

Marine Biology, in press (2007)
The original is available at www.springerlink.com
DOI: 10.1007/s00227-007-0659-3

**Density surface modelling with line transect sampling as a
tool for abundance estimation of marine benthic species: the
Pinna nobilis example in a marine lake**

Stelios Katsanevakis

Department of Zoology-Marine Biology, Faculty of Biology, University of Athens,
Panepistimioupolis, 15784 Athens, Greece
e-mail: stelios@katsanevakis.com

Abstract

Density estimation of marine benthic fauna is most often conducted with fishery surveys using dredges or trawls. These estimates are often unreliable due to low and variable efficiency and are inappropriate when dealing with rare or endangered species. In the marine Lake Vouliagmeni, a density surface modelling (DSM) approach using survey data from line transects, integrated with a Geographic Information System (GIS), was used to estimate the population density of the endangered fan mussel *Pinna nobilis*. This is the first time that such an approach has been applied for a marine benthic species. DSM was beneficial in relation to traditional distance sampling. Apart from providing a more precise total abundance estimate, it related the density of the species to spatial covariates of interest, gave a depiction of the species dispersion in the study area, and provided abundance estimates in any sub-region of the study area. In Lake Vouliagmeni, a marked zonation of *P. nobilis* distribution was revealed, with the species being restricted in the shallow peripheral zone at depths <22 m. Two density peaks were observed, a major peak at depths between 12 and 13 m and a secondary peak at ~4 m. A main hotspot of high density was also observed in the northeastern part of the lake. Total abundance of the species was estimated to be 6,770 individuals with a 95% confidence interval of 5,460 – 8,393 individuals.

Introduction

Density estimation of marine benthic fauna is commonly conducted with fishery surveys using dredges or trawls. However, efficiency and size selectivity issues with trawls and dredges are well documented. Catch rates of trawls and dredges may be quite low (e.g. Tuck et al. 1997; Hall-Spencer et al. 1999) and depend on several factors like trawling speed and smoothness of the bottom (Hall-Spencer et al. 1999), the size of the target species (McLoughlin et al. 1991), the duration of the haul (Shafee 1979), the time of the day, and the season (Chapman 1980). At shallow depths, direct visual surveys with SCUBA diving are advantageous compared to fishery surveys, as the records are not dependent on catch efficiency (Katsanevakis and Verriopoulos 2004; Katsanevakis 2005). In addition, direct visual surveys are non-destructive, which should be of concern, especially when dealing with rare or protected species or habitats.

Distance sampling (Buckland et al. 2001) is a widely used group of methods for estimating abundance and/or density of biological populations. Distance sampling has been used extensively in terrestrial ecology (mostly for birds and terrestrial mammals) and for marine mammals. Although it is the standard method for abundance estimations of many species, it is only rarely used for underwater surveys of benthic fauna (e.g. Katsanevakis 2005; 2006). The main distance sampling methods are line transects and point transects. A standardized survey is conducted along a series of lines (in line transects) or points (in point transects) searching for the animals of interest. For each animal detected, the distance from the line or point is recorded. A detection function is fitted from the set of recorded distances, which is used to estimate the proportion of animals missed by the survey and hence estimate abundance. When the detection of individuals is difficult, a distance sampling method is typically more efficient than simple strip transect sampling. This is because densities are corrected with the use of the detection function and the sample size is

larger for the same amount of effort as all detected individuals may be recorded regardless of how far they are from the line.

A topic for which significant methodological advances are expected in the near future is the integration of modelling with Geographic Information Systems (GIS) and model-based inference from sample survey data (Buckland et al. 2000; Hedley et al. 2004). Wildlife managers increasingly wish to extract more than an abundance estimate from their surveys and frequently want to relate animal density to spatial variables, reflecting various factors like topography or habitat (Hedley et al. 2004). A first step towards this direction was made by Hedley (2000) and Hedley and Buckland (2004) who developed methods for improving abundance estimation of cetacean abundance, allowing heterogeneity in the spatial distribution of cetaceans to be modeled from standard line transect data. Two approaches were suggested that enable spatial variation in animal density to be modeled using standard generalized linear modelling (GLM; McCullagh and Nelder 1989) or generalized additive modelling (GAM; Hastie and Tibshirani 1990). In the first approach (count model), the transect lines were divided into smaller discrete units, and the expected number of detections in each unit was modeled using explanatory spatial covariates. In the second approach (waiting distance model), the response was derived from the observed waiting times (or distances) between detections (Hedley 2000; Hedley and Buckland 2004; Hedley et al. 2004).

The fan mussel *Pinna nobilis* is endemic to the Mediterranean Sea. It is one of the largest bivalves of the world, attaining lengths up to 120 cm (Zavadnik et al. 1991), is long lived (up to 20 years) with very variable recruitment (Butler et al. 1993), and occurs in coastal soft-bottom areas at depths between 0.5 and 60 m. Fan mussels live partially buried by the anterior portion of the shell and attached by their byssus in the substratum. The population of *P. nobilis* has been greatly reduced during the last few decades as a result of recreational and commercial fishing for food, use of its shell for decorative purposes, and incidental killing by trawling and anchoring. Consequently, it has been listed as an endangered species in the Mediterranean and is under strict protection according to the European Council Directive 92/43/EEC. To effectively protect this endangered species there is a pressing need for better information on its population ecology and the distribution of all major local *P. nobilis* populations. In Lake Vouliagmeni (a semi-closed embayment in the Korinthiakos Gulf of Greece) (Fig. 1), a significant *P. nobilis* population has been reported (Katsanevakis 2006).

In the present work, the ‘count method’ of Hedley and Buckland (2004) was used to model the density surface of *P. nobilis* in Lake Vouliagmeni to estimate the abundance of the species, relate its density to the topography of the lake, and visually depict its dispersion in the study area. This method has never been used before for a marine benthic species and could be proved a very advantageous way of estimating abundance of benthic fauna relative to the use of trawls or dredges. Furthermore, for a protected species like *P. nobilis*, dredging or trawling is a destructive and non-acceptable means for abundance estimation; alternative methods as the one proposed here should be further developed and used by marine scientists.

