Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

StoX: An open source software for marine survey analyses

Espen Johnsen[1] | Atle Totland[1] | Åsmund Skålevik[1] | Arne Johannes Holmin[1] |
Gjert Endre Dingsør[2] | Edvin Fuglebakk[1] | Nils Olav Handegard[1]

[1]Institute of Marine Research, Bergen, Norway

[2]The Norwegian Fishing Vessel Owners Association, Bergen, Norway

Correspondence
Espen Johnsen
Email: espen.johnsen@hi.no

Handling Editor: Edward Codling

Abstract

1. Scientists across the globe conduct survey programs to monitor and characterize abundance, population structure, biodiversity and geographical distributions. To assess the state of marine fish and zooplankton, population surveys are often repeated annually using standardized sampling protocols and analysis techniques to establish trustworthy stock status. However, although transparency and repeatability are recognised as important principles of this process, it is often difficult to obtain comprehensive documentation of metadata and data processing steps. This is particularly challenging for workflows that include manual processing steps.

2. StoX was principally built to process research-vessel survey data, and we have included several standard survey estimation models. The software was developed to be robust and versatile and aimed at the open source community, such that users could easily build their own models. StoX is fully integrated with R to utilize the large number of R-packages and enable any StoX function and stock estimation model to be controlled using R.

3. There has been a large need for a freely available software for research–vessel survey estimation, and StoX is tested in surveys carried out in four continents and is the official tool for many important fish stock surveys. The basic workflow and transparency principles of StoX, together with a customizable GUI, makes StoX applicable for any geographically coded surveys.

4. Future versions of StoX will include statistical models to estimate the catch composition in commercial fisheries. In fields such as conservation management, there is also a need to document the estimation methods, and additional estimation and analyses models, including biodiversity indices are currently implemented. In parallel, we envision a closer web service integration with existing international and national data centres.

KEYWORDS

acoustic, conservation, fisheries management, marine ecology, open source, StoX, survey analyses, swept area

1 | INTRODUCTION

To monitor change in abundance, population structure, geographical distribution (Gunderson, 1993) and biodiversity (Johannesen, Høines, Dolgov, & Fossheim, 2012) of fish and zooplankton populations (Dalpadado, Ingvaldsen & Hassel, 2003), scientific acoustic and swept area trawl surveys are conducted worldwide. These surveys are often repeated annually following standard protocols to produce time series, which are used to assess the state of fish stocks (Gunderson, 1993) and the health of ecosystems (Kirkman et al., 2016). However, data analyses are carried out using a range of software applications (Excel, R, SAS, etc.) and typically involves several manual processing steps. Consequently, it is often challenging to obtain a detailed record of the complete stock estimation procedure. In addition, the lack of streamlined data structures, documentation of parameter settings and information regarding subjective user decisions, make it nearly impossible to recalculate long time series with updated parameter settings, to test the effect of new methods and carry out sensitivity tests. This bottle neck in development, and lack of transparency in the data and methods used to produce time series, slows down the implementation of new methods, especially in multinational surveys.

A frequently used tool for other types of survey, such as underwater visual census (Bozec, Kulbicki, Laloë, Mou-Tham, & Gascuel, 2011) and terrestrial life distance sampling, is the software Distance (Thomas et al., 2010). For scientific acoustic and swept area trawl surveys, obtaining an overview of the survey estimation methods used by different research institutes is challenging. There are only a few standardized software packages that are easily accessible. Such as EchoR (Doray, 2013), which is an R-package produced by IFREMER (France) for estimation of acoustic trawl surveys and was developed specifically for IFREMER surveys. Subsequently, it requires certain input data structures and is restricted by particular survey designs which do not necessarily meet the requirements of surveys carried out elsewhere. The software Beam (Totland & Godø, 2001), which was used for acoustic survey estimates in the Norwegian Sea and in the Barents Sea, is dependent on outdated SAS software with an expensive GIS module. Furthermore, the Beam does not support transect-based acoustic estimation or swept area calculations and contains no built-in variance estimation methods. Inhouse programs are available at different laboratories, but they typically lack standardization, documentation and efficient deployment outside their respective institutions.

Several research institutes working with marine surveys in the North Sea, Norwegian Sea and Barents Sea, and surveys within the FAO-Nansen Programme (http://www.fao.org/in-action/eaf-nansen/en) have requested a standard software for survey estimation. As a response, the Norwegian Institute of Marine Research (IMR) has developed the open source software package StoX.

