**ICES Journal of Marine Science: Journal du Conseil**
July 2009, Volume 66 (6) : Pages 1377-1383
http://dx.doi.org/10.1093/icesjms/fsp137






# Evaluating the uncertainty of abundance estimates from acoustic surveys using geostatistical simulations

Mathieu Woillez[1, *], Jacques Rivoirard[2] and Paul G. Fernandes[3]

[1] Département EMH, IFREMER, rue de l'île d'Yeu, BP 21105, 44311 Nantes Cedex 03, France
[2] Mines-ParisTech, Centre de Géosciences/Géostatistique, 35 rue St Honoré, 77300 Fontainebleau, France
[3] Marine Laboratory, PO Box 101, 375 Victoria Road, Aberdeen AB11 9DB, UK

*: Corresponding author : Woillez M., tel: +33 240 374123; fax: +33 240 374075, email address : mathieu.woillez@ifremer.fr

**Abstract:**

Geostatistical simulations, which can reproduce the spatial variability of a variable, are particularly helpful in estimating the uncertainty associated with the combination of different sources of variability. Acoustic surveys offer an example of such complex situations, where different data (e.g. acoustic backscatter, fish length, and fish age) must be combined to estimate abundance and its associated uncertainty. In this paper, the uncertainty of Scottish herring acoustic-survey estimates is investigated using these techniques. A specific multivariate, geostatistical model is used to describe the structural relationships, which includes highly skewed distributions of the acoustic-backscatter data and incorporates relationships between depth, mean length, and proportions-at-age. Conditional simulations, i.e. geostatistical simulations that honour the data values known at the data points, are used to generate multiple realizations of acoustic backscatter, mean length, and proportions-at-age. These are combined to produce multiple realizations of herring density over the sampled domain. Multiple realizations of total abundance and abundance-at-age are then provided. The uncertainty is assessed using basic statistics to track the significant variations of these values over the period 1989–2005. Higher coefficients of variation (*CV*s) are found on average for extreme ages (ages 1, 2i, 8, and 9+); otherwise, *CV*s are mostly around 12% for abundance-at-age and around 10% for total abundance.

**Keywords:** acoustic survey, conditional simulations, geostatistics, Scottish herring

# 1. Introduction

Scientific marine-resource surveys, such as trawl surveys or acoustic surveys, are often conducted to estimate the abundance and distribution of demersal or pelagic fish populations respectively (Gunderson, 1993). The estimation process requires the combination of spatially located data. An important component of the process is determining the uncertainty of this estimation, and at least that part of the uncertainty arising from sampling in space; in other words, accounting for those areas not sampled between trawl stations or between acoustic transects.

The existence of spatial structure in many marine resources has led to the recognition of geostatistics in fisheries as providing suitable methods for the estimation of abundance from survey data (Simard *et al.*, 1992; Petitgas, 1993; Rivoirard *et al.*, 2000), because they provide a set of coherent tools which take the autocorrelation into account explicitly. The spatial structure of fish density is captured in numerical terms through the variogram or the covariogram, either with or without a trend or drift, and an appropriate structural model can then be fitted: in geostatistical terms, this is known as the structural analysis. The estimation of global abundance can then be calculated along with its estimation variance: the latter is the variance of the error of estimation, and it can be computed from the structural model when the estimator is an arithmetic or weighted average of the data, without preferential sampling. Maps can be produced by interpolation using kriging. Finally, the spatial variability can be reproduced while honouring the values known at the data points, using conditional simulations. Non-conditional simulations, on the other hand, reproduce the geostatistical model, but do not honour the data.

In many cases, the estimation variance of abundance estimates can be computed in a straightforward manner. However, in more complex cases, for example, when acoustic- survey data are combined (i.e. acoustic backscatter, fish length, fish age), geostatistical simulations can be particularly useful, because they allow for the combination of different sources of variability. The simulations generate multiple realizations of the different variables, which are then combined to produce multiple realizations of fish density within the studied domain, hence of abundance over the domain. The uncertainty is then determined from the basic statistics (variance, standard deviation, and confidence intervals) of these realizations.

It must be noted that this uncertainty addresses only the sampling error associated with the abundance estimates. This differs from other studies that have used Monte Carlo simulations to combine uncertainties, then evaluate the measurement error. Rose *et al.* (2000) have evaluated some sources of uncertainty (e.g. target strength, detectability, species identification, and sound speed) in acoustic-survey estimates of Atlantic cod (*Gadus Morhua*) and Atlantic red fish (*Sebastes* sp.) in Newfoundland waters. O'Driscoll (2004) estimated uncertainty in acoustic surveys of hoki (*Macruronus novaezelandiae*) in the Cook Strait. The latter had an additional problem, because of the fact that hoki are transient in the survey area and, therefore there is no time at which all the fish are present together. Timing between the survey and the presence of the resources was considered as a source of uncertainty, as well as the acoustic measurements. Rare are the cases where the total random error is incorporated into biomass estimates of acoustic surveys. Demer (2004) provides an example for Antarctic krill, where spatial sampling and measurement errors were evaluated and compared.

