Hydrobiologia (2008) 612:155–169
DOI 10.1007/s10750-008-9484-z

FISH HABITAT MAPPING

Mapping abundance distribution of small pelagic species applying hydroacoustics and Co-Kriging techniques

Stratis Georgakarakos · Dimitra Kitsiou

Abstract Hydroacoustic technology provides ground tools for the estimation of abundance and spatial distribution of pelagic species. The final products of such surveys, the interpolated choropleth maps, are based on a Geostatistical analysis of the acoustic measurements to minimise, as much as possible, the interpolation error, and to quantify uncertainty. The current study is based on fisheries acoustic measurements and satellite images covering the sea area of Thermaikos Gulf over the years 1996, 1997 and 1998. Spatial interpolations describing the abundance and distribution of small pelagic species in the research area, as well as sea surface temperature (SST), Chlorophyll-a content (SSC) and average depth, were produced, based on Ordinary and Universal Kriging and Co-Kriging Geostatistical methods. The results of the Geostatistical analysis showed that the Co-Kriging spatial interpolation method produced the best results regarding fish abundance when SST and average depth variables were included in the model. The latter indicates that there is an existing spatial cross-correlation between fish abundance and the environmental variables. Consequently, the potential reduction of the overall error in the estimation process, as presented in this study, is very significant, particularly with regard to error reduction in stock assessment and management.

Guest editor: V. D. Valavanis
Essential Fish Habitat Mapping in the Mediterranean

S. Georgakarakos (✉)
Sonar Laboratory, Department of Marine Sciences, University of the Aegean, University Hill, 81100 Mytilini, Lesvos, Greece
e-mail: stratisg@aegean.gr

D. Kitsiou
GIS Laboratory, Department of Marine Sciences, University of the Aegean, University Hill, 81100 Mytilini, Lesvos, Greece

Keywords Fisheries acoustics · GIS · Co-Kriging · Small pelagic species

Introduction

Traditionally, ecologists correlate the spatial heterogeneity of fish abundance to certain physicochemical and biological parameters, which define the ecological background of the specific organisms (Hensen, 1911; Lasker, 1978; Laevastu & Hayes, 1981). These factors can influence fish activities, such as feeding, predator avoidance, migration, reproduction and habitat selection, and are therefore accountable for the spatial characteristics of their biomass distribution (Koutsikopoulos & Lacroix, 1992; Simard et al., 1992; Horne et al., 1999). However, this more or less deterministic interpretation of the effect of environmental variables is not always and entirely evident (Fréon et al., 2005), and it is supposed that the generated forces interact in such a non-linear or chaotic way that the resulting spatial fish biomass structure, on most observation scales, is stochastic (Sharp & McLain, 1993; Webster & Oliver, 2001).

How difficult is the identification or the quantitative evaluation of such interactions? There are different approaches for analysis and modelling of the relationships between environmental factors, including bottom quality and geographic peculiarity on the one side and spatial heterogeneity in biomass on the other. Most approaches emphasise the spatial autocorrelation (Maravelias et al., 1996; Paramo & Roa, 2003; Giannoulaki et al., 2005) or the multivariate character (Sullivan, 1991; Stefánsson, 1996; Swartzman et al., 1994; Maravelias, 1999) of the data at different scale levels. Since pelagic fish biomass forms, in small scale, "schooling" and "clustering" aggregation patterns (Azzali et al., 1985; Fréon & Misund, 1999), the spatial structure has also been analysed in certain studies as a spatial point process (MacLennan & MacKenzie 1988; Petitgas et al., 2001).

Classical estimation methods assume stationarity in space and time, independence among the data and identical distribution of the parameters (Rossi et al., 1992). All these conditions are very rarely met in the aquatic environment, where the spatial structure is unstable through time, data are spatially autocorrelated and the number of samples is low compared with the high variability of the structure. Unfortunately the spatial information, which is critical for evaluating trends in biomass or correlations between the latter and the relevant environmental parameters, is neglected by the most commonly used statistical methods.

Specifically for pelagic fish, hydroacoustic measurements of fish density combined with Geostatistical analysis has been recognised as the best method for modelling the spatial distribution of species biomass or the joint spatial dependence between biomass and environmental parameters (Armstrong et al., 1989; Petitgas, 1993; Simard et al., 2002). Recently, Geostatistical analysis has been applied, showing a significant effect of certain topographic characteristics on the spatial organisation of the overall small-pelagic-fish assemblage (Giannoulaki et al., 2006).

In the present study, pelagic fish density, recorded by acoustics in four different stock assessment surveys, was geostatistically analysed in order to produce choropleth maps through Kriging. The spatial characteristics of both fish density and environmental parameters acquired by in situ or satellite techniques were investigated by means of Variograms and Cross-variograms, examining possible spatial relationships among the variables. These multivariate observations were interpolated in a manner similar to Ordinary and Universal Kriging techniques, by applying Ordinary and Universal Co-Kriging. The latter is a special case of the first, introduced by Matheron (1971), where the trend is modelled as a function of coordinates (Deutsch & Journel, 1998). Finally, the developed spatial models were tested by applying the "leave-one-out cross validation" procedure (Isaaks & Srivastava, 1989), evaluating if the incorporation of the environmental variables will improve the prediction at new spatial locations, where no acoustic data are available. Furthermore, the leave-one-out cross validation residuals were used to provide statistics concerning modelling assumptions and whether standard errors estimated by the models are accurate (Isaaks & Srivastava, 1989).

Methodology

Data collection

Acoustic survey data were collected on four surveys (10–12 October 1996, 13–15 October 1996, 21–25 May 1997 and 27 April–1 May 1998) during standard fish biomass assessment programmes in the Thermaikos Gulf in the NW part of the Aegean Sea. Measurements were carried out along predetermined sampling transects, to the south until the 100 m isobath, near Cape Kassandra, in an area of about 1,600 sq.n.mi. (Tsimenides et al., 1992).

The area backscattering coefficient Sa [m^2 $n.mi^{-2}$] was estimated every n.mi using a SIMRAD BI500/EK500 echosounder (Knudsen, 1990). The system operated a 38 kHz transducer, calibrated with standard spheres (Foote, 1987), insonifying with 1 ms width pulses. The research vessel was sailing at 8 knots, and time, position and integrator values were recorded in the BI500/UNIX file system. Echograms were scrutinised to remove echo traces in the water column that were not pelagic fish. Catch data from biological sampling using a standard pelagic trawl and concurrent commercial catch data indicated that the majority of the insonified species in the study area were *Sardina pilchardus*, *Engraulis encrasicolus* and *Trachurus spp*. Other species represented less than 8% of the catches.

Temperature vertical profiles were recorded on a regular grid along survey transects by means of a calibrated Seabird S19 CTD. Surface temperature measurements were used in order to replace missing SST values, in satellite imageries, frequently observed in overcast or coastal areas. Satellite images, provided by the German Aerospace Agency's (DLR) and the Distributed Active Archive Center (NASA), were transformed into Sea Surface Temperature (SST) in Celsius and Sea Surface Chlorophyll-a concentration (SSC) using GIS tools (Valavanis et al., 2004). Salinity data have not been used, since the analysis is restricted to the two provided satellite imageries and the acoustically measured bottom depth.

Data analysis

The presented results are based on Geostatistical analysis applying ESRI's ArcGIS Geostatistical Analyst Software (GAS), which provides an extensive set of tools for performing different Kriging and Co-Kriging methods. Also other software packages (GSLIB/Fortran, WinGSLIB and GSTAT/R) were used for comparison and evaluation purposes. In certain cases the results were slightly different, indicating small variations in the implementation of the Geostatistical algorithms.

