Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 193 (2006) 615–628
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts

P. Monestiez[a,c,*], L. Dubroca[b,c], E. Bonnin[c], J.-P. Durbec[c], C. Guinet[b]

[a] *Unite de Biométrie, INRA, Domaine St Paul, Site Agroparc, 84914 Avignon Cedex 9, France*
[b] *Centre d'Etudes Biologiques de Chizé, Centre National de la Recherche Scientifique, F-79360 Villiers en Bois, France*
[c] *Centre d'Océanologie de Marseille, Université de la Méditerranée, Campus de Luminy, Case 901, 13288 Marseille Cedex 9, France*

Received 12 October 2004; received in revised form 6 August 2005; accepted 31 August 2005
Available online 21 November 2005

Abstract

Obtaining accurate maps of relative abundance is an objective that may be difficult to achieve on the basis of spatially heterogeneous observation efforts and infrequent and sparse animal sightings. However, characterizing spatial distribution of wild animals such as fin whales is a major priority to protect these populations and to study their interactions with their environment. We have associated a geostatistical model with the Poisson distribution to model both spatial variation and discrete observation process. Assuming few weak hypotheses on the distribution of abundance, we have improved the experimental variogram estimate using weights that are derived from expected variances and proposed a bias correction that accounts for the variability added by the Poisson observation process. In the same way the kriging system was modified to interpolate directly the theoretical underlying animal abundance better than noisy observations from count data. For cumulative count data of fin whales over the summers 1993–2001, the method gave a map of the relative abundance which is informative on the spatial patterns. Kriging interpolation variances were dramatically reduced – ratio from 0.015 to 0.26 – compared to usual Ordinary Kriging on raw data. Adding the hypothesis of stationarity over time the variogram estimated on cumulative data can be then used with more sparser annual data.

Keywords: Relative abundance map; Fin whale; Sightings data; Geostatistics; Kriging; Variogram estimation; Poisson distribution; Bias correction

1. Introduction

In ecology, knowledge of the spatial distribution of an animal population and characterization of its spatial structure are essential for ecologists to better understand the population's interaction with its environment,

* Corresponding author. Tel.: +33 4 32722187; fax: +33 4 32722182.
E-mail address: monestiez@avignon.inra.fr (P. Monestiez).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.08.042

especially when this environment is particularly variable and spatially structured (Wiens, 1989). Mapping spatial relative abundance of animals that cannot be captured, for example protected or marine species, is generally done from surveys where count data are collected by observers ideally following a sampling design (Forcada et al., 1996; Canadas et al., 2002; Gannier, 2002).

Processing of observational data is facilitated by geographical information system (GIS) software that allows the pooling of different sources of data and contain generally basic methods for mapping raw data (Burrough and McDonnell, 1998). More recent ones include standard geostatistical methods (Burrough, 2001), but the methodological difficulties that arise from the nature of count data are simply ignored on the grounds that linear geostatistics is distribution free (see for example Maynou et al., 1998; van der Meer and Leopold, 1995). However, more and more studies in ecology use geostatistics to map environmental characteristics (Webster and Oliver, 2001), as well as population relative abundance (Petitgas, 2001; Rivoirard and Wieland, 2001; Rufino et al., 2004; Huguet et al., 2005).

In the Mediterranean Sea, the fin whale (*Balaenoptera physalus*, Linné 1758) is the largest marine predator commonly observed (Notarbartolo di Sciara et al., 2003). Several thousand individuals were estimated to be present in the western Mediterranean Sea during summer (Forcada et al., 1995, 1996). According to Gannier (1997), several hundreds (about 800) are thought to be commonly present in the northern part of the basin in summer with an estimation of a global whale density of 0.0155 individual km^{-2}.

Precise mapping of the spatial variation of whale density with such low abundances requires the collection of a large number of observations, covering sufficiently well the area of interest. This has only been possible by pooling data from several sources during several summers. Then observations were summed on spatial cells sufficiently large to accumulate enough observations and small enough to get a good spatial resolution on abundance maps (Littaye et al., 2004; Rendell et al., 2004). However, observations remain heterogeneous with a heavy tailed distribution pattern and count distribution appears as zero-inflated.

When sampling is very heterogeneous it is necessary to divide the observed counts by the observational effort in order to map the whale distribution (Littaye et al., 2004), but this correction amplifies the asymmetry of the raw data distribution and generates very heavy tailed distribution patterns by dividing some isolated whale counts by observation times that may sometimes be really brief.

Traditionally, such distributional characteristics are particularly difficult to deal with in geostatistical procedures leading to poor variogram estimates close to a pure "nugget effect", i.e. without spatial correlation and to kriging with strong smoothing and wide predicted kriging variances. The resulting map is not biased or erroneous but poor simply because the prediction variances are too wide. The variogram, and in consequence the kriging variances, sums different sources of variability; an inhomogeneous effort distribution, a sighting process acting as a noise – although it can be naturally modelled by a Poisson distribution – and the spatial variation of the whale relative abundance. Ideally it would be preferable to use a variogram reduced to the variation of the relative abundance and to filter all other sources of noise.

A first approach to address spatial Poisson distribution was introduced by Kaiser and Cressie (1997) within the general framework of spatial autoregressive models on regular lattice and developed in the case of zero-inflated models (Agarwal et al., 2002). These models and methods generally imply Bayesian approaches with Markov Chain Monte Carlo (MCMC) for posterior distribution sampling or specific development (Griffith, 2002) due to spatial dependence. However, the spatial autoregressive framework is not well suited for irregular sampling and for interpolation objectives.

Another solution to distribution difficulties was proposed by Diggle et al. (1998) using what they called "model-based kriging", which is not exactly kriging in that they used a Bayesian framework taking into account a Poisson observation process to predict the underlying spatial distribution as a posterior random field distribution. The method is rather computer intensive and requires skill in running MCMC algorithms. With Rongelap Island radiometric data, Diggle et al. (1998) gave an example where data were rather uniformly sampled and where distributions tend to normal shape (Poisson distribution with mean 3000 and no zeros). Christensen and Waagepetersen (2002) improved the method and applied it to weed data. Wikle

Table 1
Characteristics and contribution of different categories of survey platforms

Boat category	Length range (m)	Observation height (m)	Travelling speed (knots)	Hour of observation	% of total effort
Sail boat	11–30	2–3	5–7	2169.2	65
Fish survey vessels	20–67	5–7	8–10	1000.4	30
Ferry	100–130	13–17	14–16	167.4	5

(2002) gave another example in ecology with breeding bird survey data in the USA. Although this approach seems relevant, it remains difficult for ecologists to easily implement it and for a non-statistician to interpret the results and appreciate their reliability within a Bayesian framework where you have to summarize and interpret posterior distributions of interpolated random fields with posterior distributions of parameters.

Our aim in this paper is to provide population biologists with a simple mean to include a Poisson modelling of the observation process into a classical geostatistical approach and thus to improve both variogram estimation and kriging accuracy.

