Extracting groundfish survey indices from the Ocean Biogeographic Information System (OBIS): an example from Fisheries and Oceans Canada

Daniel Ricard, Robert M. Branton, Donald W. Clark, and Peter Hurley

Ricard, D., Branton, R. M., Clark, D. W., and Hurley, P. 2010. Extracting groundfish survey indices from the Ocean Biogeographic Information System (OBIS): an example from Fisheries and Oceans Canada. – ICES Journal of Marine Science, 67: 638–645.

Scientific trawl surveys have been conducted in different regions of the world and by a variety of countries and agencies since the mid-1900s. Although the data are collected in a scientifically and statistically appropriate context and represent an important source of fishery-independent information for agency-specific stock assessments, their use and dissemination has often been limited to the agencies conducting the surveys. In recent years, Internet data portals such as the Ocean Biogeographic Information System have provided an arena for the wider distribution and use of marine fish data. Despite the increased accessibility of such data, their scientific acceptability has been limited by a lack of reproducibility in data analyses. We present a methodology for the computation of time-series of groundfish stock indices using publicly available trawl survey data derived from the Canadian Department of Fisheries and Oceans Maritimes region. Potential pitfalls associated with the computation of time-series are discussed and proper stratified random estimates of temporal abundance trends are compared with other methods for a selected subset of species. Also, the broader applicability of the methods for datasets collected under similar sampling designs is discussed, along with the reproducibility of the analyses and results.

Keywords: Ocean Biogeographic Information System, stratified random design, temporal stock dynamics, trawl surveys.

Received 20 October 2008; accepted 14 October 2009; advance access publication 13 December 2009.

D. Ricard: Department of Biological Sciences, Dalhousie University, 1355 Oxford Avenue, Halifax, NS, Canada B3H 4J1. R. M. Branton: Ocean Tracking Network, Dalhousie University, 1355 Oxford Avenue, Halifax, NS, Canada B3H 4J1. D. W. Clark: Department of Fisheries and Oceans, St Andrews Biological Station, 531 Brandy Cove Road, St Andrews, NB, Canada E5B 2L9. P. Hurley: Department of Fisheries and Oceans, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, NS, Canada B2Y 4A2. Correspondence to D. Ricard: tel: +1 902 494 2146; fax: +1 902 494 3736; e-mail: ricardd@mathstat.dal.ca.

Introduction

In the Northwest Atlantic, routine scientific trawl surveys have been conducted since the mid-1900s to provide fisheries-independent information about fish populations (Doubleday, 1981; Doubleday and Rivard, 1981). In the Scotian Shelf and Bay of Fundy region of Canada, the surveys have been conducted since 1970. Sampling activities started later off Newfoundland and in the Gulf of St Lawrence. In the United States, survey activities on Georges Bank and in the Gulf of Maine and parts of the Scotian Shelf date back to 1963. Although these surveys concentrate on commercially exploited species, they also record catch information for all species taken and provide an invaluable source of information about marine organisms.

Surveys are a major source of information for fisheries management in Canada and around the world. Agencies conduct sampling activities using a variety of gear types, vessels, and protocols. Here we concentrate on surveys conducted on the Scotian Shelf and in the Bay of Fundy region of Canada [57–68°W 43–47°N; Northwest Atlantic Fisheries Organization (NAFO) divisions 4X, 4V, and 4W; Figure 1). The sampling design was originally based on the distribution of Atlantic cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*). Survey samples use a bottom trawl and consist of 30-min tows at a speed of ~3.5 knots, giving a towed distance of 1.75 nautical miles. Beginning in 1970, tow-level data on the numbers and weights caught, and the size compositions, were recorded for all fish and some invertebrate species. Since 2000, data have been recorded for all marine species caught in the survey trawl (Tremblay *et al.*, 2007). The surveys are manned by trained scientists whose responsibilities include gathering the data within a planned sampling design and using consistent fishing gear and methods. The sampling protocol also ensures correct species identification and appropriate digital storage of the data.

For researchers outside government agencies, obtaining data from marine ecosystems often follows an *ad hoc* process: data are made available under certain conditions and analyses are run using the version received. Controversies about data interpretation often arise when analyses are not reviewed by the data custodians. In contrast, data that are made publicly available through Internet-based systems need to follow metadata standards, ensuring that data sources can be correctly referenced/cited and that analyses can be replicated.

The metadata consist of all the information necessary to understand a dataset. This includes, but is not limited to, the appropriate

