ICES CM 2008/R:25

A proposal for a consistent use of the North Sea IBTS data.

Remment ter Hofstede & Niels Daan

The ICES DAtabase for TRAwl Surveys (DATRAS) stores data from several international groundfish surveys in the north-eastern Atlantic area and is used as a major data source for many studies assessing the status of stocks and changes in the structure, functioning and diversity of commercial and non-target fish species and assemblages. Hence, it is essential that the data are of high quality and fully reliable for studies of the wider fish community. However, it has been highlighted before that DATRAS suffers from many problems associated with input errors, the misidentification of specific taxa, and inconsistent reporting at a range of taxonomic levels (species, genus or family). Besides correcting identified errors, the only way to improve the consistent use of the database among subsequent analyses is to develop a protocol for a standard correction procedure of problematic records that should be followed by all users. We provide a first proposal for a standard correction procedure of one component of DATRAS, namely the data from the International Bottom Trawl Survey covering the North Sea, Skagerrak and Kattegat (North Sea IBTS).

Keywords: data quality, misreporting, biodiversity, DATRAS, IBTS, North Sea

Contact author: Remment ter Hofstede, Wageningen IMARES, Institute for Marine Resources & Ecosystem Studies, P.O. Box 68, 1970 AB IJmuiden, The Netherlands
[tel +31 317 480900; fax +31 317 487326; e-mail remment.terhofstede@wur.nl]

Introduction

Goundfish survey data are not only becoming increasingly more important for assessing the status of commercial and non-target fish species, they also provide the major data source for large-scale spatial and temporal analyses of fish assemblages in continental shelf waters, and for the derivation of metrics with which to assess changes in the structure, functioning and diversity of these assemblages. Hence, the strict implementation of species-identification and sampling protocols of such surveys must ensure that data collection is appropriate for studies of the wider fish community.

ICES has created a common DAtabase for TRAwl Surveys (DATRAS) for storing data from several bottom- and beam trawl surveys in the north-eastern Atlantic area (Piet, 2004). Such an international database facilitates the use of the survey data by ICES working groups for stock assessments and ecosystem studies that sustain ICES advice, but they may also be used by other parties. Clearly, extensive quality control of the data entering DATRAS is an essential part of the process. All data are thoroughly checked by the national institutes before they are submitted to ICES, and upon entering the data ICES performs a comprehensive checking routine. However, as any database, DATRAS is not free of errors, especially not so in the older records. Daan (2001) highlighted potential problems associated with (a) input errors and (b) the misidentification of specific taxa. The evidence for inconsistencies in reporting by different countries has only been supported by subsequent analyses (ICES, 2006). For instance, several species or genus are sometimes reported at higher taxonomic levels (genus or family), even if there can be little doubt as to the actual species or genus involved, because others are extremely unlikely within the area surveyed. This may affect the utility of survey data for fish-assemblage studies if such discrepancies are not corrected for (e.g. in studies of biodiversity and other metrics for fish communities).

It should be emphasized that so far progress in identifying inconsistencies in the reporting of various taxa has been restricted to the International Bottom Trawl Survey covering the North Sea, Skagerrak and Kattegat (North Sea IBTS), which represents only one component of DATRAS. These studies on the quality of the North Sea IBTS dataset have provided an extensive though not exhaustive inventory of problematic records (Daan, 2001; ter Hofstede and Daan, 2006). For all other surveys, similar systematic analyses have not yet been conducted, but there is no reason to assume that the situation would be any different.

The North Sea IBTS dataset has been used at face value in a variety of studies on fish communities and their responses to human activities and global warming (e.g. Jennings *et al.*, 2002; Callaway *et al.*, 2002; Beare *et al.*, 2004). However, neglecting unmistakable errors raises questions about the reliability of their results. Analyses have sometimes been limited to only a part of the dataset that has been considered consistent and trustworthy, for example using only national data (e.g. Perry et al., 2005; Maxwell and Jennings, 2005; Greenstreet and Rogers, 2006). This is unfortunate because it undermines the idea of developing a comprehensive survey data set. More importantly, it does not solve the problem since individual countries have been shown to be inconsistent in their reporting over time (ICES, 2005, 2007). Several studies that have used the North Sea IBTS data do refer to a priori corrections (Daan *et al.*, 2005; Hiddink and ter Hofstede, 2007). However, there has been no agreed protocol for dealing with obvious mistakes, problematic taxa and unlikely records. Therefore, different authors may arrive at different results and it may be virtually impossible for other scientists to reproduce the results obtained. This is an awkward situation, particularly for ICES working groups when using these data for providing management advice.