2. Data

The fin whale sightings database used in this study merges data from the Centre de Recherche sur les Mammifères Marins (CRMM), CETUS, the Commission Internationale pour l'Exploration Scientifique de la mer Méditerranée (CIESM), Conservation Information Recherche sur les Cétacés (CIRCé), Delphinia Sea Conservation, the Ecole Pratique des Hautes Etudes (EPHE) (David, 2000; Di Miglio, 1999), the Groupe d'Etude des Cétacés de Méditerranée (GECEM), the Groupe de Recherche sur les Cétacés (GREC), the Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), the Musée Océanographique de Monaco, the Réserve Internationale en Mer Méditerranée Occidentale (RIMMO), the Supreme Allied Commander Atlantic Undersea Research Centre (SACLANTCEN), the Société de Navigation Corse Méditerranée (SNCM), the Swiss Cetacean Society (SCS) and the World Wildlife Fund (WWF, France). Available sightings data cover the period 1993–2001. Only data for which observation effort could be quantified were used.

Data were collected using three categories of opportunity survey platforms: whale counts from sailing boats, fish survey vessels and ferries. The characteristics of each category and the respective contribution of each category in terms of survey effort are summarized in Table 1. Sampling strategies were variable, some boats, such as ferries and part of the survey effort may have been non-random as these boats were travelling along predefined transects or trajectories. Conversely, photo-identification focused surveys or ecotourism surveys followed a more erratic movement pattern.

The observations were conducted only under fine weather conditions (0–3 on the Beaufort scale). Only 10% of the observations were opportunistic (i.e. observer looking continuously for whales but in all directions), the rest were done by 1–3 dedicated observers continuously watching (naked-eye) over 180° ahead of the vessel and switching every 2 h. A global positioning system (GPS) recorded the vessels' route and any change in the observation conditions (effort interruption or change in sea conditions) was recorded.

In this study we only considered the number of fin whale sightings and not fin whale numbers as group size could not be estimated properly from ferries when whales dived as some whales were likely to be still underwater and remain undetected during the ferry's passage. Sighting number was not linked with the number of observers aboard (Kendall's rank correlation test, $z = -1.6049$, $p = 0.1085$) or with the type of observation platform used (Kendall's rank correlation test, $z = 0.0991$, $p = 0.921$). Fig. 1 shows the whole zone where data were available.

All the count data were then aggregated on cells of 0.1° of longitude by 0.1° of latitude (approximately 90 km^2) in a regular grid. For each year, July and August data were assembled and we computed in each cell the total number of fin whale sightings; the observation effort was defined as the total time (in hours) spent observing in a cell. Finally the grid data was summed over the years, leading to a cumulated index over years expressed in number of sightings per unit effort.

Given the extensive spatial range of the data, whose limits are given in Fig. 1, and with the special in-

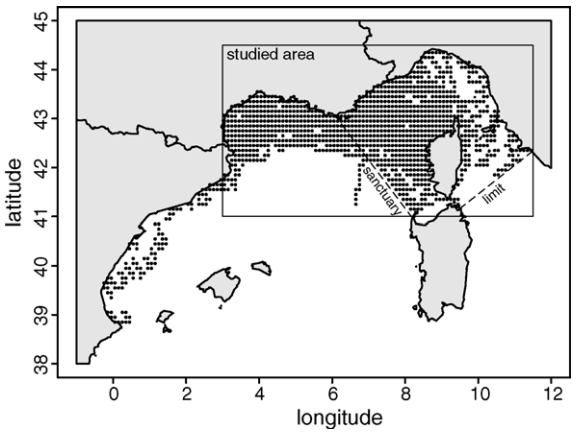

Fig. 1. Map of all available observation data. Each cell of $0.1^\circ \times 0.1^\circ$ where the sighting time was strictly positive is marked by a symbol at its center.

terest for the International Cetacean Sanctuary of the Mediterranean (Gannier, 2002), established on 25 November 1999 by the Governments of Italy, France and Monaco, we focused on a smaller area that extends from 3°E to 11.5°E and from 41°N to 44.5°N. In fact, data along the Spanish coast were only collected in 1993 and no fin whales were observed between the Spanish coast and Balearic Islands. Reducing the study to the northern part of Western Mediterranean Sea corresponds to a loss of less than 1% of the total sighting time.

Table 2 summarizes per year the observation effort and overall sightings results. Observation efforts are given in number N of cells with positive observation time and in total duration to characterize respectively how the effort was spatially spread over the area of interest and how long it was sustained. The third and fourth lines give the number of sighting events and the number of cells where some fin whales were seen. There is a high proportion of cells with zero sighting event, around 80%. So rare sightings will be generally surrounded by zeros even when their spatial concentration varies.

Observation time and fin whale count maps are given in Fig. 2. Note that spatial coverage may be very different from 1 year to another and for some years may be very sparse for some areas (Fig. 5). Conversely some areas near the Riviera coast or along the main shipping routes to Corsica are most often oversampled (Fig. 2a). Consequently, the simple map of untransformed data (Fig. 2b) reveals the fin whale occurrence as well as the observation effort density.

3. Model and methods

3.1. Model

For all sites s belonging to domain $\mathcal{D}$, we define the model for the random field $Z(s)$ by

$$Z(s)|Y(s) \sim \mathcal{P}(t(s)Y(s)) \quad (1)$$

where $Z(s)|Y(s)$ is Poisson distributed with an intensity parameter that is the product of $t(s)$ by $Y(s)$, $t(s)$ the observation time (in hours) at site s, $Y(s)$ is proportional to the animal abundance at site s and measures the expectation of sightings for a unit observation time and $Y(s)$ is a positive random field honouring order two stationarity, with mean m, variance σ_Y^2, and covariance function $C_Y(s-s')$ which depends only on the distance between s and s' noted $(s-s')$. The covariance function $C_Y(s-s') = \mathrm{Cov}[Y(s), Y(s')]$ may be replaced by the variogram function $\gamma_Y(s-s') = \frac{1}{2}\mathrm{E}[(Y(s)-Y(s'))^2]$. There is no additional hypothesis on the Y distribution

Table 2
Count data and observational effort per year: N gives per year the number of map cells with observation effort, t gives the total duration of sighting activity. R denotes the number of groups of fin whales seen during this time and N_R the number of cells with at least one fin whale group seen. The last column gives the cumulative total over the 1993–2001 period. Time is given in hours

	Year									Total
	1993	1994	1995	1996	1997	1998	1999	2000	2001	
N	303	363	475	542	282	228	300	400	327	1113
t	136.6	293.7	622.7	470.8	323.7	225.9	457.3	733.4	219.8	3337
R	1	62	160	79	32	18	69	32	37	490
N_R	1	50	106	61	22	17	48	23	33	249

