ICES Journal of Marine Science, 54: 518–523. 1997

Short communication

A new software system for the PIROP database: data flow and an approach for a seabird-depth analysis

F. Hüttmann and A. R. Lock

Hüttmann F. and Lock, A. R. 1997. A new software system for the PIROP database: data flow and an approach for a seabird-depth analysis. – ICES Journal of Marine Science, 54: 518–523.

PIROP (Programme Intégré des Recherches sur les Oiseaux Pélagiques) is a well-known and unique database for seabird observations collected from the whole eastern Canadian coast, from the Gulf of Maine to the Canadian Arctic, and some other regions, between 1966 and 1992. Although several major seabird studies have already used the PIROP data, the potential exists for further data analysis. This is particularly true for statistical analyses of seabird distribution in relation to oceanographic data. Collection of field data for PIROP has evolved and improved over a period of 20 years, and new software for data entry has been developed that allows for quick and efficient entry of data on seabird observations made at sea. Thus, advanced data analysis, including spatial analyses using Geographical Information Systems (GIS), can now be done. Data may be analysed, imported and exported, and previously compiled data-sets can also be implemented into the PIROP scheme using data conversion algorithms in the import feature of the new PIROP software system. Using the new software system, an example of an analysis of the PIROP data is presented. Positive sightings of northern gannets (*Sula bassana*) and northern fulmars (*Fulmarus glacialis*) in the month of August were extracted and analysed in relation to sea depth data derived from a digital database (ETOPO5). A Classification and Regression Tree (CART) was used as an exploratory data analysis method, indicating how a distribution model for gannets versus fulmars in relation to sea depth may be derived from these data-sets. The model suggests that fulmar observations occur north of 51.57°N latitude over both shallow and deep waters. South of 51.57°N latitude, fulmars are found over waters deeper than 203 m, while gannets are found over shallow regions less than 203 m deep. This result is probably due to the preponderance of non-breeding birds among the observed fulmars, whereas gannets are confined to the vicinity of breeding colonies, which are situated on the continental shelf.

Key words: bathymetry, CART, classification and regression tree, data entry mask, *Fulmarus glacialis*, geographical information system, GIS, seabird, software, *Sula bassana*, PIROP database.

F. Hüttmann: Atlantic Co-operative Wildlife Ecology Research Network (ACWERN), Faculty of Forestry and Environmental Management, University of New Brunswick, P.O. Box 44555, Fredericton, New Brunswick, Canada E3B 6C2. A. R. Lock: Environment Canada, Canadian Wildlife Service, Atlantic Region, Queen Square, 45 Alderney Drive, Dartmouth, Nova Scotia, Canada B2Y 2N6 Correspondence to F. Hüttmann: tel: +5064526033; fax: +5064533538; email: k9wk@unb.ca

Introduction

The Programme Intégré Recherches sur les Oiseaux Pélagiques (PIROP) was established in December 1966 as a monitoring programme for the vulnerability of seabirds to oil spills in eastern Canadian waters. The PIROP database consists mostly of counts of seabirds at sea, made in 10 min transects of undefined width from a moving platform (vessels of opportunity) (Brown *et al.*, 1975; Diamond *et al.*, 1986; Gaston *et al.*, 1987). Most of the seabird surveys in PIROP were carried out from 1966 to 1992 by R. G. B. Brown for the Canadian Wildlife Service at the Bedford Institute of oceanography in Dartmouth (Brown, 1977; Brown and Nettleship, 1984; Brown, 1986; Lock *et al.*, 1994). although some surveys were made elsewhere, for example in the

1054–3139/97/040518+06 $25.00/0/jm970260

Benguela Current (Peru) and around Tierra del Fuego (Argentina), most of the data were collected near the Canadian continental shelf and in the Canadian Arctic. In the North Atlantic, more than 180 000 observation records have been collected.

Some of the last remaining survey data (ca. 7000 observation records) have also recently been entered onto the database, as have data from the Manomet Bird Observatory. Together with the Manomet data-set, which consists of more than 10 000 observation records, PIROP now comprises ca. 200 000 records available in a digital format for analysis and research (Lock, unpublished).