Selection of the geostatistical approach

There are different groups of Geostatistical techniques where auxiliary information is used to improve spatial prediction:

- Co-Kriging (CK), if the number of auxiliary variables is low and they are not available at all grid-nodes
- Kriging with External Drift (KED), if the auxiliary information is available at all grid-nodes and correlated with the target variable (Kriging with a trend model, Deutsch & Journel, 1998; Kriging with external drift, Wackernagel, 1995). Universal Kriging (UK) can be interpreted as a special case of KED where the drift is modelled as a function of only the coordinates.
- Kriging after De-trending (KAD), if drift and residuals can also be fitted separately and then summed (Goovaerts, 1997). This technique is also known as Regression Kriging (RK) and can be combined with stratification and GAMs (McBratney et al., 2000).

UK, KED and KAD are, in fact, equivalent methods providing, under the same assumptions, the same predictions. The advantage of KED is that the equations are solved at once, while the advantage of KAD is that there is no danger of instability as with the KED system (Goovaerts, 1997).

KED techniques are not recommended for use directly on interpolated (bottom depth) data or on imputed missing values (satellite data), since the uncertainty of the predicted covariate is not taken into account and therefore the prediction variances are underestimated. Additional limitations concerning the application of the KED technique (for instance, linearity between dependent and auxiliary variables, smoothness of the external variables) are described by Goovaerts (1997) and Royle & Berliner (1999).

In our case, it is desirable to describe the spatial relationship between bottom depth, temperature, chlorophyll-a and pelagic biomass based upon the available measurements on all the variables of interest. For univariate analysis the Universal Kriging technique was chosen, since bottom depth is clearly correlated with the coordinates and partially interrelated with the rest of the covariates. The multiple surfaces described by such multivariate observations will then be interpolated in a manner similar to Universal Kriging, by applying the extended CK technique.

Stationarity test

The basic biological or environmental factors, which are involved in ecological processes, change with time and introduce non-stationary conditions into the system. Some important statistics especially correlations are reasonable under the condition of stationarity in the process. If that condition is not fulfilled, the correlation is random and any biological interpretation of it is erroneous (Bendat & Piersol, 1971). The process which generates the data is said to be second-order stationary (strict sense) if the distribution function and the joint distribution function are invariant under an arbitrary time shift. It is stationary in the wide sense if the mean and the autocorrelation are invariant (Papoulis, 1965). In Geostatistical terminology, it is required that the spatial Covariance structure is essentially identical throughout the spatial domain of interest (Journel & Huijbregts, 1978). The ARC-GIS definition requires that the mean of all locations of interest be equal and that the Covariance between any two locations depends only on the distance and direction between them and not specifically on the actual location. When it is suspected that this is not the case, more complicated models may be required. Stationarity in the present work was achieved by data transformation (logarithmic), modelling the trend and using Universal Kriging, as well as by dividing the data set into subsets.

Empirical variograms and cross-variograms

This analysis typically aims at generating an empirical variogram and then fitting a parametric model that adequately captures the structure of the empirical variogram. Data were partitioned according to the distance between distinct pairs of observation locations, so that the bins are as small as possible to retain spatial resolution, and yet large enough that the empirical variogram estimate is stable (Cressie, 1993).

The semivariance or simply variogram (γ) is estimated by the formula:

$$\gamma(h) = \frac{1}{2}\sum_{i=1}^{n}\left[Z(X_i) - Z(X_i + h)\right]^2$$

where X_{i} and $X_{\mathrm{i}} + h$ are two locations in space at a certain distance, which belongs to a binwidth class of lag h, over all directions or specific to a given direction inside the studied area. $Z(X_{\mathrm{i}})$ and $Z(X_{\mathrm{i}} + h)$ are the magnitudes of the variable at the two locations and n is the number of data pairs in each lag class. The greater the number of data points, the greater the statistical reliability in each binwidth class.

The three key parameters in the variogram, Nugget, Range and Sill, were estimated applying the spherical model (Cressie, 1993).

- The nugget parameter represents the variation at very short distances (the variation due to measurement errors or the replication variance) (Rossi et al., 1992). It is the point at which the variogram model appears to intercept the y-axis.
- The range parameter represents the maximum spatial distance, where data are effectively spatially correlated or autocorrelated.
- The sill parameter describes the maximum variance of the variogram minus the nugget effect. In some publications the maximum variance, including the nugget, is referred as the total sill.

Correspondingly, a cross-variogram model describes the co-variation (or correlation relationship) between each pair of variables (Journel & Huijbregts, 1978; Cressie, 1993; Goovaerts, 1997). The Cross-variogram ($\gamma_{1.2}$) for two variables Z_1 and Z_2 is estimated by the formula:

$$\gamma_{1.2}(h) = \frac{1}{2N}\sum_{i=1}^{N}\Big\{\left[Z_1(x) - Z_1(x+h)\right] * \left[Z_2(x) - Z_2(x+h)\right]\Big\}.$$

Kriging

Ordinary and Universal Kriging, which are used in the present study, are gradual, local, and may not be exact, perfectly reproducing the measured data. Kriging selects weights so that the estimates are unbiased and the estimation variance is minimised. Identifying the best variogram may involve running and evaluating a large number of models, a procedure supported by the ArcGIS Geostatistical software.

Let Z represent the fish density in the x, y coordinate system, for which distances are linear (expressed in n.mi). Ordinary Kriging provides a statistical model for the process $Z(x, y)$, at all points in space, (x, y), as follows:

$$Z(x,y) = u + e(x,y)$$

where u and $e(x, y)$ are the overall, large-scale mean of the process across the spatial domain and the small-scale random fluctuation of the process accordingly. Unlike Universal Kriging, Ordinary Kriging places very little emphasis on the first mean component, focusing instead on modelling the structure of the small-scale random fluctuation component. Universal Kriging is an extension of Ordinary Kriging accommodating a spatially varying trend. It can be used both to produce local estimates in the presence of trend and to estimate the underlying trend itself. By incorporating the spatial varying trend, the previous equation is modified as:

$$Z(x,y) = u(x,y) + e(x,y)$$

where the mean term is described by the model $u(x,y)$.

Co-Kriging

The multiple surfaces described by such multivariate observations can then be interpolated in a manner similar to the Kriging techniques as above. However, Co-Kriging requires much more estimations, including estimating the autocorrelation for each variable as well as all cross-correlations. If the main variable of interest is Z, both autocorrelation for Z and cross-correlations between Z and all other variable types are used to make better predictions. The functions used to model those variograms and cross-variograms must be chosen so that the variance of any possible linear combination of these variables is always positive. The linear model of co-regionalisation, described in detail by Panatier (1996), is the most commonly used method for choosing such a set of functions so that the predictor of a given covariate at a given location X is a linear combination of all covariates of interest. Finally, it should be noted that often one of the auto- or cross-models may not fit its sample variogram very well compared to the others. In this case, the overall fit should be judged accordingly since each individual model is a small part of the total model (Isaaks & Srivastava, 1989).

The resulting Co-Kriging equations are slightly more complicated than those of Ordinary Kriging, but are still relatively straightforward, based on the Cross-variogram model (Journel & Huijbregts, 1978; Cressie, 1993; Goovaerts, 1997).

Specific performance measures

Finally, the developed models were evaluated by analysing the leave-one-out cross validation residuals and their statistics are compared and tested concerning modelling assumptions and whether standard errors estimated by the model are accurate (Isaaks & Srivastava, 1989).

The Leave-one-out cross validation residuals are generated using the following procedure:

– Create an empirical variogram using all of the N available observations.
– Estimate the theoretical variogram from the empirical variogram.
– For each observation $Z(s_i)$, $i = 1,\ldots, N$ in the data set,
– Remove the observation from the data set.
– Predict the Kriged value $\hat{Z}(s_i)$at the location of the removed observation using the remaining $(N - 1)$ observations.
– Calculate the difference between the predicted value and the true value, and divide this difference by the Kriging Standard Error (Bradley and Haslett, 1992).

$$SR_i = \frac{Z(s_i) - \hat{Z}(s_i)}{KSE}$$

– Record the value SR as the standardised residual at the location of the removed observation.

Once the residual at each location has been calculated as described, their distribution is tested for normality, validating if the Kriging model assumptions were correct, and if other modelling techniques should be considered.

