ICES Journal of Marine Science, 53: 415–421. 1996

Comparison of geostatistical and random sample survey analyses of Antarctic krill acoustic data

A. W. A. Murray

Murray, A. W. A. 1996. Comparison of geostatistical and random sample survey analyses of Antarctic krill acoustic data. – ICES Journal of Marine Science, 53: 415–421.

Data from the acoustic surveys of MV SA ''Agulhas'' and FRV ''Walther Herwig'', and the 1981 RRS ''John Biscoe'' South Georgia acoustic survey were analysed by geostatistical methods. Estimates of mean density (g m^{-2}) of krill and their variances are compared with published results from statistical analyses based on random sampling theory. A further high-resolution geostatistical analysis of the MV SA ''Agulhas'' (ping-by-ping) data set of the density of each individual aggregation is also presented. These analyses illustrate the problems of applying geostatistical methods to data from highly aggregated species which can show marked skewness in their histogram of density.

Key words: acoustic survey design, Antarctic krill biomass, geostatistics, spatial pattern, stock assessment.

A. W. A. Murray: British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, England [tel: +44 1223 251400, fax: +44 1223 362616].

Introduction

Antarctic krill (*Euphausia superba* Dana), a key component of the Southern Ocean ecosystem (Laws, 1985), are widespread and locally very abundant (Miller and Hampton, 1989). Krill are also the target of a substantial commercial fishery (Everson and Goss, 1991) which is managed through the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) (Miller and Hampton, 1989). These management procedures depend on accurate estimates of krill stocks together with measures of their uncertainty. Such assessments are made by means of acoustic surveys (Miller and Hampton, 1989) and are usually analysed by random sampling theory (Jolly and Hampton, 1990) using individual transects as sampling units but making no use of the spatial structure of the data as is done in geostatistical analyses (Cressie, 1993; Petitgas, 1993a). This paper presents a first attempt at geostatistical analyses of three data sets which have been previously analysed by random sampling theory methods. Krill show an extreme degree of the aggregation behaviour commonly observed in pelagic species, both in the characteristics of the aggregations themselves and the distance between them. The occasional occurrence of exceptionally dense krill aggregations (Miller and Hampton, 1989) presents formidable problems to geostatistical methods and also has implications as to the suitability of the survey designs and analysis methods currently employed in krill stock assessment.

Materials and methods

The details of the surveys analysed here may be found in Anon. (1986) for the surveys of MV SA ''Agulhas'' (AGFX) and of FRV ''Walther Herwig'' (HEFX), and in Murphy *et al.* (1991) for the 1981 RRS ''John Biscoe'' South Georgia survey (JB03). For all three surveys echo integration was used to give mean volume backscattering strength (dB) from which krill density (g m^{-2}) has been derived using the new krill target strength (TS) found by Foote *et al.* (1990).

For the AGFX and HEFX surveys, ping-by-ping acoustic data were also collected. The processing of the high resolution ping-by-ping data from AGFX is described by Klindt and Zwack (1984). Only the AGFX data set includes data on concentrations (g m^{-3}) for individual aggregations. Concentrations were multiplied by aggregation thickness to derive a krill density (g m^{-2}) and adjusted to the new krill TS.

Random sample theory analyses for the JB03 cruise have been published by Murphy *et al.* (1991). The corresponding results for the AGFX and HEFX surveys given here are for analyses of combined day- and night-time data rather than just for day-time data, as previously published by Trathan *et al.* (1993).