This paper describes how field data can be processed for a research project. The typical procedure involves: (a) keying PIROP field data efficiently into a database; (b) browsing, querying the raw data, and visualizing the extracted data using maps; and (c) carrying out appropriate statistics and visualizing the results (graphs/maps).

For illustration, the month with the largest amount of PIROP data available, August (ca. 40 000 observation records), was selected. The northern gannet (*Sula bassana*) and northern fulmar (*Fulmarus glacialis*) were used as examples to compare abiotic factors defining the distribution of two ecologically very different species. Since two species were being compared directly, only transects that included positive sightings of one or other of these two species were included.

Methods

Data entry program

The PIROP Data Entry software was developed in FoxPro 2.6 (Lock, unpublished; Siegel, 1994). The data are stored in the relational database, PIROP, which is structured into two databases, consisting of the watch data (keyfield.dbf) and the seabird data (species.dbf). Software versions are available for IBM/WINDOWS and MAC. The PIROP Data Entry Mask is shipped on two diskettes (''program'' files for MAC and ''exe'' files for WIN3.11/WIN 95) with manuals. The standard procedures for software installation in WINDOWS were used. A first version of the new PIROP Data Entry program was used and evaluated by the first author for keying in the remaining PIROP data from the last cruises of the survey program. Few changes and updates were necessary.

Data analysis

Following Diamond *et al.* (1993), the study area includes the Canadian waters of the North Atlantic (south-west corner at 40°N 100°W, north-east corner at 80°N 40°W). The full PIROP data-set in the study area was queried for the month of August for the gannet and the fulmar using Structured Query Language (SQL). These two species serve as an example for the analytical method presented, and were selected because they breed in the study area, and thus allow comparison at this time of the year. We wished to investigate differences in distribution between gannets and fulmars in relation to depth at sea, so only positive sightings were used (i.e. presence data). An SQL query in PIROP resulted in 4716 10 min transect counts at different locations from 1966 to 1992. These SQL query results were filtered by ship speed >5 knots, view angle ca. 180°, 10 min watch types and a ''wholly reliable'' species identification. The raw data were browsed beforehand, and a data distribution was found that showed no need to exclude other data, such as poor weather conditions. Seabird sightings related to human fishery activities were retained because this is probably an important factor governing the distribution of the selected species in the study area. A total of 291 gannet sightings and 7011 fulmar sightings were extracted and used in the distribution model.

For sea depth, the ETOPO5 (National Oceanic and Atmospheric Administration 1996) data were used because these data are available in a digital format (in this case ASCII) at a high resolution, and can be downloaded conveniently from the Internet/WWW free of charge. The ETOPO5 data consist of latitude (LAT, decimal degrees), longitude (LONG, decimal degrees) and sea depth (DEPTH, m) information, mainly one measurement of depth per 5 naut. m (approximately $5' \times 5'$N/W). Finally, all data on land were excluded (DEPTH >0), resulting in 164 000 sea depth data points for the study area.

Both SQL query results, ''northern gannets'' and ''northern fulmars'', were pooled together (totalling 7302 seabird observation records). A depth value, in metres, from ETOPO5 was attached to the selected positive sightings of the gannet and fulmar (lat/long) by overlaying both data-sets in SPANS-GIS. Since PIROP provides lat/long in both formats, decimal and sexagesimal, ETOPO5 lat/long values could be used directly, without re-calculation.

The SQL query results ''northern gannets'' and ''northern fulmars'' were imported into SPANS-GIS and SPANS MAP (INTERAC TYDAC, 1992, 1993, 1995) and plotted. In order to present the seabird distribution in relation to sea depth, an ETOPO5 map was superimposed on the seabird data.