One way to improve the situation is for ICES to agree on a protocol for a standard correction procedure of problematic records that should be followed by all working groups in order to ensure consistency among subsequent analyses. This protocol should also be brought to the attention of individual scientists requesting access to DATRAS for specific analyses. Although ICES can obviously not enforce the use of a fixed protocol on such users, the guardian of DATRAS does have a responsibility for facilitating its proper use.

Elaborating on DATRAS quality issues dealt with during the ICES Workshop on Taxonomic Quality Issues in the DATRAS Database (WKTQD; ICES, 2007), we developed a first proposal for a standard correction procedure.

North Sea IBTS

The North Sea International Bottom Trawl Survey has since its start in 1965 as a specific International Young Herring Survey gradually evolved into a general demersal fish survey (Heessen *et al.*, 1997). It has been carried out annually during the first quarter up till the 1990s, on a quarterly basis in the period 1991-1996, and twice per year (1st and 3rd quarter) since 1997. During each quarter, 6-8 countries sampled a total of more than 300 hauls in the North Sea (between 51-62° latitude), Skagerrak and Kattegat, leading to a dataset with information on catches in over 24 000 hauls. The survey design is random stratified according to a grid of ICES rectangles (0.5° latitude; 1° longitude; approximately 56 x 56 km). In principle, each rectangle is sampled by two different countries during each survey. Gears have varied among years and vessels initially, but since 1983 a standard bottom-trawl net (chalut à Grande Ouverture Verticale, GOV-trawl) has been used by all countries. This gear has been designed specifically to sample fish that live close to the seabed. The catch is sampled to provide the length-frequency distributions for all fish species caught. Details of the gear and sampling strategies can be found in the manual for the North Sea IBTS, revision VII (ICES, 2006).

Methods

Analyses have been made using all available data in the DATRAS section "North Sea International Bottom Trawl Survey" (1965-2007; last modified on 11 May 2007) to get a complete list of taxa recorded. Based on previous work by Daan (2001), ter Hofstede & Daan, (2006) and WKTQD (ICES, 2007), we made a list of all taxa that have been reported inconsistently over all years and therefore need a global correction. This concerns two types of inconsistencies: i) redundant taxa, i.e. genus or families that are represented only by one species and genus, respectively, and therefore should have been reported as that particular species or genus; and ii) species of which the correct identification is highly uncertain, because this requires specialist taxonomic knowledge that is not routinely available on board.

Results

i. Consistent taxonomy

The idea of taxonomy (and its coding in one system or another) is to provide a unique interpretation of the code used. Therefore, different taxa (and codes) that lead to the same interpretation must be avoided, because they suggest a non-existent difference. If a genus is represented by a single species in a particular area, recordings at the genus level are redundant and records should only be stored in the database at the species-level. The same applies to families represented by a single genus, in which case the family name is redundant and should be considered invalid. The redundant taxa present in the North Sea IBTS dataset within DATRAS and their appropriate coding are given in Table 1. We suggest to make the appropriate changes within DATRAS rather than to the output, because there would be no loss of information involved.

Table 1. List of invalid taxa currently reported in the North Sea IBTS DATRAS and their appropriate identification.

TSN	taxon stored	TSN	proposed use
159700	*Lampetra*	159719	*Lampetra fluviatilis*
159721	*Petromyzon*	159722	*Petromyzon marinus*
160846	*Raja*	160845	Rajidae
162057	Argentinidae	162061	*Argentina*
164771	*Gadiculus*	164772	*Gadiculus argenteus*
164789	Merlucciidae	164795	*Merluccius merluccius*
165255	*Lycodes*	165284	*Lycodes vahli*
166271	Zeiformes	166287	*Zeus faber*
166309	Caproidae	166320	*Capros aper*
166438	Syngnathoidei	166443	Syngnathidae
170316	*Dicentrarchus*	170317	*Dicentrarchus labrax*
171335	Anarhichadidae	171336	*Anarhichas*
171691	Callionymidae	171692	*Callionymus*
173000	*Solea*	173001	*Solea vulgaris*
173002	*Solea solea*	173001	*Solea vulgaris*
173020	*Buglossidium*	173021	*Buglossidium luteum*
173022	*Microchirus*	173026	*Microchirus variegatus*