1054–3139/96/020415+07 $18.00/0

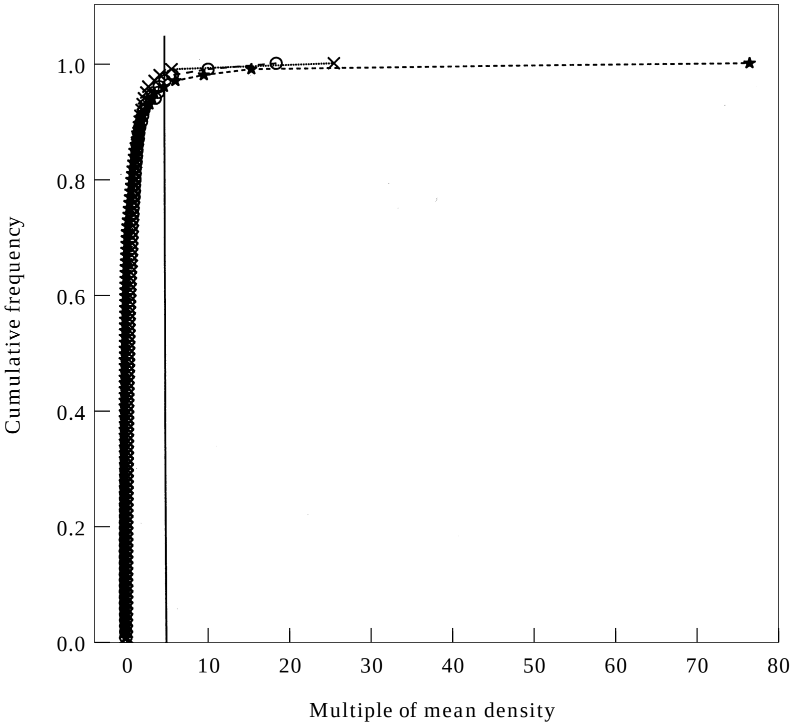

Figure 1. Quantiles of non-zero krill density (g m^{-2}) expressed as a multiple of the mean for data from the surveys of MV SA "Agulhas" (AGFX) ○, FRV "Walther Herwig" (HEFX) ★, and RRS "John Biscoe" (JB03) ×; vertical line at five times the mean.

Variograms of data were calculated using Genstat release 5.3 (Payne *et al.*, 1993) with geostatistical module. Both omni- and two-directional variograms were calculated using a lag distance equivalent to the echo-sounder distance unit (ESDU) of acoustic backscatter integration in that data set (9.26 km, 11.11 km, and 2 km, respectively for AGFX, HEFX, and JB03). For the AGFX ping-by-ping data, a lag of 0.5 km was used, which is equivalent to the 45 percentile of the distribution of the distance between aggregation centres. Spherical variogram models (e.g., see Cressie, 1993, p. 61) were fitted by maximum likelihood. Geostatistical estimation variances were calculated using the EVA software package (Petitgas and Prampart, 1993).

Results

Figure 1 shows the quantiles of the data sets plotted as a percentage of their respective means (calculated from non-zero data only). The extreme skewness of the frequency distribution of these data is very clear. Figure 2 shows the variograms (expressed as a percentage of the variance of the data) computed for the three (echo-integrated) data sets. Along-transect and across-transect variograms were also calculated, but as there was no evidence for anisotropy they are not shown here. The variograms for the AGFX and HEFX data sets show great variability and little pattern. That for the JB03 data set does conform to some extent to the expected pattern of increasing variance with increasing distance at small distances.

In light of the lack of spatial structure evident from these variograms, a more complex geostatistical model was set up as follows. $Z(\mathbf{x})$, the density (g m^{-2}) of krill at point $\mathbf{x}$, is considered as if it were a combination of two processes, $Z_1(\mathbf{x})$ and $Z_2(\mathbf{x})$ representing the process for values less than, and greater than, a cutoff level z. We also define an indicator, I, as

$$I_{Z(\mathbf{x})\geq z}=1 \text{ if } Z(\mathbf{x})\geq z$$
$$I_{Z(\mathbf{x})\geq z}=0 \text{ otherwise} \quad (1)$$

The combined process is then written thus

$$Z(\mathbf{x})=Z_1(\mathbf{x})I_{Z(\mathbf{x})<z}+Z_2(\mathbf{x})I_{Z(\mathbf{x})\geq z} \quad (2)$$