After examining the spatial distribution of the resulting data-set, it was imported into the statistical software package SPLUS (Version 3.3, WINDOWS), and a Classification and Regression Tree (CART; StatSci, 1995; Venables and Ripley, 1994; Barndorff-Nielsen *et al.*, 1993; Bell, 1996) was carried out for the columns ''seabird species id'' versus ''sea depth'' and ''latitude''. Effects of pruning and shrinking for CART were examined in SPLUS (StatSci, 1995). Venables and

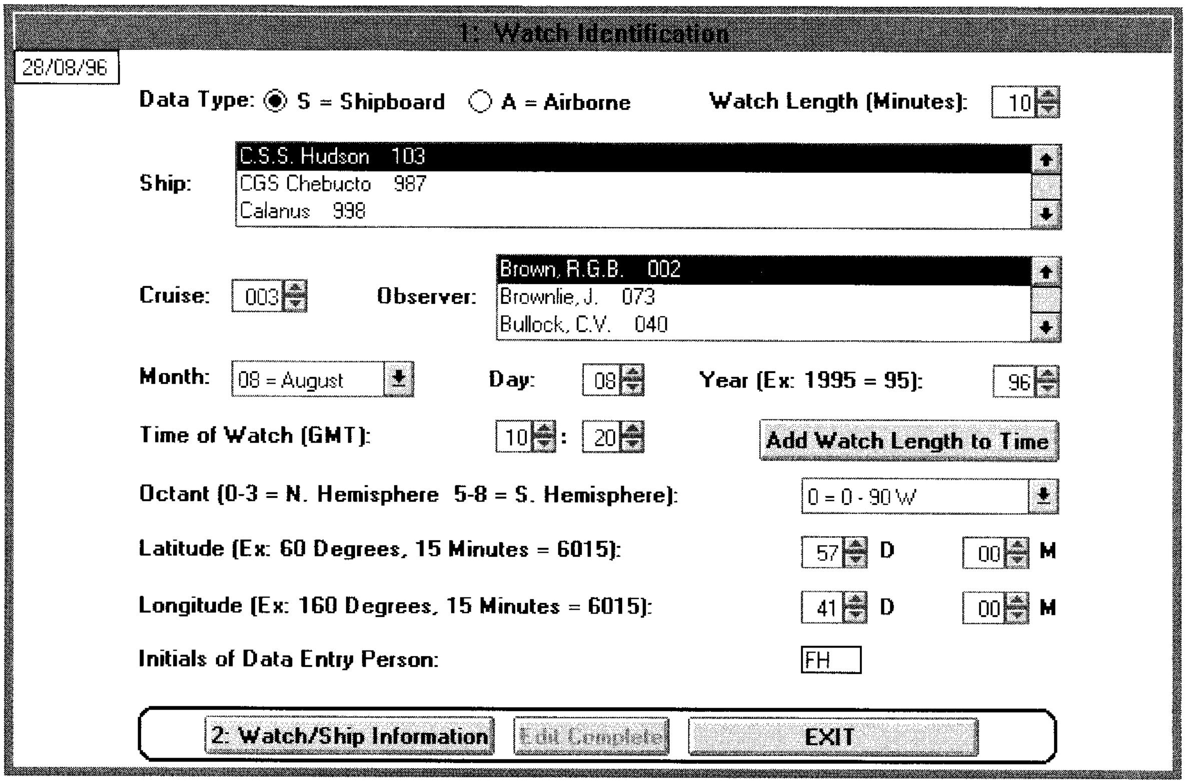

Figure 1. The first window of the PIROP Data Entry Mask.

Ripley (1994) was used in order to overcome software errors in the recent SPLUS version 3.3 for WINDOWS.

Results

Data entry program

The software proved to be a very swift and efficient tool for entering data from a seabird survey into a relational database. For example, a 5 h July transect in the Bay of Fundy with 280 records was completely encoded in ca. 1 h. The process of entering data is divided into a sequence of four windows, which includes all the necessary information for the watch conditions and the seabird information (Fig. 1). Pull-down menus, radio buttons, and predefined lists allow for entering data very efficiently. The predefined lists and pull-down menus may be modified for the user's needs. The PIROP Data Entry Mask is structured and organized logically, avoiding any redundancy in the process of entering data.

The software requires ca. 3 MB for the source code, plus the necessary space for the data on the hard disk. The program is very fast, and problems under any standard PC/MAC environment were not encountered.