To date, very few papers have tried to evaluate sampling errors from acoustic estimates, which consider the spatial structure. First, Gimona and Fernandes (2003) tried to build geostatistical simulations to estimate confidence intervals for the abundance estimates by combining length and acoustic uncertainties from an acoustic survey on Scottish herring (*Clupea harengus*). However, the transformation algorithm they used was not able to deal with the skewed distribution of acoustic

values, which are characterized by many zero values, low values, and a few very high values. The normal-score transformation ranked the zero values randomly to associate a Gaussian value to each of these, resulting in a deformation of the spatial structure and in biased results. These pitfalls were addressed by Walline (2007) using sequential Gaussian and sequential, indicator, geostatistical-simulation methods. He proposed to eliminate the bias of the transformation algorithm by simulating the spatial distribution of zeros and non-zero data separately, then carrying out the simulation only at points where eastern Bering Sea walleye pollock (*Theragra chalcogramma*) were present.

An alternative approach is proposed in this paper, based on a Gaussian model after transformation of the data, but treating the zero values in a consistent manner. The approach is illustrated in the example of the Scottish component of the North Sea herring acoustic survey (Bailey *et al.*, 1998). This survey was the subject of one of the geostatistical case studies of Rivoirard *et al.* (2000) where either the variability of acoustic component alone was determined or only the estimate of abundance-at-age was determined from the combination of trawl and acoustic data.

## 2. Materials and methods

### 2.1. Data

Acoustic surveys have been conducted in the northern North Sea (northwestern half of ICES Division IVa) in midsummer of each year since 1979 on the prespawning concentration of autumn-spawning Atlantic herring (*Clupea harengus*). The present investigation used data collected over 17 years by the research vessel Scotia around Orkney and Shetland (1989–2005). This survey is part of the larger international survey for North Sea herring. The result of the larger survey is used to tune the assessment, which ultimately aims to determine the total biomass, total numbers and numbers-at-age of the North Sea herring stock (Simmonds *et al.*, 1996).

The Scottish survey design comprises longitudinal transect lines covering a domain defined according to ICES statistical rectangles. Transects are laid down in a systematic manner with a random start point and the transect spacing is chosen according to historical levels of abundance at 30, 15 or 7.5 nautical miles. The surveyed domains are defined by the ICES Planning Group for Herring Survey, PGHERS (ICES, 2006). Calibrated acoustic backscatter data were recorded using a Simrad EK500 echosounder operating at 38 kHz and scrutinized (see Appendix 6 of ICES 1998) to estimate nautical-area-scattering coefficients attributed to herring for 15 minute (2.5 nautical miles) Equivalent Distance Sampling Units (EDSU Figure 1a). Trawl hauls were taken regularly to assist in the scrutiny process and to collect biological data, such as fish length (Figure 1b) and fish age (Figure 1c) following Simmonds and MacLennan (2005).

### 2.2. Global geostatistical model

The simulation was based on a global geostatistical model developed for the Scottish herring using the data from the surveys 1989–1994 (Rivoirard *et al.*, 2000), and applied to an extended period (1989–2005) in this paper. This model combined acoustic backscatter (recorded along transects), mean length, and proportions-at-age (measured at trawl stations). The mean length appears to increase with the depth of the seabed, and shows a spatial structure that is rather stable over years. More precisely, the variogram of the mean length computed with pairs from different years differs from the mean yearly variogram by an additional nugget effect, which can be interpreted as additional interannual variability. In short, the spatial distribution of

mean length is approximately the same every year, except for an additive shift that varies from year to year.

The kriged map of mean length for a given year was improved, first, by using depth as external drift, and second, by using the length data of other years within an adequate three-dimensional model. The structural model, inferred on the residual structure, was made up of a spatial component – a nugget effect 1.3 and a spherical model with sill 7 and range 300 nautical miles – and a time component – a nugget effect 1.6, which was the interannual variability.

The spatial distribution of age was not independent from that of the mean length. Although it presented large variations from year to year, it was closely linked to the mean length for each year separately (Fig 4.4.4 in Rivoirard *et al.*, 2000). For a given year, the relation between each cumulative proportion-at-age and the mean length could be closely fitted by a logistic relationship

age $j+ = 1/(1 + \exp(-(\text{length} - a)/b))$, (1)

with coefficients $a$ and $b$ depending on age. This allowed for the mapping of each cumulative proportion-at-age from the map of mean length, by using kriging with this regression as external drift and a small residual structure. Maps of proportions-at-age were then obtained by the difference between maps of cumulative proportions-at-age.

In the global model, the spatial distribution of age was subordinated to that of mean length for the same year, however no dependence was observed between these and the spatial distribution of acoustic backscatter. The latter were characterized by a large number of low or zero values and a few extreme values, contributing to most of the poor structure of the raw variogram. The use of a log-backtransformed variogram to infer the structure is more robust (Rivoirard *et al.*, 2000), and allows for mapping of the acoustic backscatter by ordinary kriging.