The comparison among the different developed models was carried out based on the calculation of the root-mean-squared prediction error (RMS), the average standard error (ASE) and the coefficient of determination (R^2). In case the average standard error is close to the root-mean-squared prediction error, the variability in prediction is correctly assessed. If the average standard error is greater than the root-mean-squared prediction error, the variability of the predictions is overestimated; on the other hand, if the average standard error is less than the root-mean-squared prediction error, it is underestimated.

Results

Exploratory data analysis

Since the data were autocorrelated multivariate measurements, the level of spatial influence may vary with distance and other factors making the autocorrelation function non-linear. In general, the closer the neighbours are, the greater the level of correlation, which distorts statistical tests of significance in analyses such as correlation, regression, or analysis of variance (Cliff & Ord, 1981). Therefore, the following multivariate regression analysis was of an introductory character to the Geostatistical approach and did not aspire to provide statistical models describing the relationships among the variables under study.

Simple and stepwise multiple linear regression analyses were used to investigate possible relationships between the dependent variable LnSa (logarithmic transformed area backscattering coefficient), which represents an acoustically estimated fish density index and hypothesised covariates: spatial coordinates, time, bottom depth, SST and SSC, if the latter was available. The significance of the regression models was tested by ANOVA (probability of F to enter ≤ 0.05). The stepwise multiple regression analysis exhibited in general low proportion of variation explained by the model (R^2) on all significantly different from 0 R^2 values (Table 1). The most significant models that were developed were based on the acoustic surveys 1996 A, 1996 AB (pooled both data sets) and 1998. The acoustic estimated fish density (LnSa) was related to the following potential predictors: day-night time cycle, bottom depth, the spatial coordinates (Lon, Lat), SST and SSC (available only in 1998 survey).

Furthermore, residuals were calculated, based on non-linear modelling of acoustic fish density (LnSa) versus day-night time, fitted by least squares regression, by a sine signal (Zar, 1974). The linear regression models constructed using the derived residuals versus Latitude or bottom depth were significant (Table 1). As it is expected from the geography of the surveyed area, linear regression models confirmed a statistically significant ($P < 0.1$) and relative higher relationship ($R^2 = 0.64$) between bottom depth and spatial coordinates. These results are justified in the following Geostatistical analysis, the development of Universal Kriging models, testing if there is a geographical trend in the LnSa surface that partly explains data variation.

Environmental covariates

The horizontal spatial distribution of sea temperature showed a significantly different pattern between the two survey seasons (Fig. 1A: October 1996 and April/May 1997 and 1998). Lower temperatures were observed during the May and especially the April survey, where large surfaces in southern areas were in general cooler than the rest of the picture. In addition, cold spots appeared, associated with upwelling phenomena in the north-eastern part (Fig. 1, May-1997) or riverine outflow in the south-western part (Fig. 1, April-1998). The four SST images, corrected for missing values, were used in the Co-Kriging interpolation as covariates.

The bathymetry was modelled applying Universal Kriging on all bottom depth acoustic measurements, pooled all survey data (Fig. 1B). The bottom depth, as it is shown in the interpolated map, increased in south easterly direction until the 200 m isobath.

Since satellite imagery providing chlorophyll-a concentration was only available after 1997, the Co-Kriging estimation using chlorophyll data was restricted to the 1998 acoustic survey (Fig. 1C).

Kriging and Co-Kriging interpolations

Outputs from the cross-variogram modelling, Kriging and Co-Kriging spatial interpolations for the four surveys (1996a, 1996b, 1997 and 1998) are shown in Figs. 2–5, respectively. Each figure illustrates the cross-variogram models used in the Co-Kriging interpolation (a) and the interpolated acoustic fish density (LnSa) using Kriging (b) and Co-Kriging (c) techniques. Cross-validation statistics between predicted and observed acoustic fish density (LnSa), applying the leave-one-out cross validation method, are shown in Table 2. In each survey, statistics with significant covariate main effects and interactions are included, sorted by increasing coefficient of determination (R^2).

The four interpolated maps (Fig 2B, 3B, 4B and 5B) did not show, between seasons, any marked

Table 1 Simple or stepwise multiple regression analysis

Survey	Dependent variable	Separated/No of data	Independent variables	Model R^2	F	Sig.
1996 A	LnSa	All, 260	DN, SST, Lat, Depth, Lon	0.265	18.27	0.00
1996 A	Res(LnSa $\sim$ time)	All, 260	Lat	0.382	159.80	0.00
1996 A	Res(LnSa $\sim$ time)	All, 260	Depth	0.213	70.07	0.00
1996 AB	LnSa	Night, 244	Lat, Lon, Depth, SST	0.292	24.74	0.00
1998	LnSa	441	DN, Lon, Lat, Chl, Lat, SST	0.243	27.95	0.00

Independent variable: LnSa or residuals from modelling LnSa with time as independent variable; R^2: Coefficient of determination in the leave-one-out cross validation prediction; F = ANOVA F-test, Sig = significance of the model, $P < 0.01$

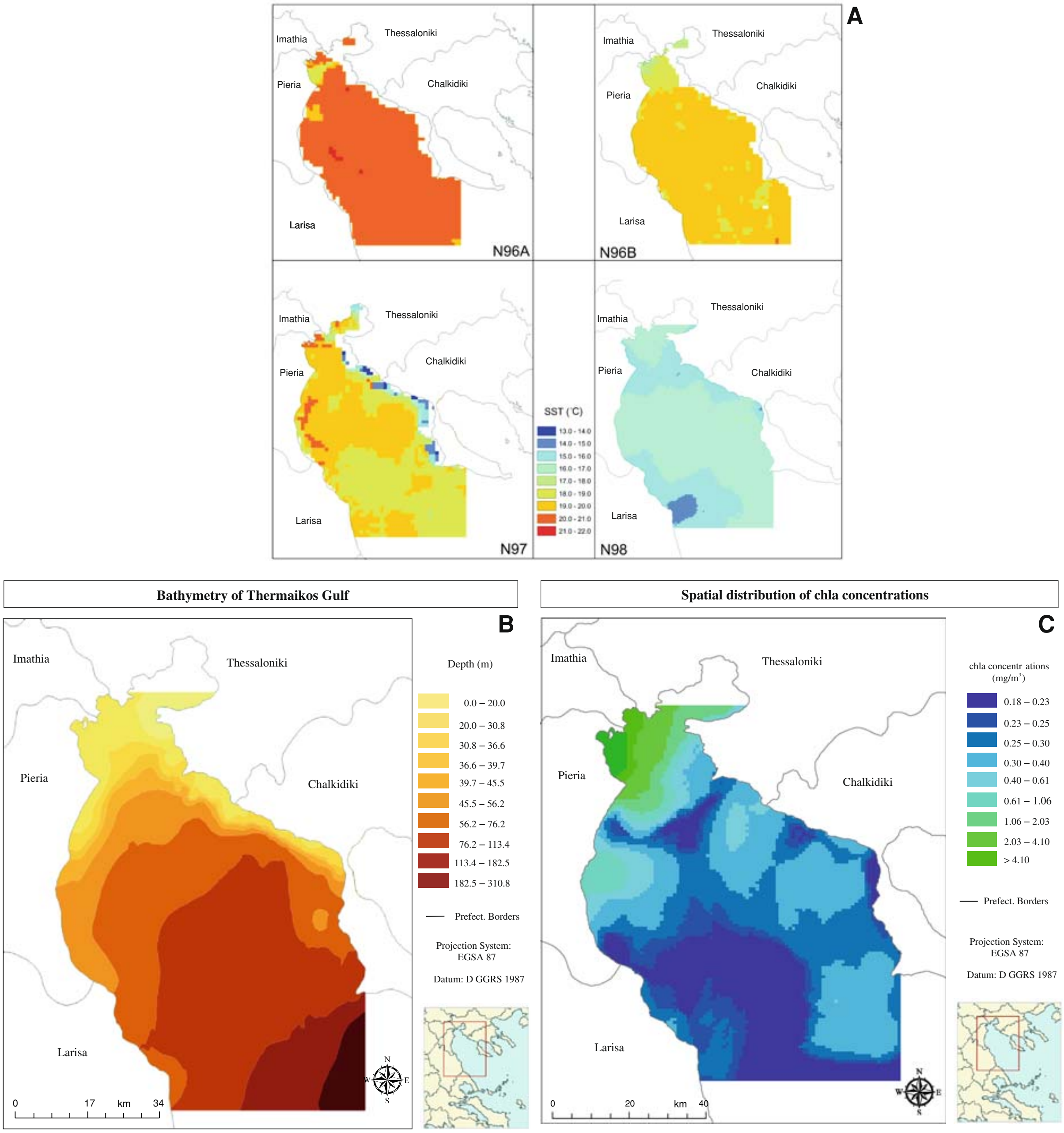

Fig. 1 (**A**) Sea surface temperature for the four sampling surveys in Thermaikos Gulf, and (**B**) Bathymetry and (**C**) Spatial distribution of chlorophyll-a concentrations

differences in their fish distribution. In contrast, within seasons, some biomass movements were reflected on the maps, concurrent to temperature changes. Considering the short time lag between the two October surveys (1–2 days), the dynamic of the spatial distribution is very high.