Data can be merged with other existing data-sets, and SQL queries may be carried out using standard database packages. This allows for further processing with analytical, statistical software packages and mapping tools, e.g. advanced GIS (see Fig. 3).

A manual accompanies the software, explaining in detail the structure of the relational DBF database and how to use and to modify the PIROP Data Entry Mask. The software is available for distribution from the Canadian Wildlife Service (see address above).

Data analysis

For 1966–1992, after filtering, PIROP August data comprised 291 gannet and 7011 fulmar observations (positive sightings) in the study area (Fig. 3). CART was used successfully to compare gannet and fulmar distributions, and allowed a model to be fitted to the data-set, using latitude and sea depth as predictors for the ''seabird id'' response variable. The CART for the full data-set using seabird ''species id'' in relation to sea depth and latitude (SPLUS command: tree (sbirdid ~ sea depth + latitude)) proved to be too large for a meaningful analysis (statistics for the ''overfitted tree'': 66 terminal nodes, residual mean deviance = 0.06141, misclassification error rate = 0.01328). Shrinking and pruning plots (SPLUS command: prune.tree(tree (sbirdid ~ sea depth + latitude), best = 3)) revealed that

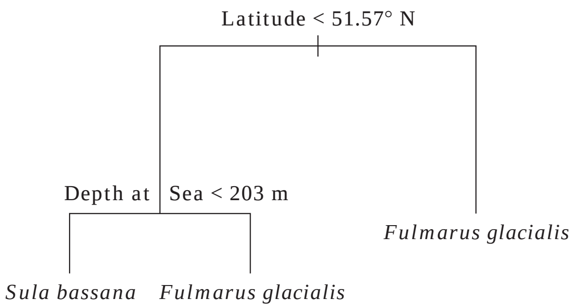

Figure 2. A CART (Classification and Regression Tree: pruned k = 40, split criteria: latitude and depth at sea, three terminal nodes, residual mean deviance = 0.1327, misclassification error rate = 0.02383) based on the PIROP August data for northern fulmars and northern gannets in relation to sea depth (from ETOPO5 database).

60% of the deviance for the distribution of gannets and fulmars was explained by the first three CART nodes (results for the ''pruned tree'': split criteria = latitude and sea depth, three terminal nodes, residual mean deviance = 0.1327, misclassification error rate = 0.02383; Fig. 2).

Latitude (51.57°N) was the principal parameter identified by CART, while depth at sea (203 m) played a secondary role in distinguishing presence of fulmar observations versus gannet observations in relation to depth at sea. The model suggests that sea depth did not affect the distribution of fulmar north of 51.57°N. The distribution map (Fig. 3) shows that fulmar observations south of 51.57°N mostly occur over waters deeper than 203 m. In the study area, this is mainly the deep sea channel (ca. 3000 m deep) between the Grand Banks and the Flemish Cap. Geographically, gannets clearly were confined to the Canadian continental shelf (boreal) in August.

Discussion

Data entry program

An efficient design of the data flow and the layout of a data entry process is very important for monitoring programs with large amounts of data. This is the interface that may act as a bottleneck for data quality. Keying in data is usually done by non-specialists, which may lead to a relatively high error rate. This needs to be avoided by making the design and layout for entering data easy to learn and as convenient and efficient as possible. We suspect that this part of the research process is usually ignored or believed to be unimportant, although data flow processes play a major role in the success of large, data-intensive projects.

The PIROP Data Entry Mask proved to be successful for keying data conveniently and efficiently into a relational database. No problems were encountered using this program. Problems in using the PIROP Data Entry Mask may occur only when the (inexperienced) user forgets to complete fields in the windows. At the outset, the amount of data that might be keyed in per seabird observation could be confusing for a new user and it might take some hours for an inexperienced user to learn to use the program efficiently. Working with the PIROP Data Entry Program is demanding, and it is necessary to focus on the data entry task. It is a general problem in such programs that no internal error-checking can compensate for missing or erroneous data.