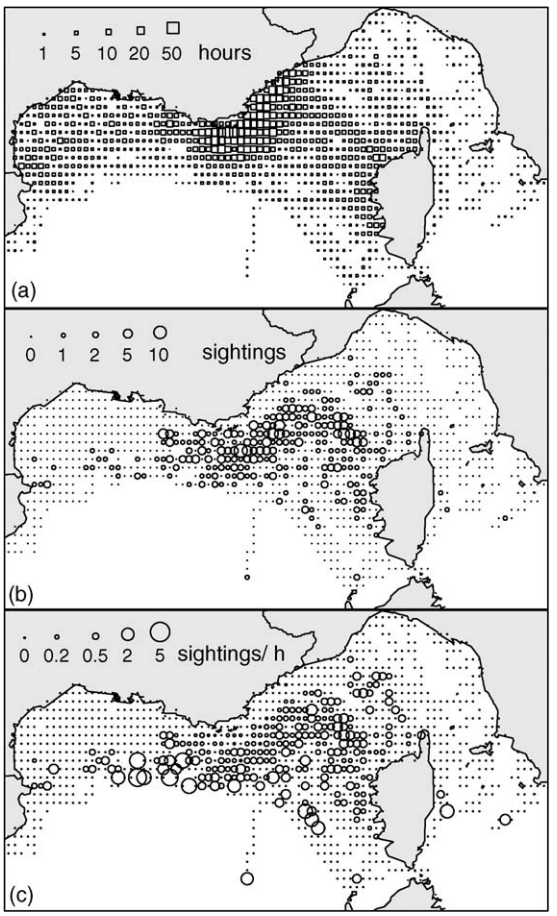

Fig. 2. Maps of raw data: (a) map of observation time, symbol area is proportional to observation time in hours, (b) map of original sightings data, symbol indicates the number of fin whale groups seen independently from the observation effort and (c) map of mean number of sightings per hour of observation effort.

except the inequality $Y \geq 0$.

Classically, to simplify notations, $Z(s)$, $Y(s)$ and $t(s)$ will be denoted in the following as Z_s, Y_s and t_s, respectively. In kriging systems, $C_{ss'}$ denotes the covariance function $C_Y(s-s')$.

3.2. Expectation and variances of Z_s

As $Z(s)|Y_s \sim \mathcal{P}(t_s Y_s)$ it follows directly that:

$$\mathrm{E}[Z_s|Y_s] = t_s Y_s, \qquad \mathrm{E}[Z_s] = t_s \mathrm{E}[Y_s] = m t_s,$$
$$\mathrm{Var}[Z_s|Y_s] = t_s Y_s, \qquad \mathrm{E}[(Z_s)^2|Y_s] = t_s Y_s + (t_s Y_s)^2,$$
$$\mathrm{Var}[Z_s] = t_s^2 \sigma_Y^2 + m t_s \tag{2}$$

For the covariance expression, the conditional independence of observations at different sites leads to:

$$\begin{aligned}\mathrm{E}[Z_s Z_{s'}|Y] &= \mathrm{Cov}[Z_s, Z_{s'}|Y] + \mathrm{E}[Z_s|Y_s]\mathrm{E}[Z_{s'}|Y_{s'}] \\ &= \delta_{ss'} t_s Y_s + t_s t_{s'} Y_s Y_{s'} \end{aligned} \tag{3}$$

where $\delta_{ss'}$ is the Kronecker delta which is 1 if $s = s'$ and 0 otherwise.

3.3. Expectation and variances of $(Z_s/t_s - Z_{s'}/t_{s'})$

In order to characterize the relationship between the variograms of Z and Y, we develop in Appendix A the expressions of the two first moments of $(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}})$. It results that:

$$\mathrm{E}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right] = 0$$

and

$$\begin{aligned}&\frac{1}{2}\mathrm{E}\left[\left(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right)^2\right] \\ &\quad = \frac{m}{2}\left(\frac{t_s + t_{s'}}{t_s t_{s'}}\right) - \delta_{ss'}\frac{m}{t_s} + \gamma_Y(s-s')\end{aligned} \tag{4}$$

Let $\gamma_Z(s-s')$ denote the non-stationary theoretical variogram corresponding to the random field (Z_s/t_s), we get for $s \neq s'$ the relationship:

$$\gamma_Y(s-s') = \gamma_Z(s-s') - \frac{m}{2}\left(\frac{t_s + t_{s'}}{t_s t_{s'}}\right) \tag{5}$$

We can check for $s = s'$ that Eq. (4) reduces to $\gamma_Y(0) = \gamma_Z(0) = 0$.

Furthermore, the conditional variance and its expectation are (Appendix A)

$$\mathrm{Var}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\middle| Y\right] = \frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}},$$
$$\begin{aligned}&\mathrm{E}\left[\mathrm{Var}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\middle| Y\right]\right] = \mathrm{E}\left[\frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}}\right] \\ &\quad = m\left(\frac{t_s + t_{s'}}{t_s t_{s'}}\right)\end{aligned} \tag{6}$$

3.4. Estimation of variogram $\gamma_Y(h)$

Let Z_α, $\alpha = 1, \ldots, n$ be the n measurements of $Z(s_\alpha)$ obtained during observation times t_α. The expression of a modified estimate of the variogram can be derived from (5) and (6).

$$\gamma_Y^*(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \frac{t_\alpha t_\beta}{t_\alpha + t_\beta} \left[\frac{1}{2} \left(\frac{Z_\alpha}{t_\alpha} - \frac{Z_\beta}{t_\beta} \right)^2 - \frac{m^*}{2} \left(\frac{t_\alpha + t_\beta}{t_\alpha t_\beta} \right) \right] \mathbb{I}_{d_{\alpha\beta} \sim h} \tag{7}$$

where $\mathbb{I}_{d_{\alpha\beta}\sim h}$ is the indicator function of pairs (s_α, s_β) whose distance is close to h, $N(h) = \sum_{\alpha,\beta} \frac{t_\alpha t_\beta}{t_\alpha + t_\beta} \mathbb{I}_{d_{\alpha\beta}\sim h}$ is a normalizing constant and m^* is an estimate of the mean of Y.

The bias correction term $-\frac{m^*}{2}\left(\frac{t_s + t_{s'}}{t_s t_{s'}}\right)$ derives directly from (5).

The weights $\frac{t_s t_{s'}}{t_s + t_{s'}}$ are introduced to homogenize the variance of differences terms $\left(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right)$ by dividing them by a weight proportional to their standard deviation $\sqrt{m \frac{t_s + t_{s'}}{t_s t_{s'}}}$ given by (6).