Materials and Methods

Study Area

Lake Vouliagmeni is located on the 'Perachora' Peninsula and is connected to Korinthiakos Gulf through a narrow (18.7 m) and shallow (1.1 m maximum depth) channel that was dredged approximately a century ago (Fig. 1). Lake Vouliagmeni has a maximum length (E-W) of 1,881 m, width (N-S) of 931 m, depth of 49 m, and a total surface area of 150.4 ha (Katsanevakis 2006). The study area was defined as the part of the lake between the coastline and the 30 m depth contour. In a previous study of the population ecology of the fan mussel in Lake Vouliagmeni (Katsanevakis 2006), the species was found mostly at depths <16 m, while no individual was found deeper than 22 m. Consequently, the study population are those individuals living at depths <30 m, which most probably includes the whole population of *P. nobilis* in the lake. SCUBA diving at depths >30 m is problematic (due to nitrogen narcosis and decompression sickness) and there was no reason to extend the line transects beyond that limit.

Line transects and field work

Eighteen line transects, perpendicular to the depth contours and extending from the shore (0 m depth) to a depth of 30 m, were defined randomly in the lake (Fig. 1). Each line transect was defined with a nylon line that was deployed using a diving reel while SCUBA diving. The line was marked with a water-resistant paint marker every meter and divided into 4-m segments, with water-resistant numbered signs. After deploying the line, all *P. nobilis* individuals within 2 m (= w) from the line were counted. For each observed individual, the following variables were recorded: the corresponding segment in which it was found, the perpendicular distance from the line (= y_i), and the shell width (= s_i). The location of an individual was defined as the mid-point of the section of the shell by the seabed plane. In addition, at the mid-point of each segment of the line transect the exact depth was measured using the electronic depth meter of a diving computer (Suunto, Vyper). All measurements were conducted during July 2006.

The shell width was measured *in situ* with aluminium vernier calipers (60 cm upper limit, 0.5 cm accuracy) for individuals >15 cm or with plastic vernier calipers (15 cm upper limit, 0.05 cm accuracy) for individuals ≤15 cm. The perpendicular distance from the line was measured with the use of a 2 m plastic rod, bearing marks every 5 cm. All data were recorded on a diving slate.

Modelling the detection function

The detection function of *P. nobilis* in the lake, $g(y)$, gives the probability of detecting an individual given that it is at distance y from the line. The detection function $g(y)$ was estimated from the distance data according to Buckland et al. (2001), using the Distance 6.0 software (Thomas et al. 2006). Specifically, two functions were considered as candidate detection functions: the one-parameter half-normal function $g(y) = \exp\left(-\frac{y^2}{2\sigma^2}\right)$ and the two-parameter hazard-rate function