Seabird abundance in relation to sea depth

Brown (1986) summarized the oceanography of the study area, but did not mention the bathymetry of this ecosystem. The bathymetry of the waters of Atlantic Canada is described using contour maps in Lock *et al.* (1994). The relationship between sea depth and seabird abundance has been studied by other authors, but only as a linear one, e.g. Cairns and Schneider (1990), Diamond and Prys-Jones (1986), Duffy (1989), Stone *et al.* (1995), Schneider (1997). A single factor approach is not likely to be an appropriate explanation for ecological problems. However, depth of sea may influence many other variables, such as water temperature, abundance of upwellings, nutrients in the water column, fish/plankton/sea mammal distribution, etc. (Shuntov, 1972; Schneider, 1997). Therefore, it may be expected to characterize the habitat for seabirds, and might also serve as a first step in explaining seabird abundance. This is especially the case for eastern Canadian waters, where depth classes vary strongly and several steep depth gradients are found (National Oceanic and Atmospheric Administration, 1996; Lock *et al.*, 1994).

Because of the difficulties of counting seabirds at sea, the explanation of 60% of the CART deviance by only three CART nodes, latitude and sea depth, can be considered as very good. The results from the CART for the selected PIROP August data-set suggest that the area of 51.57°N latitude delimits the observations of fulmars and gannets. This threshold in the model can probably be related to the breeding range of gannets, which breed south of this latitude (Brown and Nettleship, 1984; Lock *et al.*, 1994; Brown, 1986). North of 51.57°N latitude, fulmar observations occur in deep sea and shallow regions; in this region, fulmar populations will include non-breeding birds as well as birds breeding at colonies in Arctic Canada and west Greenland.

The model shows that sightings of fulmars south of 51.57°N occur at a sea depth deeper than 203 m, mainly around the area of the deep sea channel (ca. 3000 m deep) between the Grand Banks and the Flemish Cap (Brown *et al.*, 1975; Brown, 1986; Fig. 3). No breeding colonies are found in this area (Lock *et al.*, 1994) and the results of bird banding (Brewer *et al.*, in press) suggest

