como la serie temporal mínima necesaria para poder realizar análisis ecológicos temporales en este método [59,62], y más generalmente en ecología [63]. En total, se asoció cada lance de cada muestreo con cinco alertas de muestreo diferentes: 4 para el método (i) y 1 para el método (ii). Los usuarios finales pueden optar por utilizar uno de estos sistemas para estandarizar la huella espacial de los muestreos. Los diferentes métodos eliminan lances algo diferentes y resultan en conjuntos de datos básicos que son consistentes espacialmente a través del tiempo en diferentes grados Fig. 4.

Datos

La base de datos final, *FISHGLOB_data*, incluyó 216,548 lances a través de 29 muestreos, registrando la presencia, abundancia y/o peso de 2,170 taxones únicos de peces. La base de datos está disponible para descarga desde el repositorio de GitHub; también se puede descargar desde Zenodo [64]. Todas los muestreos están identificadas con un código único (‘survey’, Tabla Suplementaria S2), con la lista completa detallada en la Tabla Suplementaria S1. Todas las variables focales relevantes relacionadas con la geolocalización (‘latitude’, ‘longitude’, ‘stratum’, ‘station’, ‘stat_rec’, ‘continent’) y la fecha de muestreo (‘year’, ‘month’, ‘day’, ‘season’, ‘quarter’) se informan cuando corresponde (Tabla Suplementaria S2). Para cada muestreo se reportaron las abundancias y pesos por taxón a partir de los datos crudos y estandarizados por la duración del lance y el área barrida, cuando disponible (Tabla Suplementaria S2). La base de datos incluyó campos relacionados con la taxonomía extraída de WoRMS (‘aphia_id’, ‘accepted_name’ y clasificación relacionada), FishBase (‘SpecCode’), y los nombres y códigos reportados originalmente en los conjuntos de datos crudos (‘verbatim_name’ y ‘verbatim_aphia_id’, respectivamente).

El alcance espacial de los muestreos abarcó casi continuamente tres plataformas continentales del hemisferio norte, desde la costa noroeste del Pacífico hasta el noreste y el noroeste del Atlántico (Fig. 2). Los SBTS variaron desde series temporales relativamente largas (por ejemplo, el NS-IBTS del Mar del Norte y el NEUS del noreste de EE.UU. comenzaron en la década de 1960), hasta series temporales cortas (por ejemplo, el DFO-SOG del Estrecho de Georgia comenzó en 2012 y el NIGFS del Mar de Irlanda en 2005, Tabla

Suplementaria S1, (Fig. 2). La mayoría de los muestreos suceden de manera anual, pero en algunos casos éstos suceden cada dos años (por ejemplo, DFO-SOG, DFO-WCVI, DFO-QCS, DFO-HS, DFO-WCHG) o cada tres (por ejemplo, WCTRI). Los muestreos también variaron en su esfuerzo intra-anual; el campo 'survey_unit' (Tabla Suplementaria S2) indica qué encuestas operaron en múltiples estaciones (por ejemplo, BITS, NS-IBTS, SWC-IBTS, NEUS, SEUS, SCS, SP-ARSA). En términos de extensión espacial y duración de las series temporales, los muestreos más extensos fueron NEUS (noreste de EE.UU.), NS-IBTS (Mar del Norte), Nor-BTS (Mar de Barents, Noruega), GOA (Golfo de Alaska) y GSL-S y GSL-N (Golfo de San Lorenzo, norte y sur).

Figura. 4 Resumen de los resultados de las alertas para cada muestreo. (a) Porcentaje de taxones identificados por muestreo; (b) Porcentaje de lances identificados por muestreo relativo a los métodos de estandarización espacio-temporal; (c) porcentaje de lances identificados por método relativo a todos los muestreos. Es recomendable retirar los lances aquí identificados para tener una huella espacio-temporal limpia.

