Volume 55 • Issues 10–12 • 2007 ISSN 0025-326X

MARINE POLLUTION BULLETIN

Measuring and Managing Changes in Estuaries and Lagoons

Editors: Roberto Zonta, Stefano Guerzoni, Victor N. de Jonge and Angel Pérez-Ruzafa

Available online at www.sciencedirect.com
ScienceDirect
Marine Pollution Bulletin 55 (2007) 579–590
MARINE POLLUTION BULLETIN
www.elsevier.com/locate/marpolbul

A comparative analysis of three habitat suitability models for commercial yield estimation of *Tapes philippinarum* in a North Adriatic coastal lagoon (Sacca di Goro, Italy)

Simone Vincenzi [a,*], Graziano Caramori [b], Remigio Rossi [c], Giulio A. De Leo [d]

[a] *Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Viale G. Usberti 33A, Parma, Italy*
[b] *Istituto Delta di Ecologia Applicata S.r.l., Via Puccini, 29 Ferrara, Italy*
[c] *Dipartimento di Biologia, Università di Ferrara Via L. Borsari 46, 44100 Ferrara, Italy*
[d] *Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Viale G. Usberti 33/A, Parma, Italy*

Abstract

Habitat Suitability (HS) models have been extensively used by conservation planners to estimate the spatial distribution of threatened species and of species of commercial interest. In this work we compare three HS models for the estimation of commercial yield potential and the identification of suitable sites for *Tapes philippinarum* rearing in the Sacca di Goro lagoon (Italy) on the basis of six environmental factors. The habitat suitability index (HSI) is based on expert opinion while the habitat suitability conditional (HSC) is calibrated on observational data. The habitat suitability mixed (HSM) model is a two-part model combining expert knowledge and regression analysis: the first component of the model uses logistic regression to identify the areas in which clams are likely to be present; the second part applies the same parameter-specific suitability functions of the HSI model only in the areas previously identified as productive by the logistic component.

The HS models were validated on an independent data set and estimates of potential yield of the Goro lagoon were compared. The effectiveness of the three approaches is then discussed in terms of predicted yield and identification of suitable sites for farming.

Keywords: Habitat suitability; *Tapes philippinarum*; Clam yield; Sacca di Goro; Habitat suitability index

1. Introduction

The identification of suitable sites for farming and the estimation of the potential commercial yield are key tasks to be undertaken in any regulated aquaculture activity that aims to make compatible the achievement of highly productive harvesting activities with the long-term conservation of the natural environment. This is particularly true for coastal lagoons where extensive farming of fish (e.g. finfish and shrimps) and/or bivalves (e.g. oysters, mussels and clams) is performed as the management of aquaculture activities requires to face multiple, possibly conflicting goals, such as the preservation of crucial ecosystem services (Vives, 1996) and, at the same time, the maintenance of the environmental and economic sustainability of the farming industries. It is clear that to improve the economic efficiency of the farming activities and to insure an equitable share of exploitable areas in presence of multiple competing fishermen, the assessment of commercial yield potential and the identification of suitable sites for extensive farming is mandatory. In identifying suitable sites for farming, the core problem is to use available information on where a species occurs and on the associated habitat features to predict how likely is the species to be present also in unsampled locations. Habitat suitability (HS) modelling has

* Corresponding author. Tel.: +39 0521 905696; fax: +39 0521 906611.
E-mail addresses: svincenz@nemo.unipr.it (S. Vincenzi), grazianocaramori@istitutodelta.it (G. Caramori), remigio.rossi@unife.it (R. Rossi), giulio.deleo@unipr.it (G.A. De Leo).

0025-326X/$ - see front matter
doi:10.1016/j.marpolbul.2007.09.016

been extensively used in conservation planning (Gibson et al., 2004) on the assumption that size and spatial arrangement of suitable habitat can influence the long-term retention and persistence of faunal species (Gibson et al., 2004; Lindenmayer and Possingham, 1996). HS models are often used to predict the likelihood of occurrence and/or abundance of a species on the basis of habitat attributes that influence – or are related to – its survival, growth and reproduction (Pereira and Itami, 1991; Kliskey et al., 1999; Store and Kangas, 2001 and Gibson et al., 2004). These models are typically developed by identifying statistical relationships between the occurrence and/or the abundance of the species and the biochemical and physical properties of a given site (Store and Kangas, 2001). Several statistical models can be used for this purpose (a review of some these techniques can be found in Franklin, 1995 and in Guisan and Zimmermann, 2000) with the choice among them depending primarily upon the type of response variable that is modeled. For instance, when the response variable is binary (presence/absence) the typical approach is to use logistic regression (generalize linear model, GLM, McCullagh and Nelder, 1989) to explore the relationship between the independent variables (such as temperature, altitude and vegetation cover) and the presence of the species.

Data-driven regression analyses and the resulting model predictions are generally considered intrinsically objective, even though a number of subjective decisions usually needs to be made when developing HS models, ranging from the choice of sampling strategy to that of the predictors and of the method used to select the best model (e.g. null hypothesis-testing or information theory). The development of a data-driven, empirical HS models typically requires (i) to gather a data set describing the occurrence or the abundance of the species of interest in the studied area and another data set of candidate explanatory environmental variables (predictors) sampled in the same area; (ii) to use this information to design and calibrate on available data the mathematical/statistical model relating the species presence or abundance to the predictors; (iii) and, finally, to validate the model by assessing its predictive capabilities on a set of independent data. Unfortunately, such a procedure cannot be always accomplished for all the species of commercial interest and coastal lagoons in which extensive farming is performed. In such a case, a possible alternative to data-driven models is the use of expert knowledge. In order to predict the occurrence or abundance of the species under study, methods and instruments such as quality indexes, expert systems, multi-criteria analysis and fuzzy set theory, have been developed to frame expert knowledge into a quantitative and measurable set of rules, guidelines and/or assessment criteria (Store and Kangas, 2001).

Most of the applications of HS models have regarded terrestrial ecosystem, even though, in recent years, HS approaches have also been used for identifying appropriate sites for mollusc farming in various regions of the world (e.g. Nath et al., 2000), especially in North-America and Mexico (e.g. Kapetsky et al., 1988; Aguilar-Manjarrez and Ross, 1995) as well as in other North Adriatic lagoons. Particularly interesting in the context of the North Adriatic lagoons where mollusc farming is performed, is the commercial exploitation of the Manila Clam *Tapes philippinarum* (Adam and Reeve, 1850) in the Sacca di Goro coastal lagoon (Italy) (Fig. 1), for which two different HS models have been developed to estimate the commercial yield potential and to identify suitable sites for farming: the first one is a habitat suitability index (HSI) model that makes use of parameter-specific suitability functions for *T. philippinarum* farming in North Adriatic lagoons based on expert knowledge and of a weighted geometric mean of the suitability values to compute an overall quality index (Vincenzi et al., 2006a). The second one is a data-driven, two-part conditional (habitat suitability conditional, HSC) model estimated by using rigorous statistical analyses of data available for the Sacca di Goro lagoon (Vincenzi et al., 2006b).

In the present work we have developed a third, new habitat suitability mixed model (HSM) for *T. philippinarum* in the Sacca di Goro lagoon which is a two-part model somehow intermediate between the previous two approaches; the first component of the model uses logistic regression to identify the areas in which clams are likely to be present; the second part applies the same parameter-specific functions of the HSI model only in the areas previously identified as productive by the logistic component but, in this case, the weights of the geometric mean are estimated with a constrained linear regression.