When simplifying (7) we get for $h \neq 0$:

$$\gamma_Y^*(h) = \frac{1}{2N(h)} \times \sum_{\alpha,\beta} \left(\frac{t_\alpha t_\beta}{t_\alpha + t_\beta} \left(\frac{Z_\alpha}{t_\alpha} - \frac{Z_\beta}{t_\beta} \right)^2 - m^* \right) \mathbb{I}_{d_{\alpha\beta}\sim h} \tag{8}$$

3.5. Kriging of Y_o

The kriging of Y_o at any site $s_o \in \mathcal{D}$ is a linear predictor combining the observed data Z_α weighted by observation times t_α.

$$Y_o^* = \sum_{\alpha=1}^{n} \lambda_\alpha \frac{Z_\alpha}{t_\alpha} \tag{9}$$

In order to assure its unbiasedness, we compute its expectation:

$$\mathrm{E}[Y_o^* | Y] = \sum_{\alpha=1}^{n} \frac{\lambda_\alpha}{t_\alpha} \mathrm{E}[Z_\alpha | Y_\alpha] = \sum_{\alpha=1}^{n} \frac{\lambda_\alpha}{t_\alpha} t_\alpha Y_\alpha = \sum_{\alpha=1}^{n} \lambda_\alpha Y_\alpha,$$

$$\mathrm{E}[Y_o^*] = \sum_{\alpha=1}^{n} \lambda_\alpha \mathrm{E}[Y_\alpha] = m \sum_{\alpha=1}^{n} \lambda_\alpha$$

So the condition for unbiasedness is the usual one:

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \tag{10}$$

In the same way the variance of the error of prediction, i.e. the mean square error of prediction – MSEP – if unbiased, can be derived by application of expressions (2) and (3) to the kriging estimate (see Appendix B for details).

$$\mathrm{E}[(Y_o^* - Y_o)^2 | Y] = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta Y_\alpha Y_\beta + \sum_{\alpha=1}^{n} \frac{\lambda_\alpha^2}{t_\alpha} Y_\alpha + Y_o^2 - 2 Y_o \sum_{\alpha=1}^{n} \lambda_\alpha Y_\alpha$$

Then, when deconditioning:

$$\mathrm{E}[(Y_o^* - Y_o)^2] = \sigma_Y^2 + \sum_{\alpha=1}^{n} \frac{\lambda_\alpha^2}{t_\alpha} m + \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C_{\alpha\beta} - 2 \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} \tag{11}$$

From the unbiasedness condition (10), we have $\mathrm{Var}(Y_o^* - Y_o) = \mathrm{E}[(Y_o^* - Y_o)^2]$.

By minimizing this expression (11) on λ_i's with the unbiasedness constraint, we obtain the following Kriging system of $(n + 1)$ equations where μ is the Lagrange multiplier

$$\sum_{\beta=1}^{n} \lambda_\beta C_{\alpha\beta} + \lambda_\alpha \frac{m}{t_\alpha} + \mu = C_{\alpha o} \quad \text{for } \alpha = 1, \ldots, n,$$

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \tag{12}$$

The kriging system is often expressed with covariance $C_{\alpha\beta}$ and not variogram. From the fitted variogram

model, covariances derive from the classical relation $C(s-s') = \sigma_Y^2 - \gamma_Y(s-s')$. The kriging system may also be expressed using variogram (Chilès and Delfiner, 1999).

The expression of the minimized prediction variance resulting from this kriging system reduced after calculation to the same expression as for classical Ordinary Kriging (Appendix B)

$$\mathrm{Var}(Y_o^* - Y_o) = \sigma_Y^2 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} - \mu \qquad (13)$$

Although the expression is unchanged, kriging variances may be very different in our case because the λ_α solution of the kriging system are modified as well as the variance value σ_Y^2.

4. Results

4.1. Experimental variogram on cumulated data

We first checked that no directional effect was present in the experimental variogram on the whole data set collected over 1993–2001. We then computed the corrected experimental variogram defined by Eq. (8) under the assumption of isotropy. Fig. 3 clearly shows the effect of both corrections, data weighting and bias. Weights on pairs lead to a smoother experimental variogram with a lower sill (Fig. 3, curve b). Extreme values of differences on some pairs resulted from single whale sightings that occurred during short sighting times and the weighting system on square differences considerably reduced their importance. After application of bias correction we can guess that the underlying variable Y variogram does not exhibit nugget effect (Fig. 3, curve c) and conclude in the relative smoothness of the spatial whale abundance. A stable variogram model (Eq. 14) defined in Chilès and Delfiner (1999) was thus fitted by least squares:

$$\gamma_Y(h) = c\left(1 - \exp\left(-\left(\frac{h}{a}\right)^d\right)\right) \qquad (14)$$

with parameters $c = 0.043$, $a = 28.4$ and $d = 1.51$ (Fig. 3, curve c). The fitted model with parameter $d = 1.51$ is intermediate in smoothness between an exponential variogram ($d = 1$) and a Gaussian variogram ($d = 2$).

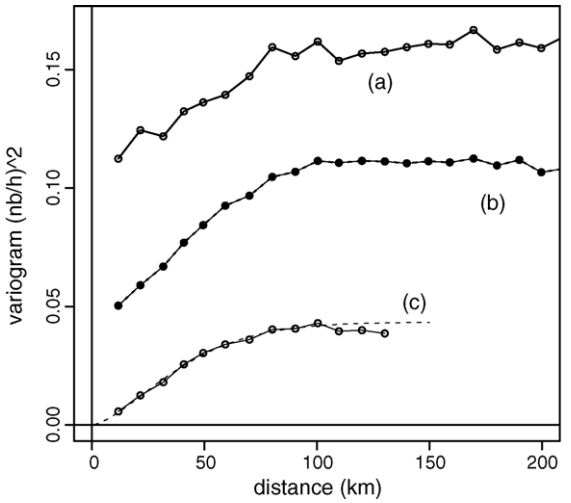

Fig. 3. Experimental variograms on corrected data over the period 1993–2001. Y-axis unit is square counts/h. (a) Classic experimental variogram computed on variable Z_α/t_α, (b) same experimental variogram when applying the weights on pairs and (c) experimental variogram from Eq. (8) including mean correction and fitted stable variogram model (dashed line).

4.2. Kriging on the 1993–2001 period data

We defined a grid for prediction of Y-values that reproduced the sample grid. The prediction grid with elementary cell of $0.1^\circ \times 0.1^\circ$ also extends from 3°E to 11.5°E and from 41°N to 44.5°N. All points beyond the coastline were removed so we were left with 2020 points to krige. Among these points, 1113 were located in a cell with observed Z and t. We used a single neighborhood for kriging, keeping all points for all predictions in establishing the kriging system (12).

Fig. 4 (left) gives the kriged map of the expectation of whale sightings per hour, that we compare with results obtained by standard Ordinary Kriging on raw data Z_s/t_s. Overall patterns are similar but our method seems to be less influenced by larger data values and exhibits more continuous spatial patterns than ordinary kriging.