Only bathymetry, SST and their interactions provided a significant effect on predicting LnSa measurements applying the leave-one-out cross validation test. Chlorophyll-a data did not improve the modelling approach. The omnidirectional variogram and cross-variogram models exhibited autocorrelation

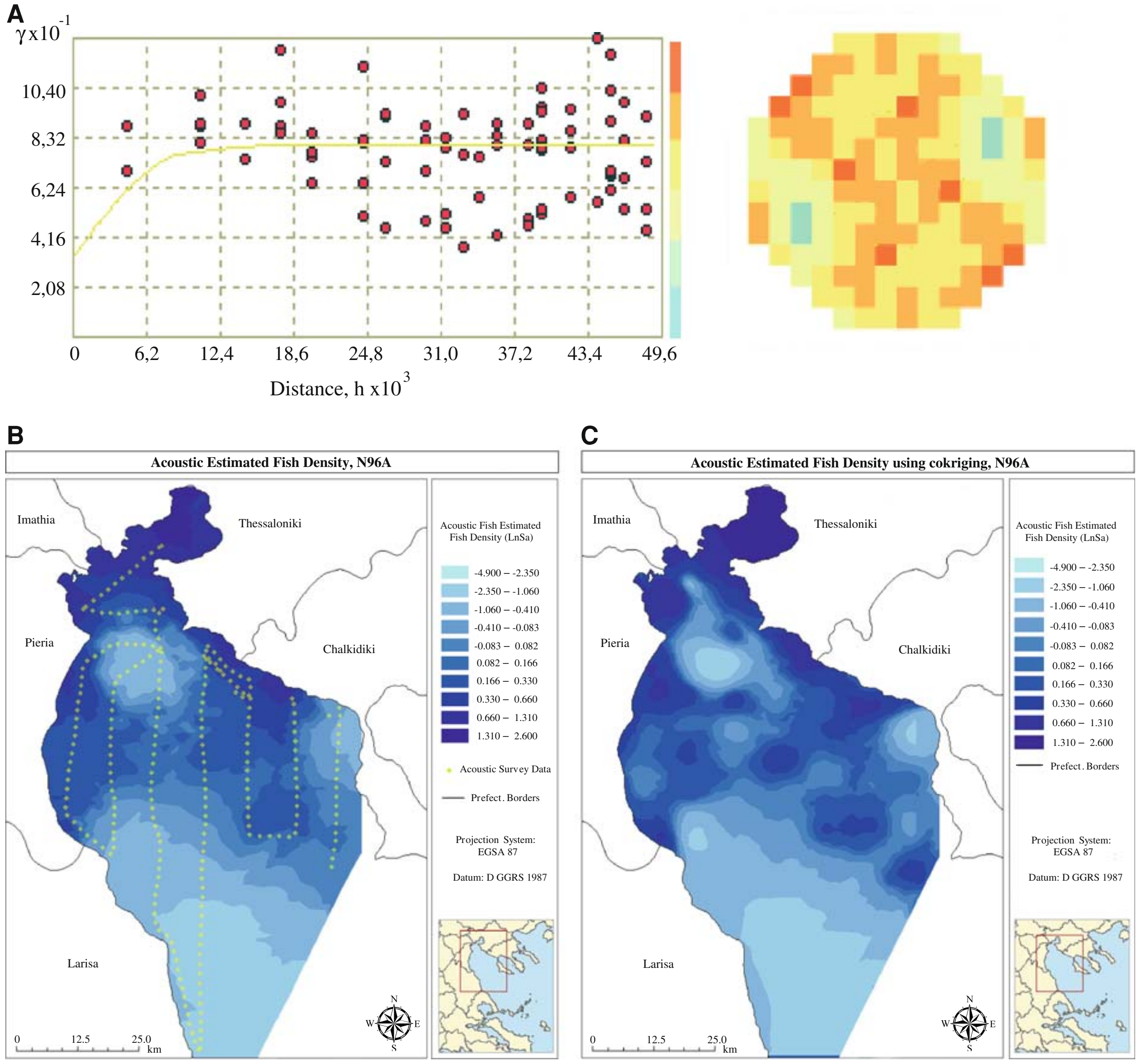

Fig. 2 Sampling survey N96A: (**A**) Cross-Variogram (omni-directional and exhaustive) of the Acoustic Estimated Fish Density (LnSa) using Co-Kriging (SST and depth), (**B**) Acoustic Estimated Fish Density (LnSa) using Kriging and (**C**) using Co-Kriging (SST and depth)

ranges of 5.7–25 km and 6 to 16 km, respectively (Table 2). Concerning the nugget effects of the cross-variograms, the October values (Figs. 2 and 3) were slighter than those of spring (Figs. 4 and 5).

Discussion

Oceanographic features such as temperature fronts, eddies, rings and upwelling areas have been related to fish biomass concentrations—at least at certain spatial scales and magnitudes (Laurs et al., 1984; Fiedler & Bernard, 1987; Chen et al., 2005)—and satellite ocean colour and sea surface temperature images have been used commercially for this purpose. In general, SST and chlorophyll-a spatial patterns are very similar, since very often, warm, nutrient-depleted water has low chlorophyll-a content and cold, nutrient-rich water has high chlorophyll-a levels.