Aim of this paper is to provide a comparative analysis of the three HS models for the estimation of commercial yield potential of *T. philippinarum* in the Sacca di Goro lagoon in terms of approach, data requirement and results and to discuss the implications of model choice. The present work is organized as follows: after a brief description of the study area and of available data, we summarize the main features of the HSI and HSC models and how they are used to obtain estimates of yield potential and we illustrate the formulation and application of the new HSM

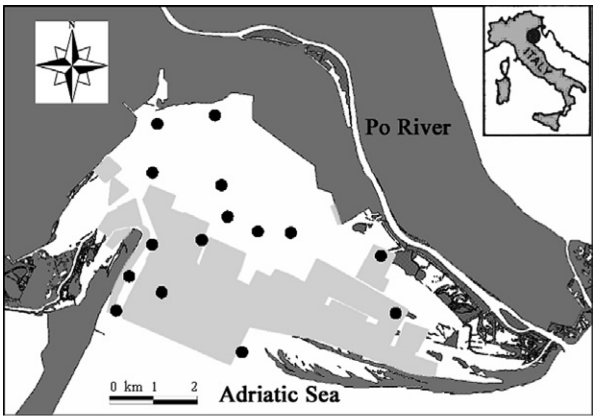

Fig. 1. Location of the rearing concessions and of the 15 sampling stations in the Goro lagoon.

model. Then, we present the application of the three HS models to the Sacca di Goro lagoon and the main differences in the results obtained by applying the three HS models. Finally, we discuss the strengths and the weaknesses on the three models.

2. Material and methods

2.1. Study area

Sacca di Goro is a shallow Northern Adriatic lagoon (surface 26 km^2, average depth 1.5 m) approximately triangular in shape located in the southern area of the Po river delta (44.78–44.83°N, 12.25–12.33°E) (Fig. 1). The lagoon is separated from the Adriatic Sea by a narrow sandy barrier with two mouths of about 0.9 km each regulating saltwater exchanges and it has four freshwater inlets, namely Po di Goro and Po di Volano rivers and Bianco and Giralda channels. *T. philippinarum*, a clam of Indo-Pacific origin, was first introduced into Italy in 1983 by a regional environmental agency aimed at evaluating its potential growth in the lagoon of Venice (Pellizzato and Da Ros, 2005) and quickly spread to the Goro lagoon (Rossi, 1989) replacing the native species *Tapes decussatus* (Paesanti and Pellizzato, 2000). To avoid the overexploitation of *T. philippinarum* and the detriment of the natural environment, a management system defined as ''culture-based fishery'' has been operating for some years in the Goro lagoon, that is, harvestable areas are assigned to clam farmers by the regulatory agency under a set of strict rules regarding the exploitation effort. The production of *T. philippinarum* in the Goro lagoon reached a peak of 16,000 t in 1991 (Rossi and Paesanti, 1992) while, at present, the limitations of exploitation activities imposed by the regulatory agency permit a annual production of about 10,000 t (Pellizzato and Da Ros, 2005).

2.2. Environmental factors affecting clam growth, survival and farming

According to a number of studies performed in the Sacca di Goro (Rossi, 1996; Paesanti and Pellizzato, 2000; Pastres et al., 2001; Solidoro et al., 2003; Melià et al., 2004), the main biogeochemical and hydrodynamic factors affecting clam growth, survival and farming are salinity, sand content in the sediment, hydrodynamism (i.e. water current), water depth, dissolved oxygen and chlorophyll ''a''. As for mean annual temperature, while it is a crucial determinant of clam growth, it does not exhibit a sufficient spatial variation in the Goro lagoon to determine variability in clam yield potential (Vincenzi et al., 2006); as a consequence, temperature has not been explicitly included in the model to assess the clam yield potential of the whole lagoon and to identify spatial differences in suitability within the lagoon. For all the three Habitat Suitability models we used the same six environmental variables as predictors of clam yield potential.

2.3. Data and sampling survey

Due to its geographic and economic importance, a number of sampling surveys have been carried out in the last years to gather information on the main biogeochemical and hydrodynamic parameters, primary productivity, water quality, etc. In particular, in the year 2003, a carefully planned sampling survey was performed to investigate also areas currently not exploited for farming. The data acquired on clam density and the environmental parameters are briefly described hereafter.

2.3.1. T. philippinarum density

Samples of *T. philippinarum* density ($n = 107$) were acquired by local expert fishermen in the 2003 by means of gear locally called ''rasca'' (Vincenzi et al., 2006). In order to collect clam samples of all sizes, *rasca* with 6 mm fine mesh net bags were used in the harvesting processes. As the same site is farmed about once a year, density samples (kg m^{-2}) were considered estimates of annual yield (kg m^{-2} $year^{-1}$). Density samples associated with habitat features ($n = 107$) were used as observational data for the calibration and validation of HSC and HSM models.

2.3.2. Biogeochemical and hydrodynamic parameters

Data on physical–chemical parameters, namely salinity, chlorophyll ''a'' and dissolved oxygen were gathered seasonally in the year 2003 in 15 different stations located within the lagoon (Fig. 1) by using a multi-parametric probe (IDRONAUT OCEAN SEVEN 301M). The probe was programmed to sample every 30 cm from surface to the bottom, recording parameter values. Observed values were then averaged to compute the annual means. Data on bathymetry were gathered by using a Eco-Sounder (248,000 points total).

Sediment sampling was performed by the Geology Department of the University of Ferrara (Simenoni et al., 2000) in order to derive the fraction of sand in the sediment. *T. philippinarum* is known to perform better in sandy sediment rather than in muddy sediment (Barillari et al., 1990; Rossi, 1996 and Paesanti and Pellizzato, 2000).

Water flow dynamics in the Sacca di Goro were acquired from the study performed by Brath et al. (2000) who calculated flow fields and flow capacity values (m^3 s^{-1} m^{-1}) by implementing the software MIKE21 (DHI, 1993). Hydrodynamism was estimated in the whole lagoon for four different conditions, namely low tide, high tide, intermediate decreasing tide and intermediate increasing tide; the average value was used in this work.

2.3.3. Data interpolation and thematic maps

Point data were interpolated *via* a *nearest neighbour* algorithm over grid of 100×88 cells (each one of 1 ha of surface) by using the software SURFER™ of Golden Software Inc. ver. 7.02 in order to produce thematic maps of the whole Goro lagoon for each of the environmental parameters considered. The mesh size identifies sites with area of 1 ha. These thematic maps are currently the official maps used by the regulatory agency for the harvesting activities in the lagoon (Province of Ferrara).

2.4. Habitat suitability models

2.4.1. HSI model

The HSI model is based essentially on expert knowledge and has been developed by Vincenzi et al. (2006) to identify suitable sites for farming in the Goro lagoon and to provide estimates of yield potential. The HSI model was derived as follows:

- The six main biogeochemical and hydrodynamic variables affecting clam presence and yield – as described in the previous section – were identified on the basis of expert knowledge (Paesanti and Pellizzato, 2000).
- Parameter-specific suitability functions (PSSFs) assessing the suitability of a given site with respect to each environmental factor were defined according to the reference manual for clam farming in North Adriatic lagoons by Paesanti and Pellizzato (2000), as shown in Fig. 2. For each variable, suitability is expressed in terms of a suitability index (SI) bounded between 0 and 1, where 0 indicates a non-suitable habitat and 1 a habitat most suitable (United States Fish and Wildlife Service, 1981).
- The parameter-specific suitabilities, obtained by applying the PSSF_i $(i = 1, \ldots, 6)$ to environmental data, were aggregated by using a weighted geometric mean to compute a site-specific habitat suitability index (HSI) for farming of *T. philippinarum*, as described here below;

$$\mathrm{HSI}(x,y) = \left(\prod_{i=1}^{6} \mathrm{PSSF}_i(x,y)^{w_i}\right)^{\frac{1}{\sum_{i=1,\ldots,6} w_i}} \tag{1}$$

where w_i are the weights defining the importance of each PSSF_i according to expert knowledge of fishermen operating in the lagoon and (x, y) the site-specific coordinates (Table 1).
- A function linking the HSI to clam yield (kg m^{-2} year^{-1}) was defined as follows:

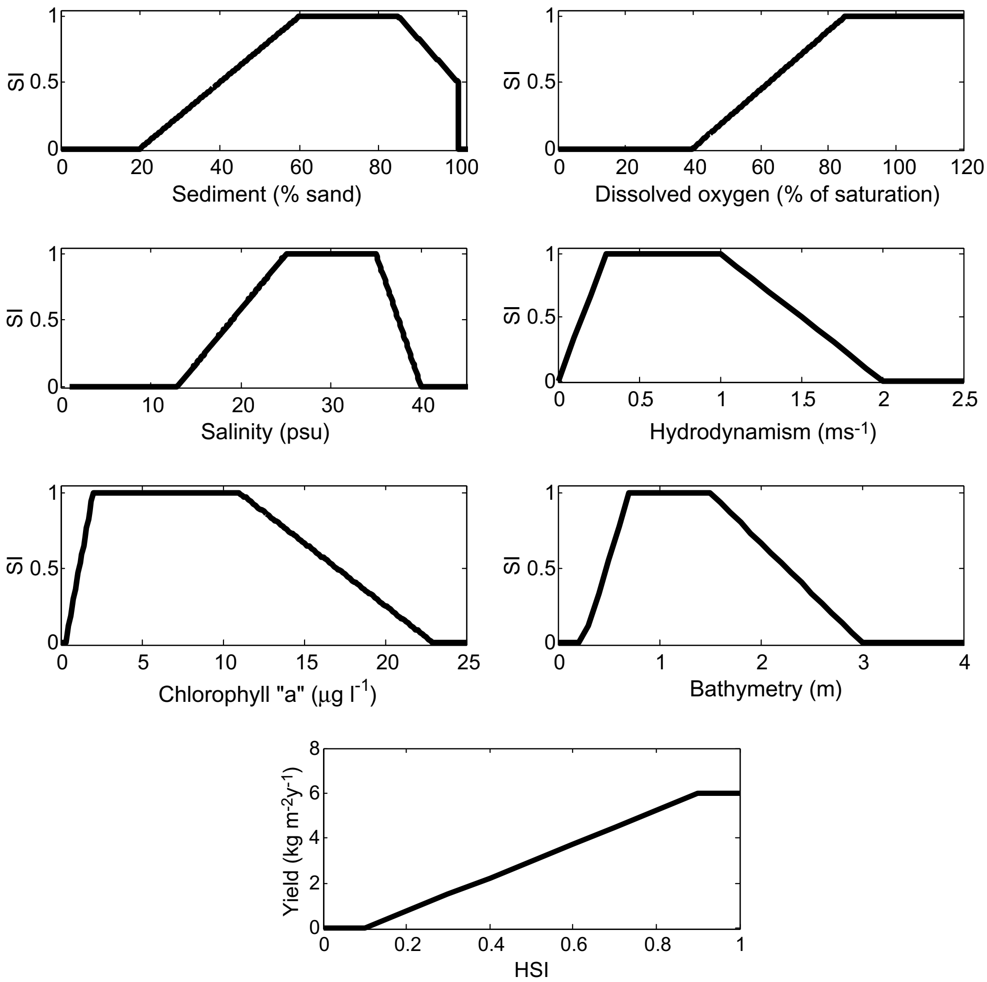

Fig. 2. Suitability graphs for the six environmental variables. Suitability is expressed in term of a suitability index (SI) bounded between 0 and 1 (habitat most suitable).

Table 1
Weights w_i of the parameter-specific suitability functions (PSSFs) used to compute the overall habitat suitability (Eq. (1))

Variable	Weights w_i	
	HSI model	HSM model
Sediment	0.5	0.5
Hydrodynamism	1	1
Bathymethry	0.2	0
Salinity	0.1	0.3
Oxygen	0.1	0
Chlorophyll "a"	0.1	0

In the HSI model, weights have been set through expert opinion, while in the HSM model through linear regression on the log-transformed values of the habitat suitability index and of the PSSFs (Eq. (6)). For comparison, weights have been standardized to 1.

$$E(Y) = \begin{cases} 0 & \text{if HSI} \leqslant 0.1 \\ 6.25\text{HSI} - 0.625 & \text{if } 0.1 < \text{HSI} < 0.9 \\ 6 & \text{if HSI} \geqslant 0.9 \end{cases} \quad (2)$$

where we omitted (x, y) for simplicity. According to field observations, expert knowledge (Edoardo Turolla, personal communication) and findings of Rossi (1996) and Melià et al. (2004) the maximum attainable yield for *T. philippinarum* in the Goro lagoon is about 6 kg m^{-2} year^{-1}.

The model was then applied to the Sacca di Goro lagoon to identify suitable sites for clam farming and the commercial yield potential.

2.4.2. HSC model

The HSC model is a data-driven two-part conditional model (Fletcher et al., 2005) developed by Vincenzi et al. (2006b) to take into account the relatively high presence of zero yield data in the observational data set (22%). The HSC model was derived by separately estimating a logistic model in which the response variable was presence and absence of *T. philippinarum* and an ordinary regression model where the response variable was the log-transformed clam yield for those sites in which *T. philippinarum* was present. To estimate the two sub-models, first the observational data set ($n = 107$) was randomly splitted in two parts: the calibration data set ($n = 80$) and the validation data set ($n = 27$). Then, two further data sets were derived from the calibration data set, one in which the response variable was presence/absence of clam (presence data set) and the other in which the response variable was *T. philippinarum* annual yield (kg m^{-2} year^{-1}) when *Tapes* was present (yield data set). Logistic regression was used to explore the relationship between presence/absence of *T. philippinarum* and the six environmental variables and ordinary regression when the response variable was clam yield. For both the sub-models, we started from the full model incorporating also all 2-way interactions among the six environmental variables; the best model was selected by a backward stepwise selection procedure based on the AIC value of the model (Akaike, 1974). The logistic and ordinary regression models were then combined to estimate the potential commercial yield [in kg m^{-2} year^{-1}] of *T. philippinarum* in the Sacca di Goro lagoon, as described hereafter. Let $Y(\mathbf{w})$ be the yield of *T. philippinarum*, where $\mathbf{w}$ is the vector of explanatory variables (that is, of the variables affecting yield abundance) and let $Z(\mathbf{x})$ be a binary variable – equal to 1 when *Tapes* is present and 0 when not – where $\mathbf{x}$ is the vector of explanatory variables (that is, of the variables affecting the probability of clam presence). The expected value of Y is given by:

$$\begin{aligned} E(Y) &= Pr(Z = 1)E(Y|Z = 1) + Pr(Z = 0)E(Y|Z = 0) \\ &= Pr(Z = 1)E(Y|Z = 1) = \pi\mu \end{aligned}$$

As showed by Stefansson (1996), Welsh et al. (1996) and Fletcher et al. (2005), the estimate of the expected yield of *T. philippinarum* is computed as follows:

$$\widehat{E}(Y) = \hat{\pi}\hat{\mu} \quad (3)$$

where

$$\hat{\pi} = Pr(Z = 1) = \exp(\mathbf{x}'\hat{\beta})/\{1 + \exp(x'\hat{\beta})\} \quad (4)$$

and

$$\hat{\mu} = E(Y|Z = 1) = \exp(\mathbf{w}'\hat{\theta} + \hat{\sigma}^2/2) \quad (5)$$

are the estimates of π and μ computed from the two regression models. Thus, $\hat{\beta}$ is the vector of the estimates of the coefficients in the logistic regression, $\hat{\theta}$ is the vector of the estimates of the coefficients in the ordinary regression and $\hat{\sigma}^2$ is the residual mean square in the ordinary regression model for the positive values, as described by Crow and Shimizu (1988).