Las bases de datos integradas también incluyen indicadores taxonómicos y espacio-temporales para ayudar a los usuarios a realizar estudios espacio-temporales con taxones consistentes, así como huellas espaciales regionales coherentes (ver los campos al final de la Tabla S2 del material suplementario). En promedio, el 7,8% de los taxones de cada muestreo fueron identificados como potencialmente problemáticos (Fig. 4a). En algunos muestreos se identificó una proporción relativamente alta de taxones problemáticos, especialmente en series temporales más cortas. Este fue el caso del muestreo del Mar de Irlanda (NIGFS), que cuenta con 16 años de muestreo y en el cual 58% de los taxones fueron identificados como potencialmente problemáticos (56 de los 96 taxones muestreados). Sin embargo, varios estudios presentaron menos del 1% de ese tipo de taxones, como fue el caso de la plataforma de Scotian Shelf, el norte del Golfo de San Lorenzo, el este del mar de Bering, el Banco Porcupine, la costa oeste de Escocia, el mar de Barents, el sureste de EE.UU. y las regiones de estudio del mar Céltico y el golfo de Vizcaya (SCS, GSL-N, EBS, SP-PORC, SWC-IBTS, Nor-BTS, SEUS y EVHOE, respectivamente).

Se detallaron los resultados de los indicadores espacio-temporales utilizando dos métodos diferentes y distintas escalas espaciales para el mar Céltico y el golfo de Vizcaya (EVHOE) como ejemplo de opciones para recortar lances y garantizar una huella espacial de estudio consistente a lo largo del tiempo (Fig. 3). Cuando se aplicó un método de estandarización a dicho estudio, algunos lances fueron identificados debido a que ciertas celdas de la cuadrícula y años no fueron muestreados de manera consistente en el tiempo o el espacio. Con el primer método de recorte, se identificó el 14,7% de todos los lances del estudio EVHOE bajo un umbral del 0% de datos faltantes, y el 3,3% bajo un umbral del 2% en la resolución de celda de cuadrícula más gruesa (Fig. 3d,g). A una resolución espacial más fina, el porcentaje de lances identificados fue mayor: 35,9% bajo el umbral del 0% y 14,4% bajo el umbral del 2% (Fig. 3e,h). El segundo método de recorte desarrollado para la base de datos BioTIME resultó en 6,4% de los lances identificados para su eliminación (Fig. 3i).

El porcentaje de lances identificados debido a la cobertura espacio-temporal, que no fue constante, varió entre muestreos y métodos, lo que podría influir en los estudios de biodiversidad que utilizan métricas sensibles a la geo-localización y a las fluctuaciones temporales del estudio (Fig. 4b,c). El muestreo de invierno de la costa oeste de Escocia (SWC-IBTS, cuarto trimestre) tuvo la mayor proporción de lances marcados, lo que indica que su huella espacio-temporal no es muy constante (se identificaron entre el 23,1% y el 59% de los lances, dependiendo del método). Por el contrario, el estudio del este del mar de Bering (EBS) ha seguido un diseño de muestreo estricto con una regularidad notable, siendo que los métodos de estandarización identificaron muy pocos lances (de 0% a 8,9%). El enfoque más liberal fue el recorte con la resolución de celda de cuadrícula más gruesa (7) y la inclusión de un 2% de celdas-año faltantes (6,6% de lances identificados en promedio entre unidades de estudio, Fig. 4c), mientras que el más restrictivo fue el de mayor resolución de celda de cuadrícula (8) y el umbral del 0% (17,9% de lances identificados en promedio entre unidades de estudio).

Validación técnica

Pasos generales de CC/AC. A todos los SBTS se les hicieron controles de calidad (CC) y aseguramiento de la calidad (CA) que incluyeron la verificación de la coherencia taxonómica y pasos de homogenización