A map of herring densities was obtained by combining the acoustic and the length maps, using the relationship between target strength and length for North Sea herring (ICES, 2006). This was disaggregated into maps of herring densities-at-age using the maps of proportions-at-age. Finally, the abundance-at-age and the total abundance estimates were obtained by summing the herring densities-at-age and total herring densities times the grid node size (set to 2.5 × 2.5 nautical miles) over the domain.

In this paper, repeated conditional simulations have been used instead of kriging to give repeated simulations of abundance estimates, conditional on the different types of data, with the associated uncertainty. Kriged estimates of the different variables (acoustic backscatter, mean length, and proportions-at-age) were also produced to verify that they were close to the average of the simulations. Therefore, no measurement errors (e.g. target strength, ageing, maturity staging, or others) will be considered here and the evaluated uncertainty will concern spatial-sampling errors only (i.e. conditional to our model and assumptions, one of which is that the “target strength to length relationship” is taken as deterministic).

### 2.3. Conditional simulation

The Gaussian, random-function model, defined by all its univariate and multivariate distributions being Gaussian, is especially well suited to conditional simulations (Lantuéjoul, 2002). In the stationary case, the univariate distribution, as represented by the histogram of values, is supposed to be Gaussian. Very often the Gaussian assumption is not acceptable on the original variable (e.g. a skewed histogram), but is acceptable after transformation, or anamorphosis. Therefore, the conditional simulation is made on the Gaussian variable, and is then backtransformed to obtain a simulation of the original variable.

In the case of mean length and cumulative proportions-at-age, the Gaussian, conditional-simulation technique was applied without the need for any transformation,

because their distribution was close to Gaussian (working implicitly with the residuals from drifts). For the acoustic variable, because of the skewness of its distribution, a transformation was necessary. However, because of the large proportion of zeroes, the values cannot be directly transformed into values of a Gaussian random function, hence the following developments.

The model for the acoustic variable $Z(x)$ is supposed to stem from a standard Gaussian random function $Y(x)$ using a Gaussian anamorphosis $\Phi$: $Z = \Phi(Y)$. The anamorphosis is a non-decreasing function that characterizes the histogram of $Z$, and therefore the cumulative distribution functions $F(z) = P(Z<z) = P(Y<y) = G(y)$. This allowed for the association of any value $Y$ to the value $Z$ having the same cumulative probability. The reverse would hold if the anamorphosis were considered invertible. This is generally the case when $Z$ takes positive values with no spike. However, this is not the case when the histogram of $Z$ values contains a large proportion of zeroes. For instance, if the values of $Z$ have 50 % zeroes, any negative value for the standard Gaussian $Y$ can be associated to a zero value of $Z$. More generally, if the proportion of zeroes is $p_0$, any value of $Y$ lower than $y_c$ can be associated with a zero value of $Z$, where the threshold $y_c$ is determined by the proportion of zeroes $p_0 = P(Z = 0) = P(Y<y_c) = G(y_c)$.

Hence, it was not possible to transform the original zero data values of $Z$ directly into Gaussian values. Because the Gaussian values of $Y$ are not determined at all data points, the above conditional-simulation method cannot be directly applied; moreover, the variogram of the Gaussian variable to be simulated cannot be computed directly. Therefore, some adaptations were needed. First, we demonstrated how to inform the $Y$ values at data points where $Z$ is zero, while conforming both to the Gaussian distribution and to the variogram of $Y$. Then we demonstrated how this variogram could be determined beforehand.

Consider the first point. Data on $Y$ are either determined where $Z$ >0, or undetermined, but below $y_c$, where $Z = 0$. A Gibbs' sampler was used to simulate the $Y$ values at data points where $Z$ is zero, conditional on their being lower than $y_c$ and on the $Y$ values at the other data points (Lantuéjoul, 2002). It is an iterative process on the set of samples to be modified, where the $Y$ value at such a point is simulated from the $Y$ values at all data points using the variogram model of $Y$.

The determination of this variogram relies on the fact that the variogram of any transformation of the Gaussian variable is a function of the variogram (or autocorrelation ρ(h)) of $Y$:

$$\gamma_{f(Y)}(h) = \mathrm{var}\left[f(Y)\right] - \int \left[f(t) - E\left[f(Y)\right]\right]\left[f(u) - E\left[f(Y)\right]\right] g_{\rho(h)}(t,u)\,dtdu \text{ , (2)}$$

where $g$ is the bivariate, standard, normal probability density function (pdf):

$$g_\rho(t,u) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{t^2 - 2\rho tu + u^2}{2(1-\rho^2)}\right). \text{ (3)}$$

This can be used to determine the variogram of $Y$ indirectly from the variogram of a transformed variable known at all data points. For this, the lower-cut Gaussian variable (Woillez, 2007)

$$Y^+ = y_c.1_{Y \le y_c} + Y.1_{Y > y_c} = y_c.1_{Z=0} + Y.1_{Z>0} \text{ (4)}$$

is considered equal to $Y$ where $Z$ is positive and to the continuous limit $y_c$ where $Z$ is zero.