2.4.3. HSM

While the strength of the HSI model lays in the use of widely accepted expert knowledge to define the PSSFs (Paesanti and Pellizzato, 2000), its main weakness is that the weights used in Eq. (1) to define the relative importance of the environmental variables have been defined on purely subjective judgment of fishermen operating in the lagoon, with no reference in international or "grey" literature. To overcome this weakness, we have decided to derive a new model (the HS Mixed model, HSM) that combines the strength of the HSI model (i.e., the use of widely accepted PSSFs) with a rigorous calibration procedure to estimate the weights of the geometric mean. The HSM model is also a two-part model in which the logistic component of the HSC model (Eq. (4)) is used to identify the areas of the lagoons where clams are likely to be present and the potential yield [kg m^{-2} year^{-1}] is then estimated by using the same PSSFs of the HSI model but with the weights of the geometric mean calibrated on available data through constrained linear regression on the log-transformed version of Eq. (1). As for the logistic component, the Receiver Operating Characteristic (ROC) curve (Fielding and Bell, 1997) was used to determine the optimal cutoff value to transform the continuous response variable of the logistic model (the probability of presence) in a binary output indicating presence (1) or absence (0) of *Tapes*. The accuracy of a classifying mode, in fact, depends on how well the model discriminates the data into true positives and true negatives. If the costs of misclassifying positive and negative cases are the same – as in our case –, the optimal cut-off is the ROC threshold value that provides the optimal trade off between sensitivity (true positive rate) and specificity (true negative rate). True and false positive rates (1-specificity) at incremental cut-off scores were plotted in a ROC graph to identify the best cut-off value. Accuracy is usually measured by the area under the ROC curve (usually referred to as AUC), as it provides a single measure of overall accuracy that is not dependent upon a particular threshold (Deleo, 1993): an area of 1 represents a perfect model for the data investigated while an area of 0.5 represents a worthless model. Once the optimal cut-off has been determined, the potential yield is set to zero in the areas of the lagoon in which the probability of occurrence is below the cut-off value. In the areas of the lagoon where the probability of occurrence is above the cut-off, the potential yield is estimated by using Eqs. (1) and (2), where the weights of the geometric mean of Eq. (1) are estimated as follows. Yield data were transformed into HSI values by using the inverse function of Eq. (2). Eq. (1) was log-transformed so as to obtain the following linear relationship:

$$\log \mathrm{HSI}(x,y) = \sum_{i=0}^{6} w_i \log \mathrm{PSSI}_i(x,y) \qquad (6)$$

Weights w_i were estimated by solving a constrained linear least-square problem with $\sum_{i=0}^{6} w_i = 1$ and $w_i \geqslant 0$. To account for the logarithmic transformation of the HSI and PSSI values, only data for which $\mathrm{HSI}(x,y) > 0$ and $\mathrm{PSSFs}(x,y) > 0$ were used for the calibration ($n = 78\%$, 73% of total observations).

2.4.4. Model validation

The HSI and the HSC models were validated on the same independent validation dataset ($n = 27$) extracted from the observational data set. We could not validate the HSM model because all the suitable data from the original data set were used to calibrate the model. In fact, at least 10 observations are needed for each predictor included in the model (Harrell, 2001), so all the 78 observations of the original data set for which the corresponding $\mathrm{PSSFs}(x,y)$ and $\mathrm{HIS}(x,y)$ values were larger than zero have been used to calibrate the weights w_i (Eq. (6)) by solving the constrained linear least-squares problem.

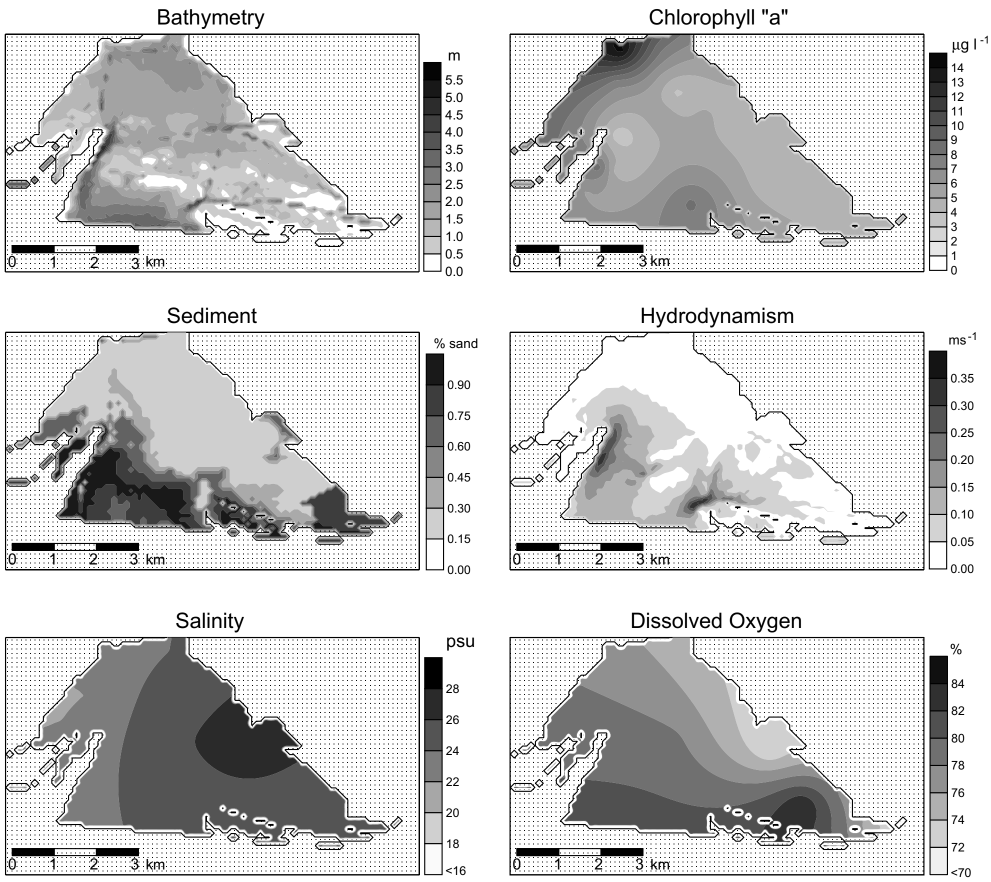

Fig. 3. Thematic maps of the Goro lagoon representing bathymetry (m), chlorophyll "a" (µg l^{-1}), sediment (% of sand), hydrodynamism (m s^{-1}), salinity (psu) and dissolved oxygen (% of saturation).

3. Results

Parameter-specific thematic maps of the whole Goro lagoon are provided in Fig. 3. According to the ROC curve, the best cutoff value for the logistic model (Eq. (4)) was 0.5. AUC of the logistic model for *T. philippinarum* was 0.97, which indicated an excellent discrimination between true positives and true negatives. For the HSC and HSM models, share of sand in the sediment, dissolved oxygen and hydrodynamism were the only significant variables in determining presence of *Tapes*, with no interaction between variables being significant. Table 2 reports the parameters estimates for the logistic component of both the HSC and HSM models (Eq. (4)) and for the ordinary regression component in the HSC model (Eq. (5)). In Fig. 4 and Fig. 5 we reported the clam probability of presence (Eq. (4)) and the yield potential (Eq. (3)) as a function of the environmental factors included in the logistic model (incorporated in both the HSC and HSM models) and in the ordinary regression model (yield part of the HSC model), respectively. Weight estimates for the HSM model are reported in Table 1. For only three environmental parameters, namely share of sand in the sediment, salinity and hydrodynamism, the estimated weights were significantly larger than zero. Estimates of yield potential for the whole Goro lagoon and total area suitable for farming for the three Habitat Suitability models are reported in Table 3, while model-specific maps of predicted yield of *T. philippinarum* in the whole Goro lagoon and a map reporting the range of predicted yields are reported in Fig. 6. The cumulative frequency distribution (CDF) of

