Ecological Modelling 222 (2011) 1471–1478

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy

Simone Vincenzi[a,*], Matteo Zucchetta[b], Piero Franzoi[b], Michele Pellizzato[c], Fabio Pranovi[b], Giulio A. De Leo[d], Patrizia Torricelli[b]

[a] *Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Viale G. P. Usberti 33/A, I-43125 Parma, Italy*
[b] *Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari Venezia, Castello 2737/B, 30122 Venezia, Italy*
[c] *AGRI.TE.CO sc Ambiente Progetto Territorio, Via Carlo Mezzacapo 15, 30175 Marghera, Italy*
[d] *Dipartimento di Scienze Ambientali, Università degli Studi di Parma, Viale G. P. Usberti 11/A, I-43125 Parma, Italy*

ARTICLE INFO

Article history:
Received 25 June 2010
Received in revised form 20 January 2011
Accepted 6 February 2011
Available online 4 March 2011

Keywords:
Ruditapes philippinarum
Venice lagoon
Random Forest
Yield
Habitat suitability

ABSTRACT

We present a modelling framework that combines machine learning techniques and Geographic Information Systems to support the management of an important aquaculture species, Manila clam (*Ruditapes philippinarum*). We use the Venice lagoon (Italy), the first site in Europe for the production of *R. philippinarum*, to illustrate the potential of this modelling approach. To investigate the relationship between the yield of *R. philippinarum* and a set of environmental factors, we used a Random Forest (RF) algorithm. The RF model was tuned with a large data set ($n = 1698$) and validated by an independent data set ($n = 841$). Overall, the model provided good predictions of site-specific yields and the analysis of marginal effect of predictors showed substantial agreement among the modelled responses and available ecological knowledge for *R. philippinarum*. The most influent environmental factors for yield estimation were percentage of sand in the sediment, salinity, and water depth. Our results agree with findings from other North Adriatic lagoons. The application of the fitted RF model to continuous maps of all the environmental variables allowed estimates of the potential yield for the whole basin. Such a spatial representation enabled site-specific estimates of yield in different farming areas within the lagoon. We present a possible management application of our model by estimating the potential yield under the current farming distribution and comparing it to a proposed re-organization of the farming areas. Our analysis suggests a reduction of total yield is likely to result from the proposed re-organization.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The Manila clam *Ruditapes philippinarum* (Adam and Reeve, 1850), which is of Indo-Pacific origin, was introduced in the Venice lagoon (Fig. 1) in the 80s as a culture species (Cesari and Pellizzato, 1985) and radically changed the exploitation of living resources in the lagoon. Within a few years, *R. philippinarum* became the most important exploited species in the lagoon, with a production reaching a peak of over 40,000 t y^{-1} at the end of the 90s, estimated from various sources and using expert knowledge by Pellizzato and Da Ros (2005). No official fishery landings data for the whole lagoon is available and yield potential is largely unknown, despite the relevant social, economic and environmental consequence of the exploitation activities.

Since the introduction, the exploitation of *R. philippinarum* has been carried out in a regime of free access. In 1999, the Province of Venice began a gradual shift to a concession regime, i.e., to a system where harvesting areas are divided by the regulatory agency among a number of concessions, each managed by local clam fishermen under a strict set of rules on access limitation and exploitation effort. Technically, concessions are divided in farming (i.e., where clams are seeded) and fishing (i.e., where clams are naturally recruited) areas. In the following, we will use the word concession without further differentiating between farming and fishing areas. In 2007, about 42 km^2 of the Venice lagoon were given in concession to fishermen for harvesting of *R. philippinarum* (Fig. 2a and Table 1).

However, the transition from uncontrolled fishing to a "culture-based fishery" based on correct and sustainable rearing procedures, while being successful in reducing production of *R. philippinarum*, revealed to be more complex than expected and cannot be considered successfully completed (Pellizzato and Da Ros, 2005). The Province of Venice is willing to reduce the number of fishermen operating in the lagoon (from about 900 to 600) and to remodel and reduce the areas given in concession to clam fishermen (G.R.A.L., 2006, 2009; Province of Venice, 2009) to reduce

* Corresponding author. Tel.: +39 0521 905696; fax: +39 0521 906611.
E-mail address: simone.vincenzi@nemo.unipr.it (S. Vincenzi).