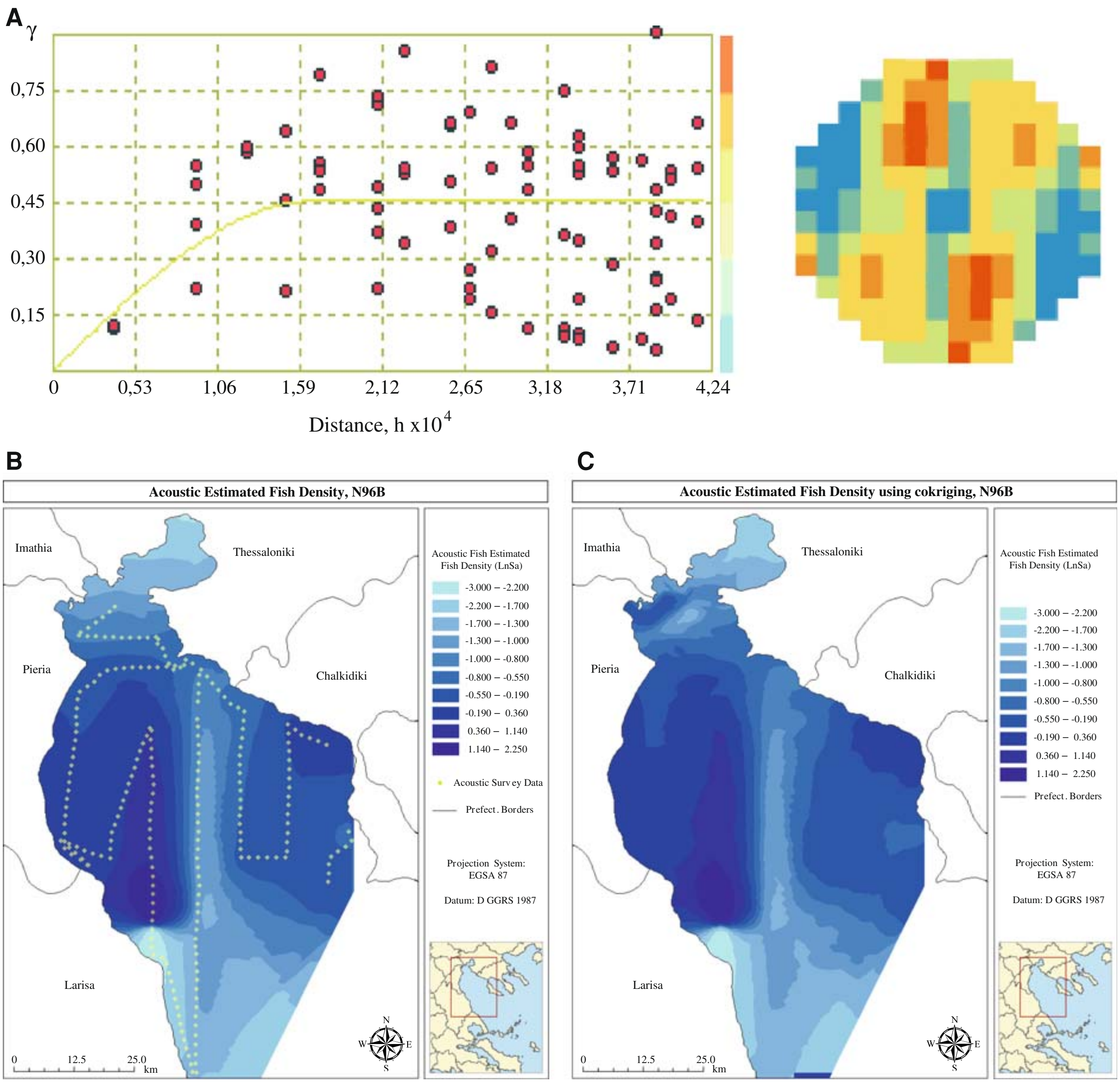

Fig. 3 Sampling survey N96B: (**A**) Cross-Variogram (omni-directional and exhaustive) of the Acoustic Estimated Fish Density (LnSa) using Co-Kriging (SST and depth), (**B**) Acoustic Estimated Fish Density (LnSa) using Kriging and (**C**) using Co-Kriging (SST and depth)

The relationship between sea surface satellite imagery and species spatial distribution depends, among other things, upon the number of linkages between phytoplankton and the given species trophic level. Some species, such as anchovy and sardine, which are closer to phytoplankton at some time periods of their life cycle, may exhibit a stronger linkage (Ware & Thomson, 2005) than other species which are at a higher trophic level.

However, this more or less expected spatiotemporal relationship between environment and fish density is not always and clearly evident (Fréon et al., 2005). For instance, the biomass structure in areas developing lower magnitudes of the oceanographic features with a smaller spatiotemporal scale may be affected in a stochastic and less predictable way (Sharp & McLain, 1993; Webster & Oliver, 2001).

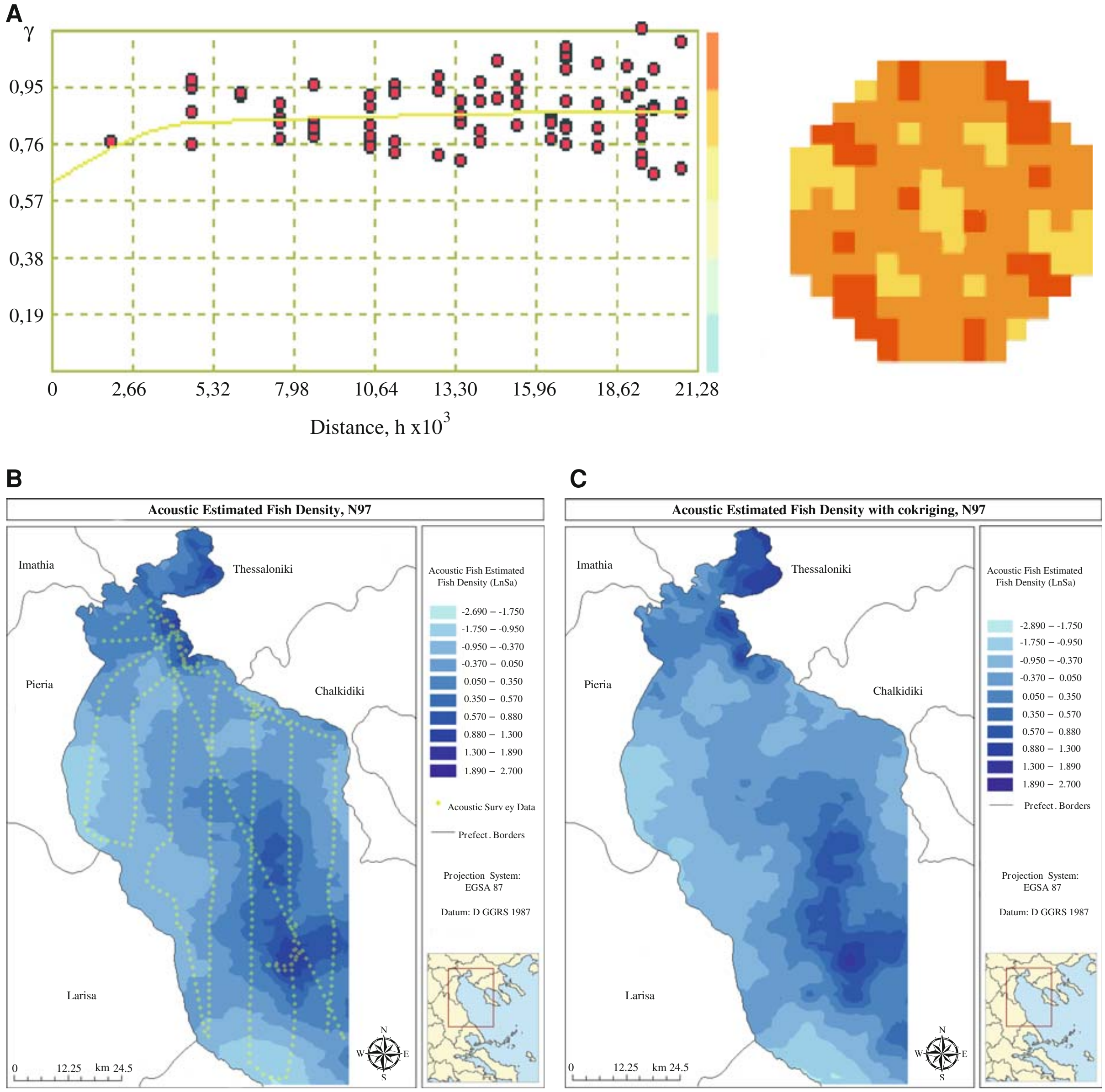

Fig. 4 Sampling survey N97: (**A**) Cross-Variogram (omnidirectional and exhaustive) of the Acoustic Estimated Fish Density (LnSa) using Co-Kriging (SST), (**B**) Acoustic Estimated Fish Density (LnSa) using Kriging and (**C**) using Co-Kriging (SST)

The Thermaikos Gulf, the study area in the present paper, is a relatively small region (30 nm wide × 100 nm long), developing a mesoscale environmental dynamic. Its mean depth varies from 40 m in the northern inner continental shelf (Thermaikos Bay) until the south tip of the Kassandra Peninsula, facing the open NW Aegean Sea, at the 200 m isobath. Freshwater from four rivers (Aliakmon, Pinios, Loudias and Axios) runs in the western coast of the Thermaikos Gulf resulting in low salinity and high turbidity in the sea water. About 40% of the total water volume of the Gulf has its origin from the rivers (Poulos et al., 2000). Another source of cold and low-saline waters is provided by the Black Sea inflow and other coast rivers emptying the northern Aegean Sea. This nutrient input can reach the Thermaikos Gulf, as it is clearly seen in surface chlorophyll-a satellite imagery (Agostini & Bakun, 2002).