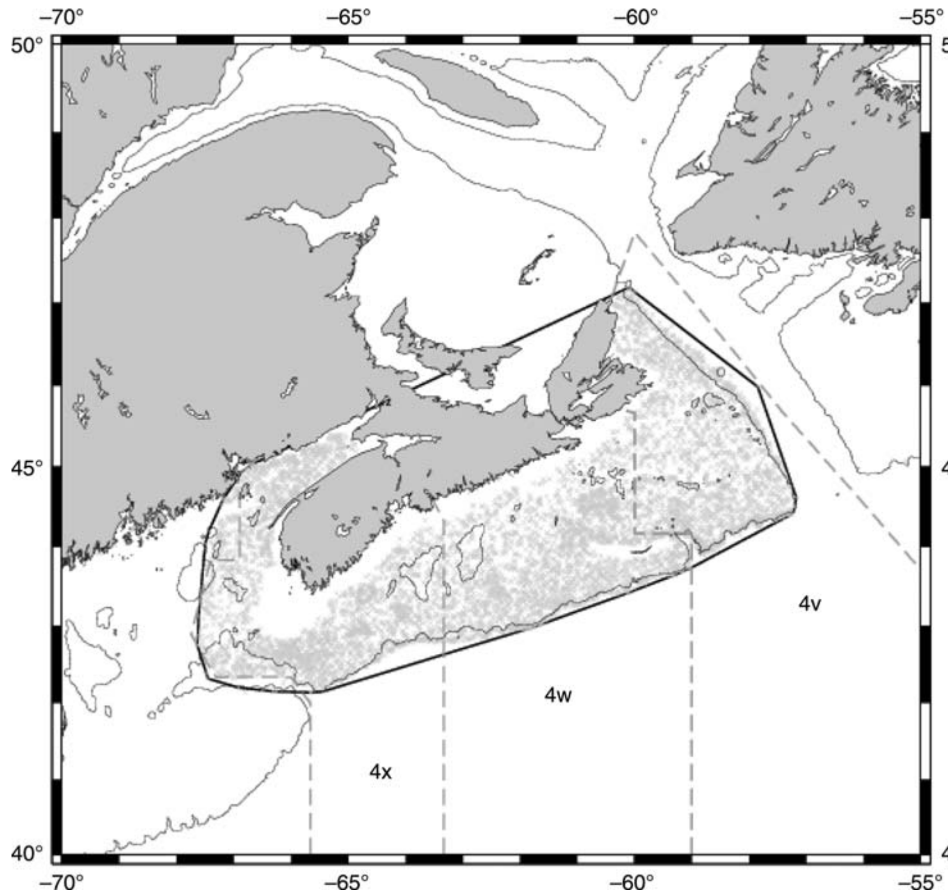

Figure 1. The Scotia–Fundy region of the Northwest Atlantic showing NAFO divisions 4X, 4W, and 4V. The solid black polygon shows the extent of DFO Maritimes' SUMMER and SUMMER_TELEOST surveys. Fishing tow locations are plotted as tiny grey crosses, and the 200-m isobath is also shown.

citation of the data, and their spatial and temporal coverage. Good metadata give credibility to publicly available datasets and foster proper interpretation of biogeographic data. Several digital standards exist to capture metadata, and being tasked with authoring them can be daunting; discussion of the merits and pitfalls of the different metadata standards, however, is beyond the scope of this paper.

The Internet has changed the way that data are shared and obtained by researchers, though in fields of marine science such as physical oceanography, datasets have historically been shared among scientists. Examples of such datasets include bathymetric grids, conductivity–temperature–depth (CTD) profiles of the water column, infrared imagery (sea surface temperature products), and ocean colour imagery (chlorophyll *a* products). As such, the computational tools and standards required to store, query, extract, and analyse the data are mature and available. In contrast, marine biogeographic datasets such as those derived from scientific surveys are often held in disparate formats by various agencies. The standards and tools required to share biogeographic information were only recently formulated, and are now beginning to gain acceptance within the biogeographic research community.

Data portals exist that provide access to a wide variety of data that can be used for scientific analyses of fish population dynamics and fish diversity. It is hoped that the increasing quantity of publicly available biogeographic data will foster novel research initiatives that will utilize the data in a context broader than that for which they were collected. For example, the Census of Marine Life's Ocean Biogeographic Information System (OBIS) provides access to more than 20 million records from almost 700 different data sources (OBIS, 2009a). The OBIS data portal provides a centralized location to access data from a multitude of sources. It also provides visualization tools and the possibility of downloading data in OBIS schema format. The OBIS schema format (OBIS, 2009b) is derived from the Darwin Core 2 specification for exchange of information on the geographic occurrence of living organisms (Taxonomic Diversity Working Group, 2009). Although the number of records on the OBIS portal is staggering and making maps of the data is encouraged, it is important for such systems to go beyond map production. To gain acceptance in the wider ecological community, data portals need to provide information that can be used in a broader context, such as the analysis of temporal and spatial dynamics of marine populations (Myers, 2000).

Here we describe the steps and methods involved in making the Department of Fisheries and Oceans (DFO) Maritimes survey data available on OBIS. This includes creating effective metadata and generating a properly formatted version of the dataset. We also demonstrate how the data can be used and misused by comparing computation methods that generate abundance time-series with various degrees of knowledge on the sampling design of the trawl surveys. We also discuss the temporal changes observed in the survey-derived abundance estimates of eight fish species caught commercially. We show that a naïve interpretation of the publicly available data yields improper time-series of species abundance, but that appropriate temporal trends can be computed using additional information about the sampling design.

The steps detailed here can be applied to other datasets generated in scientific surveys conducted by other fisheries agencies, and we hope therefore that the methods described to document and encode survey data collections can and will be repeated. This will ensure that erroneous interpretations of trawl survey data are minimized while still promoting wider dissemination of data through portals such as OBIS.