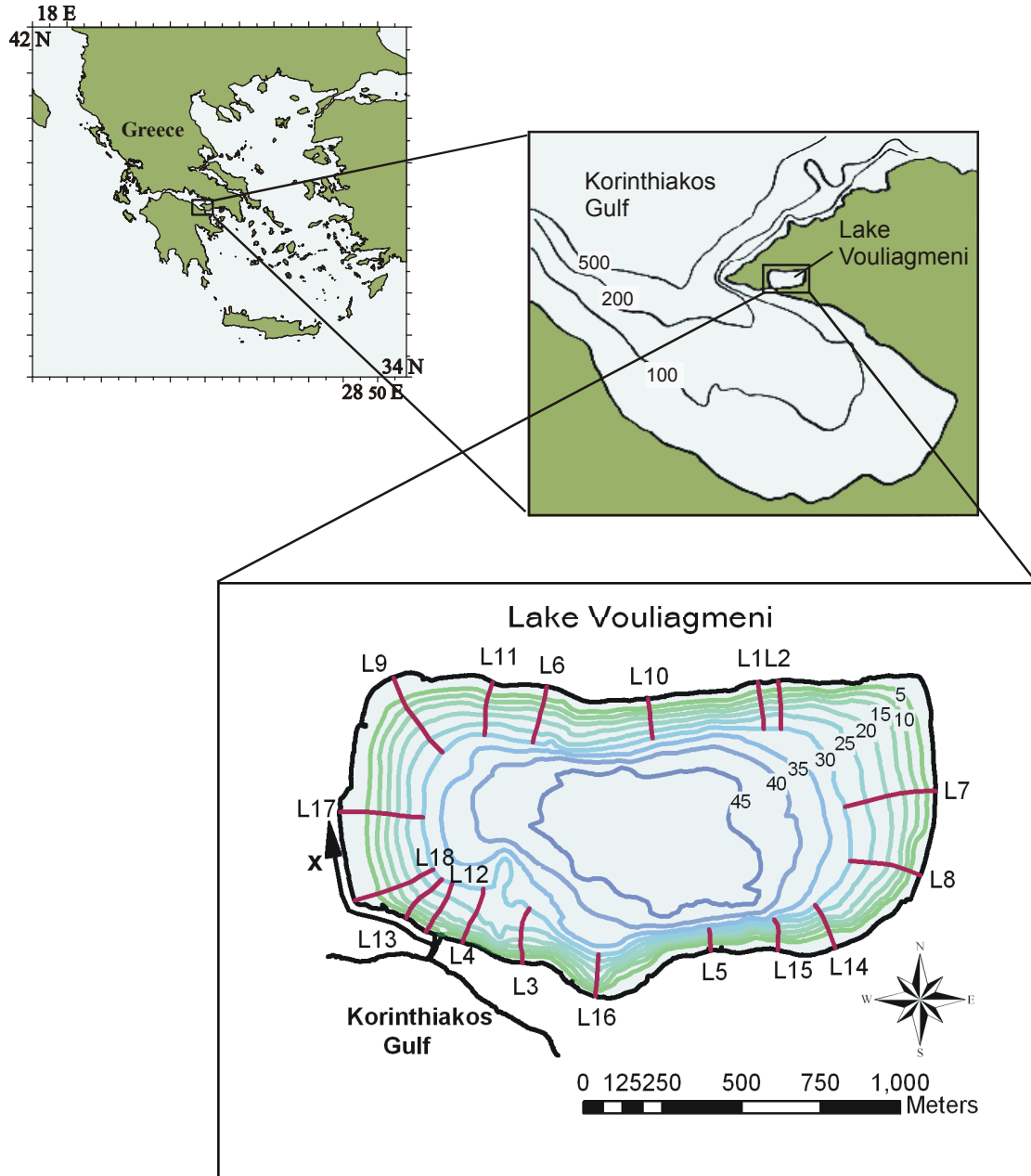


Fig.1: Map of Lake Vouliagmeni. The positioning of the eighteen line transects (L1 to L18) and the bathymetric contours of the lake (at 5-m intervals) are given.

$g(y)=1 - \exp\left[-\left(\frac{y}{\sigma}\right)^{-b}\right]$, where σ is a scale parameter and b a shape parameter. Four

different models g_i , $i = 1$ to 4, were fitted to distance data. In g_1 and g_2 , the half normal and the hazard rate function, respectively, were used and the scale parameter σ was considered constant. In g_3 and g_4 , the half normal and the hazard rate function were again used and the size s of *P. nobilis* (as expressed by the shell width) was included as a covariate, because shell size might significantly affect detectability (Katsanevakis 2006). Specifically, it was assumed that the covariate affects detectability via the scale term σ , according to the relationship $\sigma = \exp(\beta_0 + \beta_1 \cdot s)$ (Marques and Buckland 2004).

The small-sample, bias-corrected form AIC_c (Hurvich and Tsai 1989) of the Akaike Information Criterion (Akaike 1973; Burnham and Anderson 2002), was used for model selection among the set of candidate models. Specifically, $AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$ and $AIC = -2\ln(\mathcal{L}(\hat{\theta}|\text{data})) + 2K$, where $\ln(\mathcal{L}(\hat{\theta}|\text{data}))$ is the numerical value of the log-likelihood (natural logarithm) at its maximum point and $\hat{\theta}$ is the vector of the model's estimated parameters. The model with the smallest AIC_c value ($AIC_{c,\min}$) was selected as the 'best' among the models tested. The AIC_c differences, $\Delta_i = AIC_{c,i} - AIC_{c,\min}$, were computed over all candidate models. Goodness-of-fit of the best model was assessed with Q-Q plots and the weighted (to give higher weight to distances near zero) Cramér-von Mises test, as described by Burnham et al. (2004). Only the best model was used for density surface modelling.

Density surface modelling (DSM)

Two spatial covariates were used in modelling the density surface of *P. nobilis* in the lake. These two variables were depth d and the variable x representing the distance from the channel connecting the lake with the open sea, when moving clockwise along the shoreline (Fig. 1). The study area (i.e. between the shoreline and the 30-m contour) was actually a narrow strip with x axis in the large dimension and a q axis perpendicular to x . The q variable was highly correlated with depth. Multicollinearity in the covariates indicates that a substantial part of the information in one or more of the predictor variables is redundant, which makes it difficult to separate the effects of the different variables on the response variable. Because depth is quite an important covariate for *P. nobilis* distribution (Katsanevakis 2006), d was chosen over q , which was not considered as a covariate in density surface modelling.

The surveyed line transects were divided into T 4-m ($= l$) segments and the number of *P. nobilis* individuals detected within each segment was denoted by n_i , $i = 1, \dots, T$. The probability of detecting animal j on segment i was symbolized by \hat{p}_{ij} and was obtained from the estimated detection function. The total number of individuals within segment i was obtained using the Horvitz-Thompson-like estimator

$\hat{n}_i = \sum_{j=1}^{n_i} \frac{1}{\hat{p}_{ij}}$ (Hedley et al. 2004). In the current context, in all candidate models for the

detection function, detectability was considered independent of any spatial covariates and thus the i subscript in \hat{p}_{ij} may be eliminated, i.e. $\hat{p}_{ij} = \hat{p}_j$. Additionally, only shell size was considered as a covariate in two of the candidate models. Thus the

estimator of total number of individuals may be written as $\hat{n}_i = \sum_{j=1}^{n_i} \frac{1}{\hat{p}_j}$, if shell size is

considered as a covariate and thus probability of detection is individual-dependent, or

as $\hat{n}_i = \frac{n_i}{\hat{p}}$ if probability of detection, \hat{p} , is common for all individuals.

The expected values of abundance in each segment were related to the spatial covariates using GAMs (Hastie and Tibshirani 1990), according to the formulation

$f\left(\frac{E[\hat{n}_i]}{2wl}\right) = c + \sum_m s_m(z_{mi})$, where f is the link function, $s_m(\cdot)$ is the one-

dimensional smooth function for spatial covariate m , z_{mi} is the value of spatial

covariate m for segment i , and $a = 2wl$ is the covered area of the segment. The logarithmic link (which ensures positive values of the mean response) and a quasipoisson error distribution were used. With a logarithmic link the above equation becomes $f(E[\hat{n}_i]) = c + \sum_m s_m(z_{mi}) + f(a)$. The smooth function $s_m(\cdot)$ was represented using penalized regression splines (cubic splines with basis dimension $q = 10$), estimated by penalized iterative least squares (Wood 2006). The optimum degree of smoothing was defined by generalized cross validation (GCV), increasing the amount that each model effective degree of freedom counts in the GCV score by a factor $\gamma = 1.4$. The GCV method is known to have some tendency for occasional overfitting, and it has been suggested that using $\gamma \approx 1.4$ can largely correct this without compromising model fit (Kim and Gu 2004). In addition, the smooth for the spatial covariate x was a cyclic regression spline, meaning that the smooth function had both the same value and first and second derivatives at its upper and lower boundaries (because the shoreline of the lake is a closed loop).