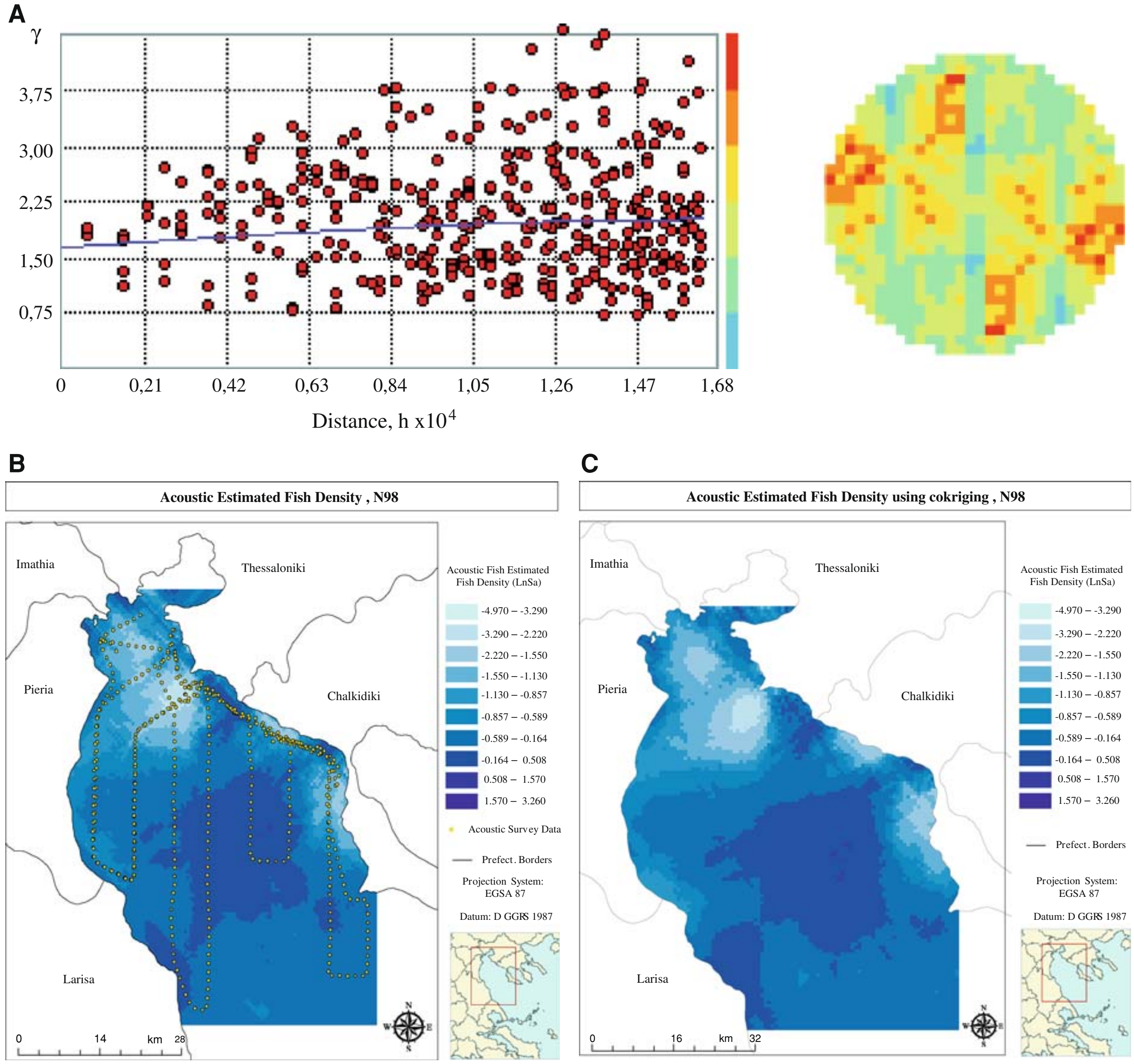

Fig. 5 Sampling survey N98: (**A**) Cross-Variogram (omnidirectional and exhaustive) of the Acoustic Estimated Fish Density (LnSa) using Co-Kriging (depth and SST), (**B**) Acoustic Estimated Fish Density (LnSa) using Kriging and (**C**) using Co-Kriging (depth and SST)

The properties of the water masses within the Gulf vary seasonally. The water column is homogenised during winter and stratified between spring and autumn (Tragou et al., 2005). According to Zervakis et al. (2005) the vertical water structure in the Gulf, during September and October, is characterised by a two-layer system, with the less saline water mass occupying the upper layer.

The SST images during the first trip in October (Fig. 1, N96 A) showed two surface regions: a large area of a relatively warm water (WW) mass (>20.0°) dominating in the Gulf, and some middle temperature water (MW) masses, (16.0°–20.5°) in the north-western coast, originating from riverine discharges, flowing southwards (Kontoyiannis et al., 2003). The second trip of October (Fig. 1, N96 B) was undertaken after a rapid weather change in the surveyed area, dominated by a strong northern wind which lowered the SST values over all of the Gulf and especially in its south-eastern part. It is proposed that during northerly

Table 2 Cross-validation statistics between predicted and observed acoustic fish density (LnSa), applying the leave-one-out cross validation method

Model		Range (Km)	RMS	ASE	R^2
1996a	LnSa	12.5	0.791	0.897	0.378
	LnSa & Depth	8.0	0.699	0.759	0.514
	LnSa & Depth & SST	12.0	0.684	0.696	0.534
1996b	LnSa	16.0	0.265	0.271	0.892
	LnSa & Depth & SST	16.0	0.265	0.260	0.982
1997	LnSa	25.0	0.853	0.918	0.265
	LnSa & SST	6.0	0.834	0.909	0.299
1998	LnSa	5.7	1.077	1.273	0.288
	LnSa & Depth	6.3	1.070	1.363	0.297
	LnSa & Depth & SST	6.8	1.062	1.350	0.307

First model represents the Kriging results, followed by the Co-Kriging with related covariates

winds a 2-gyre system is established in the area (Poulos et al., 2000). Zervakis et al. (2005) studied in detail the gyre structure of the Gulf in two hydrographic cruises during September and October. During September, a strong cyclonic gyre was presented in the southern part of the Gulf, whilst north of the gyre the circulation was anticyclonic. During October, the Gulf developed a slightly more complicated system; its central part was divided into two-gyre systems: a cyclonic in the west and an anticyclonic in the east.

The surface temperature pattern was different in May 1997 and April 1998. In May 1997, the rivers contributed warm waters forming a strong surface front, which is confirmed by hydrographic surveys carried out during the same period (Kontoyiannis et al., 2003). The MW surface mass (16.0°–20.0°) dominated most of the Gulf area, whilst along the north-eastern coasts cold waters (CW, 13.0°–16.0°) appeared, associated with the upwelling phenomena (Fig. 1, N97). In April 1998, the surface temperature was very low over the entire Gulf and especially in its south-western part, where the riverine outflow remained cooler compared to the rest of the sea waters (Fig. 1, N98).

Due to the riverine outflow and the Black Sea waters, the Thermaikos Gulf is more productive than the oligotrophic Aegean Sea, representing a rich fishing area both for trawlers and purse seiners (Stergiou et al., 1997). Although the species composition inside the Gulf may reveal seasonal variations, we decided to analyse the pelagic acoustic data as a whole, in order to avoid the introduction of species identification errors, by manually scrutinising the echograms or applying more sophisticated techniques (Haralabous and Georgakarakos, 1996). Since sardine appears to be by far the most abundant species in the area, possible drawbacks from pooling all species together are minimised.