The objective of this paper is to (a) outline the general design of StoX; (b) present the organization of data files and storage of user settings ensuring transparency and repeatability; (c) present the interactions between various software components in use; (d) demonstrate (by examples) how standard annual survey estimates
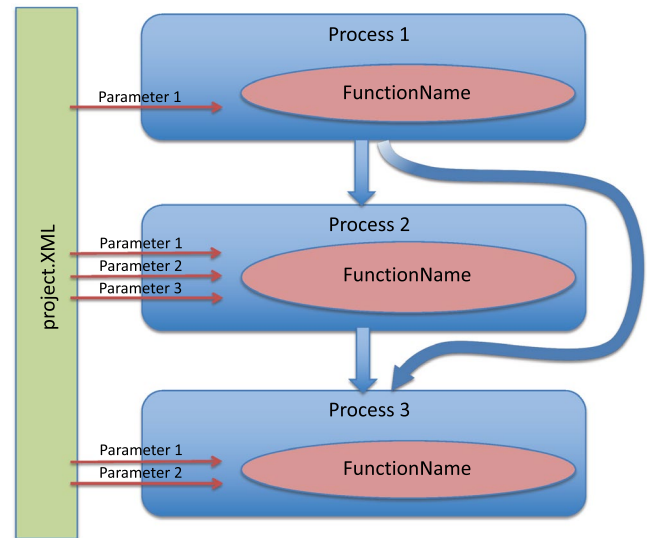

FIGURE 1 Schematic presentation of a StoX model with processes that use functions and their associated parameters. The blue arrows indicate the output and input data for each process

are calculated, and (e) demonstrate how more advanced users can utilize the software for scientific analyses and sensitivity testing. Finally, we discuss present and future development of StoX and give an overview of upcoming modules.

2 | THE MODULAR SOFTWARE DESIGN OF StoX

StoX is designed as a tool for transparent and reproducible survey estimation across nations and survey objectives. All methods, user settings and links to input data are documented in a description file named "project.xml". To enable method flexibility, users can create calculation models (hereafter referred to as StoX models) consisting of sequentially structured StoX processes (hereafter referred to as processes), which are documented in the project.xml file. A process is a user-defined call to one function available in a library of functions defined by StoX. Any StoX model can be modified by changing parameters of the processes, adding or removing processes and by rearranging their order of execution. Different StoX functions require one or more input datasets, which can be data files or output from previously executed processes (Figure 1).