Methods

The metadata record for the OBIS version of the DFO Maritimes Research Vessel Trawl Surveys Fish Observations data was created in Directory Interchange Format (DIF; NASA, 2008) and made available and discoverable on NASA's Global Change Master Directory (GCMD, http://gcmd.nasa.gov/) metadata portal. The metadata records contained in the GCMD also appear on Canada's equivalent metadata discovery site named Geodiscover (http://geodiscover.cgdi.ca/). These metadata records provide information about the surveys' spatio-temporal coverage, and their citation details, and also provide links to the data available on the OBIS portal. Additionally, the OBIS portal provides an extended metadata record containing additional information about the dataset. The url to both the GCMD and the OBIS metadata records is listed in the References (Clark and Branton, 2007a, b).

To be made accessible to the OBIS portal, the data need to be formatted to follow the OBIS schema. The groundfish survey data from DFO Maritimes are stored in a relational database management system at the Bedford Institute of Oceanography in Dartmouth, NS, Canada. The database contains all information recorded during the surveys and in subsequent post-survey analyses, such as the age records of fish determined from otoliths. Using the Structured Query Language (SQL), the data are formatted to follow the OBIS format and made available to the data portal through a Distributed Generic Information Retrieval (DiGIR) server, also located at the Bedford Institute of Oceanography.

For the DFO Maritimes Research Vessel Trawl Surveys Fish Observations data (Clark and Branton, 2007a), only "valid" tows are included in the OBIS version. This means that fishing tows that did not meet the requirements for acceptance (accurate duration and functioning of the gear, not crossing a stratum boundary, etc.) are removed. Additionally, the catch data are normalized for the distance towed and can be used directly as an indicator of abundance and biomass. An example OBIS record from the survey data is provided in Table 1.

The OBIS version of the data used in this paper was obtained in ASCII format from the OBIS data portal. We used the Advanced Search facility of the portal to obtain the data from the DFO Maritimes Research Vessel Trawl Surveys Fish Observations dataset. The data available from OBIS contain many different survey series conducted throughout the year with different vessels and gears. For the current analysis, we concentrated on the SUMMER and SUMMER_TELEOST survey series, which consist of data collected during July and August in the Bay of Fundy and on the Scotian Shelf from 1970 to present. Over the time frame of available data, the SUMMER and SUMMER_TELEOST series follow a consistent sampling design. Additionally, we only include data from stratum 440 to stratum 495, because they are the most consistently sampled over the duration of the survey series.

Table 1. Example of a single OBIS record from the SUMMER survey of the DFO Maritimes Research Vessel Trawl Surveys Fish Observations dataset.

Field name	Value
res_name	DFOgfsDBfish
scientificname	*Gadus morhua*
institutioncode	BIO
catalognumber	TEM2008830-178-10-1
latitude	44.34
longitude	−61.9
collectioncode	SUMMER
datelastmodified	2007-07-13T18:24:50Z
yearcollected	2008
monthcollected	7
daycollected	31
minimumdepth	163
maximumdepth	164
slatitude	44.35767
slongitude	−61.90883
elatitude	44.33
elongitude	−61.89983
Class	Actinopterygii
kingdom	Animalia
ordername	Gadiformes
phylum	Chordata
family	Gadidae
genus	*Gadus*
species	*morhua*
scientificnameauthor	Linnaeus, 1758
collector	TEM
fieldnumber	TEM2008830-178
locality	462
observedindividualcount	1
observedweight	0.345
samplesize	1.71 nautical miles × 41.0 ft

Before 1982, fishing was carried out with a Yankee 36 otter trawl, but is now conducted with a Western IIa trawl. There were also vessel changes during the survey series, CCGS "A.T. Cameron" being the main survey platform from 1970 to 1981, CCGS "Lady Hammond" from 1982 to 1991, and CCGS "Alfred Needler" from 1983 to today. However, for technical reasons, the CCGS "Alfred Needler" was substituted by its sister ship CCGS "Wilfred Templeman" in 2008 and by CCGS "Teleost" in 2004 and 2007. For purposes of illustration, abundance time-series were generated for eight species caught commercially that have >2000 records in the SUMMER and SUMMER_TELEOST series of the OBIS dataset (Table 2). These species cover a wide range of taxa, abundance, and exploitation history, and provide a base case to evaluate the methods presented here.

Following recommendations made by Branton and Ricard (2007), individual tows i in the OBIS version of the DFO dataset can be identified by the value in the fieldnumber field of the OBIS schema. Similarly, stratum s is identified by the locality field, the swept-area of the trawlnet by the samplesize field, and the vessel by the collector field. The OBIS version of the dataset consists of observations only, and all sampling locations (including

Table 2. Species used in the analysis and the number of records from the SUMMER and SUMMER_TELEOST survey data (for stratum 440 to stratum 495) available in the OBIS dataset.

Common name	Scientific name	Number of records in OBIS
American plaice	*Hippoglossoides platessoides*	5 049
Atlantic cod	*Gadus morhua*	4 613
Haddock	*Melanogrammus aeglefinus*	4 373
Silver hake	*Merluccius bilinearis*	3 470
Redfish species	*Sebastes* spp.	3 108
White hake	*Urophycis tenuis*	2 790
Herring	*Clupea harengus*	2 262
Pollock	*Pollachius virens*	2 143

those where a species is not caught) need to be used to determine observations of zero catch. We assigned values of zero catch to tows where a species was not observed. For each combination of year y, stratum s, and tow i, we used the OBIS data to create observations of fish abundance for the eight species of interest. Each abundance observation is either the value observed in the "observedindividualcount" field of the OBIS data record, or zero if a tow did not have a catch record for a given species. In other words, for each species we generated observations $a^0_{y,s,i}$. For illustration purposes, we also created catch records $a_{y,s,i}$ that included only the observations and that did not account for records of zero catch.