para garantizar que los nombres taxonómicos fueran válidos y lo más precisos posible. El proceso de CC/AC incluyó la consulta de manuales y expertos para incorporar el conocimiento específico de cada estudio en la base de datos. Esta incorporación se realizó de manera iterativa a medida que se identificaban problemas, como en el caso de la investigación de los indicadores taxonómicos. Los indicadores taxonómicos han sido un método útil para identificar divisiones y agrupaciones taxonómicas a lo largo del tiempo, así como otras inconsistencias en la identificación de especies en campo. Sólo se incluyeron lances con artes de pesca consistentes dentro de cada región de muestreo y dentro de la(s) temporada(s) o mes(es) del año mejor muestreados (Tabla S1 del material suplementario). Otros controles estándar incluyeron verificar que las abundancias y los pesos fueran nulos, positivos o registrados como datos faltantes, pero nunca negativos. Los valores atípicos y los rangos de valores no fueron restringidos en la base de datos para maximizar su uso, ya que los usuarios pueden estar interesados en una parte específica de la base de datos y aplicar sus propios filtros. Sin embargo, los rangos de valores y los valores atípicos pueden identificarse en los resúmenes de los estudios descritos a continuación.

Pasos de específicos CC/AC por estudio. Se llevaron a cabo varios pasos específicos de control de calidad para los SBTS obtenidos de ICES. Los aspectos clave de validación técnica incluyeron:

- Comparación de la estimación de área muestreada con la desarrollada por ICES (65) (Figura S1 del material suplementario). Se observaron algunas diferencias donde se identificaron ciertas incertidumbres (especialmente en el Mar Báltico, BITS). Sin embargo, los valores de área muestreada en general se mantuvieron dentro del mismo rango en ambas fuentes, validando los métodos.
- Comparación de las abundancias y pesos estimados a partir del cálculo de abundancia por talla con las abundancias y pesos reportados a nivel de taxa-lance. Se encontraron correlaciones muy altas para las abundancias (todas cercanas a 1), mientras que los pesos mostraron algunas diferencias. Estas diferencias pueden deberse a inconsistencias no resueltas en los conjuntos de datos, como unidades mal reportadas para pesos o mediciones de longitud. ICES ya ha identificado este problema en los datos históricos del Mar Báltico y del Mar del Norte (BITS y NS-IBTS, respectivamente), y sigue siendo un problema en estos conjuntos de datos. Por esta razón, se decidió incluir los pesos re-calculados a partir de la abundancia por talla en los conjuntos de datos de ICES.

Resúmenes de los muestreos. Se generaron resúmenes utilizando RMarkdown para cada muestreo de arrastre de fondo mostrando las características del muestreo, la distribución de variables y la estandarización de los resultados. Estos resúmenes se emplearon para realizar controles de calidad visuales en los datos generados en cada muestreo y fueron revisados cuidadosamente antes de finalizar la base de datos. Este paso de control de calidad complementó los filtros de calidad aplicados al inicio (ver sección de *Estandarización de Datos*), y estos resúmenes deberían ser examinados por los usuarios finales para evaluar la idoneidad de los datos para sus necesidades.

Cada resumen incluye información clave del muestreo (proveedor de datos, período temporal, región de estudio) y el código en R utilizado para la limpieza de datos, garantizando transparencia en la documentación de metadatos y el procesamiento de los estudios. Asimismo, se proporciona una visión general de la información del muestreo que incluye:

1. Un fragmento de la tabla de datos del muestreo
2. El número de lances por año para verificar y evaluar la consistencia en el muestreo desde el inicio hasta su final
3. La distribución de las variables de muestreo por año (por ejemplo, área barrida, duración del lance, profundidad de muestreo) para verificar la consistencia en el rango de la variable de muestreo a lo largo de los años
4. Las variables biológicas por año (por ejemplo, abundancia y peso, junto con sus unidades estandarizadas como CUPA y CPUE)
5. La distribución de los valores de abundancia y peso por año para identificar posibles valores atípicos en los datos a lo largo de los años muestreados