Three different GAMs h_i , $i = 1$ to 3, were fitted. In h_1 only d was used as a spatial covariate ($m = 1$, $z_{1i} = d_i$), in h_2 only x was used ($m = 1$, $z_{1i} = x_i$), while in h_3 one-dimensional smooths of both d and x were included ($m = 2$, $z_{1i} = d_i$, $z_{2i} = x_i$). Model selection among the set of the three density surface models was conducted based on their GCV score (Wood 2006). For generalized additive modelling, the package *mgcv* (Wood 2000; 2006) was used in R v.2.3.1 (R Development Core Team 2006).

Abundance estimation

Using ArcMap 8.1 software (ESRI), a prediction grid was created on the map of the study area by dividing it to 8670 square cells of 88.54 m². The abundance of *P. nobilis* in the study area was estimated as the sum of $E[\hat{n}_r]$ at each cell r of the prediction grid, i.e. $\hat{N} = \sum_r E[\hat{n}_r]$, where $E[\hat{n}_r]$ were predictions according to the selected spatial density model h_i (where in the offset term the covered area of the segment $a = 2wl$ was replaced by the surface of the prediction cell). To predict $E[\hat{n}_r]$, the values of the covariates d and x at each cell r were estimated using both ArcMap 8.1 (ESRI) and the bathymetric surface of the lake produced by Katsanevakis (2006). The above abundance estimation was conducted using the DSM analysis engine of the Distance 6.0 software (Thomas et al. 2006). Based on the predictions $E[\hat{n}_r]$, a distribution map of *P. nobilis* in Lake Vouliagmeni was produced using ArcMap 8.1.

There are two variance components to the final abundance estimate, one component arising from the detection function model and a second one from the density surface model. It was assumed that these components were independent, and the delta method (Seber 1982) was used to estimate the total variance, according to the relationship $[cv(\hat{N})]^2 = [cv(\hat{p})]^2 + [cv(\hat{N}_{DSM})]^2$, where $cv(\hat{p})$ is the coefficient of variation of the estimator of detection probability and $cv(\hat{N}_{DSM})$ is the coefficient of variation related to density surface modelling. The first component was obtained analytically according to Buckland et al. (2001). For the second component, a parametric bootstrap approach was followed (Efron and Tibshirani 1993), with sample size $B = 999$. The bootstrap approach used a moving block of m segments as a sampling unit, large enough so that segments more than m units apart would be

independent and small enough so that spatial correlation present in segments less than m units apart would be preserved. The block size m was defined by examining the variogram of the deviance residuals (geoR package in R). Unmodelled spatial autocorrelation results in a variogram that increases sharply before eventually reaching a plateau; the start point of the plateau would define the block size m .

Comparison of DSM and traditional Distance Sampling

To compare estimations with DSM and traditional distance sampling, the abundance of *P. nobilis* in the study area was also estimated with the traditional line transect sampling approach using the same detection function as before. Non-parametric bootstrap was used for standard error and confidence interval estimations (resampling transects with sample size $B = 9,999$).

Standard distance sampling analysis methods assume that all points within the study area have an equal probability of being included in the sample region (i.e. there is uniform coverage probability). In the design of the present study the coverage probability was slightly non-uniform in the study area, increasing towards the centre of the lake. Thus, to correctly estimate abundance the coverage probability also had to be estimated and included in the analysis. The study area was divided into the following thirteen bathymetric zones: <1, 1 – 3, 3 – 5, ..., 23 – 25 m. The area of each zone was denoted A_z ($z = 1$ to 13), l_{qz} was the part of the q -th line transect ($q = 1$ to 18) that belonged to the z zone and p_j was the detection probability of individual j . Assuming that within each bathymetric zone there is a uniform coverage probability p_{cz} , and using a Horvitz-Thompson-like estimator:

$$\hat{N} = \sum_{z=1}^{13} \sum_{\substack{\text{all } j \\ \text{within } A_z}} \frac{1}{p_{cz} p_j} = \sum_{z=1}^{13} \sum_{\substack{\text{all } j \\ \text{within } A_z}} \frac{1}{\left(2wA_z^{-1} \sum_{q=1}^{18} l_{qz} \right) p_j}, \quad \text{where } A_z \text{ was taken from}$$

Katsanevakis (2006), l_{qz} was estimated with Arcmap 8.1, and p_j was estimated from the best detection function among the set of candidate models. The standard error and confidence intervals of the above abundance estimate were estimated with non-parametric bootstrap (resampling transects with sample size $B = 9,999$).

Results

Modelling the detection function

The best model of the detection function among models tested was the half-normal function with constant scale parameter (i.e., no covariate) (Table 1). This model provided a good absolute fit, as assessed by the Q-Q plot and the weighted Cramér-von Mises test ($p = 0.3$). The two models that included a size covariate had less support from the data.

The best model was given by the equation $g(y) = \exp(-0.199y^2)$ (Fig. 2), and detectability was constant and common for all individuals after integrating out distance. The corresponding probability of detection (\pm SE) was $\hat{p} = 0.788 \pm 0.083$ and assumed no effect of size.

Table 1: Summary of the set of four models of the detection function, including the bias-corrected form AIC_c of the Akaike Information Criterion and AIC_c differences Δ_i . Models are sorted from ‘best’ to ‘worst’. k : number of model parameters.