The following assumptions are also made: (i) $Z_1(\mathbf{x}),Z_2(\mathbf{x})$, and I are spatially independent; (ii) the spatial structure of the indicator, I, is pure nugget, so that the subdomains $[\mathbf{x}|Z(\mathbf{x})\geq z]$ and $[\mathbf{x}|Z(\mathbf{x})<z]$ are only known through their respective probabilities within the whole domain, allowing estimation of the mean and

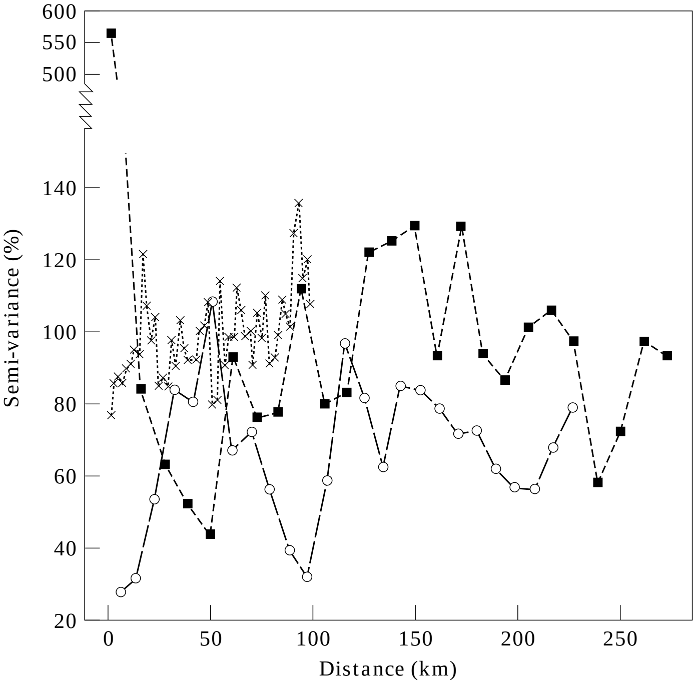

Figure 2. Variograms (semi-variances expressed as a percentage of the variance of the data) of the full data sets from the surveys of MV SA ''Agulhas'' (AGFX) ○, FRV ''Walther Herwig'' (HEFX) ■, and RRS ''John Biscoe'' (JB03) ×.

variance of $Z_i(\mathbf{x})$ over the whole domain; (iii) we do not here take into account any uncertainty in the probability of $Z(\mathbf{x}) \geq z$.

Further variograms were calculated for data sets truncated at various levels. An arbitrary cutoff level, z, of five times the mean of non-zero data was chosen and applied to all three echo-integrated data sets and the resulting variograms are shown in Figure 3. Parameters of spherical models fitted to these are given in Table 1 (see also Fig. 3). None of the models, except perhaps that for JB03, was a particularly good fit (Table 1). Estimation variances for these data were calculated using the modelled spherical variograms (see Table 2).

Variograms of the indicator I were calculated and are shown in Figure 4. These were too variable to be successfully modelled, so a pure nugget effect is assumed.

In the absence of any apparent spatial structure for the values above the cutoff, the variance of these data was calculated as if they were independent observations.

Figure 5 shows the variogram for the AGFX ping-by-ping data using the same cutoff as for the AGFX echo-integrated data. This is fairly flat and exhibits only a weak tendency to increasing semi-variance at small scales as distance increases.

Overall mean densities (g m^{-2}) of krill and associated variances (Table 2) were calculated by combination of the means and variances of the above-cutoff and below-cutoff subsets of the data using their respective proportions and squared proportions as weights. Estimation variances, and percentage coefficients of variation (CV), calculated by random sampling theory are also given in Table 2, which shows that the (geostatistically derived) estimation CV for the majority (96–99%) of the data below the cutoff is lower than that derived from random sampling theory for the full data sets. When combined with the variance estimate for the upper tail of the histogram, the overall estimation variance (CV%) is about the same as the random sampling theory variance for AGFX, just a little increased for JB03 and more than doubled for HEFX. This reflects the different degree of skewness in the various data sets.