6. La relación entre las variables biológicas y el área barrida para verificar/ser conscientes del efecto del área barrida en la abundancia y/o biomasa a lo largo de los años muestreados
7. Las tendencias de abundancia o peso de los seis taxones más abundantes para verificar/ser conscientes de la tendencia de abundancia o peso. Estas tendencias se verificaron en comparación con tendencias conocidas de evaluaciones de stock disponibles u otras plataformas existentes que proporcionan tendencias temporales específicas por especie para las mismas regiones de muestreo y conjuntos de datos (<https://james-thorson.shinyapps.io/FishViz/>)
8. El mapa de las ubicaciones de los lances del muestreo para verificar/ser conscientes de la distribución espacial a lo largo de los años muestreados
9. Los resultados del método de marcado taxonómico para verificar/ser conscientes de los taxones que no se identifican como muestreados de manera consistente a lo largo de los años muestreados
10. Los resultados del método de marcado espacio-temporal para verificar/ser conscientes de la consistencia en la cobertura de datos espacio-temporales

Se generó un resumen general de *FISHGLOB_data* para verificar la distribución de algunas variables a lo largo de los muestreos. En particular, dicho resumen expone la distribución de las variables muestreadas por año. Esta visualización ayudó a comprobar si las variables se mantuvieron relativamente constantes durante el período de muestreo. Por ejemplo, no se esperaba un cambio significativo en el área barrida, la profundidad de muestreo o la duración del lance a lo largo del tiempo. Todos los resúmenes están disponibles en: https://github.com/AquaAuma/FishGlob_data/tree/main/summary y proporcionan una visión general de cada muestro. Se utilizaron estos resúmenes en las primeras etapas para detectar y corregir problemas de control de calidad en la base de datos procesado. Se recomienda que los usuarios consulten estos resúmenes antes de utilizar los datos en sus propios análisis.

Notas de uso

La base de datos espacio-temporal de comunidades de peces *FISHGLOB_data* fue diseñado para maximizar su utilidad para una variedad de propósitos. Sin embargo, una mala interpretación o un uso inadecuado de este conjunto de datos pueden generar resultados no deseados. Los usuarios deben ser plenamente conscientes de sus limitaciones. Se recomienda prestar especial atención a las siguientes categorías y puntos:

Elección de variables focales:

- **Variables de abundancia o peso:** No todos los muestreos incluyen datos de abundancia y peso. Elegir una u otra opción eliminará automáticamente varios muestreos. Los usuarios deben considerar cuidadosamente qué medida es más relevante para sus preguntas. Por ejemplo, indicadores de biodiversidad a nivel de comunidad pueden arrojar resultados significativamente diferentes si se calculan con abundancia en lugar de biomasa.
- **Estandarización de variables focales:** Los usuarios deben tener en cuenta que los métodos para calcular el área barrida en ICES DATRAS SBTS representan sólo una forma de hacerlo, y que desde el desarrollo de este conjunto de datos se han desarrollado otros métodos que están disponibles (<https://www.ices.dk>).
- **Presencia-ausencia de taxones:** Cada conjunto de datos de encuesta puede transformarse en un conjunto de datos de presencia-ausencia de taxones. Esto significa que cualquier taxón puede considerarse como ausente mismo que éste aparezca a lo largo del muestro pero que no se encuentre en un lance (aunque se presente en otra parte del muestreo). Esta suposición sólo es válida para los taxones identificados de manera consistente en todos los años. Es decir, un taxón descrito recientemente no puede considerarse ausente en años anteriores.