Model	k	AIC_c	Δ_i
g_1 (half-normal)	1	100.94	0.00
g_2 (hazard rate)	2	101.32	0.38
g_3 (half-normal + size covariate)	2	102.46	1.52
g_4 (hazard rate + size covariate)	3	103.50	2.57

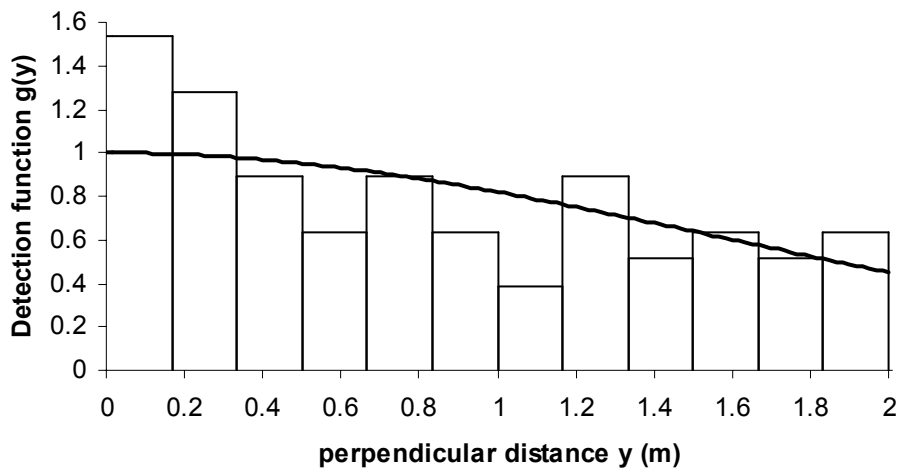


Fig.2: Histogram of the distance data (grouped into distance bins and scaled) and the corresponding ‘best’ model among the set of candidate models (half-normal).

Table 2: Evaluation of the three density surface models, based on their GCV (generalized cross validation) score. The percentage of deviance explained by the model and the corresponding estimated degrees of freedom (edf) are also given.

Model	GCV score	Deviance explained (%)	edf
h_1	0.554	28.4	6.13
h_2	0.725	6.4	6.08
h_3	0.527	33.5	10.91

Density surface modelling

The best density surface model was h_3 , which included univariate smoothes of both d and x (Table 2). The expression of h_3 was $\log(E[\hat{n}_i]) = c + s_1(d) + s_2(x) + \log(a)$, where $a = 16 \text{ m}^2$, $c = -7.08 \pm 0.53$ (\pm SE), and the smooth functions s_i are given in Fig. 3. The density of *P. nobilis* in Lake

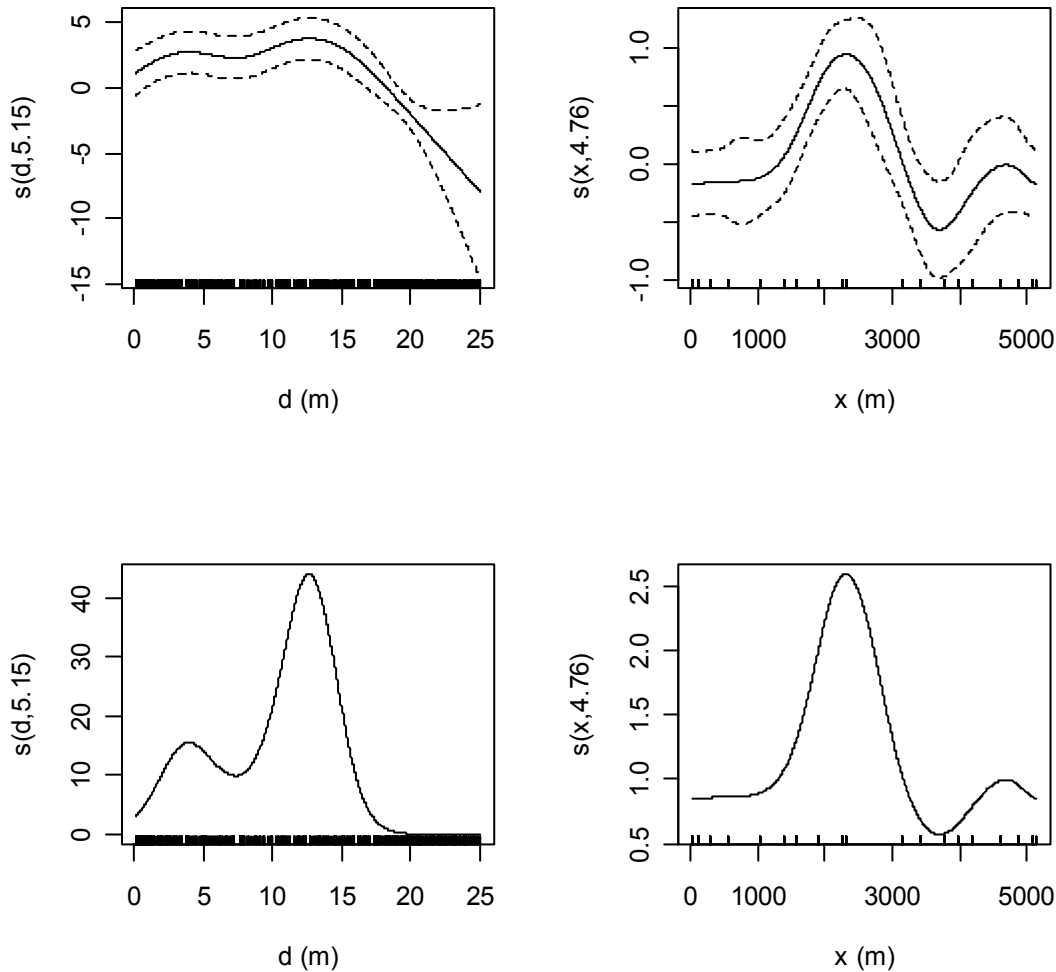


Fig.3: Estimated smooth terms for the h_3 model of *Pinna nobilis* abundance in 4m x 4m quadrats in Lake Vouliagmeni. In the top panel, the smooth terms are given in the linear predictor scale and the corresponding 95% confidence intervals are given with dotted lines, while in the bottom panel the smooth terms are given in the response scale (i.e. smooth terms were exp-transformed). The number *edf* in $s(z_m, edf)$ is the estimated degrees of freedom of the smooth term. A one-dimensional scatterplot is given at the bottom of each graph, using a vertical bar as the plotting symbol, to illustrate the distribution of available data.

