ICES Journal of Marine Science, 58: 1208–1218. 2001
doi:10.1006/jmsc.2001.1107, available online at http://www.idealibrary.com on IDEAL®

Weighting and smoothing of stomach content data as input for MSVPA with particular reference to the Barents Sea

Tatiana Bulgakova, Dmitri Vasilyev, and Niels Daan

Bulgakova, T., Vasilyev, D., and Daan, N. 2001. Weighting and smoothing of stomach content data as input for MSVPA with particular reference to the Barents Sea. – ICES Journal of Marine Science, 58: 1208–1218.

Multispecies Virtual Population Analysis (MSVPA) is based on parameterization of the average relative food compositions for all possible prey-age predator-age combinations in the model by year and quarter. This sets high demands on stomach sampling programmes in terms of spatial coverage of the predator population and of sampling intensity for individual cohorts. In practice, there are many sources of error in the input data and large variances, which call for a smoothing procedure to avoid outliers in the MSVPA output. We investigate the potential of geostatistics (specifically kriging) in improving (1) estimates of total and partial stomach content weights from spatially non-uniformly distributed samples and (2) smoothing of the average stomach content weights over the two-dimensional input array of predator age and years. The examples shown indicate that kriging provides an efficient method to deal with geographical variability in food composition and predator abundance as well as to fill gaps and make extrapolations within a two-dimensional array in temporal space characterized by many empty cells or cells not sufficiently well sampled.

Keywords: MSVPA input, (partial) stomach content weights, missing data, kriging, cod, Barents Sea.

Received 8 February 2000; accepted 29 March 2001; electronically published 22 August 2001.

T. Bulgakova and D. Vasilyev: Russian Federal Research Institute of Fisheries and Oceanography (VNIRO), 17, V. Krasnoselskaya, Moscow, 107140, Russia; e-mail: tbulgakova@vniro.ru; N. Daan: Netherlands Institute for Fisheries Research, PO Box 68, 1970 AB IJmuiden, The Netherlands; e-mail: niels@rivo.wag-ur.nl

Introduction

Multispecies Virtual Population Analysis (MSVPA) relies heavily upon stomach sampling programmes for the estimation of the contribution of individual prey age groups to the food composition of predator age groups (Hislop, 1997). However, the excessive number of combinations of prey categories and predator age groups often prohibits the reliable estimation of each data point based on field sampling alone. In particular, the presence of empty cells as well as exceptional values caused merely by variance in the sampling may cause irregularities in the results. Thus, the ability of the model to provide a realistic description of the historic development and its reliability as a prognostic tool depend critically on statistical methods to fill the holes in the arrays and to smooth the data. Within the North Sea application, emphasis has been put on smoothing the output of MSVPA to (e.g. partial predation mortality) rather than the input (ICES, 1994; Daan, 1997). However, the irregularities in the input data for the Barents Sea are very much larger than for the North Sea, because the stomach sampling programmes were not dedicated to MSVPA development. For instance, they were not concentrated in particular years to the same extent as they have been in the North Sea (ICES, 1994) and they did not always cover the entire area. Although a very large Russian–Norwegian Stomach Content Data Base for cod (*Gadus morhua*) has been built up over the last 25 years (Mehl and Yaragina, 1992), the data for individual years are often restricted and characterized by high variances. Thus, elaboration of statistical methodologies for the reliable estimation of the food composition by prey category by year, quarter, and predator age group represents an essential component of multispecies modelling in relation to the Barents Sea fish community.

To obtain a reliable estimate of the average composition of the stomach contents of each predator age group over the full geographical range of the entire population, the following aspects have to be taken into account: (i) stomachs samples relate to predator sizes and age-length keys (ALK) are required to translate these into predator ages. As a consequence, the data are

1054–3139/01/061208+11 $35.00/0

not uniformly distributed over the age range of the predator, because there is often no information for the older and less abundant age groups. By analogy, the same applies to individual prey species. The ultimate effect is that many cells of the potential prey age–predator age combinations are empty as a direct consequence of limited sampling. (ii) The point samples collected over the area reflect different concentrations of the predator, which must be taken into account to obtain an average stomach content composition weighted by population abundance. Although the data base includes catch per unit effort (cpue) during surveys, the use of cpue as weighting factors is not straightforward because the trawl stations are not uniformly distributed over the geographical range. (iii) Interannual differences exist in the overlap of spatial distributions of the predators and the individual prey categories.

Thus, advanced statistical methods are required to determine the representative weight of each sample relative to the entire stock, to make reliable extrapolations to the under-sampled older ages and to smooth (interpolate) the entire predator-prey input array. We address these issues with particular reference to the following topics:

(1) Estimation of average total and partial stomach content weights from spatially non-uniformly distributed samples;
(2) Smoothing of total and partial stomach content weights over the entire input array.

The purpose is to elaborate specific examples with a view of investigating the potential use for application in MSVPA rather than to provide a coherent analysis over the entire array of years.

Material and methods

Primary data handling procedures

MSVPA (e.g. Sparre, 1991) is aimed at improving fish stock assessment by including intra- and interspecific predation rates and therefore requires input files with information on the average total stomach content weight by year and quarter and the corresponding average partial stomach content weights by age group of prey. The question addressed here is how these average values can best be calculated, taking into account spatial differences in feeding and abundance of predators.

The source of information is the joint Russian–Norwegian Barents Sea Fish Stomach Content Data Base. The data base consists of records describing stomach content weights observed in individual predators by prey category, which represents a combination of species, length class, and degree of digestion. The number of records per stomach reflects the number of prey categories observed. Prey age is not included in the primary data and therefore a special algorithm has been developed. The program uses age-length keys to assign length classes to age classes for seven prey species (cod; capelin *Mallotus villosus*; shrimp, *Pandalus borealis*; herring, *Clupea harengus*; haddock, *Melanogrammus aeglefinus*; polar cod, *Boreogadus saida*; and redfish, *Sebastes* spp.) and writes these ages in an additional field. The Appendix provides a description of the algorithm as well as details of other processes used to prepare input for two applications of geostatistics: (1) total or partial weights for individual stomachs by year and quarter and co-ordinates of the station where the stomach was collected; (2) average stomach contents weighted by cpue for each prey age predator age combination by year and quarter.