Filtros adicionales para variables de muestreo

- **Muestreos y estaciones:** Algunos muestreos incluyen diferentes estaciones (reflejadas en las variables ‘month’, ‘season’, ‘quarter’, ver Tabla Suplementaria S2), por lo que los usuarios deben decidir si considerar e incluir todas las estaciones para todos los muestreos. Los muestreos no suceden en las mismas estaciones en todos los años, necesariamente. Esto significa que la fenología del muestreo puede cambiar con el tiempo. Para algunos análisis, estos cambios fenológicos pueden ser problemáticos. Además, algunos muestreos se realizan en la temporada de invierno, lo que implica que el muestreo ocurre al final de un año X y al comienzo del año X+1. En tales casos, la variable ‘year’ no es la mejor manera de diferenciar los años de muestreo
- **Otras variables temporales:** El momento del día en que se realiza el muestreo puede afectar las comunidades de especies registradas. Los usuarios deben considerar si esto es relevante cuando utilizan la base de datos, especialmente en relación con la migración vertical diurna y las diferencias en las comunidades entre el día y la noche. Sin embargo, este aspecto no ha sido examinado en detalle en la base de datos, y la información sobre si el muestreo ocurrió de día o de noche no siempre está disponible.
- **Área espacial consistente:** Si el usuario requiere consistencia e a lo largo del tiempo, se recomienda filtrar los datos utilizando uno de los métodos de estandarización espacial incluidos en la base de datos (ver sección *marcado de estandarización opcional*) si el usuario requiere consistencia e a lo largo del tiempo. La variación en la huella espacial del muestreo puede afectar la composición de la comunidad observada, la distribución y la abundancia de especies y otras métricas biológicas calculadas a partir de SBTS. Sin embargo, también recomendamos que los usuarios consulten con expertos de cada muestreo para contextualizar mejor los cambios en la huella espacial con el tiempo, dependiendo del caso de uso, y asegurarse de que los filtros sean interpretados adecuadamente. Los contactos de expertos regionales se pueden encontrar en Maureaud et al. (13)
- **Extensión temporal entre encuestas:** Los muestreos no cubren series temporales idénticas, por lo que los usuarios deben explorar la extensión temporal de cada muestreo y decidir cuáles años y muestreos utilizar para sus preguntas de investigación. Por ejemplo, los años podrían limitarse a la serie temporal más reciente (2006-2020) para incluir todos los muestreos (26,28). Además, algunas regiones cuentan con años que no fueron muestreados, mientras que otros muestreos se realizaron en intervalos variables a lo largo del tiempo. Por ejemplo, el muestreo DFO-WCVI sólo se lleva a cabo cada dos años, mientras que el NS-IBTS se realiza dos veces al año (ver Tabla Suplementaria S1, Fig. 2).
- **Características del muestreo:** Todas las variables de muestreo pueden restringirse aún más, a ciertos rangos y filtrarse. Por ejemplo, es común excluir profundidades de muestreo menores a 20 metros para eliminar las comunidades muy cercanas a la costa (19).

Filtros y consideraciones taxonómicas adicionales

- **Taxones pelágicos:** Los muestreos SBTS capturan tanto taxones pelágicos como demersales, por lo que los investigadores deben considerar si incluir taxones pelágicos tiene sentido para sus preguntas. Los taxones pelágicos pueden identificarse vinculando FISHGLOB_data con FishBase (49) mediante el campo ‘SpecCode’ (ver Tabla Suplementaria S2). La capturabilidad de los peces pelágicos es mucho menor que los demersales en los muestreos de arrastre de fondo (54). Como alternativa, los investigadores podrían aplicar factores de corrección de capturabilidad por hábitat en la columna de agua para obtener estimaciones de abundancia o peso más consistentes dentro de los muestreos (27).
- **Informes taxonómicos inconsistentes/incompletos:** Algunos muestreos parecen no reportar mayoría de los taxones de peces de manera consistente a lo largo del tiempo, como PT-IBTS (bajo número de taxones por año), SP-ARSA (con bandera en el año 2018), SP-NORTH (1990-1992) y SP-PORC (año 2014).
- **Rango taxonómico:** Los taxones de peces incluidos en el conjunto de datos abarcan niveles taxonómicos superiores al nivel de especie (por ejemplo, familia, género), por lo que los usuarios deben decidir entre conservar todos los niveles o restringirse a ciertos rangos taxonómicos.

- **Taxones señalados:** Estos podrían filtrarse si los usuarios consideran que no deberían incluirse en la base de datos. Consultar con los contactos de los muestreos (13,46) también puede ser útil. Se recomienda abrir nuevos *issues* en el repositorio de GitHub (https://github.com/AquaAuma/FishGlob_data) (64) para documentar nueva información sobre taxones a incluir, excluir o marcar. Dependiendo de la fuente y el muestreo, los tratamientos taxonómicos recomendados no son los mismos, por lo que los usuarios deben verificar cuidadosamente el tratamiento taxonómico adoptado para su propio uso (23,68,69).
- **Nombre aceptado:** En algunos casos, el ‘verbatim_name’ o el ‘verbatim_aphia_id’ reportado en un muestreo puede no coincidir con el ‘accepted_name’, debido a cambios en los tratamientos taxonómicos o a errores de identificación en los muestreos, puede generar taxones duplicados. Por ejemplo, en los datos de SBTS del Mar del Norte, se asignó el género *Argentina* en lugar de los nombres reportados *Argentina Sphyraena* o *Argentina silus*, ya que se considera que la distinción entre estas especies no es suficiente para garantizar una identificación precisa. Sin embargo, el ‘verbatim_aphia_id’ si incluye la información taxonómica original reportada.