## 3. Results

The different simulation steps are now presented in detail in an example year (2005), the last in the time-series. The acoustic backscatter was characterized by a highly skewed distribution with a large number of small values, approximately 57 % zero

values, and only a few large values (Table 1). The positive raw data were first transformed into normally distributed values above the threshold $y_c$ = 0.185. The Gaussian anamorphosis function was determined by plotting the raw data values against these values (Figure 2a). It was modelled by the constant value 0 over the interval [-∞; $y_c$] corresponding to the zero values, and by an increasing cubic smoothing spline for the non-zero values. The level of smoothing was adjusted until the fit was reasonably good. This model was used at the end of the simulation process to perform the backtransformation of Gaussian realizations into raw realizations.

The values of the Gaussian variable were not known for data points where the acoustic backscatter was zero, but the values of the lower-cut Gaussian variable $Y^+$ were known for all data points. The experimental variogram of the lower-cut Gaussian variable $Y^+$ is presented in Figure 2d. It was indirectly fitted by considering that the Gaussian variable had a nested structure with a nugget of 0.23, a spherical component of 0.40 with a range of 10 nautical miles, and another spherical component of 0.45 with a range of 65 nautical miles.

The simulation of acoustic backscatter was performed in two parts. First, Gibbs' sampling was used to simulate values of the Gaussian variable $Y$ at the data points where the acoustic backscatter was zero (Figure 2c and 2d). Then, using the classical conditional- simulation method, realizations of the Gaussian variable $Y$ were produced, which were finally turned into realizations of raw acoustic backscatter (Figure 3a).

Mean length and proportions above age were simulated using the classical conditional simulation method with the geostatistical model presented above. Realizations of mean length are presented in Figure 3b. Realizations of proportions above age, and by difference, those of proportions-at-age were also produced. One realization of proportion at age 4, which presents a good spread over the domain, is illustrated in Figure 3c.

A total of 250 realizations of acoustic backscatter, mean length, and proportions-at-age was produced for the year 2005. The conditional simulations respected the statistics of the data, as well as the mean of the kriged values (Table 2). Total abundance was estimated by combining the first two simulated variables, whereas the abundances-at-age were estimated by disaggregating the total abundance according to simulated proportions-at-age. The histograms of the 250 realizations of acoustic data and mean length averaged over the field, as well as the ones of 250 simulated abundances, were fairly symmetrical (not presented here), except for some abundances-at-age (age 8 and 9+). Uncertainties associated with the different abundance values are given in Table 3. The coefficients of variation (CVs) of the abundances-at-age (the standard deviation of the simulated abundances divided by their mean) ranged from 11 % to 30 %. High CVs were found for extreme ages (age 1, 2i and 9+) that are present in the borders of the domain. Otherwise, CVs were mostly approximately 14 %, on average. The CVs of the total number were lower than CVs of numbers-at-age, around 11 %.

This simulation process was performed on the entire time-series (1989–2005). Models of acoustic backscatter fitted for the other years also had nested structures with a nugget component and, most of the time, two spherical components, a short range one and a long range one. For each year of the studied period, the kriged estimates lay within the frequency distribution of the simulation estimates. The average CVs for abundance estimates were coherent with the results of year 2005 (Table 3). Indeed, the higher CVs were found on average for extreme ages (age 1, 2i, 8 and 9+). These extreme ages also presented the greatest variability in time. CVs were mostly around 12 % for abundance-at-age estimates and around 10 % for total abundance estimates.

## 4. Discussion

In this approach, the simulation of acoustic data was performed on the whole dataset, without splitting the data into zeroes and non-zeroes. A property of the transformed Gaussian model is that, on average, it passes through medium values when going from zero values to high values; there is a border effect in the sets of positive values corresponding to herring patches. This can be assessed with

$$E\left[Y^+(x+h)-Y^+(x)\,|\,Y^+(x+h)>y_c, Y^+(x)=y_c\right]=E\left[Y^+(x+h)-y_c\,|\,Z(x+h)>0, Z(x)=0\right]$$

, (5)

which is the average increment of $Y^+$ between one point with zero $Z$ value and another point with positive $Z$ value, as a function of distance $h$. An increase in that quantity over the first distances indicates a border effect, as illustrated in Figure 2e in the example year 2005.

Another significant development is the use of a Gibbs' sampler to simulate the Gaussian values consistently at data points where $Z$ is zero. This must be done carefully, because the rate of convergence towards the conditional distribution is not well known (Galli and Gao, 2001; Lantuéjoul, 2002), and the required number of iterations on the entire set may be large. In our case, this number was set to 1000, with a check of the resulting histogram and variogram.

By combining various sources of uncertainty, one can consider their relative importance. This could be achieved by fixing one variable, for instance the mean length, and letting the acoustic variable vary. In a simpler manner, we can compare the CVs of the acoustic abundance index as computed in Rivoirard *et al.* (2000) with our CVs of total abundance. Over the period 1989–1994, they equal on average 11.8 % and 9.33 % respectively. The order of magnitude is similar, meaning that the uncertainty of the abundance estimates is largely driven by the acoustic component. Note, however, that there is no reason why one CV should be larger than the other, because of the non-linear combination of acoustic and length data.