Geographical aspects of sampling

For evaluating spatial variations in parameters, geostatistical methods (e.g. kriging) are generally considered appropriate (Petitgas, 1993, 1996). Essentially, these methods have been developed for data which are time invariant (specifically geological data; Cressie, 1993), whereas in fisheries data this is not the case: sampling the same station at different times may result in large differences, even when the interval between sampling times is relatively short. Besides, redistribution is random within the timeframe of a survey; hence, there will be no bias in the first moment of the distribution to be estimated, only the variance will be under-estimated. However, spatial differences do play a significant role at the scale of an area such as the Barents Sea, where the individual predator and prey categories may be concentrated in relatively small areas. Therefore, it would seem reasonable to implement geostatistical methods in this case.

Kriging represents a method for the calculation of a function Z(x,y) of two geographical variables over the nodes of a regular grid (x,y) based on estimates of this function in a (limited) number of grid points. The objective is to obtain the average value of the total or partial stomach content weights (W) of the "average member" of a cohort at a particular age. Therefore, the procedure must be applied both to the spatial distribution of stomach content weights and to the cpue. For a given year (Y), quarter (Q) and age group (a) of predator the following steps were taken:

(1) A smooth spatial distribution of cpue(Y,Q,a,x,y) was obtained by kriging and the integral Icpue(Y,Q,a) of this function was calculated over the entire distribution area.
(2) The same procedure was applied to W(Y,Q,a,x,y). An unweighted and spatially averaged estimate of the total stomach content weights can be obtained by dividing the integral by the area of the base.

(3) The two arrays were also combined in a new function:

$$\Phi(Y,Q,a,x,y) = \{W(Y,Q,a,x,y)*cpue(Y,q,a,x,y\}/Icpue(Y,Q,a)$$

The integral of this function provides the spatially averaged and simultaneously cpue weighted estimate of the mean stomach content of the average predator.

Separate variograms were built (Rodionov *et al.*, 1987) in longitudinal and latitudinal directions to investigate potential direction-dependent differences. Variogram γ(h) provides a measure of mean variability of a random function Z(x) between two points (x) and (x+h) in n-dimensional space as a function of distance h between all points:

$$\gamma(h) = 0.5E[Z(x+h) - Z(x)]^2$$

To exclude "false" anisotropy caused by variable distances when measured in degrees longitude, the co-ordinates of all grid points were rescaled to kilometres, relative to the position with the lowest values of longitude and latitude. The GEO-EAS program package was used for building the variograms and their approximation was done in EXCEL; the kriging procedure was carried out in WINSURF.

In analysing the variograms (Rodionov *et al.*, 1987), the range value was estimated approximately as the distance at which the variogram as a function of distance does not increase any further. Scale was estimated by the value of the variogram corresponding to the estimated range. Search radius was taken to be equal to range, search method was taken as quadrant and the option "no drift" was applied. The possible nugget effect was also estimated.

Partial stomach content weights for individual prey age groups were treated in the same fashion as total stomach content weights. Input data include large numbers of zeros because the distribution of predator age groups and prey age groups may not overlap. All zeros were treated as "structural".

Gaps in the input arrays for MSVPA

As a rule, reliable representative information is available in the data base for cod up to 8–10 years old, whereas older age groups are often badly represented. The latter may be represented by only a few samples, often with empty stomachs or stomachs containing an exceptional amount of a single prey category. The large variation in food composition and feeding rate between individuals requires fair sample sizes to provide reliable estimates. Using limited sample sizes directly in MSVPA may distort the results. The distortions may become particularly important when investigating potential effects of reductions in fishing effort because lower mortality results in higher abundance estimates for old cod.

Thus, extrapolation techniques are required for improving estimates of both total and partial stomach content weights. The key problem is to determine the age dependence of these parameters. Unfortunately, these relationships are not defined *a priori*. Nevertheless, the following not too controversial assumptions may be made: (i) total stomach content weight increases with body weight. Because there is evidence of reduced growth in older fish, the function of stomach weight with age should slow down, possibly as some asymptotic function. The parameters of this function might be different from the growth function, because changes in the energy budget of older fish might cause faster saturation of the curve. (ii) Average stomach content weight is also a function of the feeding conditions, causing quarterly and yearly variations in the parameters. Parameter uncertainty may increase when attempting to formalize age dependence of partial stomach contents because food preference is likely to change with age and therefore food conditions may be different for different age groups in different years.

On this basis, an attempt was made to describe age dependence of total stomach content weight by a logistic curve for each quarter and year separately. Unfortunately, however, in most cases the parameters could not be estimated with sufficient precision, because the data describe mainly the initial part of the curve and provide little information about its asymptotic properties. More stable estimates were observed when processing two-dimensional arrays composed of total stomach content weights for all predator ages and all years for the particular quarter selected. In contrast to generally established procedures, kriging was in this case applied to search a two-dimensional distribution of stomach content weight as a function of two variables in time (year and predator age) rather than space. The underlying assumptions are that (1) for older predators observations of any single year-age-quarter combination contain large error, and (2) observations from adjacent years and ages contain information useful for estimating the mean diet, if the potentially large sampling errors are independent. Thus, the problem was reduced to the approximation of a two-dimensional array by a surface, which is sufficiently flexible to reflect age- and year-dependent features and robust enough to overcome unlikely tendencies.

The available data are indicated in Figure 1. Sample sizes of less then five stomachs were arbitrarily excluded. A preliminary analysis indicated that a linear model better approximated the variograms after log transformation of the stomach content weights. Therefore, it was decided to base the analyses on log transformed values and to apply exponential transformation

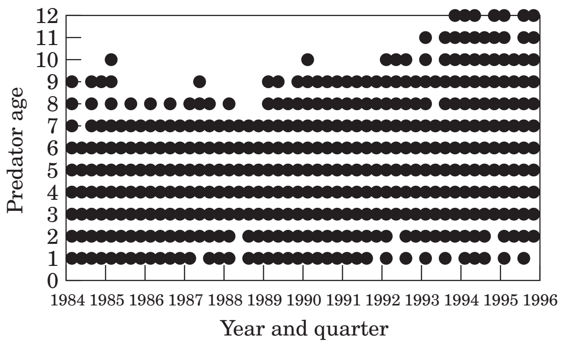

Figure 1. Availability of stomach content data (n ≥ 5 stomachs sampled) by year, quarter, and cod age group.

to the resulting points to obtain estimates of stomach content weights.