3 | INPUT AND OUTPUT DATA

A StoX project is a collection of files organized in a folder structure with three subfolders (Figure S1, Appendix S1), one for input data files, one for output files and one for the subfolder holding the project.xml file. In addition, StoX projects may require resource files such as strata polygons, which are shipped with StoX and can be
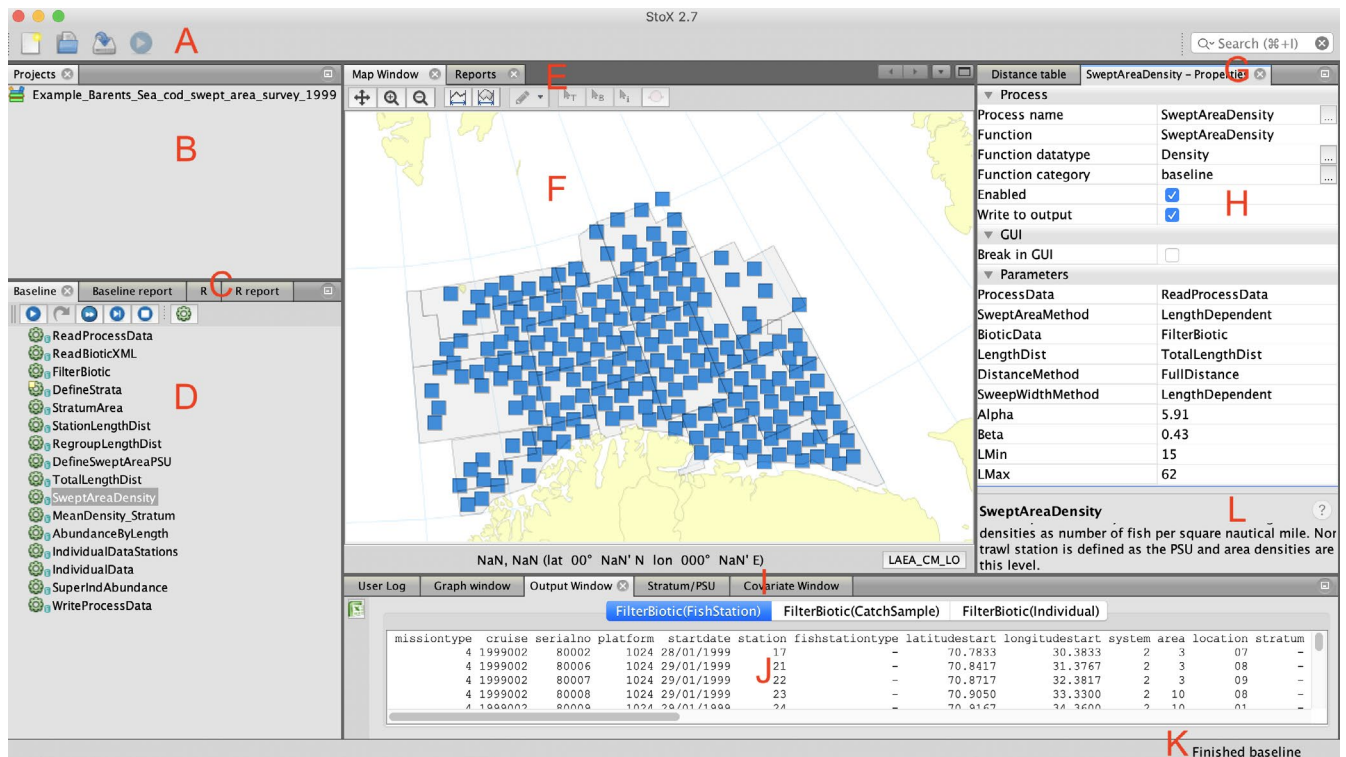
FIGURE 2 The StoX user interface. (A) Main Menu; (B) Project window; (C) Model Menu; (D) Model Window; (E) Map and Report Menu; (F) Map and Report Window; (G) Process Configuration and Distance Menu; (H) Process Configuration and Distance Window; (I) User Interface Menu, (J) User Interface Window; (K) Status Bar, and (L) Help Window

complemented by the user. StoX projects and resource files are stored by default in the folder "workspace"> "stox", in subfolders named "project" and "reference", respectively.

Currently, the input data are XML files of three different data types; biotic, acoustic and landing files (Figure S1, Appendix S1). Acoustic and biotic input XML files (see Appendix S2, Supporting Information for data formats) are available from the International Council for the Exploration of the Sea (ICES) Data Portal, Acoustic Trawl Surveys Database (http://acoustic.ices.dk/submissions). The acoustic XML files are also provided as output from the LSSS software (Korneliussen et al., 2016), and Echoview (https://www.echoview.com) has compatible output with the ICES Acoustic Trawl Survey Database. The "Landing" input folder (Figure S1, Appendix S1) is designated for commercial fisheries landings data, which will be used in future versions to estimate catch at age from fish populations (Hirst, Storvik, Aldrin, Aanes, & Huseby, 2005). Future versions of StoX may also read oceanographic variables, for example, conductivity, temperature and fields from ecosystem models.

4 | THE StoX USER INTERFACE

The StoX user interface is a platform independent tool currently building on Java. For details on installation and dependencies see Appendix S1, Supporting Information. The StoX user interface consists of several menus, windows and tabs organized to support the

basic structure of the software (Figure 2). To support constructing StoX models and setting parameters, StoX provides a specialized user interface for manual steps such as tagging of acoustic positions. The main tasks of each user interface component are described in detail in Appendix S3, Supporting Information.

5 | THE StoX - R INTERFACE

The StoX functionality can be accessed from R through the Rstox package, and all StoX projects can be executed using R-coding. New StoX projects can also be created using Rstox, however, some settings may depend on interactive tagging using the map StoX user interface map (Figure 2). The R-interface enables users to have easy access to input and output data, and all intermediate steps and parameters of a StoX project. Detailed description of the Rstox package is outside the scope of this paper.