Using the two types of record generated from the OBIS data ($a^0_{y,s,i}$ and $a_{y,s,i}$), we used four computation methods to generate annual time-series that corresponded to varying degrees of understanding about the dataset. The methods rely on the assumptions made about the data and how they were collected. The different computations presented below range from naïve interpretations of the data to appropriate estimation of catch rates with knowledge of the sampling design.

The first method does not take into account the stratification scheme of the survey and also does not take the zeroes into account. The annual estimate of population abundance for each species is calculated as the mean abundance for that year, excluding the zero observations. We refer to this time-series as OBIS raw:

$$\hat{a}_y = \frac{\sum_{s=1}^{S} \sum_{i=1}^{n_{y,s}} a_{y,s,i}}{n_y}, \tag{1}$$

where $n_{y,s}$ is the number of catch records from the OBIS dataset in year y and stratum s, n_y the total number of catch records in year y ($n_y = \sum_{s=1}^{S} n_{y,s}$), and S the number of strata sampled in year y.

This method can result in overestimation of abundance, because samples with zero catch are not included in the calculations. The annual estimates for the second method take the zeroes into account. We refer to this time-series as OBIS with zeroes:

$$\hat{a}^0_y = \frac{\sum_{s=1}^{S} \sum_{i=1}^{n^0_{y,s}} a^0_{y,s,i}}{n^0_y}, \tag{2}$$

where $n^0_{y,s}$ is the number of catch records from the OBIS dataset in year y and stratum s, including catches with zero individuals, and n^0_y the total number of catch records, including catches with zero individuals, in year y ($n^0_y = \sum_{s=1}^{S} n^0_{y,s}$). Note that, while accounting for observations of zero catch, the time-series generated from Equation (2) still does not account appropriately for the sampling design, because it assumes that each sample was independent.

To attempt to account for the stratum and year effects, a third time-series was generated using a generalized linear model (GLM) with negative binomial error and a log link using strata (s) and year (y) as factors. In other words, population abundance $a^0_{y,s} = \sum_{i=1}^{n^0_{y,s}} a^0_{y,s,i}$ is assumed to follow a negative binomial distribution of mean μ, and the linear predictor (LP) of μ is

$$\log(\mu_{y,s}) = \mathrm{LP}_{y,s} = \alpha + \beta_y + \gamma_s, \tag{3}$$

where α is the overall mean, β_y the year effect, subject to $\sum_{y=1}^{Y} \beta_y = 0$, and γ_s is the strata effect, subject to $\sum_{s=1}^{S} \gamma_s = 0$. The fitted model was then used to predict an annual time-series that we refer to here as OBIS GLM.

Finally, the fourth time-series was generated by computing the annual stratified random estimates of species abundance using strata statistics obtained from the OBIS Canada site (OBIS Canada, 2009). These statistics are required for computation of stratified random estimates of abundance. The additional information required consisted of the surface area of each stratum s, which was divided by the swept-area of the gear to obtain the number of towable units in each stratum (N_s). The data from OBIS were used to compute annual estimates of fish abundance for the different species of interest. Following the methods documented in Smith (1996) and Lohr (1999), we computed the stratified mean for each year and refer to the time-series as OBIS stratified:

$$\bar{a}_y = \sum_{s=1}^{S} \frac{N_s}{N} \bar{a}_{y,s}, \tag{4}$$

where $s = 1, 2, 3, \ldots, S$ are the different strata, $\bar{a}_{y,s}$ the catch sample mean for stratum s in year y ($\bar{a}_{y,s} = \sum_{i=1}^{n^0_{y,s}} a^0_{y,s,i}$), N_s the number of towable units in stratum s, and $N = \sum_{s=1}^{S} N_s$ the total number of towable units in the area surveyed.

Although useful, time-series of population abundance need to be interpreted carefully, and the uncertainty associated with the estimates needs to be provided. This can be done readily using the OBIS data. Following from Equation (4), the estimated stratified variance for each year can be calculated from

$$\hat{V}(\bar{a}_y) = \sum_{s=1}^{S} \left(1 - \frac{n_s}{N_s}\right) \left(\frac{N_s}{N}\right)^2 \left(\frac{\nu_s}{n_s}\right), \tag{5}$$

where n_s is the number of tows in stratum s, and ν_s the catch sample variance for stratum s.