A similar but more difficult problem emerges when considering prey composition, because in this case it would seem reasonable to compose a two-dimensional array, for each quarter of each year and each prey species, of partial stomach content estimates as a function of predator age and prey age. The logic of such a procedure would lie in the expectation that the gaps in the information on consumption of a particular prey age group by the given predator age group would be partially compensated by information on consumption of this prey category by other predator age groups as well as by information on consumption of other age groups of the same prey by the same age group of predator.

Direct estimates of partial stomach content are often zero. These require careful examination, since they may represent the true situation (e.g. young cod do not eat old cod), or result from low, or non-representative, sampling intensity. In the first case, zero values should obviously be maintained, while in the second situation some inter-/extrapolation routine should be used to obtain realistic estimates.

It is not straightforward to establish a marginal number of samples that would guarantee that partial stomach contents are representative. Tentatively, predator age groups with a number of stomachs of less than five times the number of age groups distinguished for a given prey species were excluded (e.g. 30 stomachs per quarter per predator age group were used as a minimum sample size for capelin).

Results

Geographic variation in stomach contents

The described procedure has been applied to the first quarter of 1990, using predator age group 3 as the example. This year was characterized by a relatively high stock of capelin, the preferred prey species of cod

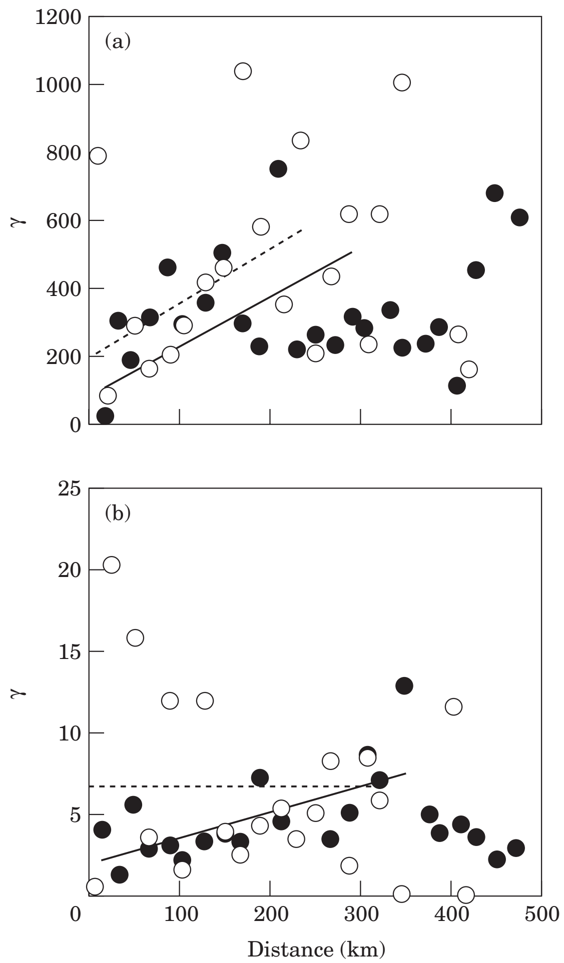

Figure 2. Variograms γ(h) and fitted linear relationship (h<range) in two geographical directions (longitude and latitude) for age 3 cod in the first quarter of 1990. (a) stomach content weight; (b) cpue (N per hour). Filled symbols, solid line ($y=ax+b$), longitude; open symbols, broken line ($y=a'x+b'$), latitude.

(Bogstad and Tjelmeland, 1992). For comparison, similar analyses were done for the same age group of cod during the first quarter of 1987, when the lowest stock size of capelin was recorded, but estimates are only given in tabular form.

The variograms derived from the cpue and the total stomach content weights for 1990 are shown in Figure 2 and estimated parameters for both years are given in Table 1. The contour map of total stomach content weights (with indication of stations; Figure 3) indicates that areas of high stomach contents were found in the centre of the area surveyed. The surface of the spatial distribution of cpue (Figure 4a) does not match the contour map of stomach contents. As a consequence, the weighted surface of cpue∗stomach weight (Figure 4b) is substantially different from Figure 3.

Table 1. Values (in km) of range, scale and nugget effect for total stomach content weight (Wtot) and number of cod (age group 3) per hour fishing (cpue) in the first quarter of 1987 and 1990, measured in two directions (longitude and latitude).

Year	Direction	Wtot Range	Wtot Scale	Wtot Nugget	cpue Range	cpue Scale	cpue Nugget
1987	Longitude	100	35	11.8	100	28.5	28.5
1987	Latitude	150	55	23.6	100	64	64
1990	Longitude	300	550	81	350	7.9	2.0
1990	Latitude	250	600	195	320	6.7	6.7