0304-3800/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2011.02.007

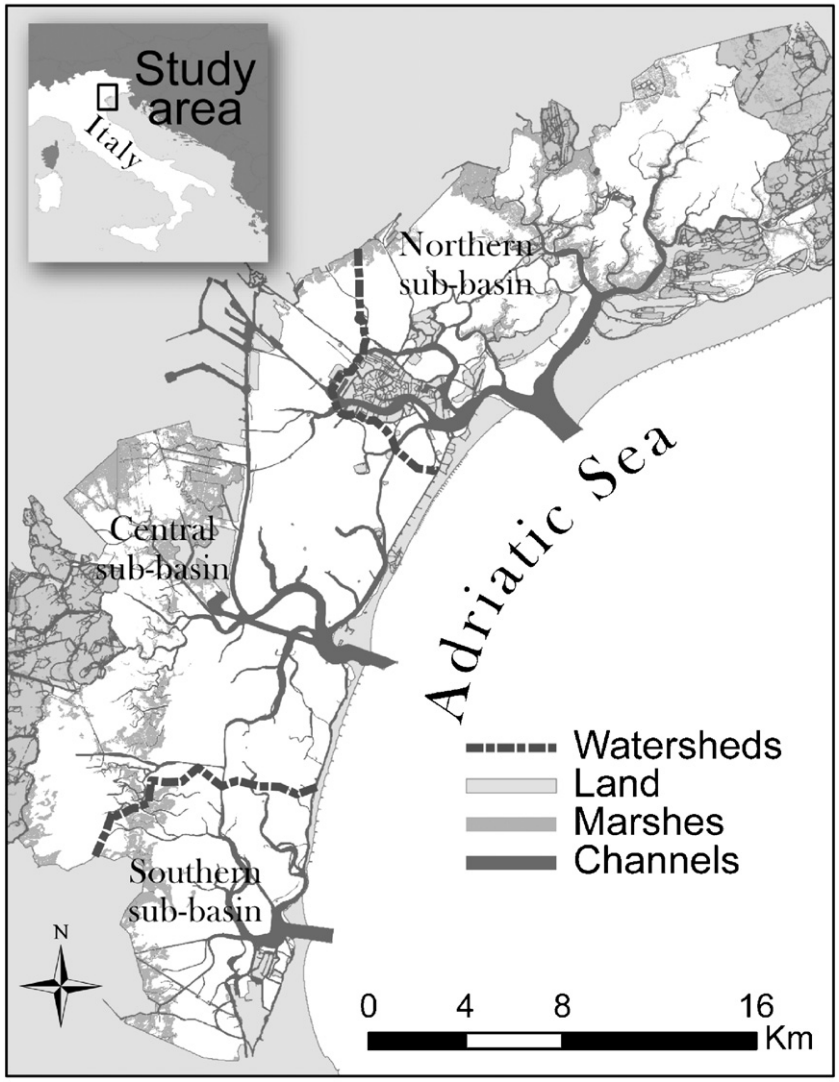

Fig. 1. Map of the Venice lagoon (Italy). The basin can be divided by two watersheds in three main sub-basins, Northern, Central and Southern.

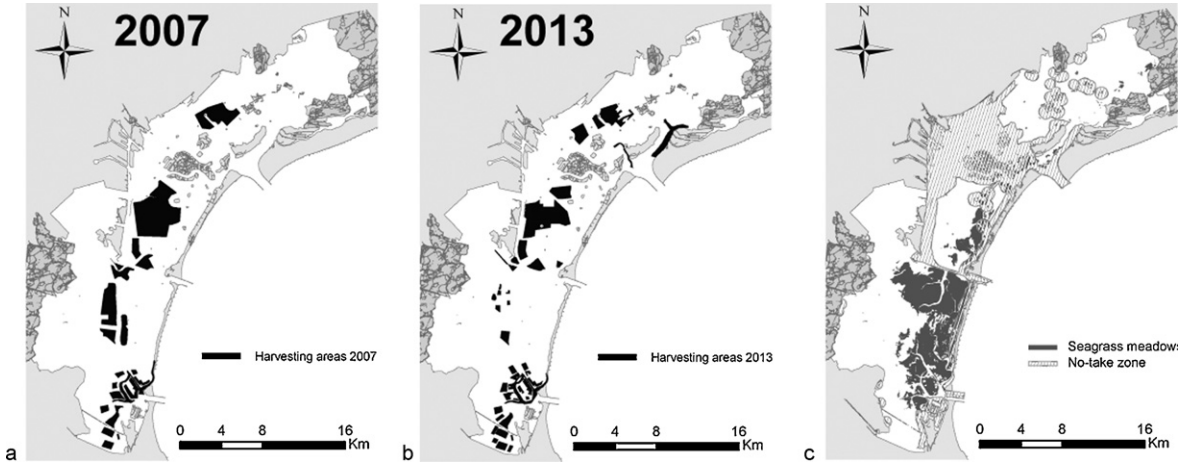

Fig. 2. Areas given in concession by local authorities in 2007 (Panel a) and, according to the reshaping plan proposed by the Province of Venice, in 2013 (Panel b). Panel c shows no-takes zones for sanitary reasons (sites highly polluted) and sites of conservation concern (seagrass meadows).

health risks linked to industrial pollutants or urban waste, minimize the environmental impacts of bottom dredging, such as the loss of sediments (e.g., Molinaroli et al., 2007), increase of water turbidity and movements of nutrients and pollutants (Pranovi et al., 2004; Sfriso et al., 2005), protect habitats of conservation concern, such as seagrass meadows (Fig. 2c), and maximize production in order to minimize fishing effort, both in space and time.

Table 1
Yield predicted by the Random Forest model for the whole lagoon of Venice, the three sub-basins and concession areas for 2007 and 2013. Total yield for areas given in concession is expected to decrease after the remodelling plan (2013).

	Total yield ($t\,y^{-1}$)	Area (km^2)	Yield per unit area ($g\,m^{-2}\,y^{-1}$)
Whole basin	69,883	216.30	323.08
Northern sub-basin	14,743	83.60	176.35
Central sub-basin	41,679	99.00	421.00
Southern sub-basin	13,461	33.70	399.44
Harvesting areas 2007	22,731	42.59	533.72
Harvesting areas 2013	18,277	35.52	514.54
Harvesting areas 2007 North	1,117	5.98	186.71
Harvesting areas 2007 Central	17,706	26.56	666.65
Harvesting areas 2007 Southern	3,908	10.05	388.87
Harvesting areas 2013 North	3,316	8.38	395.75
Harvesting areas 2013 Central	11,759	16.31	720.97
Harvesting areas 2013 Southern	3,862	10.45	369.57

In this context, the identification of suitable harvestable grounds and a reliable estimation of site-specific commercial yield potentials are necessary to guarantee a sustainable fishery, to improve economic efficiency of clam farming, ensure an equitable share of exploitable areas to competing subjects interested in the exploitation of *R. philippinarum* and to foster transparency in the decision making process aimed at planning the future exploitation activities.