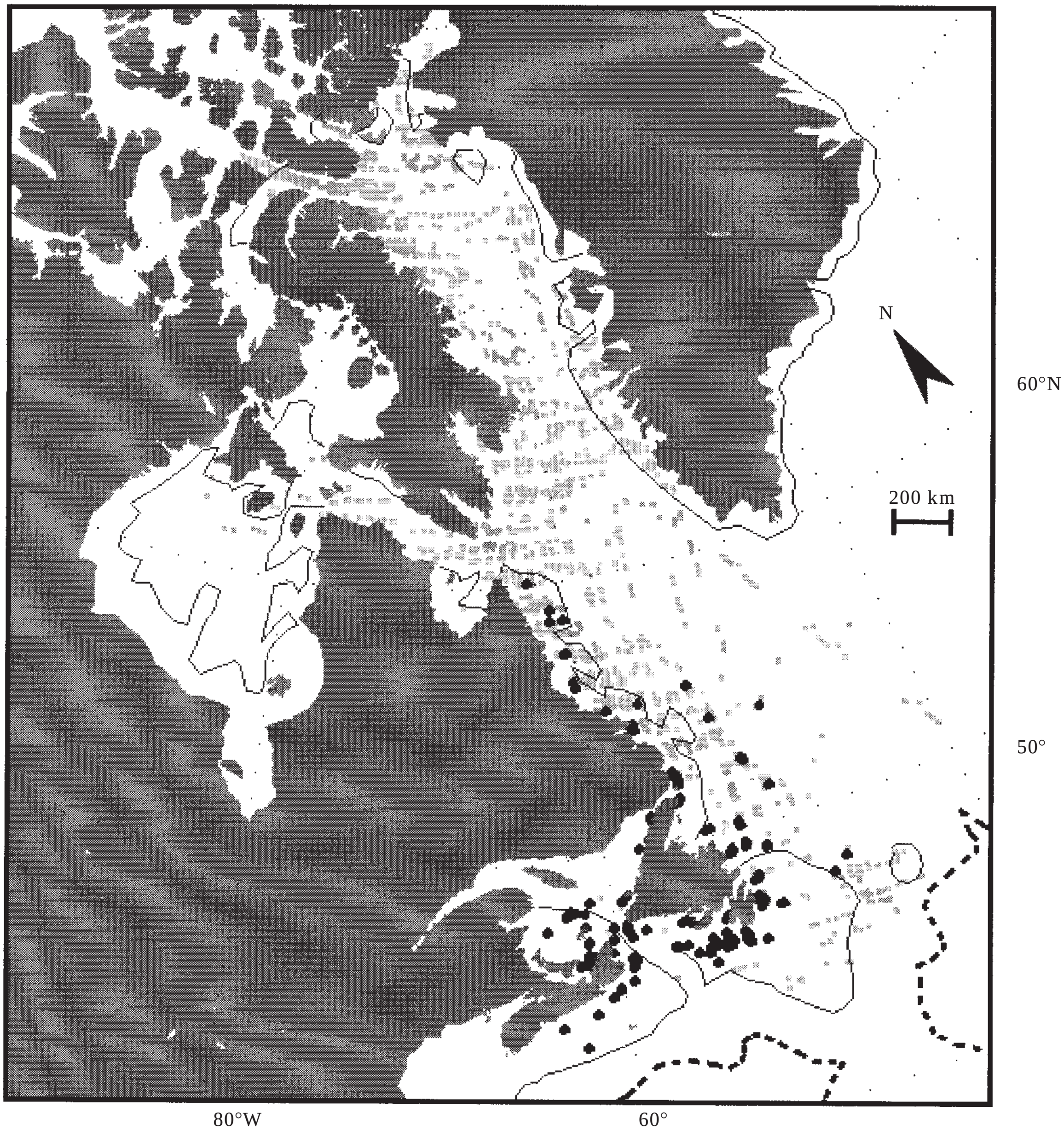

Figure 3. Positive sightings of northern gannet and northern fulmar in eastern Canadian waters, derived from PIROP data for August, in relation to sea depth from ETOPO5, 200 m and 4500 m depth contours. Dark circles indicate northern gannet; light squares indicate northern fulmar; the solid line is the 200 m contour line and the dashed line is the 4500 m contour line.

that non-breeding birds from European waters are abundant in this area. A horizontal and vertical funnel effect, driven mostly by the cold Labrador current, probably concentrates food here, or creates other attractive habitat features for fulmars. Following Cairns and Schneider (1990), this area might be considered as a ''hot spot at sea'', at least for fulmars. More data-sets need to be included to confirm these findings, although for the study area fulmars occur in highest numbers at the 1000 m depth class (Hüttmann, unpublished), as Zonfrillo (pers. comm.) has found for the North Sea.

Observations of gannets were recorded south of 51.57°N and over areas shallower than 203 m. The results suggest that gannets are still restricted to the waters around their breeding colonies in August, whereas most fulmars in this area will be non-breeders that are not constrained by the need to return to breeding colonies.

Although the original objective of PIROP was to monitor the effects of oil spills on seabirds (Brown *et al.*, 1975), the approach of overlaying the PIROP data with other digital data-sets, such as ETOPO5, in order to explore a much wider variety of questions proves to be convenient, efficient and productive. Following the structure of the PIROP database, the technique presented encourages the use of the same approach for other digital data-sets, overlayed with PIROP data. In addition, CART together with the use of a GIS presents a strong, appropriate and attractive way to improve the analysis of large seabird distribution data-sets.

Acknowledgements

We thank R. G. B. Brown, G. Forbes, and J. Linke. We thank in particular A. W. Diamond for suggestions and various helpful comments. Jim Reid and an anonymous referee also made helpful comments on the paper. This is ACWERN publication no. 3.

References

Barndorff-Nielsen, O. E., Jensen J. L., and Kendall, W. S. (Eds) 1993. Networks and chaos – statistical and probabilistic aspects. Monographs on statistics and applied probability 50. Chapman and Hall, London. 307 pp.

Bell, J. F. 1996. Application of classification trees to the habitat preference of upland birds. Journal of Applied Statistics, 23: 349–359.