Escala de análisis

- **Escala espacial:** Los usuarios pueden trabajar a diferentes escalas, como las ubicaciones exactas del muestro a nivel de lance, la agregación dentro de una cuadrícula (dos resoluciones espaciales están disponibles en la base de datos) o la agregación a nivel de región de muestreo indicada en la columna ‘survey’.
- **Comparación entre muestreos:** Debido a las diferencias en las herramientas de muestreo y el diseño de los mismos, así como en la capacidad de captura de diferentes especies (por ejemplo, la capturabilidad varía), los datos de abundancia y peso no son directamente comparables entre muestreos, incluso después de aplicar correcciones por unidad de esfuerzo o por unidad de área.

Advertencias específicas por muestreo

Se ha creado una lista de aspectos específicos de cada muestro para advertir y asesorar a los usuarios que trabajan con estos datos. Estas notas son particularmente útiles durante el proceso de integración y pueden ayudar a los usuarios a comprender mejor la fuente de datos para su propio uso. Están disponibles en línea en: https://github.com/AquaAuma/FishGlob_data/tree/main/metadata_docs

Guía general

La base de datos presentada aquí constituye tan solo una de las formas de procesar los datos de SBTS, con el objetivo de maximizar su utilidad para diversos propósitos. Sin embargo, sigue siendo responsabilidad del usuario asegurarse de utilizar los datos de manera adecuada. Se ha buscado mantener la base de datos sin filtrar, lo que maximiza su aplicabilidad, pero no corrige las inconsistencias inherentes a los datos originales.

Se recomienda a los usuarios y expertos a proporcionar retroalimentación sobre el tratamiento de los datos para mejorar tanto la base como la metodología integrada a nivel intercontinental. Esto se puede hacer contactando directamente a los autores o enviando comentarios a través de *GitHub Issues*. Además, en GitHub se encuentra un descargo de responsabilidad con lineamientos sobre el uso de estos datos.

Citas y reconocimiento

Recomendamos encarecidamente a los usuarios que citen este artículo (Maureaud et al 2024), junto con las fuentes primarias de los estudios incluidos en SBTS, para reconocer el trabajo de las instituciones y expertos que publicaron las versiones originales de las bases de datos. Todas las fuentes están citadas en la Tabla Suplementaria S1. Además, este artículo debe ser citado para referenciar los métodos de integración desarrollados a fin de reunir las encuestas regionales.

Disponibilidad del código

El código utilizado para generar FISHGLOB_data a partir de las bases de datos originales puede consultarse en GitHub: https://github.com/AquaAuma/FishGlob_data; también se encuentra archivado en Zenodo para su descarga: <https://zenodo.org/records/10218308>.

Referencias

Ver referencias en texto original.

- **Forma de Citar:** Maureaud, A. A., Palacios-Abrantes, J., Kitchel, Z., Mannocci, L., Pinsky, M. L., Fredston, A., Beukhof, E., Forrest, D. L., Frelat, R., Palomares, M. L. D., Pecuchet, L., Thorson, J. T., Denderen, P. D. van and Mériqot, B., 2024. FISHGLOB_data: an integrated dataset of fish biodiversity sampled with scientific bottom-trawl surveys. *Scientific Data*, 11 (1), 24.
- Traducido por Juliano Palacios. Revisión técnica de Guillermo Palacios. *ChatGPT* y *Google Translator* fueron utilizados en la traducción de este artículo del inglés al español.