Habitat suitability (HS) models or models predicting species distribution (the two definitions will be used interchangeably) constitute good tools supporting decision-making within the framework of applied biology. HS models have been often used to improve our understanding of species–habitat relationship in space and time and to predict the likelihood of occurrence and abundance of a species using habitat attributes affecting its survival, growth and reproduction (e.g., Guisan and Thuiller, 2005; Hirzel et al., 2006; Santos et al., 2006). Habitat suitability approaches have also been used for identifying appropriate sites for mollusk farming in North-America and Mexico (e.g., Kapetsky et al., 1988; Aguilar-Manjarrez and Ross, 1995). Vincenzi et al. (2006a,b, 2007) developed simple HS models for the estimation of yield potential of *R. philippinarum* in the Sacca di Goro lagoon (North Adriatic, Italy) by using semi-empirical and zero-inflated regression models. In recent years, machine learning methods, such as classification and regression trees (Džeroski and Drumm, 2003; Seoane et al., 2005) artificial neural networks (ANN, Pearson et al., 2002; Dedecker et al., 2004) and Random Forests (Benito Garzón et al., 2007, 2008) have been proposed for the development of spatial distribution models. Machine learning methods are capable of detecting complex relationships among model variables without making a priori assumptions about the type of relationship, such as a linear dependence on predictors, and are able to process complex and noisy data (Recknagel, 2001).

In this work, we used a Random Forest algorithm (Breiman, 2001) to explore the relationship between the yield potential of *R. philippinarum* and several environmental factors deemed important for the occurrence and abundance of the species. Several studies have shown that Random Forest (RF) models, based on an automatic combination of tree predictors, often reach top predictive performances compared to other methodologies (e.g., Prasad et al., 2006; Cutler et al., 2007). Our paper is organized as follows: after a brief description of the study area, of the environmental factors linked to the occurrence and abundance of *R. philippinarum* and of available data, we briefly illustrate the main features of the Random Forest model and proceed with the calibration and validation of the model by using two independent data sets relative to year 2007. Then, we apply the Random Forest model to the Venice lagoon to obtain estimates of potential yield inside and outside the areas given in concession in 2007. In addition, we predict the yield potential of *R. philipparum* in areas that, according to the remodelling plan proposed by local authorities, will be given in concessions in 2013. Finally, we discuss the relevant features, limitations and further development of the Random Forest approach.

2. Materials and methods

The resolution (i.e., operational scale) chosen for the study was 100×100 m cells (site), for a total of 45,443 cells.

2.1. Study area

The Venice Lagoon is located in the Northern Adriatic Sea and is the largest lagoon in the Mediterranean basin, with an area of about 550 km^2 including emerged lands (Fig. 1). The Venice lagoon is a shallow coastal ecosystem (average depth 1.2 m, Molinaroli et al., 2007), where large areas, covering about 75% of the total surface (Molinaroli et al., 2009), are connected by a network of channels, whose depth is mostly <2 m (Solidoro et al., 2002). Deeper channels connected to three wide mouths (Lido, Malamocco and Chioggia) maintain the lagoon–sea communication and allow tidal flows to enter the lagoon, within a range of ±50 cm during spring tides (Umgiesser et al., 2004). The basin can be divided by two watersheds in three main sub-basins (Northern, Central and Southern, Fig. 1) (Solidoro et al., 2004). The bottom sediments of the basin consist mainly in clayey silt, with a mean mud content of about 80% in dry weight, showing a north–south decreasing pattern in mud content (Molinaroli et al., 2007). Water salinity is influenced by freshwater inputs.

2.2. Yield data

The latest available commercial yield data in the Venice lagoon were relative to year 2007 (Fig. 2a) for ~65% of areas in concession for exploitation of *R. philippinarum* (Michele Pellizzato, unpublished data). While there exists a substantial variation in productivity within an harvesting area, yield information is usually aggregated at the concession level. Therefore, the spatial distribution of clam yield within a concession area was derived by using biomass data of *R. philippinarum* gathered in a number of independent studies described in G.R.A.L. (2009, and references therein) as a proxy indicator of yield, as described hereafter. Mean size of areas given in concession in 2007 ranged from 2.5 to 5.5 km^2. In concession areas for which total landings in year 2007 and biomass data were available ($n = 13$, out of the total 24 concession areas), a relative biomass index was computed by dividing the biomass of each 100×100 m cell by the total biomass of the area. The yield distribution for each 100×100 m cell within a concession was then obtained by multiplying the annual production for the area by the relative biomass index.

2.3. Environmental factors

As reported by Paesanti and Pellizzato (2000), *R. philippinarum* is quite tolerant to mid-high variations of relevant habitat variables typical of coastal lagoons, such as salinity, temperature, dissolved oxygen and turbidity. In the present study, we chose a set of nine environmental variables that are known to affect clam abundance and yield (Barillari et al., 1990; Paesanti and Pellizzato, 2000; Vincenzi et al., 2006b, 2007), namely: percentage of sand in the sediment (Sand), dissolved oxygen (DO), salinity (Sal), water speed (Speed), chlorophyll "*a*" (Chla), turbidity (Turb), residence time (RT), water temperature (T) and water depth (WD).

Several studies, sampling surveys and monitoring programs have been carried out in the Venice lagoon in the last five years

to gather information on the physical, chemical and hydrological characteristics of the lagoon. Point values of sand content ($n = 150$) provided by the Venice Water Authority (MAG ACQUE – SELC, 2005; MAG ACQUE – Thetis, 2005) were interpolated by using ordinary kriging on the grid chosen (100 × 100 m cells), after fitting the best model on the experimental variogram, using the 'gstat' package (Pebesma, 2004) for the software R (R Development Core Team, 2009). Data of water quality ($n = 15$) obtained from the Venice Water Authority (MAG ACQUE – SAMA – Thetis, 2007) were interpolated by using ordinary kriging to create maps of water temperature, turbidity, salinity, chlorophyll-"*a*", and dissolved oxygen. Available data for 2007 were interpolated to generate monthly maps of these environmental factors, then the map of yearly mean values was computed for each factor by averaging monthly data. Hydrodynamism for a typical tidal cycle was acquired from thematic maps with categorical discretization provided by Molinaroli et al. (2007), while residence time data were given by Cucco et al. (2009).

2.4. The calibration and validation datasets

For 2539 cells within the harvesting areas we had information about both yield for year 2007 and the nine environmental variables listed in Section 2.2. The dataset was randomly split in a calibration dataset (CD, $n = 1698$) and a validation dataset (VD, $n = 841$). The relationship between yield data and the environmental variables was modelled by using the calibration dataset and the quality of predictions was then assessed by using the validation dataset, as described in Section 2.4.