The main outcome of the whole process is the numbers in total and at age. Given their uncertainties, it is possible to follow the significant variations of these numbers over time. Thus, in Figure 4a, the total number of Scottish North Sea herring decrease significantly from 1989 to 1992, stagnates for 1993–1995, increases globally from 1995 to 2003 (except in 1999), then decreases suddenly in 2004. Moreover, age 1 has, in general, a lower abundance than age 2i and it is associated, as previously noted, with higher CVs. This age-class component of the North Sea herring acoustic survey is not precise, because the youngest fish are either close to the coast or in the southern North Sea, where this survey does not go, and they aggregate in schools close to the surface, where the ship-based echosounder samples less effectively. In the case of this particular survey, age 2i might be more representative of the recruitment strength than age 1. Thus, in Figure 4b, the studied period presents two different levels of recruitment; weak for most of the first part of the series and higher for the second part, with four strong year classes (1997, 1998, 2001 and 2003 at age 2i), but with an uncertainty less marked for 2003.

Differences were observed between the simulations, the kriged estimates, and those reported by the PGHERS (Table 3). The differences between simulations and kriging are reasonable and the kriged estimates were always included within the frequency distribution of the simulations. Furthermore, the differences were more marked for abundances-at-age, illustrating differences in the proportion-at-age variables. This might stem from the different weighting systems being used in the kriging and simulation processes.

The difference between the simulations and PGHERS estimates might be explained by slight differences in the areas used, and in the method. The PGHERS

estimates are the arithmetic means of all acoustic values weighted by the duration of the EDSU in each quarter statistical rectangle. Then, for biological data measured at trawl stations, post-stratification is performed based on a Kolmogorov–Smirnov test (Simmonds and MacLennan, 2005). As a result, the length distributions, mean lengths and age–length keys are averaged over large areas (five strata for 2005) and are used to determine the estimated numbers by quarter statistical rectangle, which gives total numbers and numbers-at-age.

## 5. Conclusion

A method of conditional simulation is proposed to evaluate the spatial sampling uncertainty of abundance estimates from acoustic surveys. This method can be extended, assuming appropriate geostatistical models, to other important fish stocks, as illustrated by studies done on anchovy in the Bay of Biscay (Woillez, 2007) and the Peruvian anchoveta (Simmonds *et al.*, 2009). It could also be extended to the entire North Sea herring stock by evaluating separately the uncertainty of abundance estimates of each survey, assuming that the geostatistical model remains valid. An interesting aspect of such simulation method is the possibility of optimizing the survey design, that is to say, the allocation of transects and trawl hauls, to obtain a more consistent data collection and estimations that are more precise.

## Acknowledgements

This study was conducted with the financial support of the European Union (project FISBOAT, 2004-2007, Fisheries Independent Survey-Based Operational Assessment Tools, Contract 502572). The analysis was conducted using R routines developed by Didier Renard, Mathieu Woillez, Nicolas Bez and Jacques Rivoirard (Mines-ParisTech, Centre de Géosciences/Géostatistique, France).

## References

Bailey, M. C., Maravelias, C. D., and Simmonds, E. J. 1998. Changes in the distribution of autumn spawning herring (*Clupea harengus L.*) derived from annual acoustic surveys during the period 1984–1996. ICES Journal of Marine Science,. 55: 545–555.

Demer, D. A. 2004. An estimate of error for the CCAMLR 2000 survey estimate of krill biomass. Deep Sea Research Part II, 51: 1237–1251.

Galli, A., and Gao, H. 2001. Rate of convergence of the Gibbs sampler in the Gaussian case. Mathematical Geology, 33(6): 653–677.

Gimona, A., and Fernandes, P. G. 2003. A conditional simulation of acoustic survey data: advantages and potential pitfalls. Aquatic Living Resources, Volume 16, Issue 3, Acoustics in Fisheries and Aquatic Ecology. Part 2, July 2003, Pages 123–129.

Gunderson, D. R. 1993. Surveys of fisheries resources, Wiley, New York.

ICES. 2006. Report of the Planning Group for Herring Surveys (PGHERS), 24–27 January 2006, Rostock, Germany. ICES Document CM 2006/LRC: 04. 239 pp.

Lantuéjoul, C. 2002. Geostatistical simulation: models and algorithms. Springer. Berlin. 256 pp.

O'Driscoll, R. L. 2004. Estimating uncertainty associated with acoustic surveys of spawning hoki (*Macruronus novaezelandiae*) in Cook Strait, New Zealand. ICES Journal of Marine Science, 61: 84–97.