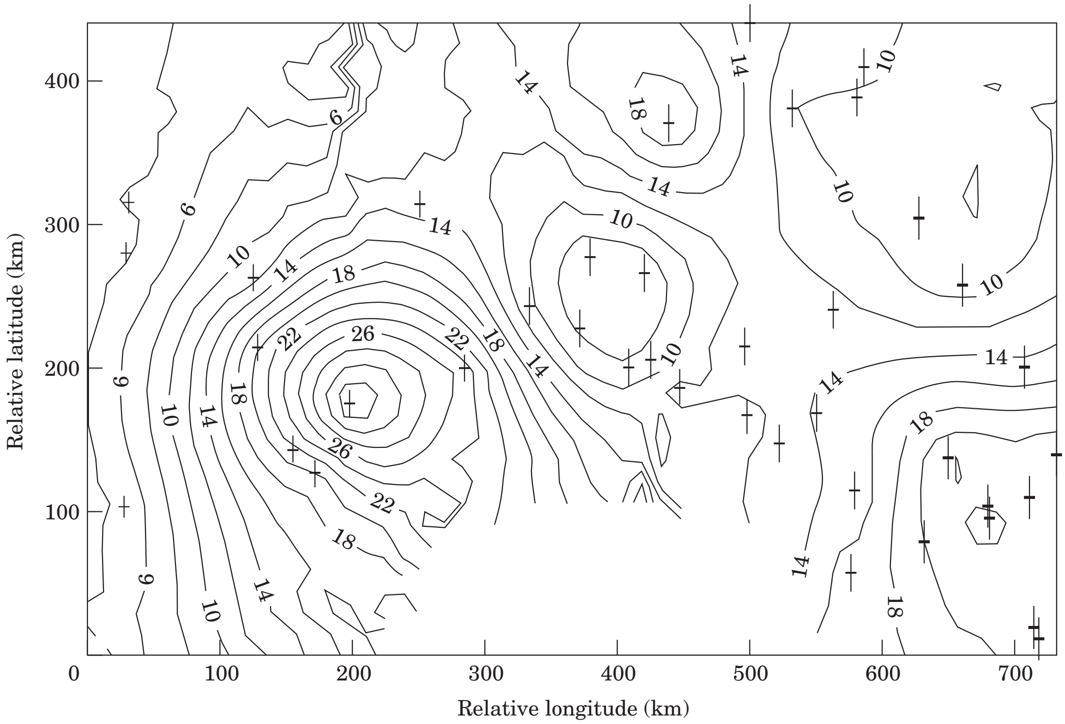

Figure 3. Contour map by relative latitude and longitude (0,0: 69°52′N 16°65′W; km scale; shore line blanked) of total stomach content weight and location of stations for age 3 cod in the first quarter of 1990.

Table 2 contrasts the arithmetic average stomach content weights, the unweighted ones smoothed by kriging, and those weighted by cpue and smoothed by kriging. Introduction of the spatial variation in the process of estimating stomach contents may result in either an increase or a decrease in the estimate. Although the effect of weighting and kriging does not change the order of magnitude of the values obtained, the absolute estimates may deviate substantially (difference between arithmetic values and after kriging ranged between −50% and +45%). Approximations of the confidence limits of the arithmetic averages are tentatively included for comparison, although their use is questionable because the underlying assumptions regarding the distribution may not be fulfilled.

To illustrate the approach followed for partial stomach contents, cod of age 3 was selected as predator and capelin of age 3 as prey, also during for first quarter of 1990. The variogram (Figure 5) indicates a range value of 130 km in both longitudinal and latitudinal direction. The linear approximations of the variograms (h<range) yield estimates of 130 for scale and 4 for nugget. The corresponding contour map (Figure 6a) shows that consumption of this prey is geographically very much restricted to the central part of the region sampled and along the coast. The surface of the partial stomach contents corrected for cpue (Figure 6b) largely follows the contour plot, but the eastern part is more pronounced owing to higher abundance of cod (cf. Figure 4a). The resulting estimates of average weight of

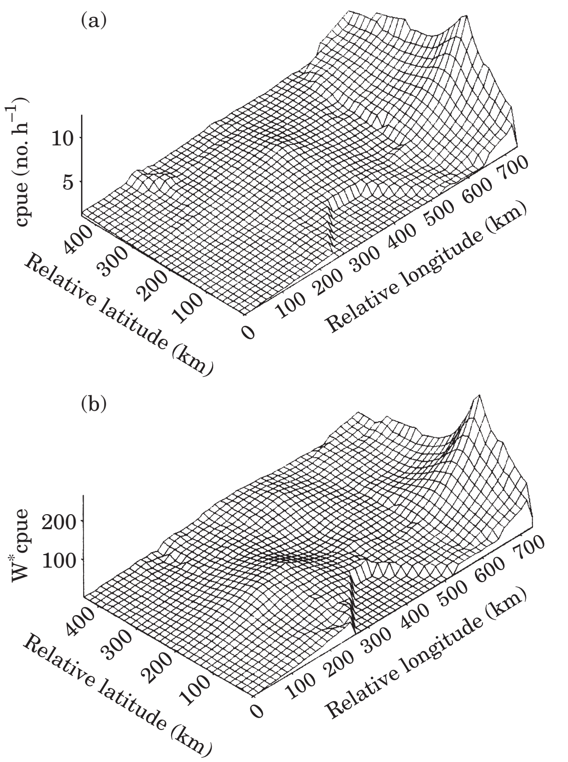

Figure 4. Plotted surface by relative latitude and longitude (0,0: 69°52′N 16°65′W; km scale; shore line blanked) for age 3 cod in first quarter of 1990: (a) cpue (N per hour); (b) product of total stomach content weight and cpue.

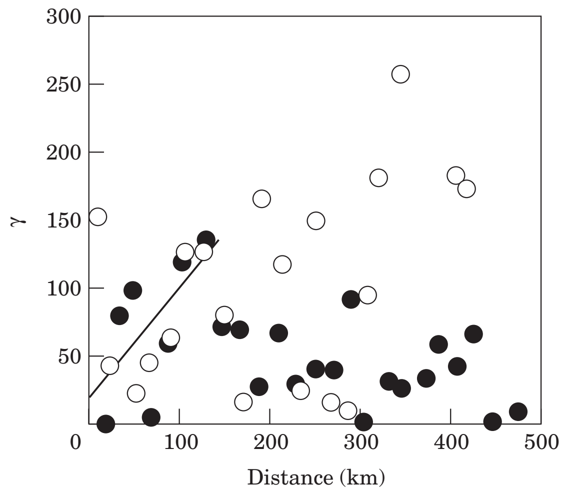

Figure 5. Variograms γ(h) and fitted linear relationship in two geographical directions [longitude (filled symbols) and latitude (open symbols)] of partial stomach content of age 3 capelin for age 3 cod in the first quarter of 1990.

capelin age 3 in the stomach contents and the estimates obtained by arithmetic averaging are given in Table 2, indicating that the differences can be large (up to 50%).