2.5. Statistical methods

Collinearity among environmental variables was tested by hierarchical cluster analysis using squared Spearman correlations (ρ^2) as similarity measure.

To model the relationship between the nine environmental variables and site-specific yield of *R. philippinarum* in the Venice lagoon, we used the Random Forest algorithm (RF, Breiman, 2001) implemented in the "randomForest" package (Liaw and Wiener, 2002) within the R environment (R Development Core Team, 2009).

RF is an ensemble learning technique developed by Breiman (2001) based on a combination of a large set of decision trees. Each tree is trained by selecting a random set of variables and a random sample from the training dataset (i.e., the calibration data set). Three training parameters needs to be defined in the Random Forest algorithm: *ntree*, the number of bootstrap samples for the original data (the default value is 500); *mtry*, the number of different predictors tested at each node (which, in this specific case, can be 9 at most, i.e., as many as the environmental covariates); *nodesize*, the minimal size of the terminal nodes of the trees, below which leaves are not further subdivided.

As the response variable (yield of *R. philippinarum*) was numerical, we confine our attention to regression Random Forest models. Among the predictors, only hydrodynamism and residence time entered the model as categorical variables, as they were acquired from thematic maps in Molinaroli et al. (2007) and Cucco et al. (2009). The algorithm performs as follows (for full details see Breiman, 2001):

(1) *ntree* bootstrap samples X_i (i = bootstrap iteration) are randomly drawn with replacement from the original dataset (training dataset), each containing approximately two third of the elements of the original dataset X (in our case approximately 1132 elements out of 1639). The elements not included in X_i are referred to as out-of-bag data (*OOB*) for that bootstrap sample.
(2) For each boostrap sample X_i an unpruned regression tree is grown. At each node, rather than choosing the best split among all predictors as done in classic regression trees, *mtry* variables are randomly selected and the best split is chosen among them.
(3) New data (out-of-bag elements) are predicted by averaging the predictions of the *ntree* trees, as explained below.

Out-of-bag elements are used to estimate an error rate, called the out-of-bag (OOB) estimate of the error rate (ERR_{OOB}), as follows:

(i) At each bootstrap iteration, the out-of-bag elements are predicted by the tree grown using the bootstrap samples X_i.
(ii) For the ith element (y_i) of the training dataset X, all the trees are considered in which the ith element is out-of-bag. On average, each element of X is out-of-bag in one-third of *ntree* iterations. On the basis of the random trees an aggregated prediction g_{OOB} is developed. The out-of-bag estimate of the error rate is computed as $ERR_{OOB} = (1/ntree)\sum_{i=1}^{ntree}[y_i - g_{OOB}(X_i)]^2$.

The ERR_{OOB} help prevent overfitting and can also be used to choose an optimal value of *ntree* and *mtry*, by selecting *ntree* and *mtry* that minimize ERR_{OOB}. Therefore, we first chose the optimal values of *ntree* and *mtry* which minimize ERR_{OOB} and then we proceeded to develop the Random Forest model.

The "randomForest" package can also produce a measure of variable importance by looking at the deterioration of the predictive ability of the model when each predictor is replaced in turn by random noise. The resulting deterioration is a measure of predictor importance. The most widely used score of importance of a given variable in regression RF models is the increasing in mean of the error of a tree (mean square error, MSE). In addition, partial plots provide a way to visualize the marginal effect of environmental variables in Random Forests estimates of potential yield.

As ERR_{OOB} is an unbasied estimate of the generalization error, in general it is not necessary to test the predictive ability of the model on an external data set (Breiman, 2001). However, we preferred to use an independent dataset (the VD data set with 841 measures of yield and environmental variables) to perform an external validation of the predictive capabilities of the RF model.

2.6. Predictive maps

Once calibrated and validated, the resulting RF model was applied to the entire lagoon of Venice to obtain an estimate of the potential yield of *R. philippinarum* for the whole basin. The potential yield of sub-basins and of all harvesting areas given in concession in 2007 were also computed (Fig. 2a). The RF model was finally used also to estimate yield potential of the areas (Fig. 2b) where clam harvesting will be allowed starting from 2013 (G.R.A.L., 2009).

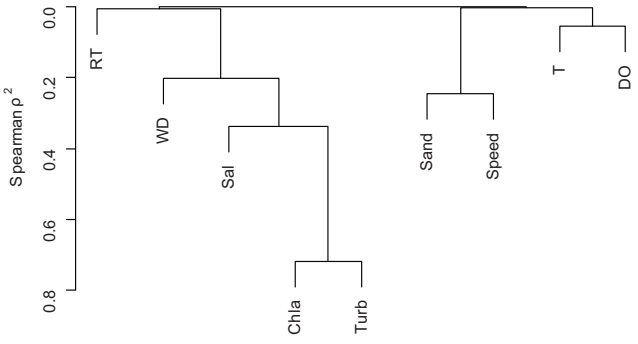

Fig. 3. Hierarchical clustering using squared Spearman correlation (ρ^2) of environmental variables as similarities. Sand, share of sand in the sediment; Sal, salinity; WD, water depth; DO, dissolved oxygen; Turb, turbidity; RT, residence time; Chla, chlorophyll-"a"; T, water temperature; Speed, water speed.