It is common to report the standard error of an estimator, which is the square root of the estimated variance. In our case, the standard error (s.e.) is

$$\text{s.e.}(\bar{a}_y) = \sqrt{\hat{V}(\bar{a}_y)}. \tag{6}$$

The four time-series generated from the publicly available data (OBIS raw, OBIS with zeroes, OBIS GLM, and OBIS stratified) were compared with each other to identify differences in interpretation associated with each methodology. The stratified random

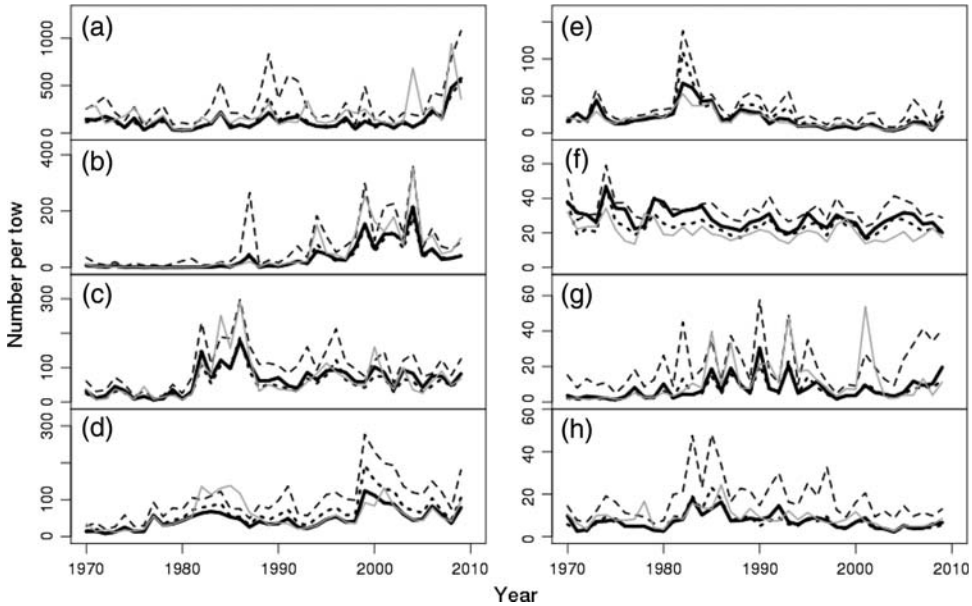

Figure 2. Abundance time-series for (a) redfish (*Sebastes* spp.), (b) Atlantic herring (*C. harengus*), (c) silver hake (*Merluccius bilinearis*), (d) haddock (*M. aeglefinus*), (e) Atlantic cod (*G. morhua*), (f) American plaice (*H. platessoides*), (g) pollock (*P. virens*), and (h) white hake (*Urophycis tenuis*) obtained using the OBIS version of the DFO's SUMMER and SUMMER_TELEOST survey data from 1970 to 2009 in the Scotia–Fundy region of the Northwest Atlantic (NAFO divisions 4X and 4VsW). The heavy black line is the OBIS stratified reference time-series, the dashed line the OBIS raw time-series, the dotted line the OBIS with zeroes time-series, and the grey line the OBIS GLM time-series.

estimate of abundance was used as the reference case because it accurately accounted for the survey design (Smith, 1996). We were interested in seeing how well the different methods estimated the abundance trends of the eight species of interest.

For comparing the different methodologies used to derive the time-series, we reported Spearman's and Pearson's correlation coefficients between the reference case (OBIS stratified) and OBIS raw, OBIS with zeroes, and OBIS GLM. These measures of standard correlation provide an indication of the association between a computation method and the reference case, but they do not account for systematic differences. If two methods are identical in their measurement properties, we would expect them to follow a linear relationship of unit slope and zero intercept. The intraclass correlation coefficient, ICC(1,1), as defined in Shrout and Fleiss (1979), measures deviations from the unit slope and zero intercept (an ICC value of near 1 would indicate values that fall close to this ideal).

All data handling was carried out using a PostgreSQL relational database management system (PostgreSQL Global Development Group, 2009). Statistical analyses and generation of plots were conducted with the R Environment for Statistics and Graphics (R Development Core Team, 2009) using the packages beanplot (Kampstra, 2008), irr (Gamer *et al*., 2007), MASS (Venables and Ripley, 2002), and RODBC (Ripley and Lapsley, 2009). The map in Figure 1 was generated using the Generic Mapping Tools (Wessel and Smith, 1991).

Results

The four abundance time-series generated using the OBIS data plotted for the eight species of interest are presented in Figure 2. The stratified random estimate of the number of fish per tow in a given year ranged from a maximum of 571.07 per tow for redfish species (*Sebastes* spp.; Figure 2a) in 2009 to a minimum of 0.13 fish per tow for herring (*Clupea harengus*; Figure 2b) in 1978. Species such as pollock (*Pollachius virens*; Figure 2g) showed a high variability in annual estimates. Other species, such as haddock (*M. aeglefinus*; Figure 2d) and redfish, showed an increasing trend in abundance over the time-series.

For the eight species of interest, the Spearman's correlation coefficients, Pearson's correlation coefficients, and intraclass correlation coefficients, ICC(1,1), between the three methods used with the OBIS data and the reference methodology (OBIS stratified) are listed in Table 3. In general, the ICC value was lower than that of either the Spearman's or Pearson's coefficients, indicating that although the time-series were highly correlated, there were discrepancies between the reference case and the other methods.

An example of the standard error being added to the time-series, to provide an estimate of the uncertainty associated with the stratified random mean for Atlantic cod (*G. morhua*), is given in Figure 3.

Discussion

The GCMD and Geodiscover websites provide access to the metadata record for the DFO Maritimes Research Vessel Trawl Surveys Fish Observations. The clear documentation of the dataset through a proper metadata record provides authority and citability to the data. However, the DIF format used by GCMD is limited in its ability to describe the details associated with the data collection in full. An alternative metadata standard that can be used is the Ecological Metadata Language (EML; EML Project, 2008), and we hope to provide the DFO Maritimes Research Vessel Trawl Surveys Fish Observations as an EML record in future.