6 | CURRENT USE OF StoX

Until now, StoX have mainly been used to estimate stock abundance of adult fish and several official international and national survey estimates of commercially and ecologically important fish stocks are produced using Stox. These include North Atlantic stocks such as herring, sprat, blue whiting, cod, haddock, lesser sandeel, boerfish, horse

mackerel and mackerel (c.f. Appendix S4, Supporting Information). StoX is also used to analyse abundance of fish larvae and demersal fish assemblages, and StoX is currently being tested for surveys in Argentina, Sri Lanka, South Africa, Angola and in the Baltics.

7 | StoX MODEL EXAMPLES

The workflow and StoX processes for the swept area abundance estimate of Barents Sea cod in the 1999 winter survey, and the acoustic trawl abundance estimate of herring the 2016 International ecosystem survey in the Nordic Seas are presented in Table 1.

From the output of SuperIndAbundance, the abundance estimate can be split by (or a combination of) population parameters such length, weight, sex, age etc. In Figure 3, the abundance estimate of the surveys are presented by age and length groups.

TABLE 1 Overview of the workflow and processes for the StoX models used to estimate the abundance of cod for the 1999 Barents Sea winter survey (A), and the acoustic trawl abundance estimate of herring in the Nordic Sea in 2016 (B)

Workflow description	Processe names
Read the project.xml file, and read and filter biological data	ReadProcessData ReadBioticXML FilterBiotic
Read strata polygons, calculate polygon areas and define Primary Sampling Units (PSU)	DefineStrata StratumArea DefineSweptAreaPSU
+Read and filter acoustic data, and define depth layers for aggregation of acoustic densities	+ReadAcousticXM +FilterAcoustic +NASC
+Define acoustic PSUs by stratum and calculate mean acoustic density by PSU	+DefineAcousticPSU +MeanNASC
+Assign weighted biotic stations to acoustic PSUs	+BioStationAssignment +BioStationWeighting
Calculate length distributions per biotic station (biotic PSU), and a weighted average of assigned length distributions (total length distribution for acoustic PSUs	StationLengthDist RegroupLengthDist TotalLengthDist
Calculate density (number fish by length group per nmi^2) by PSU, and calculate abundance by stratum	*SweptAreaDensity +AcousticDensity MeanDensity_Stratum AbundanceByLength
Identify biotic stations used in the abundance estimation and assign individuals from these stations to a proportion of the calculated abundance by length group. Each of these individuals represent an estimated abundance and is referred to as super individuals.	IndividualDataStations IndividualData SuperIndAbundance

Note: Processes only used for A or B, respectively, are marked with (*) and (+). Details about the StoX models and survey maps are presented in Appendix S5, Supporting Information.

8 | DISCUSSION

The rapid adoption of StoX as the official software for the assessment of many fish stocks in the North Atlantic (Table S2, Appendix S4) is indicative of the pressing need for a robust and freely available software that documents the process end-to-end. In parallel with the development of StoX, ICES has established databases to store acoustic and biotic trawl data to support data input files for StoX. This provides complete transparency from data input to the survey estimates used in stock assessment models, and StoX will be integrated into the new Transparent Assessment Framework that "is a framework to organize data, methods, and results used in ICES assessments, so that they are easy to reference and re-run (http://ices. dk/marine-data/assessment-tools/Pages/transparent-assessment-framework.aspx). A similar structure is in place at IMR for national and Norwegian–Russian surveys conducted in the Barents Sea. The use of StoX for stock assessment typically uses a filter to focus on a single species, but it may be appropriate to consider all or a range of species forbiodiversity indicators. These indicators are important for ecosystem monitoring, and there is ongoing work to implement StoX for biodiversity surveys. StoX is also being implemented by the EAF Nansen Program for several metrics, including stock abundance and biodiversity.

The output from the intermediate steps of a StoX project is available either directly from the software as text files or can be extracted using Rstox. This facilitates the reading and filtering of acoustic, biotic and/or landing data for analysis. The separation of the Java-based StoX application and the Rstox package, which contains a copy of the StoX function library, enables continuous
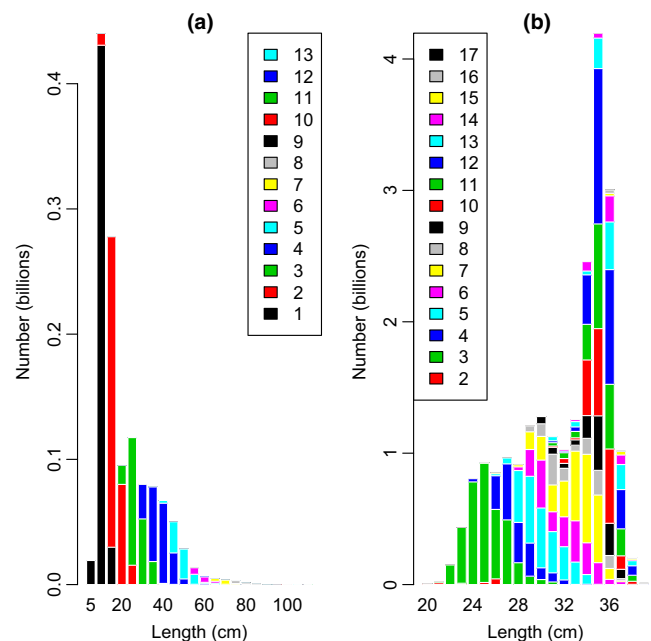