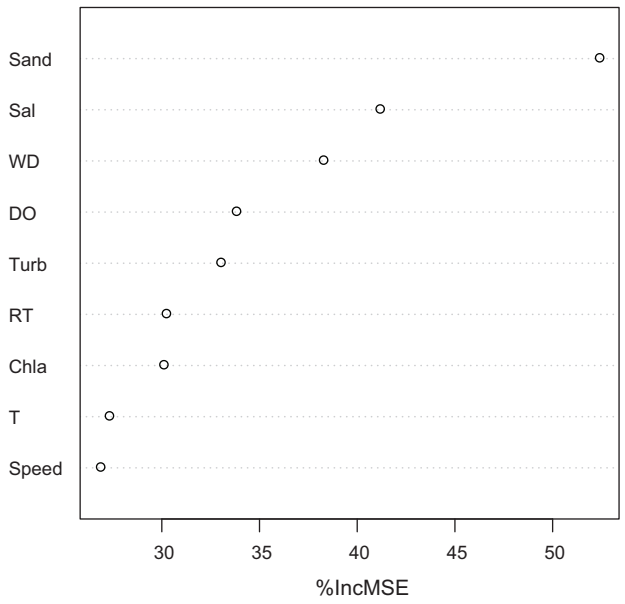

Fig. 4. Variable importance plot generated by the random forest algorithm included in the randomForest package for R software. The plot shows the variable importance measured as the increased mean square error (%IncMSE), which represents the deterioration of the predictive ability of the model when each predictor is replaced in turn by random noise. Higher %IncMSE indicates greater variable importance.

3. Results

A strong correlation was found between chlorophyll "*a*" and turbidity (Fig. 3).

The out-of-bag estimates of the error rate (ERR_{OOB}) were used to select the optimum Random Forest parameters (*mtry* = 3, *ntree* = 700, *nodesize* = 5). For the calibration dataset (CD), the Random Forest was able to explain a large proportion of the variance of yield of *R. philippinarum* ($r^2_{CD} = 0.99$). The out-of-bag validation results were examined, for which $r^2_{OOB} = 0.93$.

Fig. 4 shows the ranking of predictors by their importance. Only few of the descriptors contributed noticeably to the estimation of yield of *R. philippinarum*, namely percentage of sand in the sediment (Sand), salinity (Sal) and water depth (WD). In decreasing order of importance the other predictors included in the RF model were: dissolved oxygen, turbidity, residence time, chlorophyll-"*a*", temperature and current speed.

Partial plots representing the marginal effect of single variables included in the RF model on estimates of yield of *R. philippinarum* are shown in Fig. 5.

The Random Forest model provided a good prediction of production values in the validation dataset ($r^2_{VD} = 0.74$, Fig. 6). The results of the RF model tended to overestimate yield in low yield sites (Fig. 6). For the whole 216 km^2 of Venice lagoon (excluding emerged lands, seagrass meadows and areas in which harvesting is forbidden for sanitary reasons) the RF model estimated a potential yield of about 70,000 t y^{-1} (Table 1), with an average yield of 321 g m^{-2} y^{-1} (Table 1).

Maximum yield potential was predicted in the central part of the Central sub-basin, while the Northern sub-basin presented the lowest yield per unit area (Fig. 7 and Table 1). The estimated yield for the areas harvested in 2007 was about 22,700 t (Table 1), with an average yield of 567 g m^{-2} y^{-1} (Table 1). 77% of the total yield of the areas given in concession in 2007 was located in Central sub-basin, where only 63% of the surface of the harvesting areas were located (Table 1). The proposed reshape of concessions led to similar level of average yield potential for the Central and the Southern sub-basins, and to an increase of average yield potential in the Northern sub-basin (Table 1). Due to a reduction of harvesting surface areas, total yield will decrease to about 18,300 t, with an average yield of 514 g m^{-2} y^{-1}(Table 1).

4. Discussion

We showed that the application of a Random Forest model provides an effective methodology for identifying suitable sites and quantifying site-specific yields for the exploitation of an aquaculture species. Random Forests, both classifier and regression, have

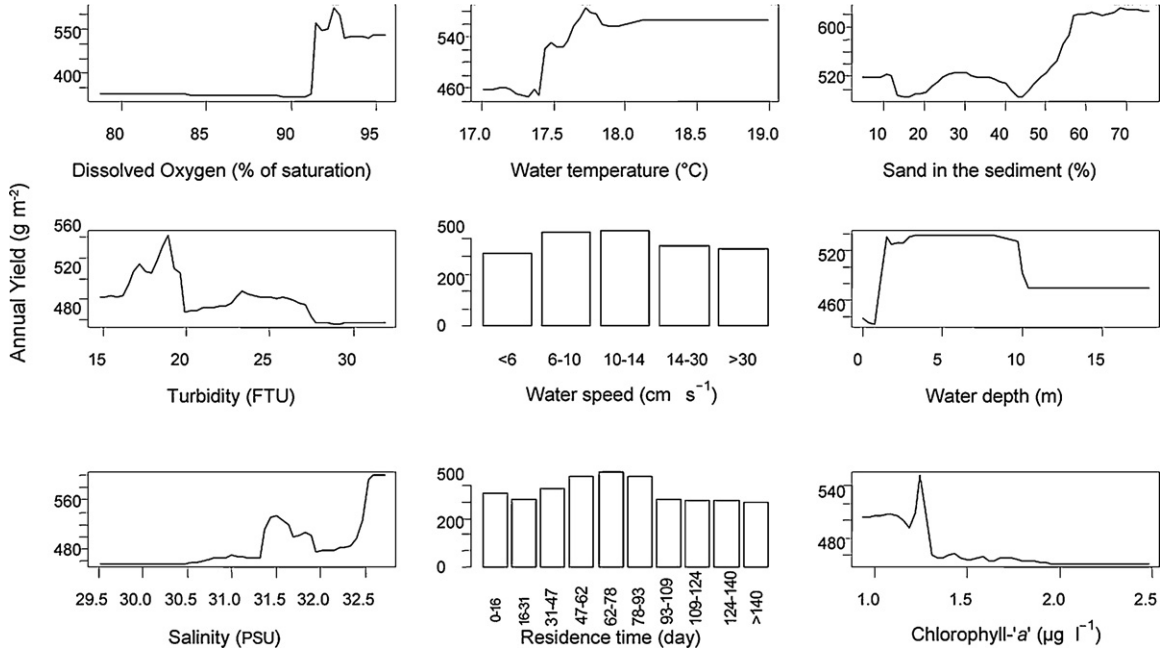

Fig. 5. Partial plots representing the marginal effect of single variables included in the RF model on estimates of yield of *R. philippinarum* while averaging out the effect of all the other variables. In a partial plots of marginal effects, only the range of values (and not the absolute values) can be compared between plots of different variables.

