ELSEVIER

Data Article

Fisheries-dependent and -Independent data used to model the distribution of diadromous fish at-sea

Sophie A.M. Elliott [a,e,g,*], Noémie Deleys [a,b], Laurent Beaulaton [a,f], Etienne Rivot [a,e], Elodie Réveillac [d], Anthony Acou [a,c]

[a] *Management of Diadromous Fish in their Environment OFB, INRAE, Institut Agro, UPPA, 35042 Rennes, France*
[b] *French Research Institute for the Exploration of the Sea (IFREMER) Vigies, 44311, Nantes, France*
[c] *Centre d'expertise et de données sur le patrimoine naturel, Patrinat OFB-MNHN-CNRS-IRD, Station marine du Museum National d'Histoire Naturelle, 35800 Dinard, France*
[d] *Littoral, Environnement et Sociétés (LIENSs), UMR 7266, La Rochelle Université-CNRS, 17000 La Rochelle, France*
[e] *DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, Ifremer, INRAE, Rennes, France*
[f] *Service Conservation et Gestion Durable des Espèces Exploitées, DRAS, OFB, U3E, 35042 Rennes, France*
[g] *Game & Wildlife Conservation Trust, Salmon & Trout Research Centre, East Stoke, Wareham BH20 6BB, UK*

ARTICLE INFO

ABSTRACT

A database of 168 904 hauls covering the period from 1965 to 2019, from 46 surveys containing both fisheries-dependent (fishing vessels) and -independent data (scientific surveys) were collated from across the eastern Atlantic (Greater North Sea, Celtic Sea, Bay of Biscay and Iberian coast) and Metropolitan French Mediterranean waters. Data on diadromous fish (the European sturgeon (*Acipenser sturio*), allis shad (*Alosa alosa*), twait shad (*Alosa fallax*), Mediterranean twaite shad (*Alosa agone*), European eel (*Anguilla anguilla*), thinlip mullet (*Chelon ramada*), river lamprey (*Lampetra fluviatilis*), sea lamprey (*Petromyzon marinus*), smelt (*Osmerus eperlanus*), European flounder (*Platichthys flesus*), Atlantic salmon (*Salmo salar*) and the sea trout (*Salmo trutta*)) presence-absence was extracted and cleaned. The gear type and gear category which caught these species, their spatial location, and the date of capture (year and month), were also

cleaned and standardised. Very little is known about diadromous fish at-sea and modelling data-poor and poorly detectable species such as diadromous fish is challenging for species conservation. Furthermore, databases which contain both scientific surveys and fisheries-dependent data on data-poor species at the temporal and geographical scale of this database are uncommon. This data could therefore be used to improve knowledge of diadromous fish spatial and temporal trends, and modelling techniques for data-poor species.

Specifications Table

Subject	Biological science: Biodiversity
	Environmental science: Ecological modelling, Nature and Landscape Conservation
Specific subject area	A database of fisheries and scientific survey data was collated to analyse the distribution of diadromous fish at-sea and their bycatch within western European waters.
Type of data	CSV file
	Table
How the data were acquired	Data were acquired by contacting data holders (International Council for the Exploration of the Sea (ICES) Database of Trawl Surveys (DATRAS) the Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) and the French Ministry of Fisheries and Aquaculture (DGAMPA) for catch data on diadromous fish at-sea. To clean the data, R and QGIS were used. As a result of the content of fishing activity data, fisheries-dependent data have been provided on an ICES sub-rectangle scale as requested by the DGAMPA.
Data format	Raw
	Analysed
Description of data collection	This database was collated from fisheries-dependent and -independent data. It contains the spatiotemporal locations of the haul and gear type of each haul and the presence-absence of each species. Data were cleaned using descriptive statistics and existing literature on the species to see whether the species migration (distance from the coast over the course of the year and their length) matched that of existing literature.
Data source location	North-eastern Atlantic waters
	Metropolitan French Mediterranean waters
	For raw data sources see Table S1
Data accessibility	Repository name: SEANOE
	Direct URL to data: https://doi.org/10.17882/91719
Related research article	S.A.M. Elliott, A. Acou, L. Beaulaton, J. Guitton, E. Réveillac, E. Rivot, 2023. Modelling the distribution of rare and data-poor diadromous fish at sea for protected areas management. Progress in Oceanography. 210: 102924. https://doi.org/10.1016/j.pocean.2022.102924

Value of the Data

- Much information exists on diadromous fish during their freshwater stages, yet precise information on their at-sea habitat occupancy is still missing. This dataset provides valuable temporal and spatial information on diadromous fish during their at-sea life-history stages.
- Scientific survey and onboard fisheries observer data have often been analysed separately. Collating this data and standardising them into a single database represents unprecedented progress.