Petitgas, P. 1993. Geostatistics for fish stock assessments: a review and an acoustic application. ICES Journal of Marine Science, 50: 285–298.
ICES. 1998. Report of the Study Group on Echo Trace Classification. ICES Document CM 1998/B: 1. 58 pp.
Rivoirard, J., Simmonds, J., Foote, K. G., Fernandes, P., and Bez, N. 2000. Geostatistics for estimating fish abundance. Blackwell Science, Oxford. 206 pp.
Rose, G., Gauthier, S., and Lawson, G. 2000. Acoustic surveys in full monte: simulating uncertainty. Aquatic Living Resources 13, 367–372.
Simard, Y., Legendre, P., Lavoie, G., Marcotte, D. 1992. Mapping, estimating and optimising sampling of spatially autocorrelated data: case study of the northern shrimp. CJFAS, 49, 1, 32–45.
Simmonds, E. J., Toresen, R., Corten, A., Pederson, J., Reid, D. G., Fernandes, P. G., and Hammer, C. 1996. 1995 ICES Coordinated Acoustic Survey of ICES Divisions IVa, IVb, VIa and VIIb. ICES Document CM. 1996/H: 08.
Simmonds, E. J., and MacLennan, D. N. 2005. Fisheries acoustics: theory and practice. Second edition. Blackwell Publishing, Oxford, 456 pp.
Simmonds, E. J., Gutierrez, M., Chipolini, A., Gerlotto, F., Woillez, M., and Bertrand, A. 2009. Acoustic surveys for Peruvian anchoveta: Optimisation of survey design. ICES journal of Marine Science, (SEAFACTS edition). 66, 1341-1348.
Walline, P. D. 2007. Geostatistical simulations of eastern Bering Sea walleye Pollock spatial distributions, to estimate sampling precision. ICES Journal of Marine Science, 64: 559–569.
Woillez, M. 2007. Contributions géostatistiques à la biologie halieutique. Thèse de Docteur en Géostatistique, Ecole Nationale Supérieure des Mines de Paris, France. 175 pp.

## Tables

Table 1. Descriptive statistics (minimum (min.), mean, standard deviation (s.d.), maximum (max.), percentage of zeros (% 0) and number of samples (*N*)) of the acoustic backscatter ($m^2$ nautical $mile^{-2}$), mean length (cm) and proportion-at-age 4 (%) data for the year 2005.

| Variable | min. | mean | s.d. | max. | % 0 | *N* |
|---|---|---|---|---|---|---|
| Acoustic | 0.00 | 244 | 994 | 11 863 | 57 | 940 |
| Mean length | 21.18 | 27.97 | 2.06 | 30.93 | 00 | 26 |
| Proportion-at-age 4 | 14.51 | 33.55 | 9.61 | 48.79 | 00 | 26 |

Table 2. Comparative statistics of acoustic backscatter ($m^2$ nautical $mile^{-2}$), mean length (cm) and proportion-at-age 4 (%) for year 2005. Estimates considered here are the mean and the variance of raw data weighted by their areas of influence, of kriged values, and of the simulated values; the mean and variance have been computed for each realization before averaging the whole over the 250 realizations.

| Variable | Acoustic | | Mean length | | Proportion-at-age 4 | |
|---|---|---|---|---|---|---|
| | mean | variance | mean | variance | mean | variance |
| Simulated values | 213 | 791 000 | 27.80 | 10.03 | 28.85 | 1.52 |
| Kriged values | 223 | 104 000 | 27.80 | 8.08 | 29.11 | 1.43 |
| Weighted data | 213 | 790 000 | 28.10 | 4.68 | 31.31 | 0.99 |

Table 3. Mean, standard deviation (s.d.) and the coefficient of variation (CV; %) of abundance-at-age and total abundance estimates (in millions) obtained from simulations for the year 2005. Mean CV values of abundance-at-age and total abundance estimates have been computed for the period 1989–2005. Kriging and ICES PGHERS estimates are added for comparison.

| Age | Conditional simulation | | | | Kriging estimate | PGHERS estimate |
|---|---|---|---|---|---|---|
| | mean | s.d. | CV | mean CV | | |
| 1 | 224 | 52 | 23 | 32 | 410 | 95 |
| 2i | 178 | 36 | 20 | 19 | 238 | 61 |
| 2m | 698 | 117 | 17 | 12 | 958 | 1 037 |
| 3i | 28 | 5 | 18 | 12 | 39 | 46 |
| 3m | 1 504 | 184 | 12 | 11 | 1 793 | 3 028 |
| 4 | 3 514 | 393 | 11 | 11 | 3 755 | 4 405 |
| 5 | 1 560 | 175 | 11 | 12 | 1 521 | 1 114 |
| 6 | 1 642 | 189 | 12 | 14 | 1 552 | 1 106 |
| 7 | 163 | 22 | 13 | 15 | 142 | 90 |
| 8 | 198 | 32 | 16 | 18 | 171 | 103 |
| 9+ | 248 | 75 | 30 | 22 | 188 | 96 |
| Total | 9 957 | 1 073 | 11 | 10 | 10 767 | 11 181 |

## Figures

Figure 1.