ii. Uncertain identification

Large irregularities in reporting may result from superficial distinguishing features being insufficient. Several North Sea fish species can only be identified by counting fin rays, gillrakers or scales and there is usually no time on board to do this on a routine basis. Thus, even several of the more common species such as sandeels and gobies may be reported differently by different scientists in charge. Rather than accepting reports on individual species within these groups at face value, we suggest to raise all the historically collected species information to the genus level (see Table 2). In this case, it would be appropriate not to change the data base, because a small part of the records may have been appropriately assigned to species. However, for routine community analyses as well as for determining temporal and spatial trends by species these records cannot be trusted and therefore the higher taxa should be provided in the DATRAS output.

Table 2. List of uncertain taxa currently reported in the North Sea IBTS DATRAS and their proposed identification.

TSN	taxon stored	TSN	proposed use
160240	*Mustelus asterias*[1)]	160226	*Mustelus*
160242	*Mustelus mustelus*[1)]	160226	*Mustelus*
161708	*Alosa alosa*[2)]	161701	*Alosa*
161716	*Alosa fallax*[2)]	161701	*Alosa*
572694	*Alosa agone*	161701	*Alosa*
161996	*Salmo salar*	161994	*Salmo*
161997	*Salmo trutta*	161994	*Salmo*
166463	*Syngnathus rostellatus*	166444	*Syngnathus*
166464	*Syngnathus acus*	166444	*Syngnathus*
166467	*Syngnathus typhle*	166444	*Syngnathus*
171676	*Ammodytes tobianus*	171671	*Ammodytes*
171677	*Ammodytes marinus*	171671	*Ammodytes*
171680	*Gymnammodytes semisquamatus*[3)]	171671	*Ammodytes*
171682	*Hyperoplus lanceolatus*	171681	*Hyperoplus*
171683	*Hyperoplus immaculatus*	171681	*Hyperoplus*
171978	*Pomatoschistus minutus*[4)]	171977	*Pomatoschistus*
171980	*Pomatoschistus pictus*	171977	*Pomatoschistus*

1) *The commonly used feature of white spots on the sides is extremely variable and runs from hardly visible to pronounded white stars. It is not by itself considered a discriminating feature.*
2) *These two species are known to interbreed and species identified as Alosa alosa have seldom all the discriminating features.*
3) *Including Gymnammodytes under the genus Ammodytes is a practical solution, because the two genus are difficult to distinguish, but as a group they can be readily distinguished from Hyperoplus. Formally, the taxon should be Ammodytidae excluding Hyperoplus.*
4) *Pomatoschistus minutus is likely to include locally P. microps and P. lozanoi.*

Discussion

During explorative analyses of the North Sea IBTS section in DATRAS, obvious errors and inconsistencies have been encountered that can result in a major bias on the results of various community analyses as well as on temporal trends and spatial distribution maps of the species involved. This may obviously lead to misinterpretation of the survey results. In case the reliability of specific records is mistrusted on the basis of available trusted taxonomic and biological information, these records should be adjusted before they are used, particularly in the context of providing management advice. Using the proposed protocol for reaching consistency in taxonomy (Table 1) and for avoiding the use of highly uncertain identifications (Table 2) can be considered as a first step towards a more uniform exploitation of the North Sea IBTS dataset.

A problem within DATRAS that has not been addressed here is related to unreliable subsets. Annual inconsistencies in reporting are known to occur in particular vessel-year subsets, due to insufficient experience from the responsible crew members with the identification of specific species, for instance because certain species are rarely seen in the area covered by the vessel, or

since the scientist on board (and their experience) vary over time. Daan (2001), ter Hofstede & Daan (2006), and ICES (2007) have listed many of these inconsistencies based on inter-vessel and inter-year comparisons of reported catches of several species. These errors are most difficult to deal with, because raising all species to a higher taxonomic level would result in the loss of a large amount of reliable information. The only solution to this problem is that individual countries regularly check their own data carefully against those of others fishing in the same area, make the appropriate corrections (change either the species code or use the appropriate genus code), and submit a new set of corrected data to DATRAS.

For community analyses, records referring to a higher taxon in specific subsets have to be split into the appropriate lower taxa, because both deleting and including the record would bias the results. An appropriate protocol for dealing with e.g. genus records would be to use the reliable species information available within the data base as the basis for decomposing genus records in the relevant species. To obtain the most likely species composition, it would seem appropriate to use a stepwise procedure, incorporating reliable records on catches of other vessels, in neighbouring ICES rectangles, or in preceding and following years. However, the development of such an algorithm goes beyond the scope of this paper.