Stomach contents by age group

The variograms based on the analysis of log transformed total stomach content weights during the first quarter (Figure 7) suggest that variability due to year is an order of magnitude lower than that due to age. This is reflected in the scale, which was estimated as 4 and 21 for year and age, respectively. In both year and age direction, range was estimated as 12. This value corresponds to the total range of observations and reflects that the data do not indicate a curvature of the variogram. The estimated nugget effect was zero. Kriging was carried out with the following conditions: lag=1 (year); average scale was assumed as 12.5; search method was taken as quadrant; search radius=12 (Figure 8). An attempt to introduce linear drift resulted in unrealistic (too high) estimates of stomach weight for older ages. Because decreasing the search radius may help to find some internal trend, a second variant was calculated with the same conditions except for search radius=8 (further reduction did not allow extrapolations for all older age groups). The

Table 2. Comparison of the estimates of (a) mean total stomach content weight (g) of age 3 cod during the first quarter of 1987 and 1990, and (b) partial stomach content weight (g) of age 3 capelin in stomachs of age 3 cod during the first quarter of 1990, obtained without or with kriging and without or with weighting by cpue; approximations of the confidence interval (p=0.95) of the arithmetic averages are given in parentheses.

	1987		1990	
	Unweighted	cpue weighted	Unweighted	cpue weighted
a. Total SCW				
Arithmetic	2.31 (0.59)	1.57 (0.41)	11.95 (2.5)	9.45 (2.1)
Kriging	1.21	1.12	13.34	13.71
b. Partial SCW				
Arithmetic	—	—	2.77 (1.18)	2.47 (1.18)
Kriging	—	—	1.82	1.83

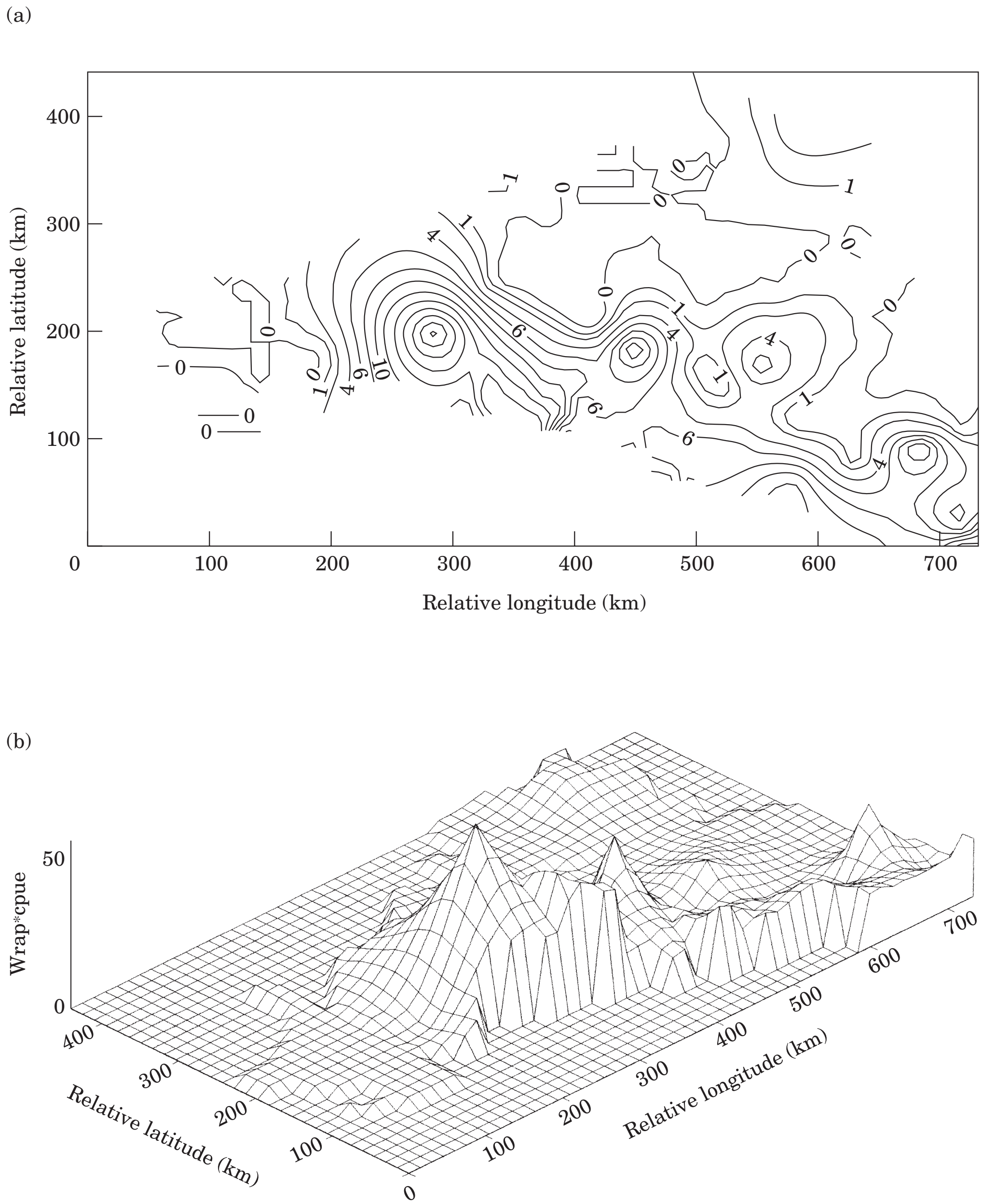

Figure 6. Spatial distribution by relative latitude and longitude of age 3 capelin for age 3 cod in the first quarter of 1990 (cf. Figure 4): (a) contour map of partial stomach content; (b) surface of product of partial stomach content weight and cpue.

resulting surface showed more pronounced annual bumps and troughs for age groups 12 and older, but the main trends were maintained (not presented).