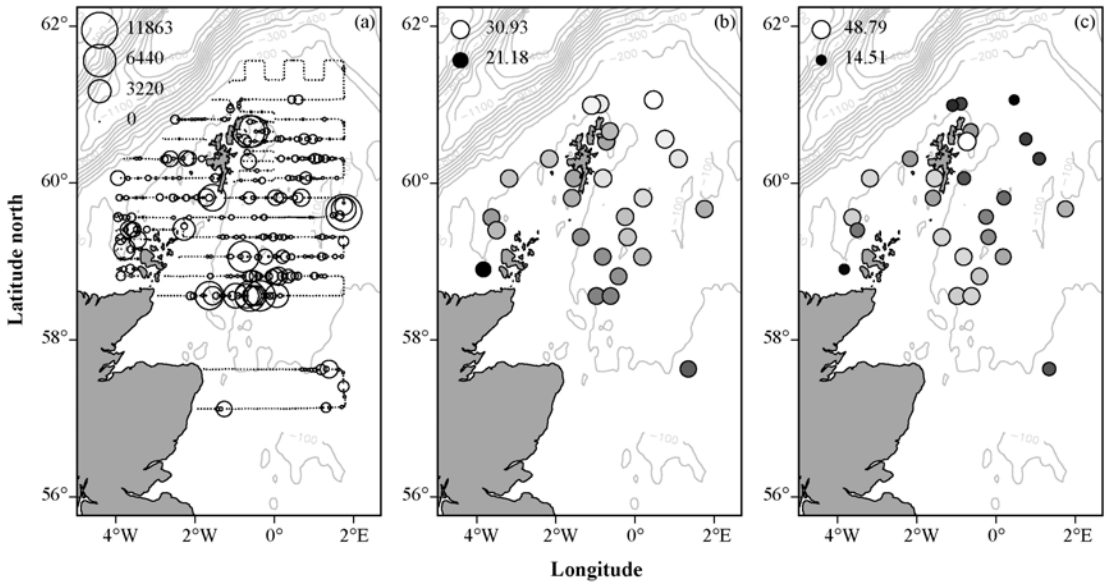


Figure 1. Proportional representation of acoustic backscatter ($m^2$ nautical $mile^{-2}$) in (a), of mean length (cm) in (b) and of proportion-at-age 4 (%) in (c) of herring from the Scottish survey around Orkney and Shetland in July 2005.

Figure 2.

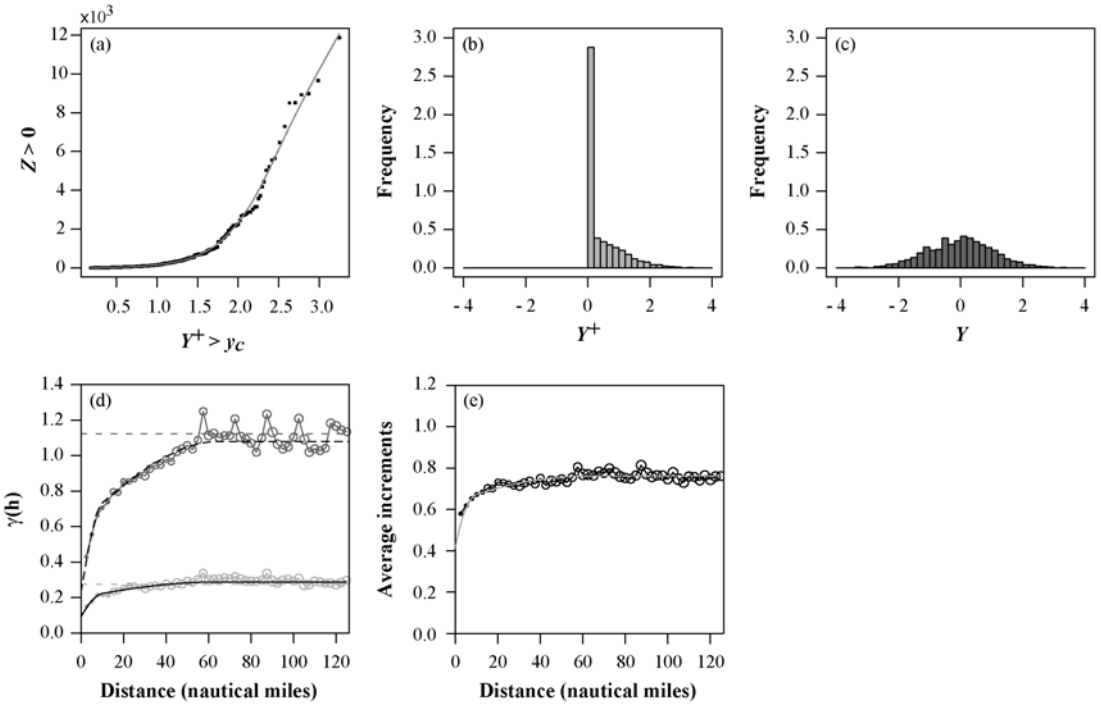


Figure 2. (a) Gaussian anamorphosis of the acoustic variable with its cubic smoothing spline model. (b) Histogram of $Y^+$. (c) Histogram of $Y$ obtained by Gibbs sampling. (d) The experimental variogram of the lower cut Gaussian $Y^+$ with circles proportional to number of data is represented in light grey. The model of the Gaussian $Y$ and the corresponding expression for $Y^+$ are represented respectively by a dashed and solid black line. The experimental variogram of the Gaussian $Y$ with circles proportional to number of values, obtained after the Gibbs sampling, is represented in dark grey and is consistent with the Gaussian model. (e) Theoretical (solid grey line) and experimental (solid black line) border effect.