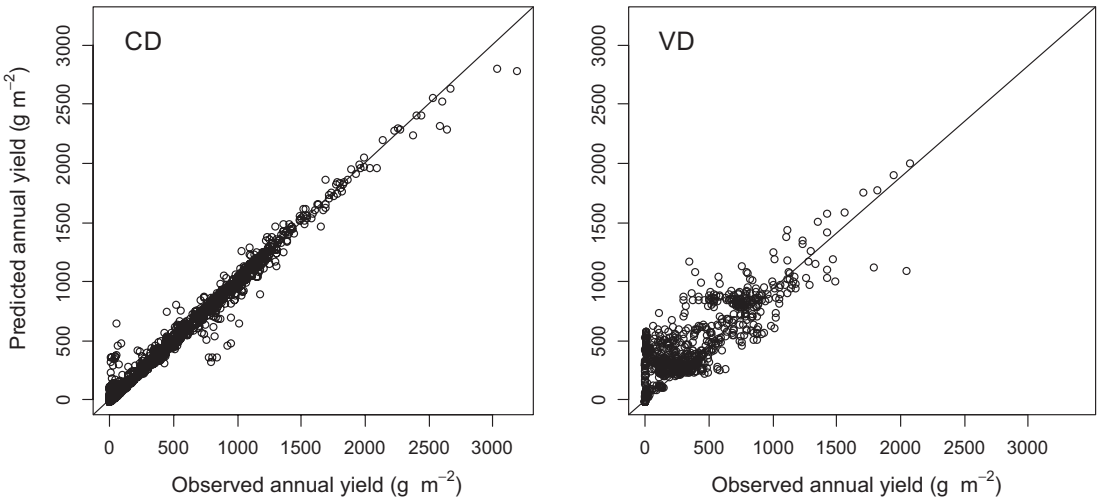

Fig. 6. Application of the RF model to the calibration data set (CD, $r^2 = 0.99$, $p < 0.01$) and the validation data set (VD, $r^2 = 0.74$, $p < 0.01$). Predictions of the RF models are more uncertain in sites with low yield potential.

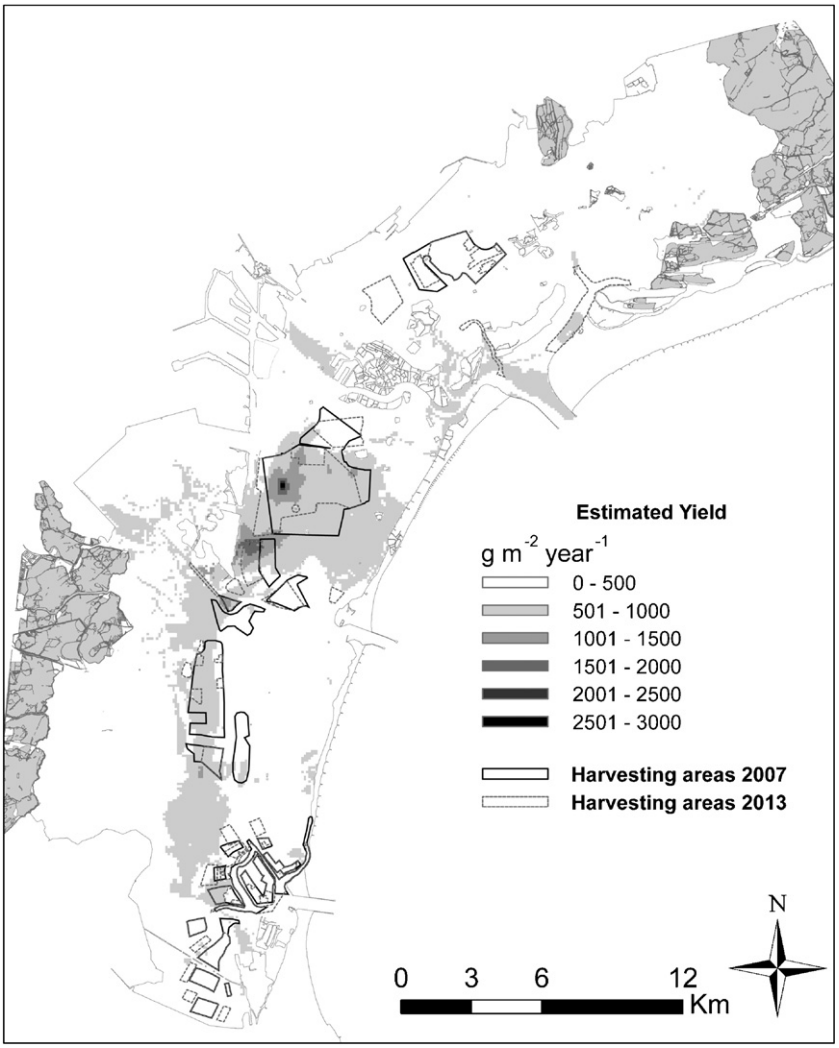

Fig. 7. Map showing the prediction of yield of *R. philippinarum* in the Venice lagoon obtained by the application of the Random Forest model. Areas given in concession in 2007 and in 2013, according to the reshaping plan proposed by the Province of Venice, are showed.

been already used in several applicative context and were recently applied to predict plant and animal habitat suitability (e.g., Iverson et al., 2005; Lawler et al., 2006, 2009; Benito Garzón et al., 2007, 2008).