Table 2
Estimates, standard errors and standardised estimates of the coefficients for the explanatory variables for the best logistic and ordinary regression models chosen using stepwise selection based on the AIC value of the model

Parameter	Ordinary regression			Logistic regression		
	Estimate	Std. error	Standardised estimate	Estimate	Std. error	Standardised estimate
H	−18.00	10.68	−1.69	3.66	1.25	2.93
Sa	−2.05	0.47	−4.36	–	–	–
Sd	81.56	17.83	4.57	6.04	3.09	1.95
B	−21.08	11.6	−1.82	–	–	–
C	−31.37	10.37	−3.03	–	–	–
O	−0.44	0.89	−0.49	0.74	0.28	2.64
Sa:C	0.42	0.07	6.00	–	–	–
Sd:C	−0.93	0.31	−3.00	–	–	–
Sd:O	−0.95	0.22	−4.32	–	–	–
C:O	0.25	0.12	2.08	–	–	–
C:H	−0.45	0.13	−3.46	–	–	–
C:B	0.93	0.28	3.32	–	–	–
O:I	0.25	0.13	1.92	–	–	–
O:B	0.23	0.14	1.64	–	–	–
H:B	1.50	0.46	3.26	–	–	–
R^2	0.82	–	–	–	–	–
σ^2	0.21 (76 d.f.)	–	–	–	–	–

The logistic model is part of the HSC and of the HSM models. H, log(Hydrodynamism); Sa, Salinity; Sd, Share of sand in the sediment; B, log(Bathymetry + 1); C, chlorophyll "a"; O, dissolved oxygen. Couples of variables indicate interaction between variables.

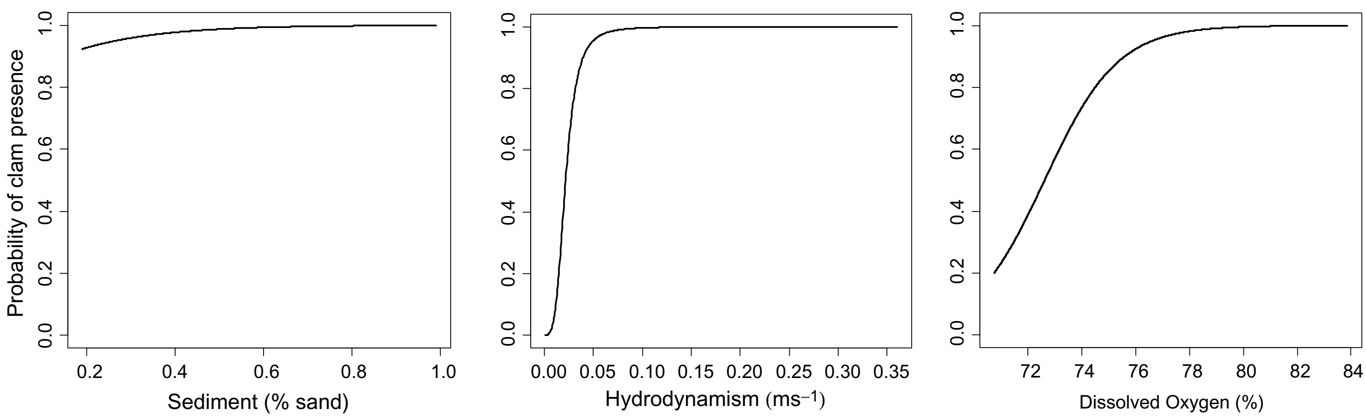

Fig. 4. Estimates of probability of presence of *T. philippinarum* plotted against share of sand in the sediment (%), hydrodynamism (m s^{-1}) and dissolved oxygen (% of saturation) for the logistic part of the HSC and HSM models. Predictions are for an average site, that is, with the other environmental parameters values set at their means in the study area.

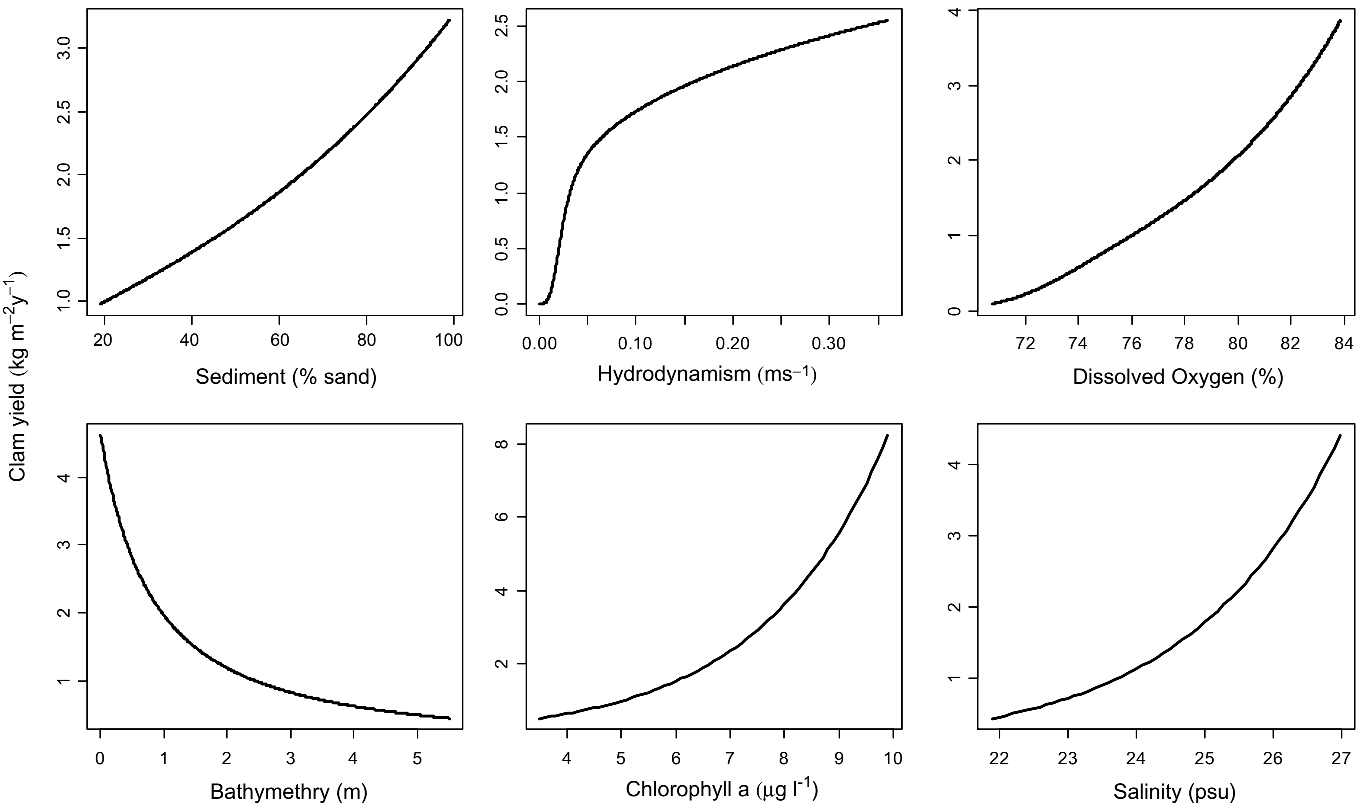

Fig. 5. Estimates of expected clam yield potential against environmental parameters values for the HSC model. Predictions are for an average site, that is, with the other environmental parameters values set at their means in the study area.

Table 3
Basic statistics relative to the estimates of the three HS models

HS model	Total yield (t year^{-1})	Farming surface (ha)
HSI	25,518	1027
HSC	32,022	1409
HSM	32,657	2111

farming areas (% of farming areas with annual yield >0) for annual predicted yields is reported in Fig. 7. As it can be noted in Fig. 8 both the HSI and HSC model provided a good prediction of the observed yield in the validation data set.