FIGURE 3 Abundance estimate by length group and age (c.f. colour legend) for cod from the 1999 Barents Sea winter survey (a) and herring from the 2016 Nordic Sea acoustic trawl survey (b). The estimates are produced using StoX

development of methods using R that can be incorporated as functionality in StoX if relevant.

Within the StoX architecture, stock estimation models are created from a set of functions. This has several advantages over other survey estimation programs, for example, Beam (Totland & Godø, 2001), which are restricted to a single estimator with little flexibility to add custom functions. Reusing functions for different types of surveys is beneficial as many processing tasks share common functions. For example, acoustic trawl surveys and swept area surveys use similar processing for trawl sampling data. By reusing functions across a range of different survey types, programming time and the risk of software errors are reduced, ensuring that similar tasks are consistently handled. This also ensures consistent use of StoX's framework for preserving parameter settings and manual annotations, which facilitates consistent recomputation and exchange of project files between researchers. In addition, the same model can be used for several surveys over a time series by changing only a part of the function parameters (e.g. the input files).

The StoX architecture enables new methods to be tested and implemented by replacing or adding new processes in a StoX estimation model. As no blueprint exists on how to perform an abundance estimate, the flexibility of StoX facilitates the adjustment of estimation models to meet different needs, such as between different species and regions. This is, in part, the reason for the uptake of StoX in several international surveys. Due to the modular design of the software, any required alterations can be implemented when needed; in future versions, the introduction of more template models will be available. StoX can also perform calculations over different vertical and horizontal resolutions, and is, in principle, only restricted by the resolution of the input data and computer memory. Therefore, it is possible to estimate fish density by depth channel, considering, for example, a depth dependent target strength (Ona, 2003). In addition, new methods in biological sampling using optical systems (e.g. Rosen, Jörgensen, Hammersland-White, & Holst, 2013) to pinpoint the catch to exact layers and depths can easily be implemented in StoX.

Modern data analysis pipelines usually deploy automated machine to machine interfaces that enable running analysis pipelines on a server. Conversely, users are typically required to manually handle data input, analysis and reporting when using traditional desktop applications. A major strength of StoX and Rstox is that they facilitate both. The "desktop approach" (StoX) is in daily use during surveys and provides the assessment model input, whereas the backend solution (run through Rstox) run on our servers provides a tool to re-analyse the survey time series without the need of a manual point and click interface. This is useful for a range of purposes. The parameters and functions can be modified through scripts, and the impact of changing parameters and functions can be tested for the entire time series. This includes sensitivity analyses and testing new estimators. Running the pipeline on the backend servers also allows us to monitor data integrity and data quality and is useful when testing new versions of the software, e.g. to see if it can re-produce the estimates derived from earlier versions. It is also possible to automatically generate the estimates and present them through web portals such as the Norwegian Marine Data Centre (https://nmdc.no/nmdc) and the ICES data centre (http://www.ices.dk/marine-data/).

The workflow and transparency principles of StoX may be useful for other types of marine and terrestrial surveys, and new functions can be added to the function library. The next steps for the StoX project are to include methods for biodiversity indices, implement methods for fishery dependent data (Hirst et al., 2005), develop web services that can run StoX and associated R-packages on a website, and ensure tighter integration with the ICES and IMR data processing pipelines. Since StoX and Rstox are fully open source, we also envision a closer interaction between other user groups and developers. We hope that the flexibility of the program will encourage others to develop their own modules and interact with existing users and developers. This will ultimately further enhance the features and functions of both the StoX and Rstox software packages.