The coastal morphology of the Gulf as well as the NW-SE bottom depth gradient dominated in the multiple regression analysis, especially after removing the time effect (Table 1). Consequently, a Universal Kriging interpolation was used, modelling fish density (LnSa) versus the spatial coordinates. The residuals of this model still included autocorrelation, which was exploited in the spatial multivariate analysis and the Co-Kriging interpolation (Table 2).

In the interpolated October choropleth maps (Figs. 2, 1996a; Fig. 3, 1996b), most of the fish abundance was observed near the shallow northern part of the Gulf, on both coastal sides, occupying warmer (WW) and middle warmer (MW) water temperatures. In contrast, during May 1997 and April 1998, the biomass was found further south in deeper areas where lower temperatures dominated (Figs. 4 and 5).

As mentioned above, the water mass is stratified between spring and autumn and consequentially the SST imagery reflects the upper sea layer condition. However, in most of the cases, the introduction of the SST covariate using Co-Kriging algorithms improved the coefficient of determination (R^2) in the cross-validation procedure (Table 2).

Bottom depth was the most important covariate and explained a portion of the cross-validation variance-additionally to the Kriging interpolation-contributing a minimum of 0.9% (1998) until a maximum of 13.6% (1996A). Bottom depth was also positively correlated with the LnSa-residuals of the time dependent model explaining 21.3% of the variability (Table 1). The encountered positive correlation between area backscattering coefficient Sa and depth can be explained by the fact that the acoustic beam volume increases with depth. However, this correlation strength varies due to the vertical heterogeneity of the pelagic biomass in the water column. Furthermore, bottom depth is partially interrelated with other environmental variables (SST, SSC) and fish density.

Co-Kriging with acoustic and SST data improved the prediction performance by 3.4% (1997), whilst expanding the bottom depth Co-Kriging by adding SST data increased the explained variance by 2.0% (1996A) and 1.0% (1998).

The observed small portions of the covariate variance are in accordance with the encountered spatial low auto- and cross-correlations (Figs. 2–5) and the expected intrinsic stochastic processes (Sharp & McLain, 1993; Webster & Oliver, 2001). Similar studies investigating the relationship between environmental factors and small pelagic fish abundance provide controversial or less evident results (Fréon et al., 2005). Positive responses are published for spawning anchovy in relation to certain environmental parameters (Castillo et al., 1996; Koutsikopoulos & Le Cann, 1996). However, the explained part of the observed sardine variance usually is small (Kerstan, 1993). GAMs applied on 11 years data from the South African sardine fishery using SST as a covariate explained only 0.5%–1.5% of the variance of the catch (Agenbag et al., 2003).

Despite the difficulty of scientifically proving relationships between environmental variables and fish density, satellite SST data are commonly used by pelagic fishing fleets searching for the best fishing areas (Santos, 2000). This is probably evidence of the importance of the spatiotemporal scale used (fishing fleets are able to access high resolution satellite images in short times) and the often greater strength of the phenomena (fronts and fish abundance) in the areas the satellite data is supplied.

The presented results show that despite the smaller spatiotemporal scale, Co-Kriging techniques can provide a useful tool for investigating possible environmental effects in the spatial distribution of small pelagic species. From the practical point of view, the resulted improvement in the accuracy, applying Co-Kriging algorithms, reduces the uncertainty in biomass estimation, which always is an important goal in stock assessment. Both the dynamic parameter (SST) and the static (bottom depth) parameter used in the analysis are the most important covariates for reducing estimation uncertainty. Bottom depth routinely is automatically measured during insonification. Therefore, bottom depth is perhaps the most useful auxiliary variable, which is also considered as a known and time invariant covariate in the surveyed area. Moreover, the SST and SSC covariates are commonly remotely assessed via satellite, requiring off-line processing.

From the aforementioned hydrographic studies, it is obvious that salinity gradient is important for understanding water stratification and potentially existed biomass distribution patterns. Certainly, increasing the number in the auxiliary variables, an improvement of the Co-Kriging efficiency could be expected. However, this paper is restricted to the methodological advantages of the simultaneous application of more than one Co-Kriging variables.

Concerning the spatial analysis presented in this paper, the incorporation simultaneously of more than one spatial covariate in the Co-Kriging technique improved the performance of the models, as it is depicted in Table 2. More advanced techniques have been proposed recently (Holland et al., 2003), modelling non-stationarity and heterogeneous covariance structures. It is suggested that these approaches, including hierarchical Bayesian techniques or Artificial Neural Networks (Georgakarakos et al., 2006), are very promising and require further research.

Acknowledgements In this work, the acoustic data collection and processing were supported by the Commission of the European Union (CLUSTER/FAIR, 1997). We would also like to thank V. Trygonis and D. Markopoulou and N. Fletcher for the initial analysis of the data.

References

Agenbag, J. J., A. J. Richardson, H. Demarcq, P. Fréon, S. Weeks & F. A. Shillington, 2003. Relating local abundance of South African pelagic fish species to environmental variables using generalized additive and linear models. Progress in Oceanography 59: 275–300.

Agostini, V. N. & A. Bakun, 2002. "Ocean trials" in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (with example application to European anchovy, *Engraulis encrasicolus*). Fisheries Oceanography 11: 129–142.

Armstrong, M., D. Renard & P. Berthou, 1989. Applying geostatistics to the estimation of a population of bivalves. ICES/CM/1989/K:37.

Azzali, M., G. Buracchi, S. Conti, S. Gambetti & M. Luna, 1985. Relationship between the forms of pelagic fish distribution and nycthemeral periods. A tentative model of behaviour. Oebalia 11: 471–488.

Bendat, J. S. & A. G. Piersol, 1971. Random data: analysis and measurement procedures. Wiley-Interscience, New York.

Bradley, R. & J. Haslett, 1992. High-interaction diagnostics for geostatistical models of spatially referenced data. The Statistician 41: 371–380.

Castillo, J., M. A. Barbieri & A. Gonzalez, 1996. Relationships between sea surface temperature, salinity, and pelagic fish distribution off northern Chile. ICES Journal of Marine Science 53: 139–146.

Chen, I. C., P. F. Lee & W. N. Tzeng, 2005. Distribution of albacore (*Thunnus alalunga*) in the Indian Ocean and its relation to environmental factors. Fisheries Oceanography 14: 71–80.

Cliff, A. D. & J. K. Ord, 1981. Spatial Processes: Models and Applications. Pion, London.

Cressie, N. A. C., 1993. Statistics for Spatial Data. John Wiley & Sons, New York, New York.

Deutsch, C. V. & A. G. Journel, 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, New York.

Fiedler, P. C. & H. J. Bernard, 1987. Tuna aggregation and feeding near fronts observed in satellite imagery. Continental Shelf Research 7: 871–881.

Foote, K. G., 1987. Fish target strengths for use in echo integrator surveys. Journal of the Acoustical Society of America 82: 981–987.

Fréon, P. & O. A. Misund, 1999. Dynamics of Pelagic Fish Distribution and Behaviour: Effects on Fisheries and Stock Assessment. Fishing News Books, Oxford.

Fréon, P., P. Cury, L. Shannon & C. Roy, 2005. Sustainable exploitation of small pelagic fish stocks challenged by environmental and ecosystem changes: a review. Bulletin of Marine Science 76: 385–462.

Georgakarakos, S., D. Koutsoubas & V. Valavanis, 2006. Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters. Fisheries Research 78: 55–71.

Giannoulaki, M., A. Machias, C. Koutsikopoulos & S. Somarakis, 2006. The effect of coastal topography on the spatial structure of anchovy and sardine. ICES Journal of Marine Science 63: 650–662.

Giannoulaki, M., A. Machias, S. Somarakis & N. Tsimenides, 2005. The spatial distribution of anchovy and sardine in the northern Aegean Sea in relation to hydrographic regimes. Belgian Journal of Zoology 135: 151–156.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Haralabous, J. & S. Georgakarakos, 1996. Artificial Neural Networks as a tool for species identification of fish schools. ICES Journal of Marine Science 53: 173–180.