Conversely, the two methods gave completely different variance maps (Fig. 4, right). The kriging variances are rather flat in the central region where data and prediction grid overlap. The variances with the Pois-

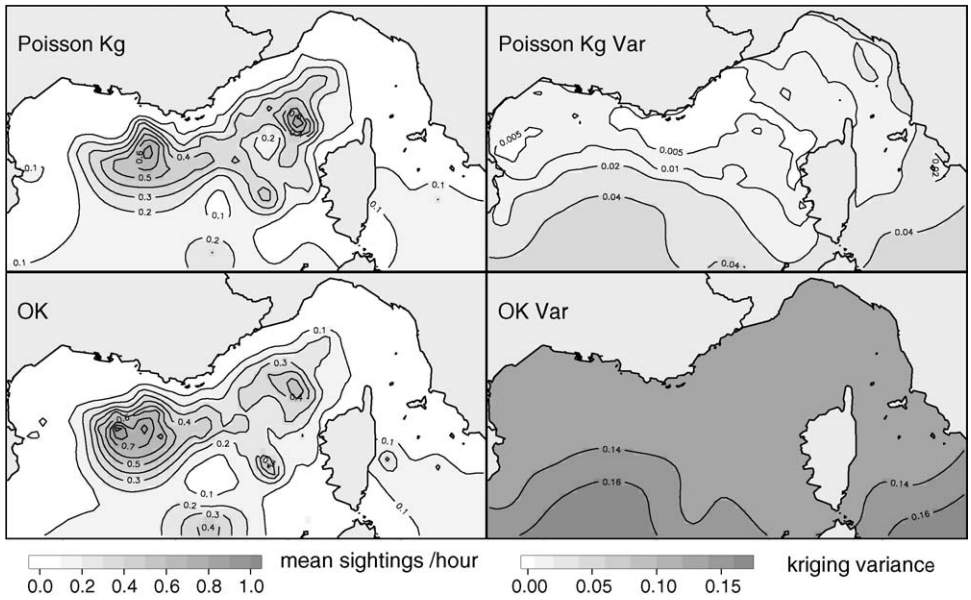

Fig. 4. Comparison of krigings. Maps of kriged values (expected number of whale sightings per hour, left) and kriging variance (square mean number of whale sightings per hour, right) for our modified kriging of Y (noted Poisson Kg, top) and for standard Ordinary Kriging of Z/t (noted OK, bottom).

son modelling are substantially less than with Ordinary Kriging, and are modulated by the observation effort. The ratio between kriging variances ranges from 0.015 for the cell with the highest observational effort to 0.26 for the area with no data around. The kriging variance does not depend solely on the geometrical configuration of data, but is rather proportional to the inverse of the observation effort when available. This results from the modification on the kriging system that filters the variance of the observation process from the variance of the spatial whale distribution itself.

4.3. Kriging annual maps

With a hypothesis of stationarity in time from 1 year to the next, we can assume that the variogram model is the same for yearly data and for cumulative data. So annual experimental variograms that would not have been correctly estimated because of a low amount of data (cells with positive observation effort) and an extremely low number of actual non-null counts, is simply modelled by the variogram fitted on cumulative data. Three years, 1994, 1995 and 2001 were taken as example and respective count-per-hour data and kriging results are given in Fig. 5. Because years 1994 and 1995 differ a lot from 2001 in their observation effort, we added annual observation effort maps and the associated kriging variance maps which take them into account (Fig. 6). Even when we limit the comparison to regions with low kriging variance for all 3 years, some changes in the spatial distribution patterns of fin whales can be noted from 1 year to another.

5. Discussion and conclusions

The proposed mapping method which is a kriging written specifically for the Poisson distribution case is no more difficult to implement than Ordinary Kriging. However, standard software must be modified, which is quite easy with some statistical software such as R (R Development Core Team, 2004) as done by the authors, but can be more problematic with some programming plugged in GIS commercial software or with user-oriented packages (Ribeiro and Diggle, 2001).

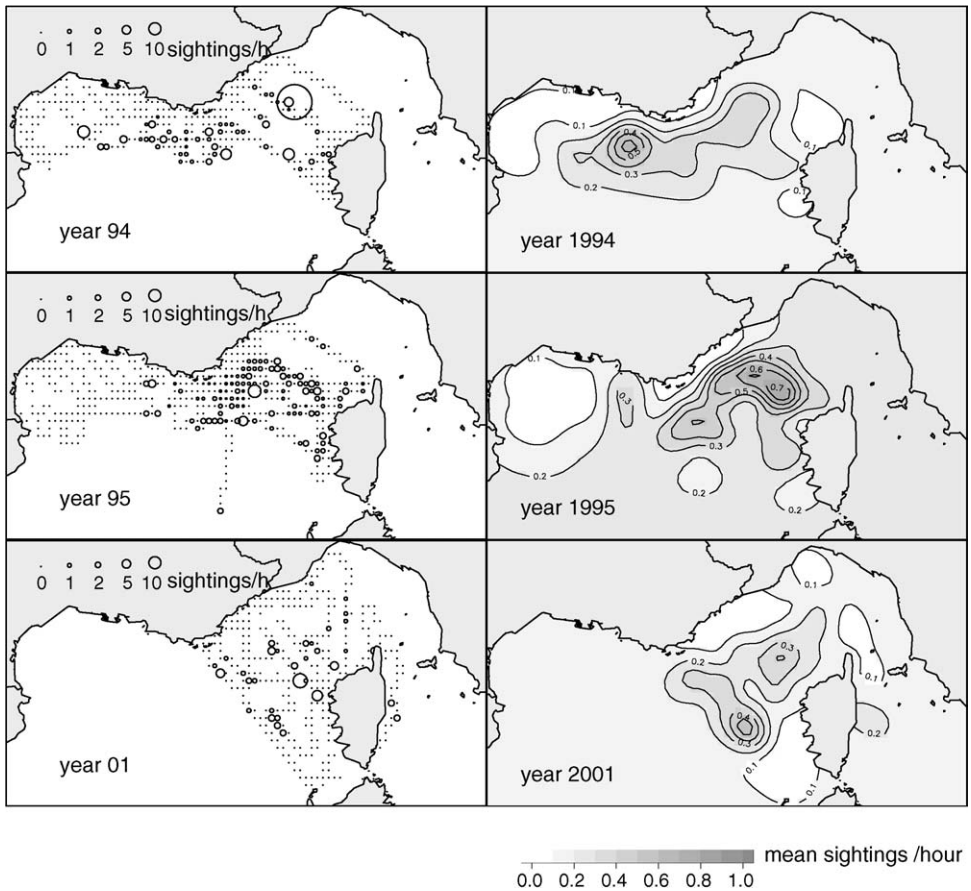

Fig. 5. Maps of sighting data (mean sightings per hour, left) and kriged expectation of whale sightings (expected number of sightings per hour) for years 1994, 1995 and 2001.

The variogram correction and the kriging based on this ad hoc modelling gave maps that were suited to ecologist's needs and consistent with other studies (Forcada et al., 1996; Gannier, 2002). The gain in variance in the error of prediction is substantial compared to Ordinary Kriging. The fin whale data set was quite extreme considering the heterogeneity of observation effort and the very low values for the expected density – the parameter of the Poisson distribution – so we believe that the proposed method will be able to give equivalent results in many other applications in ecological surveys.