AUTHORS' CONTRIBUTIONS

A.T., E.J., Å.S., G.E.D. and N.O.H. conceived and planned the design of the software. Å.S. did the Java coding. A.J.H. and E.F. contributed to the final structure of StoX. E.J., G.E.D., A.T. and A.J.H. wrote the Supplementary Material. E.J., A.T., N.O.H., A.J.H wrote the paper in consultation with G.E.D. and E.F.

DATA AVAILABILITY STATEMENT

The StoX software is available at http://www.imr.no/forskning/prosjekter/stox/nb-no. The source codes for StoX and Rstox are available at (https://github.com/Sea2Data/StoX, https://doi.org/10.5281/zenodo.3254380) and (https://github.com/Sea2Data/Rstox, https://doi.org/10.5281/zenodo.3254378), respectively. Full versions of the StoX model examples, including the relevant input data, are archived on Zenodo: https://doi.org/10.5281/zenodo.3255039.

ORCID

Espen Johnsen https://orcid.org/0000-0001-8213-6750

Atle Totland https://orcid.org/0000-0002-9904-8121

Åsmund Skålevik https://orcid.org/0000-0003-3327-5601

Arne Johannes Holmin (iD) https://orcid.org/0000-0002-1225-1144

Edvin Fuglebakk (iD) https://orcid.org/0000-0002-3192-6768

Nils Olav Handegard (iD) https://orcid.org/0000-0002-9708-9042

REFERENCES

Bozec, Y. M., Kulbicki, M., Laloë, F., Mou-Tham, G., & Gascuel, D. (2011). Factors affecting the detection distances of reef fish: Implications for visual counts. *Marine Biology*, *158*(5), 969–981. https://doi.org/10.1007/s00227-011-1623-9

Dalpadado, P., Ingvaldsen, R., & Hassel, A. (2003). Zooplankton biomass variation in relation to climatic conditions in the Barents Sea. *Polar Biology*, *26*(4), 233–241.

Doray, M. (2013). EchoR package tutorial. Retrieved from http://archimer.ifremer.fr/doc/00128/23879/

Gunderson, D. R. (1993). *Surveys of Fisheries Resources*. (pp. 248). New York: John Wiley & Sons, Inc.

Hirst, D., Storvik, G., Aldrin, M., Aanes, S., & Huseby, R. B. (2005). Estimating catch-at-age by combining data from different sources. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(6), 1377–1385. https://doi.org/10.1139/f05-026

Johannesen, E., Høines, Å. S., Dolgov, A. V., & Fossheim, M. (2012). Demersal fish assemblages and spatial diversity patterns in the Arctic-Atlantic Transition Zone in the Barents Sea. *PLoS ONE*, *7*(4), e34924. https://doi.org/10.1371/journal.pone.0034924

Kirkman, S. P., Blamey, L., Lamont, T., Field, J. G., Bianchi, G., Huggett, J. A., … Yemane, D. (2016). Spatial characterisation of the Benguela ecosystem for ecosystem-based management. *African Journal of Marine Science*, *38*(1), 7–22. https://doi.org/10.2989/1814232X.2015.1125390

Korneliussen, R. J., Heggelund, Y., Macaulay, G. J., Patel, D., Johnsen, E., & Eliassen, I. K. (2016). Acoustic identification of marine species using a feature library. *Methods in Oceanography*, *17*, 187–205. https://doi.org/10.1016/j.mio.2016.09.002

Ona, E. (2003). An expanded target-strength relationship for herring. *ICES Journal of Marine Science*, *60*(3), 493–499.

Rosen, S., Jörgensen, T., Hammersland-White, D., & Holst, J. C. (2013). DeepVision: A stereo camera system provides highly accurate counts and lengths of fish passing inside a trawl. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(10), 1456–1467. https://doi.org/10.1139/cjfas-2013-0124

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., … Burnham, K. P. (2010). Distance software: Design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, *47*(1), 5–14. https://doi.org/10.1111/j.1365-2664.2009.01737.x

Totland, A., & Godø, O. R.(2001). BEAM – an interactive GIS application for acoustic abundance estimation. InT. Nishida, P. R. Kailola, & C. E. Hollingworth (Eds.), *Proceedings of the First Symposium on Geographic Information System (GIS) in Fisheries Science*. (pp. 195–201). Kawagoe-City, Saitama, Japan: Fishery GIS Research Group. http://www.esl.co.jp/Sympo/1st/sympo7_2.htm

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.