- By combining these different sources of data into a single database enables the possibility of improved inferences on species distribution using integrated modeling approaches.
- These data could be further used to understand seasonal and long-term changes in the distribution of diadromous fish at-sea.
- The data could be used to understand spatial and seasonal variation in fishing gear.
- The data can be used by fisheries ecologists, conservation managers, and statisticians interested in analysing complex datasets containing rare species.

1. Objective

We collated and cleaned this database containing fisheries-dependent and -independent data to improve knowledge on diadromous fish at-sea which inhabit eastern Atlantic waters (Greater North Sea, Celtic Sea, Bay of Biscay, and Iberian coast) and Metropolitan French waters within the Mediterranean. These are a group of species for which very little is known about their at-sea spatiotemporal occupancy and marine migrations.

2. Data Description

The data contains long-term (1965-2019) fisheries-dependent (from fishing vessels) and -independent (scientific surveys) data on the distribution of twelve diadromous fish at-sea (*Acipenser sturio, Alosa alosa, Alosa fallax, Alosa agone, Anguilla anguilla, Chelon ramada, Lampetra fluviatilis, Osmerus eperlanus, Petromyzon marinus, Platichthys flesus, Salmo salar, Salmo trutta*) at a large spatial scale (Greater North Sea, Celtic Sea, Bay of Biscay and the Iberian Shelf, French Metropolitan Mediterranean waters).

Fisheries-dependent data came from the French fisheries onboard observer program (ObsMer), held by the French Research Institute for Exploitation of the Sea (IFREMER), and available on request and approval from the French Ministry of Fisheries and Aquaculture (DGAMPA) [1]; Table S1). The ObsMer program was set up in 2003, under European regulation 1639/2001. The aim of the observer program is to assess species caught during commercial fisheries hauls, including bycatch and discards. Table S2 provides a list of the different gear types fisheries observers have collected onboard observer data from. Such data can thus provide valuable information on diadromous fish at-sea and their bycatch since they are rarely targeted by fisheries. The observer data contains spatiotemporal information on diadromous fish caught from 2003 to 2019 and the gear types and categories (Table S2) they were caught by. Absence data are also included in the dataset. As a result of missing information on fishing vessel characteristics, it was not possible to calculate catch per unit effort from the information provided. Due to the sensitive nature of this data, precise latitude and longitudinal information were removed, including metiers used (gear and target species) and information on whether the species were targeted, landed, or discarded. It should be noted that only ~2-4% of French fishing vessels have on-board observers and the observers semi-randomly sample fishing vessels and hauls when on-board [1].

Fisheries-independent data were downloaded from the International Council for the Exploration of the Seas (ICES) Database of Trawl Surveys (DATRAS; Table S1). DATRAS data are collected for Scientific Bottom Trawl Surveys (SBTS) in the Baltic, the Greater North Sea, the Celtic Sea, the English Channel, and the Bay of Biscay and the Iberian coast (https://www.ices.dk/data/data-portals/Pages/DATRAS.aspx). DATRAS SBTSs were initially designed to quantify commercial fish stock status (e.g., herring, cod, haddock, mackerel, etc.,). These surveys, however, collect information on all catch and can therefore provide valuable data on other non-target species. 48 separate metropolitan French scientific surveys programs containing information on diadromous fish at-sea were also collated by contacting each of the staff members leading on the different IFREMER surveys (https://campagnes.flotteoceanographique.fr/; [2]; Table S1). Details of fish captured by the different surveys within [2], helped indicate survey information with potentially relevant diadromous fish data. Information from [2] also provides details of the purpose of each of the French scientific surveys. Data from all the scientific sur-

veys were collated from when they began until 2019 (Table S1). It should be noted that although some SBTS data back to the 60s, standardisation of these surveys only took place from 1977 and changes to the survey design and specific gear used has taken place over the years (http://datras.ices.dk/Home/Descriptions.aspx).