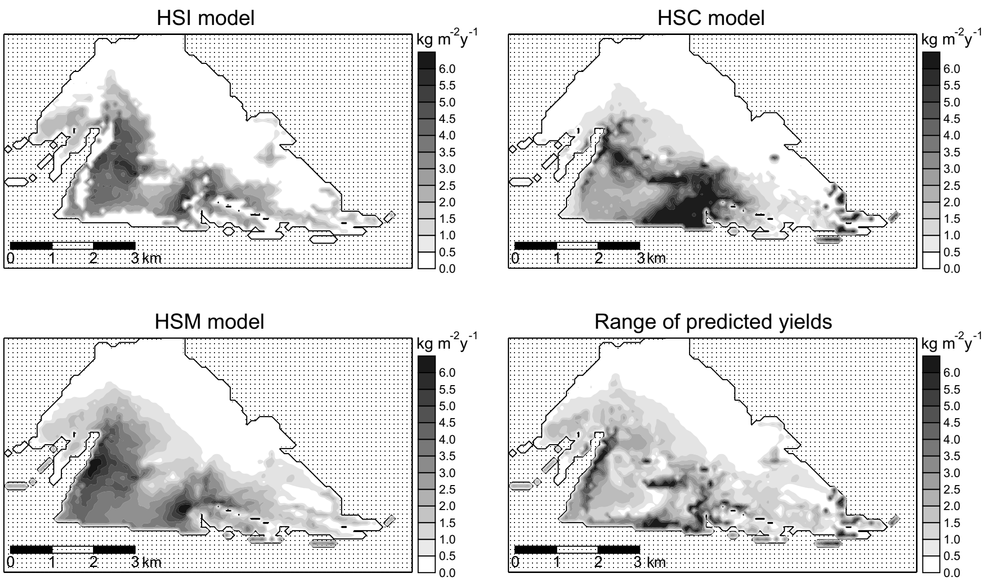

Fig. 6. Annual commercial yield of *T. philippinarum* in the Goro lagoon [kg m^{-2} year^{-1}] as estimated by the three HS models. The range of estimated yield (RP) is the difference between maximum and minimum estimates provided by the three models, namely $\mathrm{RP}(x,y) = \max(\mathrm{HSI}, \mathrm{HSC}, \mathrm{HSM})_{x,y} - \min(\mathrm{HSI}, \mathrm{HSC}, \mathrm{HSM})_{x,y}$.

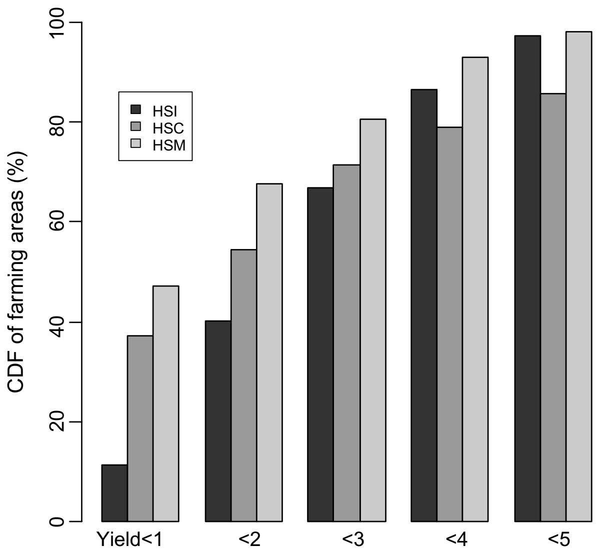

Fig. 7. Cumulative frequency distribution (CDF) of farming areas (% of total farming area) for annual commercial yield [kg m^{-2} year^{-1}] for the three HS models.

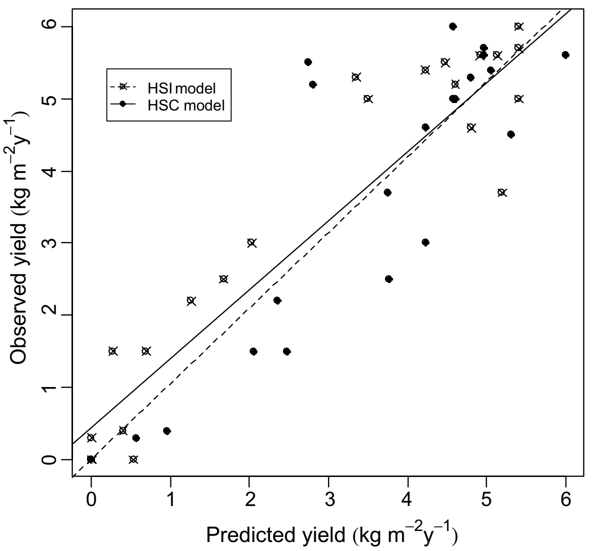

Fig. 8. Predicted vs. observed yield for the HSI model (open crossed circles) and the HSC model (black diamonds). The continuous line and the dashed line represent the linear regression between predicted and observed yield in the in the validation data set (HSI model: $p < 0.001$, $R^2 = 0.87$, intercept: 0.44 ± 0.25 and slope: 0.95 ± 0.07; HSC model – $p < 0.001$, $R^2 = 0.85$, intercept: 0.01 ± 0.30, slope: 1.05 ± 0.09). Six predicted and observed values are equal to 0 in the HSC, seven in the HSI model.

4. Discussion

In presence of multiple competing fishermen of *T. philippinarum* in the Sacca di Goro lagoon, the identification of suitable sites for farming and the corresponding yield potential is mandatory in order to enhance transparency in the subdivision and assignment of harvestable areas and to optimize the exploitation activities. In this context, habitat suitability models are useful tools in order to explore the relationship between habitat features and the distribution of the exploited species, to identify suitable sites for farming and the corresponding yield potential and to reduce the perception of subjective bias in the organization of the harvesting activities. The present analysis illustrated the structure and implementation of three habitat suitability models for the estimation of commercial yield of *T. philippinarum* in the Goro lagoon and the identification of suitable sites for farming within the lagoon. Each HS model approaches the issue of *T. philippinarum* habitat modelling in a different way. The habitat suitability index (HSI) model is based on expert knowledge and follows the U.S. Fish and Wild life Service Habitat Evaluation Procedures Program (Terrel, 1984; Bovee and Zuboy, 1988) which was primarily developed for freshwater and terrestrial species, but has also been applied to coastal areas (e.g. Brown et al., 2000). The habitat suitability conditional (HSC) model is a data-driven two-part model, with a logistic part determining presence/absence of *Tapes* and an ordinary regression part determining the yield potential when *Tapes* is present (yield part). The two-part model approach arises from the presence of zero-inflation in the calibration data set, that is a proportion of zeros so large that the response variable cannot be fit by using standard distributions (e.g. Normal, Poisson). The high prevalence of zero values is a consequence of the patchy-distribution of *Tapes* and reflects differences in suitability inside the lagoon. Conditional models for zero-inflated data have been used in a number of different context, from terrestrial (Welsh et al., 1996) to marine (Lo et al., 1992; Stefansson, 1992) settings. In particular, as in the case of the HSC model, conditional models combining the binomial distribution for the zero component and the log-normal distribution for the non-zero part have been used to model mackerel egg counts (Pennington, 1983) and the habitat suitability of common sole *Solea Solea* (Le Pape et al., 2004). Also the habitat suitability mixed (HSM) model is a two-part model, which incorporates both the data-driven approach of the HSC model and the expert knowledge approach of the HSI model.