Vouliagmeni had a bimodal distribution in relation to depth; a major peak was found at depths between 12 and 13 m and a secondary peak at ~4 m depth. In relation to the x coordinate, there is a peak of *P. nobilis* density at $x \approx 2,300$ m.

Abundance estimation

The variogram of the deviance residuals (Fig. 4) suggested no spatial autocorrelation and thus for parametric bootstrap, the segment was the sampling unit, i.e. $m = 1$. The abundance of *P. nobilis* in the study area was estimated to be $\hat{N} = 6770$ with $cv(\hat{N}) = 11.0\%$. The variation component due to uncertainty in the estimation of the detection function was $cv(\hat{p}) = 10.6\%$, while the component due to

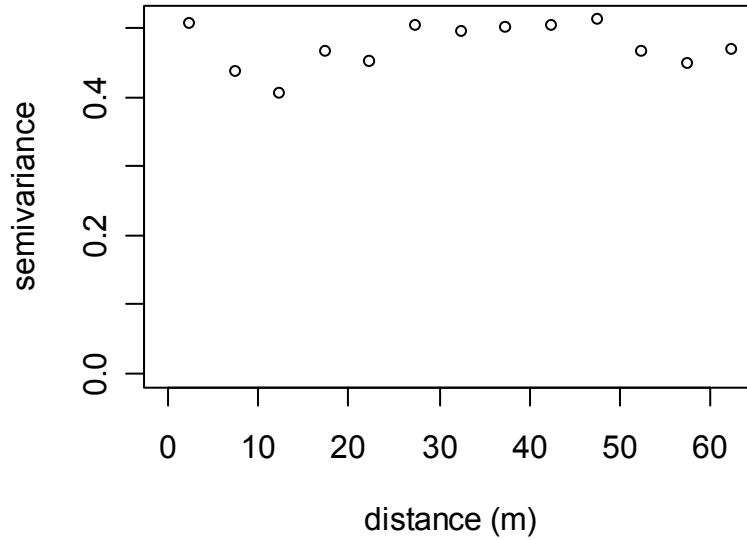


Fig.4: Variogram of the deviance residuals of the best GAM (h_3).

density surface modelling was $cv(\hat{N}_{DSM}) = 3.0\%$. The 95% confidence interval of the abundance of *P. nobilis* in the study area was 5,460 – 8,393 individuals. The distribution map of *P. nobilis* according to h_3 is given in Fig. 5. There was a marked zonation of *P. nobilis*, which was restricted in the shallow peripheral zone of the Lake. Most individuals were predicted to be in the strip between the 10 and 15 m contours, i.e. 3,316 individuals, which represented 49% of the total predicted population, while the area of the strip corresponds to ~7% of the total lake area. Only 31 individuals were predicted to be at depths >18 m and no individual at depths >22 m (hence limiting the study area at depths <30 m was sufficient to estimate *P. nobilis* abundance in the whole lake).

Comparison of DSM and traditional Distance Sampling

The point estimation of *P. nobilis* abundance with traditional distance sampling was only slightly less than that with DSM (Table 3). The coefficient of variation of the abundance estimate with traditional distance sampling was approximately double that with DSM (Table 3), as part of the unmodeled variability with the traditional approach was actually modeled with the DSM approach. Consequently, the estimated confidence interval with DSM was substantially narrower than with the traditional

Table 3: Comparison of the *Pinna nobilis* abundance estimations in Lake Vouliagmeni, based either on density surface modelling (DSM) or on traditional distance sampling (DS). In DS(1) uneven coverage probability distribution was not taken into account, while in DS(2) it was. cv: coefficient of variation, CI: confidence interval.

Abundance	Method		
	DSM	traditional DS (1)	traditional DS (2)
point estimation	6770	6254	5971
cv (%)	11.0	21.9	20.6
95% CI	5460 - 8393	3973 - 9269	3644 - 8463

approach (Table 3).

Discussion

Traditional distance sampling is founded on a combination of model-based and design-based inference. By modelling the detection function, model-based inference is drawn within the covered area. To extend that inference to the whole study area, traditional distance sampling relies on design-based methods (i.e. line transects are placed randomly ensuring that they are representative of the whole study area, so that the estimated density applies to the entire area and not just to the surveyed strips). In contrast, density surface modelling with GAMs employs model-based inference for abundance and density estimation in the study area. With DSM, abundance is related to spatial covariates and the distribution of the species may be visually depicted. The inferred relationships between population density and spatial or other ecologically meaningful covariates might have great biological significance and could be of more importance than simply abundance estimation. Density surface modelling and the inferred relations to spatial covariates might provide insights and act as the starting point for further ecological investigations and manipulative experiments seeking causal relationships between abundance and environmental covariates.

With a model-based approach like DSM, it is not required that the line transects are located according to a formal and restrictive survey sampling scheme. This is quite important, as in an underwater benthic survey like this one, it is not always easy or feasible to strictly follow a survey design that ensures uniform coverage probability. Post-hoc effort to take into account coverage probability might be quite difficult depending on the shape of the study area, and further approximations may be