Discussion

The skewness of the three krill data sets reported here is even greater than that found by Petitgas (1993b) for a Norwegian herring survey. The HEFX data set includes one ESDU with a density of 4.5 kg m^{-2}, some five times the next highest value. Not only are the values in the upper tail of these distributions many tens of times their mean, but they are also very rare (Fig. 1). Thus, the upper tail of the histogram is not well represented even in the large samples obtained from these extensive acoustic surveys. This sparsity of data implies a genuine uncertainty in knowledge of the significant contribution

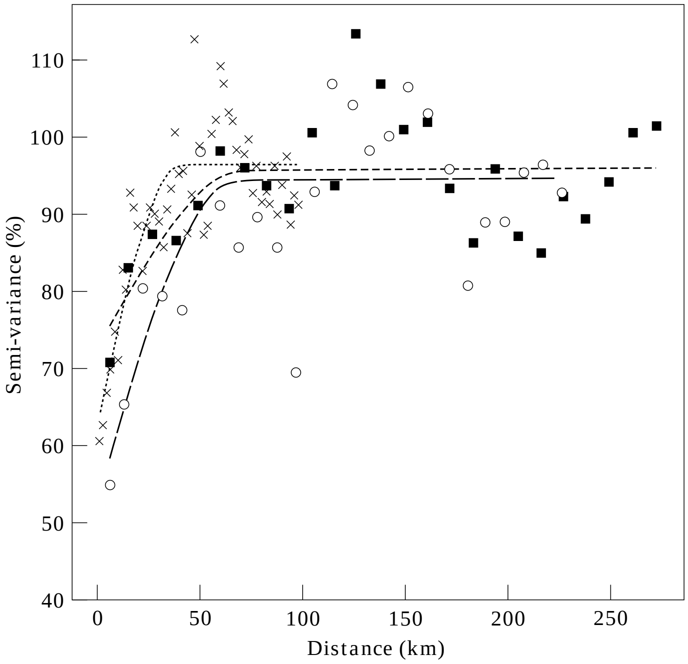

Figure 3. Variograms (semi-variances expressed as a percentage of the variance of the data) of the truncated data sets with cutoff at 5 × mean krill density (g m^{-2}). Symbols as for Figure 2. Fitted spherical models are also shown for MV SA ''Agulhas'' (AGFX) long-dashed line, FRV ''Walther Herwig'' (HEFX) short-dashed line, and RRS ''John Biscoe'' (JB03) dotted line.

Table 1. Parameters (standard errors) of variogram models for data ≤cutoff of 5 × mean (g m^{-2}) of non-zero data from the surveys of MV SA ''Agulhas'' (AGFX), FRV ''Walther Herwig'' (HEFX), and RRS ''John Biscoe'' (JB03).

	AGFX ≤ 43.03 (n = 534)	HEFX ≤ 293.5 (n = 528)	JB03 ≤ 95.31 (n = 1503)
Nugget (g m^{-2})^2	28.22	1781.0	115.1
	(8.35)	(445)	(10.0)
Sill (g m^{-2})^2	22.41	578.4	64.1
	(8.38)	(445)	(10.0)
Range (km)	67.5	73.1	42.1
	(28.9)	(42.9)	(7.75)
% variance accounted for in fit	28.3	3.6	53.1

such high values can make to the stock estimate and should be reflected in the estimation variance given for it.