It is worth noting that the goal of the HS models here presented was the planning of the concession regime, that is, the identification of areas with different degrees of suitability for clam farming in the lagoon and the corresponding yield potential under the assumption that the well-established day-to-day management and rearing practices are carried on. In contrast, the modelling approaches for extensive clam farming developed by Pastres et al. (2001), Solidoro et al. (2003) and Melià et al. (2003, 2004) were derived for purposes different from planning of the concession regime, that is to simulate clam population dynamics and to explore alternative management practices to maximize the yield of the economic values of rearing activities.

Our analyses show that potential yield predicted by the three models for the whole Goro lagoon is more than twice as much officially reported by the fishery, with the HSM model providing the highest potential yield (Table 3). This is due the fact that presently only one third of the 2600 ha of the Goro lagoon are exploited, while the models provided estimates also for areas outside the sites currently farmed. Moreover, to avoid the overexploitation of the resource, limitations also on total available catch (TAC) and on fishing effort are imposed by the regulatory agency. On the other hand, the potential yield predicted by the three models should not be taken as the maximum sustainable yield of the lagoon; in fact, to sustain the fishery in the long run and to conserve the natural environment by preserving the ecological processes and the irreplaceable ecosystem services, the regulation of the exploitation activities must account also for the ecological carrying capacity of the lagoon, as argued by Nizzoli et al. (2006). In this sense, the approaches proposed by Cellina et al. (2003), Melià et al. (2003), Pastres et al. (2001) and Solidoro et al. (2003) are probably more appropriate to address issues such as the long-term sustainability of the farming activities or the risk of dystrophic crises and of catastrophic collapse of the fishery. The three HS models illustrated in the present work simply provide a quantitative relationship between habitat features and suitability for clam farming and expected yield.

In the present work, the HSI and the HSC models have been validated on an independent dataset derived from the original study material, while it has not been possible to validate the HSM model on an independent dataset, as all the suitable observational data were necessary to calibrate the model. It is interesting to note that both the HSI and the HSC models provided a good prediction of the observed yield in the validation dataset, while the estimates of yield potential and suitable sites for farming provided by the two models are quite different (Table 3 and Fig. 6). As it can be noted from Fig. 6 the main differences in suitability and yield potential are found in the southern part of the lagoon, where the HSC model predicted a high yield potential while the HSI and HSM model only a moderate suitability for farming. This is related to conditions of low current velocity in the southern part of the lagoon (Fig. 3), which is more limiting in terms of predicted yield for the HSI and the HSM models (Fig. 2) than for the HSC model (Fig. 5). Obviously, more detailed field studies are needed for definitive conclusions on which model provides the most accurate estimate of the potential yield and site-specific suitability. The HSI model predicted a 40% of farming areas with moderate suitability for farming (yield $<2\ \mathrm{kg\ m^{-2}\ year^{-1}}$, Fig. 7), a proportion much lower of those predicted by the HSC and HSM model. Thus, while both the HSI and HSM models use the same PSSFs, the choice of weights in Eq. (1) is critical for model predictions. As it can be noted from the estimation of the logistic model (Table 2 and Fig. 4) share of sand in the sediment, dissolved oxygen and hydrodynamism play a major role in determining the occurrence of *Tapes* in the Sacca di Goro. It is clear that in other lagoons where farming of *T. philippinarum* is performed, the limiting factors of yield or the relative importance of the various habitat features could be different; it follows that while the approach for the data-driven HS models here presented remains valid also for other coastal lagoons, a re-calibration of these models on lagoon-specific data is strongly recommended. The estimates of yield potential of the HSC model presented in Fig. 5 seem to confirm the findings of Paesanti and Pellizzato (2000) on the optimal environmental conditions for clam growth and survival in the Goro lagoon (Fig. 2).

The weights of the PSSFs estimated through linear regression in the HSM model have been compared to the weights assigned by experts of clam farming in North Adriatic lagoons (Table 1); in the HSM model three weights (share of sand in the sediment, salinity and hydrodynamism) were significantly larger than zero while the other three variables (bathymetry, dissolved oxygen and chlorophyll "a") were irrelevant. This is probably linked to the limited importance, given their range of variation and the associated PSSFs, of dissolved oxygen, bathymetry and chlorophyll "a" in determining the yield potential. Nevertheless, it is worth noting that the weights estimated in the HSM model have the same relative importance of the weights assigned by experts of clam farming (Table 1), with hydrodynamism being the most important factor among the six initially considered in the estimation of the HSI and HSM models.

Which are the implications of this study in terms of modelling approach, of sampling strategy and of sampling size also when implementing similar HS models in other lagoons where extensive mollusc farming is performed? First, one of the most important issues when developing species management models is the selection of model variables. The ideal management model should rely on few key explanatory variables. Moreover, with a large suite of predictor variables, "overfitting" is likely to occur, to the extent that model predictions are very accurate on the context of the data set used to create it, but the model performs poorly when used elsewhere. The HSI model developed by Vincenzi et al. (2006) is based solely on expert knowledge and does not rely on observational study material on the distribution of the species and/or habitat features, which is usually expensive and time-consuming to collect. As rigorous data-driven HS models cannot be always readily developed for all species of commercial interest and in all sites where extensive farming is performed, the adoption of a HSI model to based on expert opinion can represent a good alternative to data-driven models, especially when solid knowledge has been gathered through practical experience and expert opinion can be considered sufficiently reliable and widely accepted. This is the case of *T. philippinarum* in the Sacca di Goro, where the parameter-specific suitability functions published by Paesanti and Pellizzato (2000) are representative of the shared knowledge of expert of clam farming in North Adriatic lagoons, and thus the HSI model represents a valid tool for a preliminary assessment of potential yield and the identification of suitable sites for farming. As for data-driven approaches like the HSC and HSM models, well designed sampling surveys are mandatory. Choosing the right sampling strategy is therefore crucial in order reduce the risk of making a poor prediction (Hirzel and Guisan, 2002), particularly for presence/absence estimates, as in the logistic part of the HSC and HSM models. As for sample size, Harrell (2001) proposed for regressions a minimum of 10 observations for each variable included in the models, with 20 observations for predictor being a good ratio for a reliable estimation of regression parameters. As it can be noted from the different prediction of the three HS models (Fig. 6 and Table 3) model choice is a critical step. Therefore, in the specific case of *T. philippinarum* in the Goro lagoon, which of the three HS models can be considered more reliable? While both the HSI and the HSC models provided a good fitting of yield data in the validation data set (Fig. 8) we believe that, given the available data, the adoption of the HSC model is presently the best option, as it is data-driven and therefore does not rely on expert knowledge as the HSI and HSM models which are subject to an intrinsic bias that cannot be quantified, even though representative of the shared knowledge of the fishermen operating in the lagoon. Moreover, as the sample size allowed for both the calibration and the validation of the model on independent data sets, all the stages in the modeling process of data-driven HS models (Rushton et al., 2004) were developed in an objective modelling framework. As for the application of the two HS models relying in part or totally on expert knowledge to other coastal lagoons (that is the HSI and the HSM models), we believe that the PSSFs proposed by Paesanti and Pellizzato (2000) can be likely considered still valid also in other North Adriatic coastal lagoons where *Tapes* is farmed (e.g. Venice lagoon). Anyway, as clam yield is heavily influenced by the trophic state of the lagoon (Pastres et al., 2001), a re-calibration of the scaling function to transform HS values into estimates of yield potential may be necessary to match the actual maximum yield when the model is implemented in a lagoon different from Goro. Moreover, in order to adopt one of the three HS modelling approaches presented in this work for yield estimation of other bivalve species and/or other lagoons, a rigorous statistical investigation is recommended in order to investigate whether other environmental factors (e.g. mean length of day light, nutrient cycle and load) play a major role in determining the suitability of lagoon area for bivalve farming.

We are confident that despite the intrinsic limitations, HS models based either on widely accepted expert knowledge or on rigorous statistical methods or combining both these two approaches can prove to be extremely useful to objectively identify the areas most suitable for aquaculture activities in coastal lagoons.