The results of the RF model tended to overestimate yield in low yield sites and were more accurate for increasing yields (Fig. 6). In order to assess if the overestimation was dependent on the particular subsets of the data used for calibration and validation, we

re-fitted the RF model using other splits, but the RF model parameters and predictions did not substantially change. As sites of low yield are obviously sites of low commercial interest, the relative underperformance of the model in those sites does not hinder the application of the model for the identification of suitable sites for clam harvesting and the estimation of site-specific yield potential. However, this could substantially overestimate yield outside concession areas, as many of the sites not harvested are not suitable for *R. philippinarum* (Fig. 7). Therefore, the prediction of 70,000 t for the whole lagoon must be taken with great caution and only represent a gross estimation of potential yield for the whole Venice lagoon.

Another important aspect of modelling involves the evaluation and biological interpretation of the results. In the case of *R. philippinarum*, the results are encouraging. The marginal effects of single environmental factors (Fig. 5) confirm the findings of Barillari et al. (1990), Paesanti and Pellizzato (2000) and Vincenzi et al. (2006a,b, 2007) on their optimal values for *R. philippinarum* growth and survival in North Adriatic lagoons.

Share of sand in the sediment play a major role in determining the yield of *R. philippinarum* in the Venice lagoon (Fig. 4). Several studies showed that greater growth rates, maximum size and successful juvenile settlement in *R. philippinarum* occur in sandy sediments than in sediments with higher fraction of silt (Barillari et al., 1990; Rossi, 1996; Melià et al., 2004). Moreover, share of sand in the sediment was the most important factor in determining both presence/absence of the species and its abundance also in the habitat suitability models developed by Vincenzi et al. (2006a,b, 2007) for *R. philippinarum* in the Sacca di Goro lagoon.

The other two most important predictors were salinity and water depth. According to Paesanti and Pellizzato (2000), the optimal values for salinity range between 25 and 35 PSU. In the RF model, the marginal effect of salinity increased for salinity values greater than 30 PSU. As for water depth, areas shallower than 0.5 m are not suitable for *R. philippinarum*, as clams could emerge with low tide and unfavorable wind conditions and are also vulnerable to predation by birds. As harvesting is carried out by mechanical dredging of the bottom, sites with water depth greater than 10 m are in general not suitable for commercial exploitation, especially in sites where clams are seeded. Vincenzi et al. (2006a) found that the three most important predictors of yield of *R. philippinarum* were share of sand in the sediment, salinity and water speed. Surprisingly, in the RF model, water speed was the least important factor in terms of marginal effect (Fig. 4). This further confirms the necessity of site-specific calibrations of correlative models of species abundance. Optimal sites for clam farming are characterized by intermediate currents, typically from 0.3 to 1.5 m s^{-1} (Paesanti and Pellizzato, 2000).

The predictive maps obtained by the application of the RF model fitted on the calibration data set (CD) showed strikingly site-specific differences in potential yield (Fig. 7). Turolla et al. (2008) estimated by using expert knowledge in c.a. 27,500 t the total production for 2007, accounting for unauthorized fishing, that is fishing occurring in polluted areas (Pellizzato and Da Ros, 2005). It is worth noting that no official landing data for the whole lagoon are available.

Considering the 16% reduction of the surface of exploited area due to environmental (i.e., presence of seagrass meadows), health-risk constraints and the attempt to maximize production for unit area given in concession, the total production estimated for the new configuration of concessions seems to be adequate with respect to local managers expectations (13,000 t, G.R.A.L., 2006). However, some areas given in concession in 2007 and also maintained in the 2013 plan, mostly in the Northern and Southern sub-basins, showed low potential yield, according to model predictions. It is clear that model predictions for areas which will be given in concession in 2013 did not take into account possible changes in environmental conditions in the lagoons, both natural and human-induced, that might alter the distribution of suitable sites within the lagoon, and further restrictions on exploitation activities (e.g., fishing days, type of harvesting tools, individual quotas, etc.). In addition, as the RF model was fitted on data from a single year (2007), additional studies should be carried out to investigate the relative importance of annual fluctuations of the biogeochemical and hydrodynamic factors included in the RF model in determining the observed inter-annual variability of the commercial yield of *R. philippinarum* in the Venice lagoon (Pellizzato and Da Ros, 2005) and to assess if the inclusion of other biogeochemical parameters could further improve the predictions of the RF model.

The application of a Random Forest model (both classifier and regression) to predict the distribution (occurrence and abundance) of a species is particular useful when there are complex interactions between predictors and response variable (in our case the yield *R. philippinarum*) and the possibility of highly correlated predictor variables. A further advantage of the RF model is that this statistical learning modelling framework does not require assumptions of normality of model variables and can deal with non-linear relationships. Here, the application of a RF model was particularly recommended, as previous models for *R. philippinarum* developed for specific application to the Sacca di Goro lagoon (Italy) clearly showed the non-linearity of the relationship between environmental factors and yield potential (Vincenzi et al., 2006a, 2007) and at least two environmental variables were highly correlated.

In the context of aquaculture, correlative approaches, in which the relationship between the presence or abundance of a species and environmental conditions is statistically analyzed, are most useful when spatially-explicit information on the occurrence or abundance of the investigated species is available and a measure of the suitability of sites for harvesting is the most important result to be obtained from the analysis.

Both mechanistic and correlative approaches are currently used to model species distribution (see Buckley et al., 2010 for a recent review). The goal of the correlative RF model presented here was the identification of areas with different degree of suitability for clam farming and the corresponding yield potential under the assumption that the well-established day-to-day management and rearing practices are carried on. On the contrary, mechanistic approaches were used by Pastres et al. (2001), Solidoro et al. (2003), Melià et al. (2003, 2004) and Spillman et al. (2008) to analyse optimal management strategies or to identify suitable rearing sites for *R. philippinarum*. These models, based on functional traits and physiological constraints and making use of complex 3D models of water circulation, although costly to design, calibrate and validate, are particularly appropriate to address issues such as long-term sustainability of exploitation activities, effects of alternative rearing strategies (e.g., seeding size and density), risk of dystrophic crises and algal blooms. For instance, in the Venice lagoon an enormous amount of seed is needed each year (7 billion individuals, Pellizzato and Da Ros, 2005), and while seed is produced naturally in high abundance, densities of juveniles are greatest in areas characterized by high organic pollution. Thus, great attention must be devoted to the implementation of efficient rearing strategies and mechanistic models can be a valuable tool for the evaluation of alternative strategies. In this case, our RF model could: (i) provide information on site-specific yield potential, especially for areas outside the concession areas and thus less investigated with dynamic models; (ii) guide the application of the mechanistic model, for instance by limiting the costly application of computing-intensive mechanistic models to sites where yield potential is above a certain threshold; and (iii) suggest particular traits or processes to include in a mechanistic model.