The variograms of the indicator for data in the upper tail of the distribution show no evidence of spatial structure (Fig. 4), nor is there any evidence from maps of the raw data that they are concentrated in any one clearly defined zone. Thus, we must assume that these very high values are random in space. Hence, their contribution to the overall variance of the mean density is simply a proportionate combination of their variance with the geostatistical estimation variance derived using the spatial model for the majority of the data. Petitgas (1993b) analysed a Norwegian herring survey by means of a disjunctive kriging model based on a four cutoff subdivision of the data. In the case of the krill data presented here, such a multi-part breakdown is unlikely to give much improvement in the modelling of the spatial properties of the histogram due to their very much more extreme skewness and the evident lack of spatial pattern for the high values.

For the JB03 survey, a comparison of Figure 2 and Figure 3 shows an apparent increase in the range of the variogram as the data are cut off at five times the mean. For AGFX and HEFX, we see a change from variograms which would be best described by a nugget only to variograms which can be fitted by a spherical model when the data are truncated. A nugget only model has

Table 2. Estimation variances for mean densities (g m^{-2}) from the surveys of MV SA "Agulhas" (AGFX), FRV "Walther Herwig" (HEFX), and RRS "John Biscoe" (JB03).

	AGFX	HEFX	JB03
Data below cutoff of 5 × mean (g m^{-2}) of non-zero data			
Mean	4.42	26.67	8.17
Estimation variance	0.198	5.419	0.350
CV%	10.1	8.7	7.2
Proportion of data	0.976	0.961	0.993
Data above cutoff of 5 × mean (g m^{-2}) of non-zero data			
Mean	85.10	852.86	226.33
Estimation variance	1936.3	807 022.6	21 950.7
CV%	51.7	105.3	65.5
Proportion of data	0.024	0.039	0.007
Combined estimate for all data (weighted by relative proportions below and above cutoff)			
Mean	6.34	58.32	9.77
Estimation variance	1.285	1188.825	1.515
CV%	17.9	59.1	12.6
Calculated by random sampling theory using transects as sampling units			
Mean	6.05	56.65	11.70
Variance	1.209	200.289	1.228
CV%	18.2	25.0	9.5

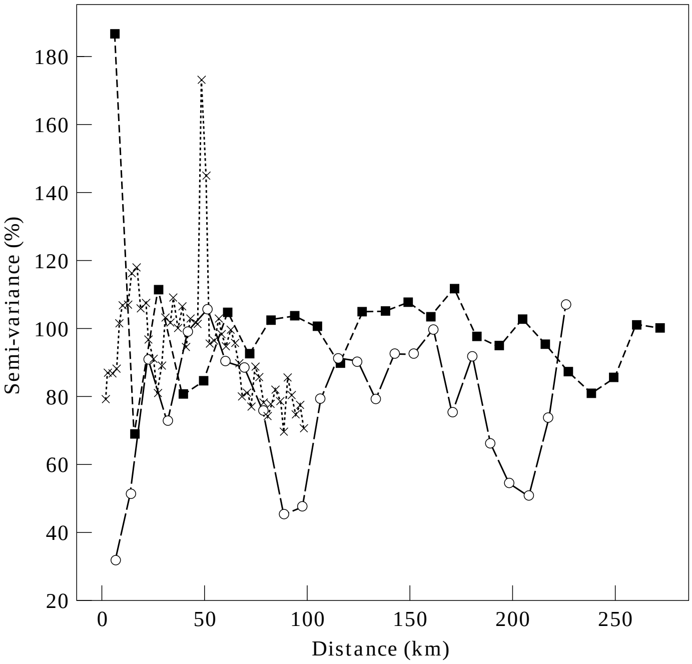

Figure 4. Variograms of the indicator (as defined in Equation (1) for data above the cutoff at 5 × mean krill density (g m^{-2}). Symbols as for Figure 2.

no range. The increasing range with the introduction of a cutoff implies that the area of influence around the higher values is very small. This provides further evidence that the highest values are unrelated to those around them.