Brewer, D., Diamond, A. W., and Woodsworth, E. J. in press. Atlas of bird banding in Canada. Canadian Wildlife Service Special Publication, Ottawa.

Brown, R. G. B. 1977. Atlas of eastern Canadian seabirds. Supplement 1 Halifax-Bermuda transects. Canadian Wildlife Service, Ottawa. 24 pp.

Brown, R. G. B. 1986. Revised atlas of Eastern Canadian seabirds. Canadian Wildlife Service, Ottawa. 110 pp.

Brown, R. G. B. and Nettleship, D. N. 1984. The seabirds of northeastern North America: their present status and conservation requirements. *In* Status and conservation of the world's seabirds, pp. 85–100. Ed. by J. P. Croxall, P. G. H. Evans, and R. W. Schreiber. ICBP, Cambridge. 778 pp.

Brown, R. G. B., Nettleship, D. N., Germain, P., Tull, C. E., and Davis, T. 1975. Atlas of eastern Canadian seabirds. Canadian Wildlife Service, Ottawa. 219 pp.

Cairns, D. K. and Schneider, D. C. 1990. Hot spots in cold water: feeding habitat selection by thick-billed murres. Studies in Avian Biology, 14: 52–60.

Diamond, A. W., Gaston, A. J., and Brown, R. G. B. 1986. Converting PIROP counts of seabirds at sea to absolute densities: Progress Notes No. 164. 21 pp.

Diamond, A. W., Gaston, A. J., and Brown, R. G. B. 1993. Studies of high-latitude seabirds. 3. A model of the energy demands of the seabirds of eastern Arctic Canada. Ed. by W. D. Montevecchi. Canadian Wildlife Service Occasional Paper, No. 77, Ottawa. 39 pp.

Diamond, A. W. and Prys-Jones, R. P. 1986. The biology of terns nesting at Aldabra Atoll, Indian Ocean, with particular reference to breeding seasonality. Journal of Zoology, London, 210: 527–549.

Duffy, D. C. 1989. Seabird foraging aggregations: a comparison of two southern upwellings. Colonial Waterbirds, 12: 164–175.

Gaston, A. J., Collins, B. T., and Diamond, A. W. 1987. Estimating densities of birds at sea and the proportion in flight from counts made on transects of indefinite width. Canadian Wildlife Service, Ottawa. 15 pp.

INTERAC TYDAC 1992. SPANS MAP User's guide. Interac Tydac Technologies Inc., Nepean/Ontario. 161 pp.

INTERAC TYDAC 1993. SPANS-GIS. Interac Tydac Technologies Inc., Nepean/Ontario. 552 pp.

INTERAC TYDAC 1995. International SPANS workshop. Technical Demonstrations Notes, August 30–September 1, 1995. Interac Tydac Technologies Inc., Ottawa. 281 pp.

Lock, A. R., Brown, R. G. B., and Gerriets, S. H. 1994. Gazetteer of marine birds in Atlantic Canada. Canadian Wildlife Service, Ottawa. 82 pp.

National Oceanic and Atmospheric Administration, 1996. Five minute gridded earth topography data. http://edcwww.cr.usgs.gov/glis/hyper/guide/etopo5.

Schneider, D. 1997. Habitat selection by marine birds in relation to water depth. Ibis, 139: 175–178.

Shuntov, V. 1972. Seabirds and the biological structure of the ocean. Dal'izdat, Vladivostok. National Technical Information Service Translation TT 74055032 (1974), Springfield, VA. 566 pp.

Siegel, C. 1994. Mastering FoxPro 2.6 -Special edition-. Sybex, San Francisco. 981 pp.

StatSci. 1995. SPLUS guide to statistical and mathematical analysis, version 3.3. MathSoft Inc., Seattle.

Stone, C. J., Webb, A., and Tasker, M. L. 1995. The distribution of auks and Procellariiformes in north-west European waters in relation to depth of sea. Bird Study, 42: 50–56.

Venables, W. N. and Ripley, B. D. 1994. Modern applied statistics with S-Plus. 2nd ed. Statistics and Computing. Springer, New York. 462 pp.