Each computation method includes a different degree of understanding of the data collection, which is reflected in how well the results compare with the reference method. In general, the OBIS raw method yielded the least agreement with the reference case. ICC values were lower than either Spearman's or Pearson's correlation coefficients, indicating that methods that seem to do a good job at estimating annual abundance are in reality rather poor. Because it takes the sampling design into account, the stratified random methodology is the most appropriate estimator of yearly abundance.

Table 3. Value of correlation coefficients between the reference time-series and the other OBIS-derived time-series.

Species	Correlation measure	OBIS raw	OBIS with zeroes	OBIS GLM
Hippoglossoides platessoides	Spearman	0.768	0.823	0.583
	Pearson	0.808	0.841	0.624
	ICC(1,1)	0.531	0.627	0.089
Gadus morhua	Spearman	0.866	0.942	0.910
	Pearson	0.899	0.922	0.930
	ICC(1,1)	0.618	0.877	0.885
Melanogrammus aeglefinus	Spearman	0.908	0.957	0.872
	Pearson	0.956	0.979	0.708
	ICC(1,1)	0.290	0.815	0.644
Merluccius bilinearis	Spearman	0.957	0.979	0.908
	Pearson	0.954	0.954	0.859
	ICC(1,1)	0.617	0.917	0.777
Sebastes spp.	Spearman	0.917	0.931	0.520
	Pearson	0.890	0.934	0.587
	ICC(1,1)	0.369	0.935	0.453
Urophycis tenuis	Spearman	0.899	0.881	0.671
	Pearson	0.886	0.863	0.705
	ICC(1,1)	0.062	0.847	0.628
Clupea harengus	Spearman	0.936	0.991	0.971
	Pearson	0.921	0.982	0.955
	ICC(1,1)	0.636	0.976	0.814
Pollachius virens	Spearman	0.853	0.917	0.745
	Pearson	0.842	0.888	0.616
	ICC(1,1)	0.197	0.867	0.457

The OBIS raw time-series tended to overestimate the average number of fish caught per tow for most years and most species, likely because the computed annual average does not include observations of zero catch. As a large proportion of the tows have zero catch (Figure 4), it is important to account for zeroes in the computation of the time-series. Note, however, that the OBIS raw time-series sometimes provided annual estimates less than the stratified random estimates (e.g. 1971, 1972, and 1980 for American plaice, *Hippoglossoides platessoides*; Figure 2f). Using raw annual averages from the OBIS data is, however, wrong, so the time-series derived should not be used for analysing trends in population abundance, especially when combining data from different sources (as was unfortunately done by NOAA, 2005a).

The OBIS with zeroes time-series more closely matched the reference time-series than the OBIS raw time-series but still did not account for the sampling design. For species with a large proportion of tows of zero catch (Figure 4), accounting for the zeroes provided time-series estimates that more closely matched the reference time-series but still did not account for the sampling design.

The OBIS GLM time-series was generated from a fitted model that aimed to account for stratum and year effects, estimating a parameter for each year and each stratum. However, the model assumes that the strata effects are the same over the whole time-series and that the year effects are the same over all strata. As the GLM uses a log link and a negative binomial error distribution it treats catches of zero differently from the other methods and also reduces the effects that large catches have on the overall annual estimate.

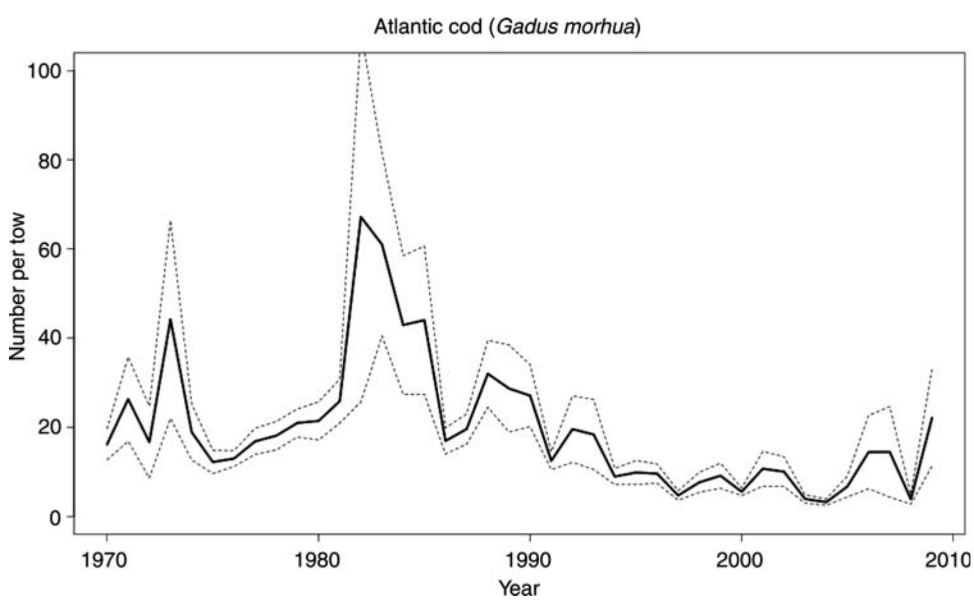

Figure 3. Abundance time-series for Atlantic cod obtained using the OBIS version of the DFO's SUMMER and SUMMER_TELEOST survey data from 1970 to 2009 in the Scotia–Fundy region of the Northwest Atlantic (NAFO divisions 4X and 4VsW), including an estimate of the uncertainty of the annual estimate of stratified random mean abundance. The mean annual estimate is plotted as a solid line along with ±1 s.e. (dotted lines).