Figure 5 shows that even at very small scales (one twentieth of the ESDU for the echo-integrated data set) the AGFX ping-by-ping data have a very substantial nugget effect equivalent to about 70% of the variance of the data and little indication that they could be modelled readily. The high nugget suggests that transitions between aggregations of widely different densities can happen at very small spatial scales (≤0.5 km). Echo-integrating over 9.26 km ESDU does tend to smooth out

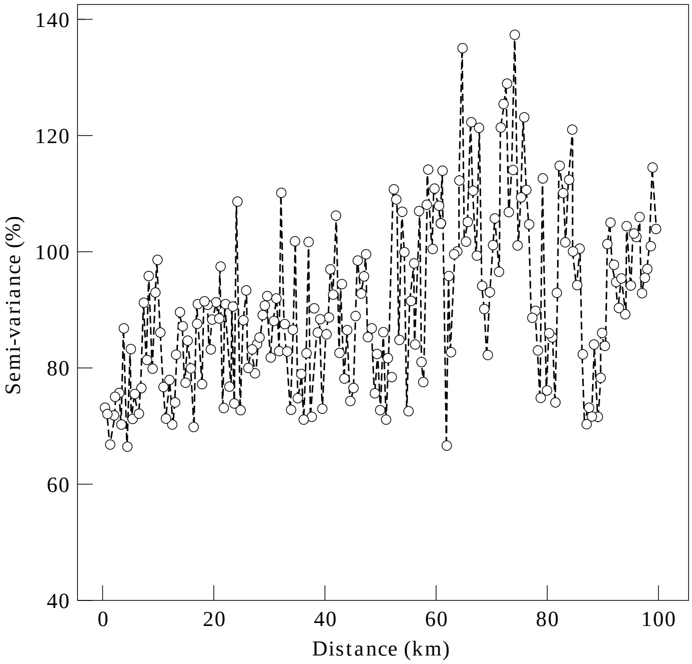

Figure 5. Variogram (semi-variances expressed as a percentage of the variance of the data) for the MV SA ''Agulhas'' (AGFX) ping-by-ping data set.

some of this variation but Figure 3 shows that the nugget is still more than 50% of the variance of the truncated AGFX data.

The HEFX survey took place at the time (austral summer 1981) and in the area (close to Elephant Island) where a so-called ''super-swarm'' was observed (Macaulay *et al.*, 1984). Such super-swarms have been reported for a number of areas, usually on the continental shelf of Antarctica or near to islands (see Miller and Hampton, 1989 for a review) although Murray *et al.* (1995) report a similar aggregation in deep water in the Bellingshausen Sea. These aggregations have been reported to be at least several kilometres in extent and to contain very high concentrations (up to hundreds of g m^{-3}) and can persist for several days. The typical inter-transect separation employed in large-scale krill surveys is usually more than 20 km. Thus, a super-swarm, potentially containing the majority of the stock in an area, could be missed by a routine survey.

The analysis presented here has shown that it is not possible to model the spatial relationships of such extreme values. Their spatial pattern appears to be entirely random and there is no information on their occurrence to be gleaned from the variogram of the data. It may well be that we must accept that estimates of krill biomass in such areas will carry a very large measure of uncertainty. Analysis methods based on transect averages as units, although statistically valid, tend to smooth out the small-to-medium scale variation which is revealed by the geostatistical analysis to dominate the pattern of variation in the data. The high geostatistical estimation variance presented here for the HEFX data set is perhaps more realistic than the estimate using whole transects as sample units because it does take some account of the uncertainty associated with the extreme high values.

It may be that it is the lack of spatial correlation of the high values rather than the extreme skewness of the histogram which is the problem in the case of krill. This could be explored by means of simulation. Future acoustic surveys for krill should investigate the spatial structure in more detail, especially in the vicinity of any super-swarms. Data should be collected using the smallest practicable ESDU (1 km or less), and transects should be closer than has been customary in the past. Stratified or adaptive survey designs may allow a more focused distribution of sampling effort.