Acknowledgement

The authors are very thankful to Matteo Parelli for assistance in the preparation of the maps presented in this work and to Edorardo Turolla for his invaluable help in providing the data and the information to calibrate the models.

References

Aguilar-Manjarrez, J., Ross, L.G., 1995. Geographical Information System (GIS) environmental models for aquaculture development in Sinaloa state, Mexico. Aquaculture International 3, 103–115.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Barillari, A., Boldrin A., Pellizzato M., Turchetto, M., 1990. Condizioni ambientali nell'allevamento di *Tapes philippinarum*. In: *Tapes philippinarum* Biologia e Sperimentazione, pp. 183–195. Regione Veneto, E.S.A.V.

Bovee, K., Zuboy, J.R., 1988. Biological Report 88 (11). In: Proceedings of the Workshop Development, Evaluation of Habitat Suitability, Criteria, US Fish & Wildlife Service.

Brath, A., Gonella, M., Polo, P., Tondello, M., 2000. Analisi della circolazione idrica nella Sacca di Goro mediante modello matematico. Studi Costieri, 105–122.

Brown, S., Buja, K., Jury, S., Monaco, M., Banner, A., 2000. Habitat suitability index models for eight fish and invertebrate species in Casco and Sheepscot Bays, Maine. North American Journal Of Fisheries Management 20, 408–435.

Cellina, F., De Leo, G., Rizzoli, A.E., Viaroli, P., Bartoli, M., 2003. Economic modelling as a tool to support macroalgal bloom management: a case study (Sacca di Goro, Po river delta). Oceanologica Acta 26, 139–147.

Crow, E.L., Shimizu, K. (Eds.), 1988. Lognormal Distributions: Theory and Applications. Dekker, New York.

Deleo, J.M., 1993. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. In: Proceedings of the Second International Symposium on Uncertainty Modelling and Analysis, pp. 318–325. College Park, MD: IEEE Computer Society Press.

DHI, 1993. MIKE21 User Guide and Reference Manual. Danish Hydraulic Institute, Hoersholm, Denmark.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24, 38–49.

Fletcher, D., Mackenzie, D.I., Villouta, E., 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. Environmental and Ecological Statistics 12, 45–54.

Franklin, J., 1995. Predictive vegetation mapping: geographical modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19, 474–499.

Gibson, L.A., Wilson, B.A., Cahill, D.M., Hill, J., 2004. Modelling habitat suitability of the swamp antechinus (*Antechinus minimus maritimus*) in the coastal heathlands of southern Victoria, Australia. Biological Conservation 117, 143–150.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.

Harrell, F., 2001. Regression Modelling Strategies. Springer.

Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling? Ecological Modelling 157, 331–341.

Kapetsky, J.M., Hill, J.M., Worthy, L.D., 1988. A geographical information system for catfish farming development. Aquaculture 68, 311–320.

Kliskey, A.D., Lofroth, E.C., Thompson, W.A., Brown, S., Schreier, H., 1999. Simulating and evaluating alternative resource-use strategies using GIS-based habitat suitability indices. Landscape and Urban Planning 45, 163–175.

Le Pape, O., Guerault, D., Desaunay, Y., 2004. Effect of an invasive mollusc, American slipper limpet *Crepidula fornicata*, on habitat suitability for juvenile common sole *Solea Solea* in the Bay of Biscay. Marine Ecology Progress Series 277, 107–115.

Lindenmayer, D.B., Possingham, H.P., 1996. Modelling the inter-relationships between habitat patchiness, dispersal capability and metapopulation persistence of the endangered species, Leadbeater's possum, in south-eastern Australia. Landscape Ecology 11, 79–105.

Lo, N.C.H., Jacobson, L.D., Squire, J.L., 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences 49, 2515–2526.

McCullagh, P., Nelder, J.A., 1989. Generalised Linear Models, 2nd ed. Chapman and Hall, London.

Melià, P., Nizzoli, D., Bartoli, M., Naldi, M.C., Gatto, M., Viaroli, P., 2003. Assessing the potential impact of clam rearing in dystrophic lagoons: an integrated oxygen balance. Chemistry and Ecology 19, 129–146.

Melià, P., De Leo, G.A., Gatto, M., 2004. Density and temperature-dependence of vital rates in the Manila clam *Tapes philippinarum*: a stochastic demographic model. Marine Ecology Progress Series 272, 153–164.

Nath, S.S., Bolte, J.P., Ross, L.G., Aguilar-Manjarrez, J., 2000. Applications of geographical information systems (GIS) for spatial decision support in aquaculture. Aquacultural Engineering 23, 233–278.

Nizzoli, D., Bartoli, M., Viaroli, P., 2006. Nitrogen and phosphorous budgets durino a farming cycle of the Manila Clam *Ruditapes philippinarum*: an in situ experiment. Aquaculture 261, 98–108.

Paesanti, F., Pellizzato, M., 2000. *Tapes Philippinarum*. Veneto Agricoltura, 73 pp.

Pastres, R., Solidoro, C., Cossarini, G., Melaku Canu, D., Dejak, C., 2001. Managing the rearing of *Tapes philippinarum* in the lagoon of Venice: a decision support system. Ecological Modelling 138, 231–245.

Pellizzato, M., Da Ros, L., 2005. Clam farming quality as a management tool: a proposal based on recent studies in northern Adriatic lagoons. Aquaculture International 13, 57–66.

Pennington, M., 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39, 281–286.

Pereira, J.M.C., Itami, R.M., 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering & Remote Sensing 57, 1475–1486.

Rossi, R., 1989. Pesca e allevamento di vongola verace nella Sacca di Goro. Mare Nostrum 3, 16–18.

Rossi, R., 1996. Allevamento di vongola verace filippina (*Tapes philippinarum*): gestione della semina e del trasferimento in banco naturale per la ottimizzazione del raccolto. Ministero delle Politiche Agricole, III Piano.

Rossi, R., Paesanti, F., 1992. Successful clam farming in Italy. Proceedings of the National Shellfish Association 23, pp. 62–68.

Rushton, S.P., Ormerod, S.J., Kerby, G., 2004. New paradigms for modelling species distributions? Journal of Applied Ecology 41, 193–200.

Simenoni, U., Fontolan, G., Dal Cin, R., 2000. inamica sedimentaria dell'area di Goro (Delta del Po). Studi Costieri, 139–151.

Solidoro, C., Melaku Canu, D., Rossi, R., 2003. Ecological and economic considerations on fishing and rearing of *Tapes phillipinarum* in the lagoon of Venice. Ecological Modelling 170, 303–318.

Stefansson, G., 1992. Notes on the stock-dynamics and assessments of the Icelandic cod stock. ICES C.M. 1992/G:71.

Stefansson, G., 1996. Analysis of ground fish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Sciences 53, 577–588.

Store, R., Kangas, J., 2001. Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modelling. Landscape and Urban Planning 55, 79–93.

Terrel, J.W. 1984., Biological Report 85 (6). In: Proceedings of the Workshop: Fish Habitat Suitability Index Models. U.S. Fish and Wildlife Service.

United States Fish and Wildlife Service., 1981. Standards for the development of habitat suitability index models.

Vincenzi, S., Caramori, G., Rossi, R., De Leo, G.A., 2006a. A GIS-based habitat suitability model for commercial yield estimation of *Tapes philippinarum* in a Mediterranean coastal lagoon (Sacca di Goro, Italy). Ecological Modelling 193, 90–104.

Vincenzi, S., Caramori, G., Rossi, R., De Leo, G.A., 2006b. Estimating clam yield potential in the Sacca di Goro lagoon (Italy) by using a two-part conditional model. Aquaculture 261, 1281–1291.

Vives, P.T., 1996. Monitoring Mediterranean Wetlands: A Methodological guide, Slimbridge, UK and ICN, Lisbon, 150 pp.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B., 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 88, 297–308.