One advantage of the proposed method compared to the model-based geostatistics proposed by Diggle et al. (1998) consists in the absence of hypothesis on the distribution pattern of the underlying animal density. There are no natural reasons for it to follow a log-normal random field as required in the Diggle et al. (1998) or Wikle (2002) approaches. Conversely, the drawback of our model's free spatial density is the possibility of obtaining negative mapped values. For fin whales this happened exceptionally in regions of lowest density with negative predictions whose absolute values were negligible. A simple way to solve the difficulty was to set at zero the rare negative predictions, so the MSEP was reduced overall. However, if such negative predictions based on positive data become more frequent, this could suggest that the chosen variogram model is wrong, especially for short distances.

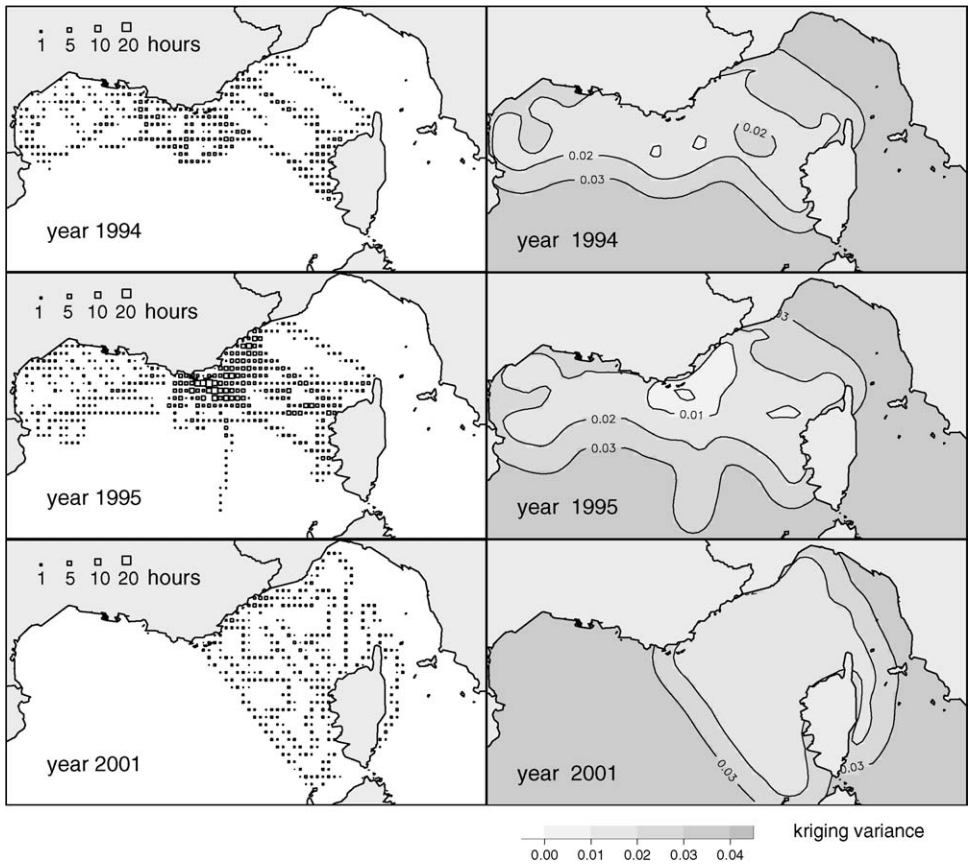

Fig. 6. Maps of observation effort (hours, left) and kriging variance (square mean number of whale sightings per hour, right) for years 1994, 1995 and 2001.

Another problem introduced by kriging is its behavior when no more data are available in a nearby neighborhood. Standard ordinary kriging gives the overall mean as prediction with an error variance equal to the data variance. Our modified kriging behaved the same way and gave prediction for southern parts of the map that are close to the annual mean with an error variance based on the fitted variogram γ_Y^*. This also explains why extrapolated areas without close data are not mapped with the same value for the different years in Fig. 5. For annual maps with sampling that spatially varies among years, we could decide that the estimated mean for 1 year may be biased or too imprecise. An alternative would be to use the mean of Y estimated from the all-year data set in each annual kriging system.

The question of stationarity must also be addressed. Kriging methods are generally robust. Discrepancies from strict stationarity do not significantly change predicted maps and only variances of prediction may be affected. In our case it is different because we use the estimated mean for correction at two stages of the method. We checked that addition or suppression of areas which are probably a poor habitat for fin whales, such as, for example, the Gulf of Lions or the eastern part of the Gulf of Genova, do not lead to systematic bias due to a overall non-stationarity. Adding areas with quasi-exclusively null count data lowers both the estimated mean used for bias correction and the experimental variogram values on raw data, so the corrected variogram remains quite unchanged and in consequence produces locally very close kriging estimates and variances. Our

method seems to be as robust as the standard Ordinary Kriging.

We considered sighting event counts rather than fin whale counts mainly because the group size estimate was not always reliable but also for consistency with the Poisson model. Groups were not so frequent but when groups were observed, whales were obviously travelling or feeding together (Notarbartolo di Sciara et al., 2003) invalidating the Poisson hypothesis of independence between individual sightings. So we chose to consider groups of fin whales and we then assumed that the observations of groups were mutually independent.

A last point about whale density maps is the proportionality between maps of relative abundance and our maps of sightings expectation. The maps we obtained are well-suited for interpretation of spatial patterns or for potential whale surveys, but are not directly useful for quantifying total abundance. To convert our kriged map, which is expressed in whale sightings per hour, to an abundance map that will be expressed in individuals by km^2 we need accurate estimates of two multiplicative coefficients: the first is the mean area covered per hour of observation which depends only on the observation process, the second is the mean size of observed groups of whales which is slightly greater than 1. With values ranging from 4 to 9 km of transect per hour and with 1.6 whales per group, the map of relative abundances in Fig. 4 is consistent with the data of Gannier (2002) who gave a summer abundance in the basin as ranging from 0.6 to 9.0 whales/100 km of transect over the period 1991–2000 (5.7 whales/100 km of transect for our dataset over the period 1993–2001). To match the overall whale density of 0.0155 individuals m^{-2} given by Gannier (1997), the first coefficient should vary from 8 to 10 km^2 observed per hour, which is uncheckable with our dataset.

Can our method be generalized? Previous works by Oliver et al. (1993) proposed adapting classical kriging methods to binomial data in the context of epidemiological studies, for example, estimation of the risk of childhood cancer. The difficulties encountered were similar; very low frequencies, one half of zeros, skewed distribution and heterogeneity of observation weights—here the binomial n parameter which plays the same role as our observation effort t. They proposed a bias correction term on the experimental variogram – this term included the weights and the binomial variance form – and a modified kriging system specific to the binomial but based on the same reasoning, which was what we did in the Poisson case. The resulting maps and the associated kriging variances were clearly improved (Oliver et al., 1998). However they did not introduce weights to homogenize variances of each squared difference in the experimental variogram. Although the expression of this last correction term has to be checked in the binomial case, it is probable that their experimental variogram estimate can be further improved.