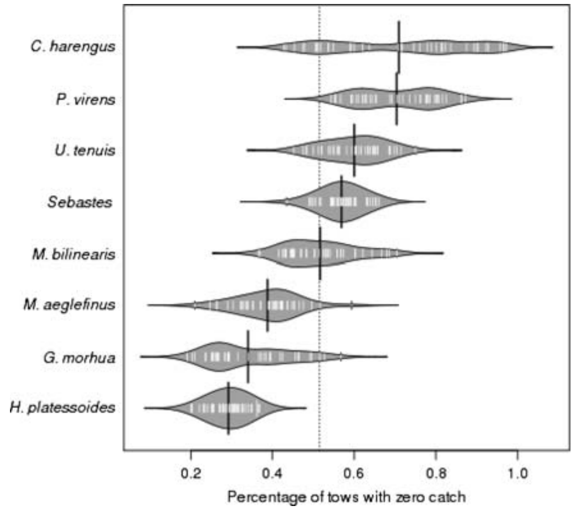

Figure 4. Proportion of tows with zero catch for the eight species of interest. The proportion of tows with zero catch is computed for each species and each year in the survey data ($n = 40$ for each species) and presented as a beanplot, showing the proportion of tows with zero catch for each year (thin vertical white lines) and the mean proportion of tows with zero catch (thick vertical black line). The dotted vertical line represents 50% of tows with zero catch.

The number of records in the OBIS datasets is a reflection of how common a species is in the survey samples. Although a species may be taken frequently, its real abundance in the samples may be less than that of species that are taken less often. This is the case for American plaice, the species with the most records in the OBIS record but with values of annual abundance far less than those of other species (Figure 2).

Agencies conducting similar surveys to those conducted by DFO Maritimes should consider making their data available on OBIS. An essential step in making data available on OBIS is to provide authoritative metadata. Additionally, using interpretations of OBIS fields similar to those used here will facilitate the wider use of groundfish survey datasets by researchers. Trawl survey scientific data from the US Northeast Fisheries Science Center were recently made available on OBIS (NOAA, 2005b), and they use similar encodings to DFO Maritimes data.

Although they provide crucial baseline information about a particular area, trawl surveys still have limited spatio-temporal coverage. Although combining survey time-series poses many challenges, doing so can augment knowledge of the spatio-temporal dynamics at scales that exceed those of individual surveys. As the different time-series computed from the OBIS data show, it is important to incorporate the sampling design into computations of abundance time-series.

To interpret correctly how trawl survey data relate to fish population dynamics, it is also important to consider seasonal migrations and other spatial shifts in the distribution of species during the time-series available. As such, the estimates derived from the survey data are just a starting point for further analysis of fish population dynamics. A stock assessment will also use data on landings and other sources of information about the species. The range of a species may also be limited to a subset of the strata, and assessment scientists need to use this knowledge to generate realistic time-series of abundance and biomass over the survey area. The methodology presented here can be used also to compute stratified random estimates over a subset of strata in the surveyed area.

We are aware that the time-series we computed provide only a measure of the number of fish caught per tow and not an estimate of total abundance over the area surveyed. Knowledge of the catchability of the fish species to the gear is also necessary in computing a value of total abundance. We only considered species abundance, but the OBIS data also contain biomass data for the surveys (stored in the "observedindividualweight" field of each OBIS record). As such, estimates of population biomass can also be computed using publicly available data.

The methods used to generate the OBIS version of the DFO Maritimes Research Vessel Trawl Surveys Fish Observations also mean that once the results of a survey have been uploaded into the production database, new data can be quickly updated in the public domain. In other words, additional catch records added to the OBIS version of a dataset are soon available for use. The data available on OBIS can also be used to conduct analyses of the diversity of fish species. The sampling protocols for fish species are consistent for the duration of the surveys, but care is necessary in accounting for the catchability of different fish species. Nevertheless, a dataset similar to that provided on OBIS has been used successfully to examine changes in fish diversity in the surveyed area (Shackell and Frank, 2003).

One of the many limitations of the OBIS data format is that it is not well suited for life-history analyses. For example, analyses of growth rate, maturation schedules, and age-specific fecundity cannot be performed using the data available on OBIS. However, such data are usually easily available from the data custodians identified in the OBIS metadata record, making the data discoverable and authoritative and providing a vehicle for wider dissemination of scientific trawl survey data.

Acknowledgements

We thank the officers and crews of Canadian Coast Guard vessels for their assistance during and efforts to complete scientific trawl surveys. We are also grateful to Lenore Bajona and Bette Hatt of DFO for help in updating the publicly available dataset and metadata records presented here, and the late Ransom Myers for his ideas and enthusiasm in developing a general framework for sharing scientific trawl survey data. Coilín Minto made useful comments on an earlier version of this manuscript, Wade Blanchard and Stephen Smith gave statistical advice, and three anonymous reviewers provided constructive comments which greatly improved the quality of this document.