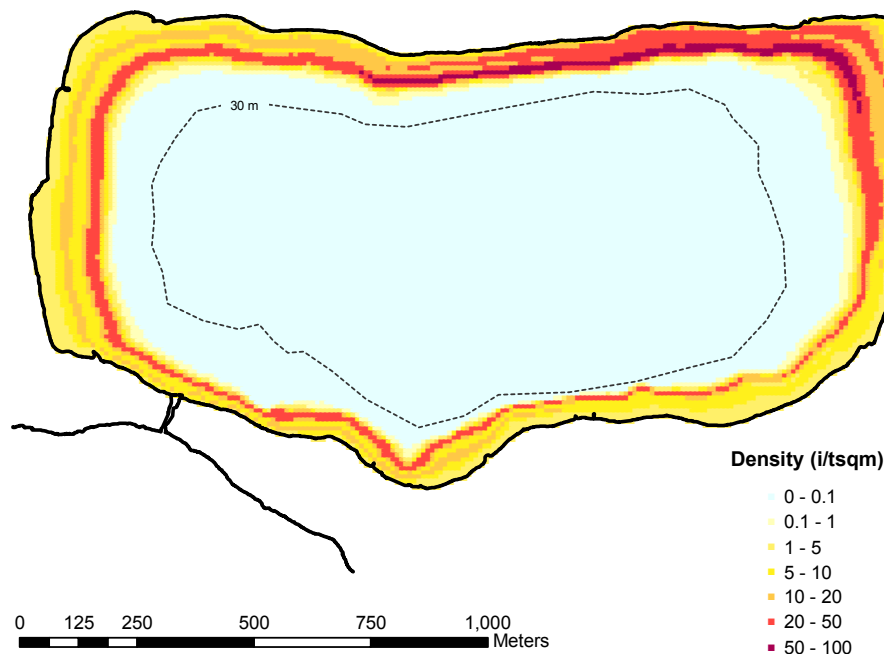


Fig. 5: Distribution map of *Pinna nobilis* in Lake Vouliagmeni. The actual study area was between the shoreline and the 30-m contour.

necessary. In this study, it was assumed that in each bathymetric zone the coverage probability was constant, which is an approximation, and the error in estimating areas A_z and lengths l_{qz} was ignored. Thus, the corresponding confidence interval of the abundance estimate based on traditional DS(2) was probably underestimated. Provided that the line transects give a good spatial coverage of the study area and the spatial model is a good approximation of reality, spatial models could offer more reliable estimation of abundance than traditional DS.

Another benefit of using a DSM approach is that abundance may be easily estimated in any subset of the study area by simply numerically integrating under the relevant section of the fitted density surface, while conventional DS restrict abundance estimations to the whole study area or to predefined (at the design stage) strata (Hedley et al. 2004).

DSM is expected to provide an improvement in the precision of the abundance estimate, as part of the spatial variability is actually modeled. Indeed, in the *P. nobilis* case in Lake Vouliagmeni, the cv of the abundance estimate under the DSM approach was nearly half of the cv under the traditional approach. However, it has to be noted that the performance of the techniques used in this study to estimate the variance of density and abundance estimates from the density surface model requires further investigation (Hedley et al. 2004). In addition, model selection in GAMs is an area of active research and several related issues (e.g. on the optimum degree of smoothing or the performance of GCV) are expected to improve in the future (Wood 2006).

Other possible methods for making spatial predictions from point measurements include geostatistical approaches, initially developed in Mining Geology (Journel and Huijbregts 1978). Geostatistical methods such as kriging have been proposed to assess the spatial structure of the populations of benthic species (e.g. Conan 1985) and have been recently applied to study commercially important species like e.g. the Norway lobster *Nephrops norvegicus* (Maynou et al. 1998). A main difference between the techniques presented here and geostatistical methodology is that with geostatistics density surfaces are estimated directly in geographic space rather than in predictor (environmental) space. With the present methodology, density surfaces may be fitted as a function of environmental predictors (which may include e.g. habitat type, sediment characteristics, depth, and every other variable of interest), and then the spatial pattern of the predictor surfaces is used to predict the response in geographic space. The main advantage is that, apart from a density map, relationships between density and the environment are described providing ecological insight. In addition, the present methodology is simpler to implement and very easy to be integrated with line transect sampling.

In the present study, a marked zonation of *P. nobilis* was observed, with a bimodal distribution with depth. Katsanevakis (2006) studied the bathymetric distribution of the species in the lake in 2004, by conducting 15 density measurements at the following depths 2, 4, 6, ..., 30 m, at each of 4 randomly chosen locations of the lake. A bimodal bathymetric distribution was found with one peak in the 12-m zone and another in the 4-m zone, exactly as in the present study, further supporting the validity of the GAM results. The complete absence of *P. nobilis* in areas deeper than 22 m might be because of the high silt content of the deeper areas (which may have negative effects on respiration and feeding), low year-round temperatures in deep areas (Katsanevakis 2006), or even the reduced food supply of deep areas that results from the intense summer thermocline. More research and experimentation is needed to interpret this restriction of the species to the shallow portion of the lake, as the result is contrary to other nearby areas in Korinthiakos Gulf, where *P. nobilis* is found

deeper than 50 m (personal observation). The reduced densities at depths < 10 m were mostly because of illegal fishing of large individuals, as deduced by monitoring a large number of marked individuals in two depth zones, at ~ 4 m and >10 m (unpublished data). Illegal fishing is conducted exclusively by skin diving, which due to high turbidity, is restricted to shallow areas. The second (minor) mode of the bathymetric distribution (at ~4 m) is mostly because of the increased recruitment of *P. nobilis* in shallow areas, especially at the north-east part of the lake (Katsanevakis 2006); those individuals are eventually fished after 2-3 years, when they have become larger (unpublished data). The reason for this increased recruitment in shallow areas (preferential settlement or decreased juvenile mortality) also requires further research.

New cohorts of *P. nobilis* settle during late summer to early autumn (Richardson et al. 1999), and given that the current study was conducted during mid summer, it was actually a pre-settlement survey. The youngest (and smallest) individuals had an age of ~ 9 to 11 mo and their mean shell length and width were ~ 10.6 and 4.3 cm respectively (Katsanevakis 2006), i.e. they were large enough to be easily detectable. This mostly explains why shell size was not found to have a substantial effect in detectability of *P. nobilis*, which might not hold true had the survey been conducted at a different time of the year.