Beyond first comparison simply based on advantages or drawbacks between model-based geostatistics proposed by Diggle et al. (1998) and the Poisson model kriging we propose, a detailed comparison on large data sets for different conditions would appear to be necessary. Another perspective is to extend the present methodology to multivariate geostatistics (Wackernagel, 1995) to address multi-specific spatial distribution as in Rueda and Defeo (2003).

Acknowledgements

We wish to thank all the people and associations who participated in the sightings data collection and collation: the Centre de Recherche sur les Mammifères Marins (CRMM), the CETUS, the Commission Internationale pour l'Exploration Scientifique de la mer Méditerranée (CIESM), Conservation Information Recherche sur les Cétacés (CIRCé), Delphinia Sea Conservation, the people who conducted cetacean surveys under the program lead bt P.C. Beaubrun, Laboratoire de Biogéographie et Ecologie des Vertébrés (BEV) at the Ecole Pratique des Hautes Etudes (EPHE) particularly L. David, N. Di-Méglio and A. Gannier, the Groupe d'Etude des Cétacés de Méditerranée (GECEM), the Groupe de Recherche sur les Cétacés (GREC) and particularly A. Gannier, the Institut Français de Recherche pour l' Exploitation de la Mer (IFREMER), the Musée Océanographique de Monaco, the Réserve Internationale en Mer Méditerranée Occidentale (RIMMO), the Supreme Allied Commander Atlantic Undersea Research Centre (SACLANTCEN), the Société de Navigation Corse Méditerranée (SNCM), the Swiss Cetacean Society (SCS) and the WWF, France.

Appendix A. Expectation and variances of $(Z_s/t_s - Z_{s'}/t_{s'})$

In order to characterize the relationship between the variograms of Z and Y and to determine adequate weights or correction terms, we developed conditional and non-conditional expectations related to the quantity $\left(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right)$ using Eq. (2).

$$\mathrm{E}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\middle| Y\right] = \frac{1}{t_s}\mathrm{E}[Z_s|Y_s] - \frac{1}{t_{s'}}\mathrm{E}[Z_{s'}|Y_{s'}]$$
$$= Y_s - Y_{s'}, \qquad \mathrm{E}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right] = m - m = 0$$

The expression of the non-conditional order-2 moment (4) is derived from Eqs. (2) and (3)

$$\mathrm{E}\left[\left(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right)^2\middle| Y\right]$$
$$= \frac{1}{t_s^2}\mathrm{E}[(Z_s)^2|Y_s] + \frac{1}{t_{s'}^2}\mathrm{E}[(Z_{s'})^2|Y_{s'}]$$
$$- \frac{2}{t_s t_{s'}}\mathrm{E}\left[Z_s Z_{s'}|Y\right] = \frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}} - 2\delta_{ss'}\frac{Y_s}{t_s}$$
$$+ (Y_s - Y_{s'})^2, \qquad \mathrm{E}\left[\left(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right)^2\right]$$
$$= \mathrm{E}\left[\frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}} - 2\delta_{ss'}\frac{Y_s}{t_s} + (Y_s - Y_{s'})^2\right]$$
$$= m\left(\frac{t_s + t_{s'}}{t_s t_{s'}}\right) - 2\delta_{ss'}\frac{m}{t_s} + 2\gamma_Y(s - s')$$

For $s \neq s'$, we also have

$$\mathrm{Var}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\middle| Y\right] = \mathrm{E}\left[\left(\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\right)^2\middle| Y\right]$$
$$- \mathrm{E}^2\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\middle| Y\right] = \frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}} + (Y_s - Y_{s'})^2$$
$$- (Y_s - Y_{s'})^2 = \frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}},$$
$$\mathrm{E}\left[\mathrm{Var}\left[\frac{Z_s}{t_s} - \frac{Z_{s'}}{t_{s'}}\middle| Y\right]\right] = \mathrm{E}\left[\frac{Y_s}{t_s} + \frac{Y_{s'}}{t_{s'}}\right]$$
$$= m\left(\frac{t_s + t_{s'}}{t_s t_{s'}}\right)$$

Appendix B. Variance of the error of kriging in Poisson case

The expression of this variance is derived from the kriging estimate definition

$$\mathrm{Var}(Y_o^* - Y_o) = \mathrm{Var}\left[\left(\sum_{\alpha=1}^{n}\lambda_\alpha\frac{Z_\alpha}{t_\alpha} - Y_o\right)^2\right],$$
$$\mathrm{E}[(Y_o^* - Y_o)^2|Y] = \mathrm{E}\left[\left(\sum_{\alpha=1}^{n}\lambda_\alpha\frac{Z_\alpha}{t_\alpha} - Y_o\right)^2\middle| Y\right]$$
$$= \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\frac{\lambda_\alpha}{t_\alpha}\frac{\lambda_\beta}{t_\beta}\mathrm{E}[Z_\alpha Z_\beta|Y] + Y_o^2$$
$$- 2Y_o\sum_{\alpha=1}^{n}\frac{\lambda_\alpha}{t_\alpha}\mathrm{E}[Z_\alpha|Y] = \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta Y_\alpha Y_\beta$$
$$+ \sum_{\alpha=1}^{n}\frac{\lambda_\alpha^2}{t_\alpha}Y_\alpha + Y_o^2 - 2Y_o\sum_{\alpha=1}^{n}\lambda_\alpha Y_\alpha,$$
$$\mathrm{E}[(Y_o^* - Y_o)^2] = \mathrm{E}\left[\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta Y_\alpha Y_\beta\right.$$
$$\left.+ \sum_{\alpha=1}^{n}\frac{\lambda_\alpha^2}{t_\alpha}Y_\alpha + Y_o^2 - 2Y_o\sum_{\alpha=1}^{n}\lambda_\alpha Y_\alpha\right]$$
$$= \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta C_{\alpha\beta} + \sum_{\alpha=1}^{n}\frac{\lambda_\alpha^2}{t_\alpha}m + \sigma_Y^2$$
$$- 2\sum_{\alpha=1}^{n}\lambda_\alpha C_{\alpha o} + 2m^2 - 2m^2 = \sigma_Y^2 + \sum_{\alpha=1}^{n}\frac{\lambda_\alpha^2}{t_\alpha}m$$
$$+ \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta C_{\alpha\beta} - 2\sum_{\alpha=1}^{n}\lambda_\alpha C_{\alpha o}$$

Then, from the unbiasedness of Y_o^*:

$$\mathrm{Var}(Y_o^* - Y_o) = \sigma_Y^2 + \sum_{\alpha=1}^{n}\frac{\lambda_\alpha^2}{t_\alpha}m + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta C_{\alpha\beta}$$
$$- 2\sum_{\alpha=1}^{n}\lambda_\alpha C_{\alpha o}$$

After minimization on λ_α, i.e. resolution of the kriging system, we obtain as kriging variance (Eq. 13):

$$\begin{aligned}\mathrm{Var}(Y_o^* - Y_o) &= \sigma_Y^2 + \sum_{\alpha=1}^{n} \frac{\lambda_\alpha^2}{t_\alpha} m + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C_{\alpha\beta} \\ &- 2\sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} = \sigma_Y^2 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} \\ &+ \sum_{\alpha=1}^{n} \lambda_\alpha \left(\sum_{\beta=1}^{n} \lambda_\beta C_{\alpha\beta} + \lambda_\alpha \frac{m}{t_\alpha} - C_{\alpha o} \right) \\ &= \sigma_Y^2 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} - \sum_{\alpha=1}^{n} \lambda_\alpha \mu \\ &= \sigma_Y^2 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} - \mu\end{aligned}$$

which is the usual expression for kriging variance.