The estimated surface fits the initial points exactly due the zero nugget effect. Thus, only the estimates for missing cells (cf. Figure 1) are important. Figure 9 shows the age dependency of total stomach content weight for selected years according to the two variants. The variation in the extrapolated annual values is considerable, but not unrealistic compared to the variation observed in the well-sampled age groups. Based on the expected increase in stomach content weight, we consider that the larger search radius provides more realistic results.

Variograms of the log transformed partial stomach content of capelin as prey in cod stomachs, weighted by cpue, during the first quarter of 1990 were calculated for predator age and prey age separately (Figure 10). Apparently, the variograms might be approximated by a

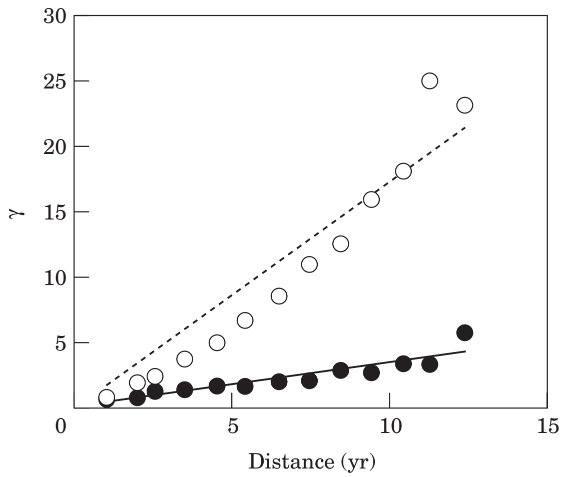

Figure 7. Variograms and fitted linear relationship in two dimensions [year (filled symbols) and predator age (open symbols)] of total stomach content weight of cod.

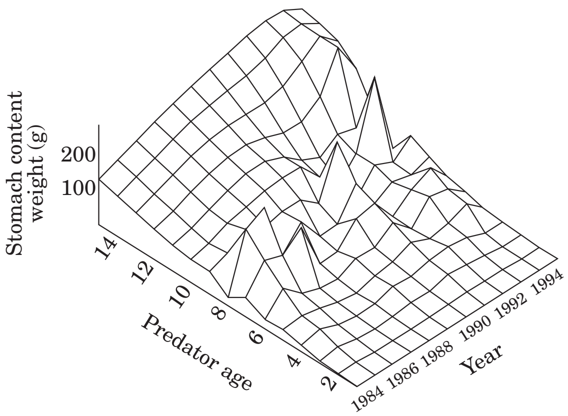

Figure 8. Plotted surface of total stomach content by year and predator age (search radius = 12).

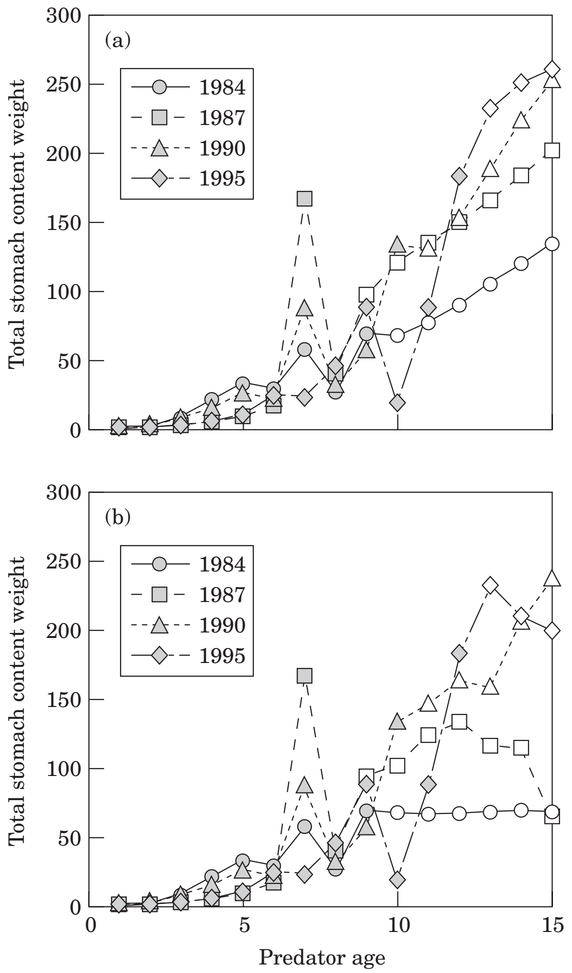

Figure 9. Observed (filled symbols) and extrapolated (open symbols) total stomach content weights of cod for four selected years: (a) search radius = 12; (b) search radius = 8.

linear model with non-zero nugget effect. However, we preferred a zero nugget effect to keep the initial values rather than introducing smoothing effects. The option "all points" had to be used because of the very low number of input points (<200). Because the maximum predator age sampled in 1990 was 8, the search radius could not be decreased. After kriging, the results were back transformed by exponentiation (Figure 11).

There were several zero estimates of capelin weight in stomach contents of older predator age groups, for which it is difficult to decide whether these are true zeros or not. We treated them as non-representative (absent information), because they drew down the extrapolated estimates. However, after kriging the original zero values were recovered in order not to overestimate partial stomach content. Figure 12 illustrates the total amount of capelin in cod stomachs by age group of cod according to the kriging approximation. The exceptionally high value for age 7 is supported by a sufficiently large number of stomachs sampled, but the other estimates follow a rather smooth curve and the extrapolated values appear to be reasonable.

Discussion

The examples shown indicate that kriging provides an efficient method for filling gaps and making extrapolations within a multidimensional data base characterized by many empty cells (or cells which are not sufficiently well sampled).