It has to be emphasized that DSM (contrary to traditional DS) has provided valuable preliminary observations regarding *P. nobilis* distribution, which may encourage ideas and further investigation towards successfully revealing ecological patterns. The methodology used in this study is a promising and beneficial method for making abundance estimations of benthic fauna and should have wider use in the near future.

Acknowledgments

I would like to thank Steve Buckland, David Borchers, Len Thomas, Louise Burt, Tiago Marques, and especially Eric Rexstad for helpful discussions and comments on the data processing and density surface modelling during the 2006 Distance Sampling Workshops in CREEM. Eric Rexstad made helpful comments in the initial version of the manuscript. I also wish to acknowledge the suggestions and comments of two anonymous reviewers, which helped to improve the quality of the manuscript.

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csàaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, pp 267–281
- Buckland ST, Goudie IJB, Borchers DL (2000) Wildlife population assessment: past developments and future directions. *Biometrics* 56:1–12
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to Distance Sampling: Estimating abundance of biological populations. Oxford University Press, London
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, Berlin Heidelberg New York
- Burnham KP, Buckland ST, Laake JL, Borchers DL, Marques TA, Bishop JRB, Thomas L (2004) Further topics in distance sampling. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) Advanced Distance

- Sampling: Estimating abundance of biological populations. Oxford University Press, New York, pp 307–392
- Butler AJ, Vincente N, Gaulejac B de (1993) Ecology of the pteroid bivalves *Pinna bicolor* Gmelin and *Pinna nobilis* L. *Mar Life* 3:37–45
- Chapman CJ (1980) Ecology of juvenile and adult *Nephrops*. In: Cobb S, Philips B (eds) *The biology and Management of Lobsters*, vol II. Academic Press, New York, pp 143–178
- Conan GY (1985) Assessment of shellfish stocks by geostatistical techniques. ICES CM 1985/K:30
- Efron B, Tibshirani RJ (1993) *An introduction to the bootstrap*. Chapman and Hall, New-York
- Hall-Spencer JM, Froggia C, Atkinson RJA, Moore PG (1999) The impact of Rapido trawling for scallops, *Pecten jacobaeus* (L.), on the benthos of the Gulf of Venice. *ICES J Mar Sci* 56:111–124
- Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models*. Chapman and Hall, London
- Hedley SL (2000) *Modelling heterogeneity in cetacean surveys*. PhD thesis, University of St Andrews, UK
- Hedley SL, Buckland ST (2004) Spatial models for line transect sampling. *J Agr Biol Envir St* 9:181–199
- Hedley SL, Buckland ST, Borchers DL (2004) Spatial distance sampling models. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) *Advanced Distance Sampling: Estimating abundance of biological populations*. Oxford University Press, New York, pp 48-70
- Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. *Biometrika* 76:297–307
- Journel A, Huijbregts C (1978) *Mining geostatistics*. Academic Press, London
- Katsanevakis S (2005) Abundance and spatial distribution of the Mediterranean scallop, *Pecten jacobaeus*, in a marine lake. *Fish Res* 76:417–429
- Katsanevakis S (2006) Population ecology of the endangered fan mussel *Pinna nobilis* in a marine lake. *Endang Species Res* 1:51–59
- Katsanevakis S, Verriopoulos G (2004) Abundance of *Octopus vulgaris* on soft sediment. *Sci Mar* 68:553–560
- Kim YJ, Gu C (2004) Smoothing spline gaussian regression: more scalable computation via efficient approximation. *J Roy Stat Soc B* 66:337–356
- Marques FFC, Buckland ST (2004) Covariate models for the detection function. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) *Advanced Distance Sampling: Estimating abundance of biological populations*. Oxford University Press, New York, pp 31–47
- Maynou FX, Sardà F, Conan GY (1998) Assessment of the spatial structure and biomass evaluation of *Nephrops norvegicus* (L.) populations in the northwestern Mediterranean by geostatistics. *ICES J Mar Sci* 55:102–120
- McCullagh P, Nelder JA (1989) *Generalized Linear Models*, 2nd edn. Chapman and Hall, London
- McLoughlin RJ, Young PC, Martin RB, Parslow J (1991) The Australian scallop dredge: estimates of catching efficiency and associated indirect fishing mortality. *Fish Res* 11:1–24
- R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>

- Richardson CA, Kennedy H, Duarte CM, Kennedy DP, Proud SV (1999) Age and growth of the fan mussel *Pinna nobilis* from south-east Spanish Mediterranean seagrass (*Posidonia oceanica*) meadows. *Mar Biol* 133:205–212
- Seber GAF (1982) The estimation of animal abundance and related parameters, 2nd edn. Griffin, London.
- Shafee MS (1979) Underwater observation to estimate the density and spatial distribution of black scallop, *Chlamys varia* (L.) in Lanveoc (Bay of Brest). *Bull Off Natl Peches (Tunisie)* 3:143–156
- Thomas L, Laake JL, Rexstad E, Strindberg S, Marques FFC, Buckland ST, Borchers DL, Anderson DR, Burnham KP, Burt ML, Hedley SL, Pollard JH, Bishop JRB, Marques TA (2006) Distance 6.0. Release Beta 1. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. <http://www.ruwpa.st-and.ac.uk/distance/>
- Tuck ID, Chapman CJ, Atkinson RJA, Bailey N, Smith RSM (1997) A comparison of methods for stock assessment of the Norway lobster, *Nephrops norvegicus*, in the Fifth of Clyde. *Fish Res* 32:89–100
- Wood SN (2000) Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal statistical society, series B*, 62:413–428
- Wood SN (2006) Generalized Additive Models: An introduction with R. Chapman and Hall/CRC, Florida
- Zavodnik D, Hrs-Brenko M, Legac M (1991) Synopsis on the fan shell *Pinna nobilis* L. in the eastern Adriatic Sea. In: Boudouresque CF, Avon M, Gravez V (eds) *Les Espèces Marines à Protéger en Méditerranée*. GIS Posidonie publ., Marseille, pp 169–178