References

Agarwal, D.K., Gelfand, A.G., Citron-Pousty, S., 2002. Zero-inflated models with application to spatial count data. Environ. Ecol. Statist. 9, 341–355.

Burrough, P., McDonnell, R., 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford, UK.

Burrough, P.A., 2001. GIS and geostatistics: essential partners for spatial analysis. Environ. Ecol. Statist. 8 (4), 361–377.

Canadas, A., Sagarminaga, R., Garcia-Tiscar, S., 2002. Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. Deep-Sea Res. I 49 (9), 2053–2073.

Chilès, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty, Wiley Series in Probability and Statistics. Wiley, New York.

Christensen, O.F., Waagepetersen, R., 2002. Bayesian prediction of spatial count data using generalized linear mixed models. Biometrics 58, 280–286.

David, L., 2002. Rôle et importance des canyons sous-marins sur la marge continentale dans la distribution estivale des cétacés de Méditerranée Nord-occidentale. Ph.D. Thesis. EPHE, Université de Montpellier II, 320 pp.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model based geostatistics. Appl. Statist. 47, 299–350.

Di Miglio, N., 1999. Distribution comparée des cétacés et des oiseaux marins de Méditerranée Nord-occidentale en période estivale. Relation avec les conditions environnementales. Ph.D. Thesis. EPHE, Université de Montpellier II, 377 pp.

Forcada, J., Agiolar, A., Hammond, P., Pastor, X., Aguila, R., 1996. Distribution and abundance of fin whales (*Balaenoptera physalus*) in the western Mediterranean Sea during the summer. J. Zool. 238, 23–24.

Forcada, J., Notarbartolo di Sciara, G., Fabbri, F., 1995. Abundance of fin whales and striped dolphins summering in the Corso-Ligurian basin. Mammalia 59, 127–140.

Gannier, A., 1997. Estimation de l'abondance estivale du rorqual commun *Balaenoptera physalus* (Linné, 1758) dans le Bassin Liguro-Provençal (Méditerranée nord occidentale). Revue d'Ecologie (Terre Vie) 53, 255–272.

Gannier, A., 2002. Summer distribution of fin whales (*Balenoptera physalus*) in the northwestern Mediterranean Marine Mammals Sanctuary. Revue d'Ecologie (Terre Vie) 57, 135–150.

Griffith, D.A., 2002. A spatial filtering specification for the auto-Poisson model. Statist. Probab. Lett. 58 (3), 245–251.

Huguet, C., Maynou, F., Abelló, P., 2005. Small-scale distribution characteristics of *Munida* spp. populations (Decapoda: Anomura) off the Catalan coasts (western Mediterranean). J. Sea Res. 53 (4), 283–296.

Kaiser, M.S., Cressie, N., 1997. Modeling Poisson variables with positive spatial dependence. Statist. Probab. Lett. 35 (4), 423–432.

Littaye, A., Gannier, A., Laran, S., Wilson, J.P.F., 2004. The relationship between summer aggregation of fin whales and satellite-derived environmental conditions in the northwestern Mediterranean sea. Remote Sensing Environ. 90, 44–52.

Maynou, F.X., Sardà, F., Conan, G.Y., 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 (1), 102–120.

Notarbartolo di Sciara, G., Zanardelli, M., Jahoda, M., Panigada, S., Airoldi, S., 2003. The fin whale *Balaenoptera physalus* (Linné 1758) in the Mediterranean Sea. Mammal Rev. 33, 105–150.

Oliver, M.A., Lajaunie, C., Webster, R., Muir, K.R., Mann, J.R., 1993. Estimating the risk of childhood cancer. In: Saores, A. (Ed.), Geostatistics Troia'92. Kluwer Academic Publishers, Dordrecht, pp. 899–910.

Oliver, M.A., Webster, R., Lajaunie, C., Muir, K.R., Parkes, S.E., Cameron, A.H., Stevens, M.C.G., Mann, J.R., 1998. Binomial cokriging for estimating and mapping the risk of childhood cancer. IMA J. Math. Appl. Med. Biol. 15 (3), 279–297.

Petitgas, P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish Fish. 1, 231–249.

R Development Core Team, 2004. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 3-900051-00-3. http://www.R-project.org.

Rendell, L., Whitehead, H., Escribano, R., 2004. Sperm whale habitat use and foraging success off northern Chile: evidence of ecological links between coastal and pelagic systems. Mar. Ecol. Prog. Ser. 275, 289–295.

Ribeiro Jr., P.J., Diggle, P.J., 2001. geoR: a package for geostatistical analysis. R News 1 (2), 14–18. http://CRAN.R-project.org/doc/Rnews/.

Rivoirard, J., Wieland, K., 2001. Correcting for the effect of daylight in abundance estimation of juvenile haddock (*Melanogrammus aeglefinus*) in the north sea: an application of kriging with external drift. ICES J. Mar. Sci. 58, 1272–1285.

Rueda, M., Defeo, O., 2003. Spatial structure of fish assemblages in a tropical estuarine lagoon: combining multivariate and geostatistical techniques. J. Exp. Mar. Biol. Ecol. 296, 93–112.

Rufino, M.M., Maynou, F., Abello, P., Yule, A.B., 2004. Small-scale non-linear geostatistical analysis of *Liocarcinus depurator* (Crustacea: Brachyura) abundance and size structure in a western mediterranean population. Mar. Ecol. Prog. Ser. 176, 223–235.

van der Meer, J., Leopold, M.F., 1995. Assessing the population size of the European storm petrel (*Hydrobates pelagicus*) using spatial autocorrelation between counts from segments of criss-cross ship transects. ICES J. Mar. Sci. 52 (5), 809–818.

Wackernagel, H., 1995. Multivariate Geostatistics. Springer.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists. Wiley, New York.

Wiens, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3, 385–397.

Wikle, C.K., 2002. Spatial modeling of count data: a case study in modelling breeding bird survey data on large spatial domains. In: Lawson, A., Denison, D. (Eds.), Spatial Cluster Modelling. Chapman & Hall, London, pp. 199–209.