With reference to the geographical variation in the north-south and east-west directions, the variogram could not always be characterized sufficiently well by the data over the full range of distances. In particular,

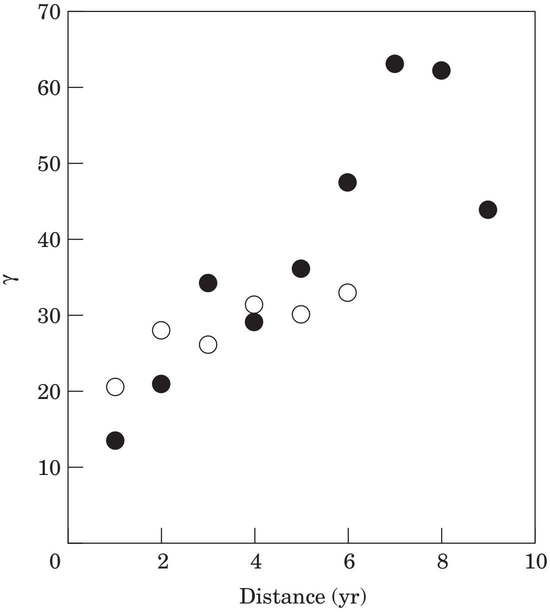

Figure 10. Variograms in two dimensions [predator age (closed symbols) and prey age (open symbols)] of capelin weight in cod stomachs.

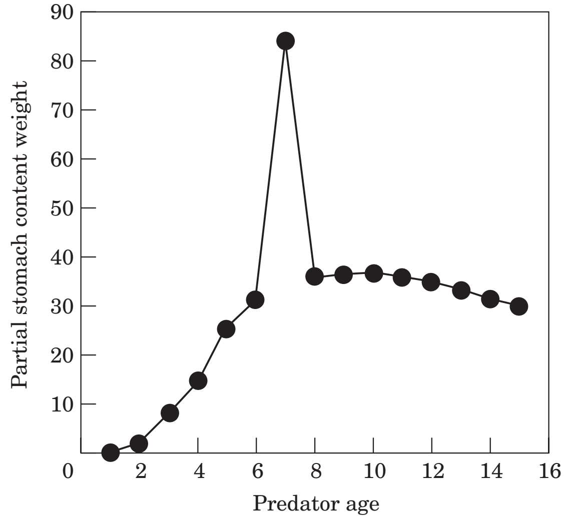

Figure 12. Fitted average partial stomach content weight of capelin by cod age.

the data rarely include points characterized by distances less than 20–30 km due to the large distances between stations, which makes it difficult to analyse the models properly. This is unfortunate because peculiarities might be revealed at distances below 20 km. For instance, Conan *et al*. (1989) estimated range for northern shrimp at only 13 km. On the other hand, given that the ultimate goal of stock assessment is to provide management advice, small scale patterns may not be that important.

The geostatistical methods applied to the distribution of stomach samples and concentration of the predator for estimation of average (partial) stomach contents yield results which lie within or close to the confidence

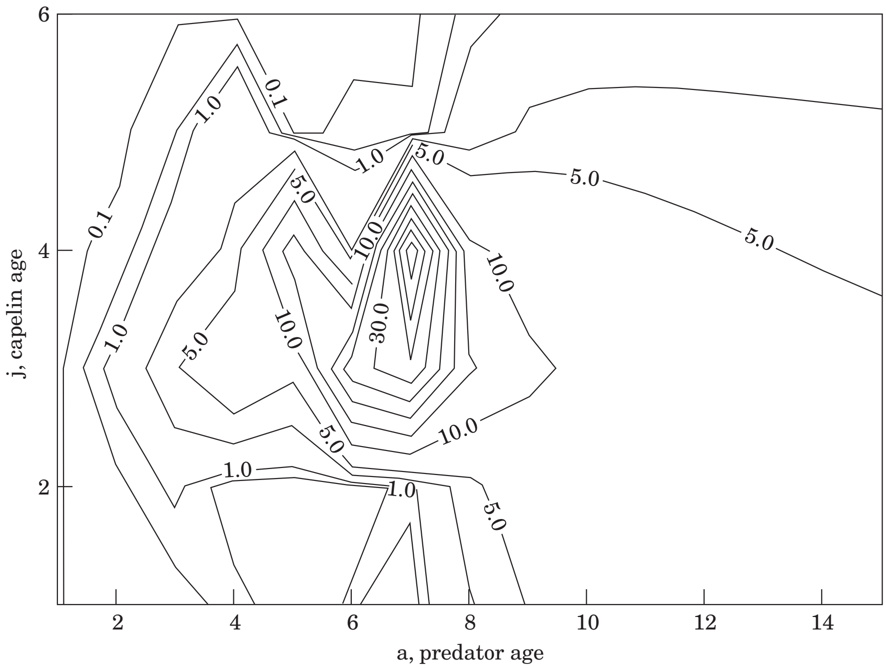

Figure 11. Contour map of capelin weight by prey age and predator age in cod stomachs.

interval for estimates obtained by means of simple weighted or unweighted arithmetic averaging. The close correspondence observed in the example of age 3 capelin may be due to the high abundance of that prey in that particular year. However, this may be an exception rather than the rule. Nevertheless, because aspects of the spatial structure of predator feeding are taken into account, kriging may provide a better estimator of the mathematical expectation of total or partial stomach content weights for the entire population, as long as sampling error is large relative to the spatial dynamics of the predator-prey interactions.

For individual prey age groups, the available data may still be insufficient to carry out this type of analysis. More extensive information on the boundaries of the appropriate ranges and overlap of predator and prey by age, quarter, and year might need to be collected and incorporated in the methods.

We have concentrated on methodical aspects, exemplified by subsets of the entire data base. While the implementation of spatial analysis for estimation of total stomach contents would be straightforward and the results could be used directly for MSVPA, the procedure for all possible combinations of partial stomach content weights by predator and prey age is exceedingly time consuming. This would require development of special software.

Geostatistical methods appear to serve as appropriate tools for the analysis of functions not only of spatial variables but also of variables of other dimensions. Application to (partial) stomach content weights as a function of predator age (or prey age), year, and quarter allows extrapolation of values to older predator ages and interpolation for age groups that are under-represented in the data base. The results indicate that it would be unsatisfactory to extrapolate the stomach content weights in the last age group sampled during a particular year and quarter for all older age groups, because of the strong age dependency. Kriging seems to provide very reasonable estimates for the many missing input data points for MSVPA, while values sampled empirically may be maintained.