References

Branton, R. M., and Ricard, D. 2007. Developing OBIS into a tool to provide reliable estimates of population indices for marine species from research trawl surveys. *In* Proceedings of "Ocean Biodiversity Informatics": an International Conference on Marine Biodiversity Data Management, pp. 19–24. Ed. by E. vanden Berghe, W. Appeltans, M. Costello, and P. Pissierssens. UNESCO/IOC, Vlaams Instituut voor de Zee (VLIZ), Bundesamt für Seeschifffahrt und Hydrographie (BSH). 192 pp.

Clark, D., and Branton, B. 2007a. DFO Maritimes Research Vessel Trawl Surveys. *In* OBIS Canada Digital Collections, version 1, Digital. OBIS Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada. http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=OBIS&OrigMetadataNode=GCMD&EntryId=OBIS.DFOgfsDBfish&MetadataView=Full&lbnode=mdlb2 (last accessed 30 August 2009).

Clark, D., and Branton, B. 2007b. DFO Maritimes Research Vessel Trawl Surveys. *In* OBIS Canada Digital Collections, version 1, Digital. OBIS Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada. http://www.iobis.org/OBISWEB/DisplayMetaData.jsp?content=meta/237.html (last accessed 30 August 2009).

Doubleday, W. G. 1981. Manual on groundfish surveys in the Northwest Atlantic. NAFO Scientific Council Studies, 2. 56 pp.

Doubleday, W. G., and Rivard, D. 1981. Bottom trawl surveys. Canadian Special Publication of Fisheries and Aquatic Sciences, 58. 273 pp.

EML Project. 2008. Ecological Metadata Language, version 2.0.1. http://knb.ecoinformatics.org/software/eml/ (last accessed 30 August 2009).

Gamer, M., Lemon, J., and Fellows, I. 2007. irr: Various Coefficients of Interrater Reliability and Agreement. R Package, version 0.70. http://www.r-project.org (last accessed 30 August 2009).

Kampstra, P. 2008. Beanplot: a boxplot alternative for visual comparison of distributions. Journal of Statistical Software, Code Snippets, 28: 1–9.

Lohr, S. L. 1999. Sampling: Design and Analysis. Duxbury Press, Pacific Grove, CA. 450 pp.

Myers, R. A. 2000. The synthesis of dynamic and historical data on marine populations and communities; putting dynamics into the Ocean Biogeographical Information System (OBIS). Oceanography, 13: 56–59.

NASA (National Aeronautics and Space Administration). 2008. Directory Interchange Format (DIF) Writer's Guide. http://gcmd.nasa.gov/User/difguide/DFI_Guide_2008.pdf (last accessed 30 August 2009).

NOAA. 2005a. Species of concern fact sheet. http://www.nmfs.noaa.gov/pr/pdfs/species/riverherring_detailed.pdf (last accessed 30 August 2009).

NOAA. 2005b. Northeast Fisheries Science Center bottom trawl survey data. NOAA National Marine Fisheries Service (NMFS) Northeast Fisheries Science Center, Ecosystems Survey Branch, Woods Hole, MA, 2005-04-30, Tabular Digital Data. http://www.iobis.org/OBISWEB/DisplayMetaData.jsp?content=meta/1435.html (last accessed 30 August 2009).

OBIS. 2009a. Ocean Biogeographic Information System. http://www.iobis.org (last accessed 30 August 2009).

OBIS. 2009b. The OBIS Schema, version 1.1. http://iobis.org/tech/provider/schemadef1.html (last accessed 30 August 2009).

OBIS Canada. 2009. Ocean Biogeographic Information System—Canadian Regional OBIS Node. DFO Maritimes survey strata statistics. http://www.marinebiodiversity.ca/OBISCanada/etc/dfo-maritimes-survey-strata-statistics (last accessed 30 August 2009).

PostgreSQL Global Development Group. 2009. PostgreSQL, version 8.3.8. http://www.postgresql.org (last accessed 30 August 2009).

R Development Core Team. 2009. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ripley, R., and Lapsley, M. 2009. RODBC: ODBC Database Access. R package version 1.3-0. http://CRAN.R-project.org/package=RODBC (last accessed 30 August 2009).

Shackell, N., and Frank, K. T. 2003. Marine fish diversity on the Scotian Shelf, Canada. Aquatic Conservation: Marine and Freshwater Ecosystems, 13: 305–321.

Shrout, P. E., and Fleiss, J. L. 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 2: 420–428.

Smith, S. J. 1996. Assessment of groundfish stocks based on bottom trawl survey results. NAFO Scientific Council Studies, 28: 25–53.

Taxonomic Diversity Working Group. 2009. Darwin Core. http://rs.tdwg.org/dwc/index.htm (last accessed 30 August 2009).

Tremblay, M., Black, G., and Branton, R. 2007. The distribution of common decapod crustaceans and other invertebrates recorded in annual ecosystem surveys of the Scotian shelf 1999–2006. Canadian Technical Report of Fisheries and Aquatic Sciences, 2762. iii + 74 pp.

Venables, W. N., and Ripley, B. D. 2002. Modern Applied Statistics with S, 4th edn. Springer, New York.

Wessel, P., and Smith, W. H. F. 1991. Free software helps map and display data. EOS Transactions of the American Geophysical Union, 72: 441.

doi:10.1093/icesjms/fsp275