Hensen, V., 1911. Das Leben im Ozean nach Zahlungen seiner Bewohner. Übersicht und Resultaten der quantitativen Untersuchungen. Ergebn. Plankton Expdn. der Humboldt Stiftung V.

Holland, D., W. M. Cox, R. Scheffe, A. Cimorelli, D. Nychka & P. Hopke, 2003. Spatial Prediction of Air Quality Data. Environmental Manager: 31–35.

Horne, J. K., P. E. Smith & D. C. Schneider, 1999. Comparative examination of scale-explicit biological and physical processes: recruitment of Pacific hake. Canadian Journal of Fisheries and Aquatic Sciences 56: 1–10.

Isaaks, E. H. & R. M. Srivastava, 1989. Applied Geostatistics. Oxford University Press, New York.

Journel, A. G. & C. J. Huijbregts, 1978. Mining Geostatistics. Academic Press, London.

Kerstan, S., 1993. Relationships between pilchard distribution and environmental factors. Marine science for a sustainable future. In Proceedings of the 8th South African Marine Science Symposium, Langebaan.

Kontoyiannis, H., V. H. Kourafalou & V. Papadopoulos, 2003. The seasonal characteristics of the hydrology and circulation in the Northwest Aegean Sea (Eastern Mediterranean): observations and modelling. Journal of Geophysical Research-Oceans 108: 3302.

Knudsen, H. P., 1990. The Bergen echo integrator: an introduction. Journal du Conseil Internation pour l'Exploration de la Mer 47: 167–174.

Koutsikopoulos, C. & N. Lacroix, 1992. Distribution, abundance of sole (*Solea solea* L.) eggs, larvae in the Bay of Biscay between 1986 and 1989. Netherlands Journal of Sea Research 29: 81–91.

Koutsikopoulos, C. & B. Le Cann, 1996. Physical processes and hydrological structures related to the Bay of Biscay anchovy. Scientia Marina 60: 9–19.

Laevastu, T. & M. L. Hayes, 1981. Fisheries Oceanography and Ecology. Fishing News Books, Oxford.

Lasker, R., 1978. The relation between oceanographic conditions and larval anchovy food in the California Current: identification of factors contributing to recruitment failure. Rapport Réunion du Conseil International pour l' Exploration de la Mer 173: 212–230.

Laurs, R. M., P. C. Fiedler & D. R. Montgomery, 1984. Albacore tuna catch distributions relative to environmental features observed from satellites. Deep-Sea Research 31: 1085–1099.

MacLennan, D. N. & I. G. MacKenzie, 1988. Precision of acoustic fish stock estimates. Canadian Journal of Fisheries and Aquatic Sciences 45: 605–616.

Maravelias, C. D., 1999. Habitat selection and clustering of a pelagic fish: effects of topography and bathymetry on species dynamics. Canadian Journal of Fisheries and Aquatic Sciences 56: 437–450.

Maravelias, C. D., D. G. Reid, E. J. Simmonds & J. Haralabous, 1996. Spatial Analysis and Mapping of Acoustic Survey Data in the presence of High Local Variability: Geostatistical Application to North Sea Herring (*Clupea harengus*). Canadian Journal of Fisheries and Aquatic Sciences 53: 1497–1505.

Matheron, G., 1971. The theory of regionalized variables and its applications. Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.

McBratney, A. B., I. O. A. Odeh, T. F. A. Bishop, M. S. Dunbar & T. M. Shatar, 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97: 293–327.

Panatier, Y., 1996. Statistics and Computing. VARIOWIN Software for Spatial Data Analysis in 2D. Springer-Verlag, New York.

Papoulis, A., 1965. Probability Random Variables and Stochastic Processes. MacGraw Hill, Kogakusha, Tokyo.

Paramo, J. & R. Roa, 2003. Acoustic-geostatistical assessment and habitat-abundance relations of small pelagic fish from the Colombian Caribbean. Fisheries Research 60: 309–319.

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

Petitgas, P., D. Reid, P. Carrera, M. Iglesias, S. Georgakarakos, B. Liorzou & J. Masse, 2001. On the relation between schools, clusters of schools, and abundance in pelagic-fish stocks. ICES Journal of Marine Science 58: 1150–1160.

Poulos, S. E., G. T. Chronis, M. B. Collins & V. Lykousis, 2000. Thermaikos Gulf Coastal System, NW Aegean Sea: an overview of water/sediment fluxes in relation to air-land-ocean interactions and human activities. Journal of Marine Systems 25: 47–76.

Rossi, R. E., D. J. Mula, A. G. Journel & E. H. Franz, 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecological Monographs 2: 277–314.

Royle, J. A. & L. M. Berliner, 1999. A Hierarchical Approach to Multivariate Spatial Modeling and Prediction. Journal of Agricultural, Biological and Environmental Statistics 4: 29–56.

Santos, A. M. P., 2000. Fisheries oceanography using satellite and airborne remote sensing methods: a review. Fisheries Research 49: 1–20.

Sharp, G. D. & D. R. McLain, 1993. Fisheries, El Nino-southern oscillation and upper-ocean temperature records: an Eastern Pacific example. Oceanography 6: 13–22.

Simard, Y., D. Lavoie & F. J. Saucier, 2002. Channel head dynamics: Capelin (*Mallotus villosus*) aggregation in the tidally-driven upwelling system of the Saguenay-St. Lawrence Marine Park's whale feeding ground. Canadian Journal of Fisheries and Aquatic Sciences 59: 197–210.

Simard, Y., P. Legendre, G. Lavoie & D. Marcotte, 1992. Mapping, estimating biomass, and optimizing sampling programs for spatially autocorrelated data: case study of the northern shrimp (*Pandalus borealis*). Canadian Journal of Fisheries and Aquatic Sciences 49: 32–45.

Stefánsson, G., 1996. Analysis of ground fish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science 53: 577–826.

Stergiou, K. I., E. D. Christou, D. Georgopoulos, A. Zenetos & E. Souvermezoglou, 1997. The Hellenic Seas: physics, chemistry, biology and fisheries. Oceanology and Marine Biology: An Annual Review 35: 415–538.

Sullivan, P. J., 1991. Stock abundance estimation using depth-dependent trends and spatially correlated variation. Canadian Journal of Fisheries and Aquatic Sciences 48: 1691–1703.

Swartzman, G. L., W. Stuetzle, K. Kulman & M. Powojowski, 1994. Relating the distribution of pollock schools in the Bering Sea to environmental factors. ICES Journal of Marine Science 51: 481–492.

Tragou, E., V. Zervakis, E. Papageorgiou, S. Stavrakakis & V. Lykousis, 2005. Monitoring the physical forcing of resuspension events in the Thermaikos Gulf - NW Aegean during 2001–2003. Continental Shelf Research 25: 2315–2331.

Tsimenides, N., G. Bazigos, S. Georgakarakos, A. Kapantagakis, 1992. Distribution of acoustic pelagic fish populations in the northern Aegean Sea. In Proceedings of the 1st World Fisheries Congress, Athens, 5: 33–42.

Valavanis, V. D., S. Georgakarakos, A. Kapantagakis, A. Palialexis & I. Katara, 2004. A GIS environmental modelling approach to essential fish habitat designation. Ecological Modelling 178: 417–427.

Wackernagel, H., 1995. Multivariate Geostatistics. Springer-Verlag, Berlin, Germany.

Ware, D. M. & R. E. Thomson, 2005. Bottom-Up Ecosystem Trophic Dynamics Determine Fish Production in the Northeast Pacific. Science 308: 1280–1285.

Webster, R. & M. A. Oliver, 2001. Geostatistics for Environmental Scientists. Wiley, Chichester.

Zar, J. H., 1974. Biostatistical Analysis. Prentice-Hall International, London.

Zervakis, V., A. P. Karageorgis, H. Kontoyiannis, V. Papadopoulos & V. Lykousis, 2005. Hydrology, circulation and distribution of particulate matter in Thermaikos Gulf (NW Aegean Sea), during September 2001-October 2001 and February 2002. Continental Shelf Research 25: 2332–2349.