For both fisheries-dependent and -independent data, presence and absence, gear type (Table S2), the spatial locations, year and month of capture were combined into a single database and formatted for comparability. Given the little information that currently exists on diadromous fish, this collated and cleaned dataset provides valuable information on these species at-sea life history traits and information on their bycatch risk. Having this data in a single database facilitates the analysis of species distribution using integrated modelling methods [3]. Furthermore, this is an interesting dataset for data scientists interested in finding new ways to model data-poor species.

3. Experimental Design, Materials and Methods

Following access to the fisheries-dependent and -independent data, all data files were reformatted and merged into a single comprehensive CSV file. This included joining station data (spatial, temporal and gear type information), with biological information (number of fish caught and their length) through a unique identification code, and standardising variables (i.e., species lengths, numbers captured, and gear types which were displayed in various ways) to the same units and categories. Data cleaning undertaken included removing points on land, converting gear types into Fisheries Agricultural Organisation (FAO) subcategory gear types (second tier) and creating gear categories, based upon the water column the vessels fish and whether they are static or mobile gears (Table S2; [5]). Existing FAO first tier gear categories group all trawls into the same category [5]. We found this unhelpful given most fish occupy different water layers and the impact of the different gear types can be very different depending on water-layer the haul is undertaken. Static gear types were not, however, separated into different water layers because they can capture much fewer species relative to trawls gear types.

For fisheries-independent data mean haul latitude and longitude was selected from the start and end of location of the haul. Erroneous negative distances (between the start and end of the haul) and overly long hauls (>100 000m in 30min) were removed from the fisheries-independent data. To ensure only fully marine data was kept for the entire dataset, all hauls below the mean low water tide were kept by clipping the data with shapefiles on freshwater and transitional waterbodies. ICES statistical divisions, squares and sub-divisions were also added to the dataset to aid with spatial data organisation and simplified visualisation. The numbers of fish caught per haul were converted to presence-absence information for distribution modelling purposes. It should also be noted, that due to the different abilities of gear types to capture diadromous fish, presence- absence information is more reliable than the numbers of individuals caught. Diadromous fish length information was standardised to the same unit, including correcting for erroneous length unit (mixing of millimetres and centimetres).

Diadromous fish biological data was cleaned by descriptive analyses prior to statistical analysis [4,6] to see whether their stage specific migration (distance from coast over the course of the year and their length) matched that of existing literature (e.g., [4]). Where it was certain that an identified species did not occur within the area found or was much larger than previously found, it was changed to its paired species or removed. Data that were modified included:

1) *Alosa alosa* and *Alosa fallax* which do not occur within the Mediterranean Sea. They were therefore converted to the *Alosa agone* (40 *A. alosa* presences and 76 *A. fallax* presences, 4% of the data) [6–8].
2) *A. fallax* maximum size is thought to be 50 cm [9,10]. Individuals larger than 55 cm were therefore converted to the larger *A. alosa* (74 presences, 4%) [6].

3) *Osmerus eperlanus* maximum size is 30 cm [11–13]. Individuals greater than 30 cm occurring far from the coast were removed (15 individuals) [6].

This collated and cleaned dataset containing 168 414 hauls from both fisheries-dependent and -independent data on twelve diadromous fish presence and absence is the largest in terms of spatio-temporal information that exists to date on diadromous fish. Although miss-identification may still be present within the dataset, extensive cleaning has taken place to try and improve accuracy of this data for modelling purposes [6].

Ethics Statements

The manuscript adheres to ethics in publishing standards. All data holders provided their written consent to have their database published.

CRediT Author Statement

Sophie Elliott: Data Curation, Formal analysis, Investigation, Methodology, Writing; **Noemie Deleys:** Data collation and curation; **Etienne Rivot:** Funding acquisition, Project administration, Scientific supervision; **Anthony Acou:** Funding acquisition, Project administration, Scientific supervision; **Elodie Reveillac:** Reviewing; **Laurent Beaulaton:** Funding acquisition, Project administration, Scientific supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Diadromous fish at sea database (Reference data) (Seanoe)

Acknowledgments

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2023.109107.