Figure 3.

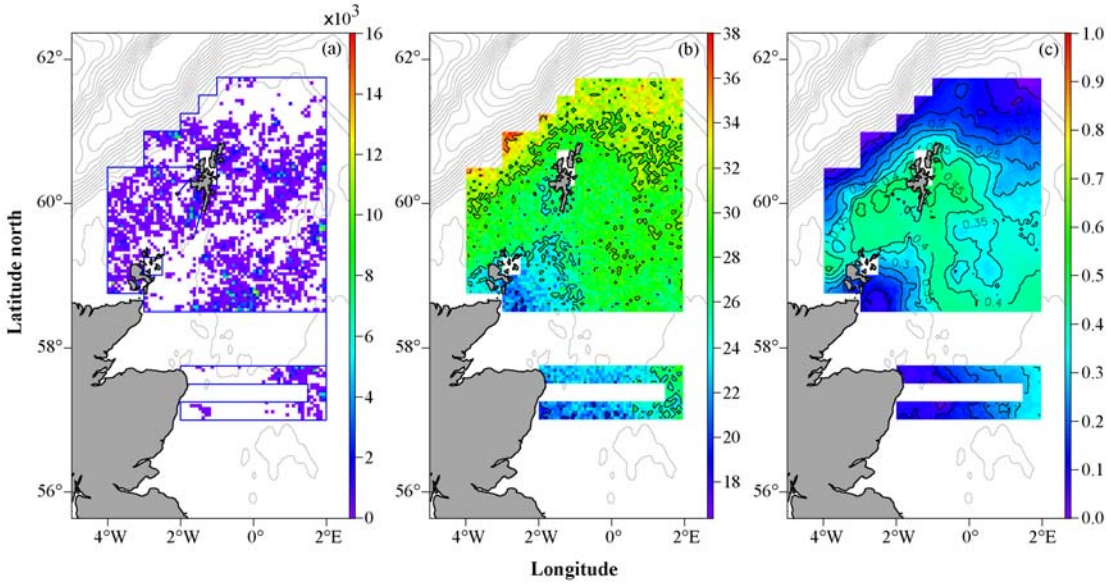


Figure 3. (a) One realization of the raw acoustic backscatter variable ($m^2$ nautical $mile^{-2}$), where only simulated values above zero have been coloured. (b) One realization of mean length variable (cm), with contour lines 20, 25, 30 and 35 cm. (c) One realization of proportion-at-age 4 (%), with contour lines every 0.05.

Figure 4.

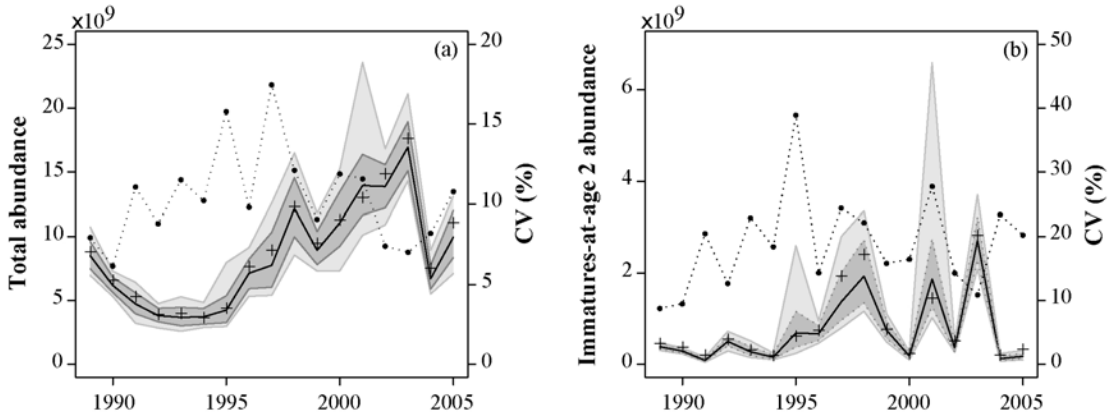


Figure 4. Distribution of abundance estimates (number of individuals) and coefficients of variation (black dots linked by a dashed line) by year obtained from 250 realizations of the total abundance (a) and the “immatures-at-age 2” abundance (b). Their distributions are summarized using some basic statistics (minimal and maximal values in light grey line, 5 % and 95 % quantiles in dark grey line and mean in black line). Kriging estimates by year are represented by a black cross.