The procedure of analysis of partial stomach content by quarter for a given prey species might be made more consistent and faster by applying a three-dimensional analysis (predator age, prey age, and year). However, this has not been pursued, because there is no assurance that this would yield realistic results because of large fluctuations in prey abundance between years.

The methodology presented may be also useful for preprocessing other arrays of data. For example, in both single species and multispecies VPA, weights-at-age in the sea are often used to obtain the appropriate biomasses. These arrays suffer from similar deficiencies as the stomach content data, with missing information for older age groups as well as for prerecruit year classes. The methods can in principle also be applied to improve input estimates in these cases.

Acknowledgements

We thank Dr Henrik Gislason, Dr Jake Rice and an anonymous referee for constructive comments and advice. We are also grateful to PINRO (Murmansk, Russia) for providing the cod stomach data base and additional biological information. The work was funded by INTAS/RFBR project 95-1204.

References

Bogstad, B., and Tjelmeland, S. 1992. A method for estimation of predation mortality on capelin using a cod-capelin model for the Barents Sea. *In* Interrelations Between Fish Populations in the Barents Sea. Proceedings of the fifth PINRO-IMR Symposium, Murmansk 12–16 August 1991, pp. 111–130. IMR, Bergen, Norway.

Bulgakova, T. I., Vasilyev, D. A., Dolgov, A. V., and Frolova, Z. N. 1998. Input information preparing for multispecies model of commercial community of the Barents Sea fishes. *In* Theory and Practice of Commercial Fisheries. Behaviour of Hydrobionts in the Gears Action Zone. Collected Proceedings, pp. 88–96. VNIRO Publishing, Moscow (in Russian).

Conan, G. Y., Parsons, D. G., and Wade, E. 1989. Geostatistical analysis, mapping and global estimation of harvesting resources in a fishery of northern shrimp (*Pandalus borealis*). ICES CM 1989/D: 1.

Cressie, N. A. C. 1993. Statistics for Spatial Data. Wiley & sons, New York. 900 pp.

Daan, N. 1997. Multispecies assessment issues for the North Sea. *In* Global Trends: Fisheries Management, pp. 126–133. Ed. by E. L. Pikitch, D. D. Huppert, and M. P. Sissenwine. American Fisheries Society Symposium, 20, Bethesda, Maryland, USA.

Hislop, J. R. G. (ed.) 1997. Database report of the Stomach Sampling Project 1991. ICES Cooperative Research Report, 219: 422 pp.

ICES. 1994. Report of the Multispecies Assessment Working Group, Copenhagen, 23 November–2 December 1993. ICES CM 1994/Assess: 9, 177 pp.

Mehl, S., and Yaragina, N. A. 1992. Methods and results in the joint PINRO-IMR stomach sampling program. *In* Interrelations Between Fish Populations in the Barents Sea, pp. 5–16. Proceedings of the Fifth PINRO-IMR Symposium, Murmansk 12–16 August 1991. IMR, Bergen, Norway.

Petitgas, P. 1993. Geostatistics for fish stock assessment: a review and an acoustic application. ICES Journal of Marine Science, 50: 185–298.

Petitgas, P. 1996. Geostatistics and their application to fisheries survey data. *In* Computers in Fisheries Research, pp. 113–142. Ed. by B. Megrey, and E. Moksness. Chapman and Hall, London.

Rodionov, D. A., Kogan, R. I., and Golubev, V. A. 1987. Reference Book on Mathematical Methods in Geology. "Nedra", Moscow., 335 pp. (in Russian).

Sparre, P. 1991. Introduction to multispecies virtual analysis. ICES Marine Science Symposia, 193: 12–21.

Appendix

Because prey age is not given in the joint Russian–Norwegian Barents Sea Fish Stomach Content Data Base, an algorithm has been developed to assign the appropriate age to all prey belonging to the seven species presently incorporated in MSVPA. In many cases prey length has not been recorded and therefore, the first step is to derive the appropriate length class. Using the entire data base, ''reconstruction'' coefficients for digestion stages 2–4 have been determined from records with length information by dividing weights of prey in stage 1 (fresh) by the corresponding weights in the respective digestion stages for the same length class (Bulgakova *et al*., 1998). Statistically reliable estimates were obtained for capelin and these have been applied to all fish species. For shrimp, no reliable coefficients could be obtained. The program calculated the weight at ingestion for all prey (except shrimp) and the associated length class based on species-specific length–weight relationships. Obviously, this is only possible if the number of prey and degree of digestion (five stages) are known and if the food was not completely digested (stage 5). If these conditions were not met, length class was considered unknown.

Food composition by age group was determined by means of year- and quarter-specific age-length keys. Appropriate procedures were developed to account for discrepancies in length class definitions used in stomach sampling and age determination. More importantly, stomach samples contain discrete numbers of organisms and therefore the appropriate fractions of the different age groups within a length class according to the key cannot be maintained in individual samples. Therefore, the algorithm dynamically calculates the sums of squares of the residuals of the relationship between the age fractions in the key and the cumulative fractions of previously assigned ages in the samples, and selects the variant giving the smallest current sum of squared residuals.

The contents of prey of ''unknown'' age were redistributed over the respective age groups for that particular species according to the average fractions for the ''known'' ages for the given predator age group, quarter, and year. Subsequently, the algorithm redistributes prey categories of various degrees of indeterminacy (''undetermined food'', ''undetermined fish'', ''undetermined gadoids'') to higher classification levels (all encountered prey, fish species, gadoid species, respectively) according to the calculated average proportions of the latter (by age group where appropriate).

In a few instances, only records may be available, for a given predator age and prey combination, with weights in the ''unknown'' prey age category. There is no formal solution to this problem. When the information is not used, the contribution of that particular prey to the total food is underestimated. Alternatively, personal judgement may be used to modify MSVPA input data manually to take these unknown-age quantities into account.

If the cpue value for a sample was not known, an arbitrary value of 1 has been assumed to reflect that the existence of a sample indicates that there were at least some predators around. The haul position in terms of latitude and longitude was changed to a decimal system.

The program was written in PAL and executed as a script within a PARADOX environment. The output served as input for geostatistical analysis.