References

[1] A. Cornou, M. Quinio-scavinner, D. Delaunay, J. Dimeet, N. Goascoz, B. Dube, L. Fauconnet, M. Rochet, Observations à bord des navires de pêche professionnelle, Bilan de l'échantillonnage 2014 (2015), doi:10.13155/39722.

[2] A. Régimbar, J. Guitton, O.Le Pape, Zones fonctionnelles pour les ressources halieutiques dans les eaux sous souveraineté française, Deuxième partie : Inventaire. (2018) https://hal.science/hal-02293032.

[3] N.J.B. Isaac, M.A. Jarzyna, P. Keil, L.I. Dambly, P.H. Boersch-Supan, E. Browning, S.N. Freeman, N. Golding, G. Guillera-Arroita, P.A. Henrys, S. Jarvis, J. Lahoz-Monfort, J. Pagel, O.L. Pescott, R. Schmucki, E.G. Simmonds, R.B. O'Hara, Data integration for large-scale models of species distributions, Trends Ecol. Evol. 35 (2020) 56–67, doi:10.1016/j.tree.2019.08.006.

[4] S.A.M. Elliott, N. Deleys, E. Rivot, A. Acou, E. Réveillac, L. Beaulaton, Shedding light on the river and sea lamprey within western European marine waters, Endanger. Species Res. 44 (2021) 409–419, doi:10.3354/esr01113.

[5] P. He, F. Chopin, P. Suuronen, R. Ferro, J. Lansley, Classification and illustrated definition of fishing gears, FAO (2021), doi:10.4060/cb4966en.

[6] S.A.M. Elliott, A. Acou, L. Beaulaton, J. Guitton, E. Réveillac, E. Rivot, Modelling the distribution of rare and data-poor diadromous fish at sea for protected area management, Prog. Oceanogr. 210 (2023) 102924, doi:10.1016/j.pocean.2022.102924.

[7] P.G. Bianco, The status of the Twaite Shad, Alosa agone, in Italy and the Western Balkans, Mar. Ecol. 23 (2002) 51–64, doi:10.1111/j.1439-0485.2002.tb00007.x.

[8] J.-L. Baglinière, P. Elie, Les aloses (*Alosa alosa* et *Alosa fallax* spp.) ecobiologie et variabilité des populations, Quae, Institut National de la Recherche Agronomique, Paris. (2000) https://www.quae.com/produit/438/9782759210763/les-aloses-alosa-alosa-et-alosa-fallax-spp.

[9] E.I. Kukuev, A.M. Orlov, New Subspecies of Twaite Shad *Alosa fallax balticus* (Clupeidae), Inl. Water Biol. 11 (2018) 407–416, doi:10.1134/S1995082918040119.

[10] C. Taverny, P. Elie, Répartition spatio-temporelle de la grande alose *Alosa alosa* (Linne, 1766) et de l'Alose feinte *Alosa fallax* (Lacépède, 1803) dans le Golfe de Gascogne, Bull. Français La Pêche La Piscic. (2001) 803–821, doi:10.1051/kmae:2001020.

[11] D. Doherty, T.K. McCarthy, The ecology and conservation of European Smelt (*Osmerus eperlanus* L.) from Waterford Estuary, in Southeastern Ireland, Biol. Environ. Proc. R. Irish Acad. 104 (2004) 125–130, doi:10.3318/BIOE.2004.104.2.125.

[12] M. Power, M. Attrill, Temperature-dependent temporal variation in the size and growth of Thames estuary smelt Osmerus eperlanus, Mar. Ecol. Prog. Ser. 330 (2007) 213–222, doi:10.3354/meps330213.

[13] D.T.G. Quigley, F. Igoe, W. O'Connor, The European Smelt *Osmerus eperlanus* L. in Ireland: general biology, ecology, distribution and Status with conservation recommendations, Biol. Environ. Proc. R. Irish Acad. 104 (2004) 57–66, doi:10.3318/BIOE.2004.104.3.57.