The analyses performed so far have by no means been exhaustive and we realise that the proposed corrections given here may not solve all problems of the North Sea IBTS data within DATRAS. Still, accepting the proposed protocol would mean an important step forward to reach a uniform use of the dataset, and our propositions as given in Tables 1 and 2 should clearly lead to more consistency in the output from DATRAS. Therefore, we plea for endorsement of the suggested protocol by ICES, both regarding the use of DATRAS by ICES working groups and as a recommended guideline for external users.

Our propositions can also be used to amend other North Sea datasets, such as those from the beam-trawl surveys, but the appropriate set of invalid taxa for other areas still needs to be determined. We recommend that initiatives should be taken within ICES to carry out comprehensive checks on data sets for other regions or survey types as well, because it seems only likely that the problems identified are not restricted to the North Sea IBTS.

Acknowledgements

Special thanks goes to the participants of the ICES Workshop on Taxonomic Quality Issues in the DATRAS Database (WKTQD), from which workshop this proposed protocol is a spin-off.

References

Beare, D.J., Burns, F., Greig, A., Jones, E.G., Peach, K., Kienzle, M., McKenzie, E., and Reid, D.G. 2004. Long-term increases in prevalence of North Sea fishes having southern biogeographic affinities. Marine Ecology Progress Series, 284: 269-278.

Callaway, R., Alsvaag, J., de Boois, I., Cotter, J., Ford, A., Hinz, H., Jennings, S., Kroencke, I., Lancaster, J., Piet, G., Prince, P., and Ehrich, S. 2002. Diversity and community structure of epibenthic invertebrates and fish in the North Sea. ICES Journal of Marine Science, 59(6): 1199-1214.

Daan, N. 2001. The IBTS database: a plea for quality control. ICES CM 2001/T:03. 19 pp.

Daan, N., Gislason, H., Pope, J.G., and Rice, J.C. 2005. Changes in the North Sea fish community: evidence of indirect fishing? ICES Journal of Marine Science, 62: 177-188.

Greenstreet, S.P.R., and Rogers, S.I. 2006. Indicators of the health of the North Sea fish community: identifying reference levels for an ecosystem approach to management. ICES Journal of Marine Science, 63(4): 573-593.

Heessen, H.J.L., Dalskov, J., and Cook, R.M. 1997. The international bottom trawl survey in the North Sea, the Skagerrak and Kattegat. ICES CM 1997/Y:31. 25 pp.

Hiddink, J.G., and ter Hofstede, R. 2007. Climate induced increases in species richness of marine fishes. Global Change Biology. accepted.

ter Hofstede, R., and Daan, N. 2006. Quality check surveys: DATRAS North Sea IBTS. In: ICES. 2006. Report of the International Bottom Trawl Survey Working Group (IBTSWG), 27-31 March 2006, Lysekil, Sweden. ICES CM 2006/RCM:03, Ref. ACFM, 298 pp.

ICES. 2006. Report of the International Bottom Trawl Survey Working Group (IBTSWG), 27-31 March 2006, Lysekil, Sweden. ICES CM 2006/RCM:03, Ref. ACFM, 298 pp.

ICES. 2007. Report of the Workshop on Taxonomic Quality Issues in the DATRAS Database (WKTQD), 23-25 January 2007, ICES, Copenhagen. ICES CM 2007/RMC:10. 46 pp.

Jennings, S., Greenstreet, S.P.R., Hill, L., Piet, G.J., Pinnegar, J.K., and Warr, K.J. 2002. Long-term trends in the trophic structure of the North Sea fish community: Evidence from stable-isotope analysis, size-spectra and community metrics. Marine biology, 141(6): 1085-1097.

Maxwell, D.L., and Jennings, S. 2005. Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. Journal of Applied Ecology, 42: 25-37.

Perry, A.L., Low, P.J., Ellis, J.R., and Reynolds, J.D. 2005. Climate Change and Distribution Shifts in Marine Fishes. Science, 308: 1912-1915.

Piet G. (2004) Development of a central database for European trawl survey data (DATRAS). Final report. DGXIV(European Commission, Brussels) EC project QLRT-2001-00025. 70 pp.