Acknowledgements

I thank Denzil Miller for supplying the AGFX aggregation data set, and Mark Thorley and Dr Phil Trathan for the AGFX and HEFX data and analyses. I thank Dr Jonathan Watkins and Dr Pierre Petitgas for helpful discussions. I also thank the two anonymous referees for very helpful comments.

References

Anon. 1986. Report on Post-FIBEX Acoustic Workshop, Frankfurt, 3–14 September 1984. BIOMASS Report Series, 40: 1–126.

Cressie, N. A. C. 1993. Series in probability and mathematical statistics: statistics for spatial data, 2nd edn. Wiley, New York. 900 pp.

Everson, I. and Goss, C. 1991. Krill fishing activity in the southwest Atlantic. Antarctic Science, 3: 351–358.

Foote, K. G., Everson, I., Watkins, J. L., and Bone, D. G. 1990. Target strengths of Antarctic krill (*Euphausia superba*) at 38 and 120 kHz. Journal of the Acoustical Society of America, 87: 16–24.

Jolly, G. M. and Hampton, I. 1990. A stratified random transect design for acoustic surveys of fish stocks. Canadian Journal of Fisheries and Aquatic Sciences, 47: 1282–1291.

Klindt, H. and Zwack, F. 1984. A method for acoustic estimation of krill (*Euphausia superba* Dana) abundance applied to FIBEX data. Archiv für Fischereiwissenschaft, 34: 121–144.

Laws, R. M. 1985. The ecology of the Southern Ocean. American Scientist, 73: 26–40.

Macaulay, M. C., English, T. S., and Mathisen, O. A. 1984. Acoustic characterization of swarms of Antarctic krill (*Euphausia superba*) from Elephant Island and Bransfield Strait. Journal of Crustacean Biology, 4 (Special Issue no. 1): 16–44.

Miller, D. G. M. and Hampton, I. 1989. BIOMASS Scientific Series, No. 9: Biology and ecology of the Antarctic krill (*Euphausia superba* Dana): a review. SCAR, Cambridge, UK. 166 pp.

Murphy, E. J., Everson, I., and Murray, A. W. A. 1991. Analyses of acoustic line-transect data from the waters around South Georgia: estimation of krill (*Euphausia superba* Dana) biomass. *In* Selected Scientific Papers (SC-CAMLR SSP/8), pp. 225–241. CCAMLR, Hobart.

Murray, A. W. A., Watkins, J. L., and Bone, D. G. 1995. A biological acoustic survey in the marginal ice edge zone of the Bellingshausen Sea. Deep-Sea Research Part II – Topical Studies in Oceanography, 42: 1159–1175.

Payne, R. W., Lane, P. W., Digby, P. G. N., Harding, S. A., Leech, P. K., Morgan, G. W., Todd, A. W., Thompson, R., Tunicliffe Wilson, G., Welham, S. J., and White, R. P. 1993. Genstat 5 Release 3 Reference Manual, 1st ed. Oxford University Press, Oxford, UK. 796 pp.

Petitgas, P. 1993a. Geostatistics for fish stock assessments: a review and an acoustic application. ICES Journal of Marine Science, 50: 285–298.

Petitgas, P. 1993b. Use of a disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. Aquatic Living Resources, 6: 201–209.

Petitgas, P. and Prampart, A. 1993. EVA (Estimation VAriance): a geostatistical software on IBM-PC for structure characterization and variance computation. ICES CM 1993/D: 65, 32 pp.

Trathan, P. N., Agnew, D. J., Miller, D. G. M., Watkins, J. L., Everson, I., Thorley, M. R., Murphy, E. J., Murray, A. W. A., and Goss, C. 1993. Krill biomass in area 48 and area 58: recalculations of FIBEX data. *In* Selected Scientific Papers (SC-CAMLR-SSP/9), pp. 157–181. CCAMLR, Hobart.