Acknowledgment

The authors thank the Venice Water Authority (Magistrato alle Acque di Venezia) for providing water quality and sediment data. Simone Vincenzi concluded this work while visiting the Center for Stock Assessment Research (CSTAR), a partnership between the Fisheries Ecology Division, Southwest Fisheries Science Center, NOAA Fisheries and the University of California Santa Cruz, supported by a research grant provided by "Fondazione Luigi e Francesca Brusarosco".

References

Aguilar-Manjarrez, J., Ross, L.G., 1995. Geographical information system (GIS) environmental models for aquaculture development in Sinaloa state, Mexico. Aquacult. Int. 3, 103–115.

Barillari, A., Boldrin, A., Pellizzato, M., Turchetto, M., 1990. Condizioni Ambientali Nell'Allevamento Di Tapes Philippinarum, Tapes Philippinarum Biologia E Sperimentazione. E.S.A.V, Regione Veneto (In Italian).

Benito Garzón, M., Sánchez De Dios, R., Sainz Ollero, H., 2007. Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. Ecography 30, 120–134.

Benito Garzón, M., Sánchez De Dios, R., Sainz Ollero, H., 2008. Effects of climate change on the distribution of Iberian tree species. Appl. Veget. Sci. 11, 169–178.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J., Sears, M.W., 2010. Can mechanism inform species' distribution models? Ecol. Lett. 13, 1041–1054.

Cesari, P., Pellizzato, M., 1985. Insediamento nella Laguna di Venezia e distribuzione adriatica di Rapana venosa (Valenciennes) (Gasteropoda Thaididae). Lavori - Soc. Ven. Sci. Nat. 10, 3–16 (In Italian).

Cucco, A., Umgiesser, G., Ferrarin, C., Perilli, A., Canu, D.M., Solidoro, C., 2009. Eulerian and lagrangian transport time scales of a tidal active coastal basin. Ecol. Mod. 220, 913–922.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.

Džeroski, S., Drumm, D., 2003. Using regression trees to identify the habitat preference of the sea cucumber (Holothuria leucospilota) on Rarotonga, Cook Islands. Ecol. Mod. 170, 219–226.

Dedecker, A.P., Goethals, P.L.M., Gabriels, W., De Pauw, N., 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders Belgium). Ecol. Mod. 174, 161–173.

G.R.A.L., 2006. Piano d'uso sostenibile delle aree in concessione per veneri coltura. Venice (In Italian).

G.R.A.L., 2009. Adeguamento del Piano d'uso sostenibile delle aree in concessione per veneri coltura (In Italian).

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Hirzel, A.H.G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Mod. 199, 142–152.

Iverson, L.R., Schwartz, M.W., Prasad, A.M., 2005. Potential colonization of newly available tree-species habitat under climate change: an analysis for five eastern US species. Landscape Ecol. 19, 787–799.

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

Lawler, J.J., White, D., Neilson, R.P., Blaustein, A.R., 2006. Predicting climate-induced range shifts: model differences and model reliability. Glob. Change Biol. 12, 1568–1584.

Lawler, J.J., Shafer, S.L., White, D., Kareiva, P., Maurer, E.P., Blaustein, A.R., Bartlein, P.J., 2009. Projected climate-induced faunal change in the Western Hemisphere. Ecology 90, 588–597.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2, 18–22.

MAG ACQUE – SELC, 2005. Studio B.12.3/III. La funzionalità dell'ambiente lagunare attraverso rilievi delle risorse alieutiche, dell'avifauna e dell'ittiofauna. Erodibilità del fondale e fattori di disturbo: Rilievi dell'erodibilità del fondale. Consorzio Venezia Nuova (In Italian).

MAG ACQUE – Thetis, 2005. Programma generale delle attività di approfondimento del quadro conoscitivo di riferimento per gli interventi ambientali. 2° stralcio triennale (2003–2006) "Progetto ICSEL". Attività A (In Italian).

MAG ACQUE – SAMA – Thetis, 2007. Progetto MELa 4. Attività A. Campagne periodiche di misura, negli anni 2007 e 2008, della qualità delle acque in collaborazione con SAMA. Consorzio Venezia Nuova 2009 (In Italian).

Melià, P., Nizzoli, D., Bartoli, M., Naldi, M., Gatto, M., Viaroli, P., 2003. Assessing the potential impact of clam rearing in dystrophic lagoons: an integrated oxygen balance. Chem. Ecol. 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. Mar. Ecol. Prog. Ser. 272, 153–164.

Molinaroli, E., Guerzoni, S., Sarretta, A., Cucco, A., Umgiesser, G., 2007. Links between hydrology and sedimentology in the Lagoon of Venice, Italy. J. Mar. Syst. 68, 303–317.

Molinaroli, E., Guerzoni, S., Sarretta, A., Masiol, M., Pistolato, M., 2009. Thirty-year changes (1970 to 2000) in bathymetry and sediment texture recorded in the Lagoon of Venice sub-basins, Italy. Mar. Geol. 258, 115–125.

Paesanti, F., Pellizzato, M., 2000. Tapes Philippinarum. Veneto Agricoltura (In Italian).

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. Ecol. Mod. 138, 231–245.

Pearson, R.G., Dawson, T.P., Berry, P.M., Harrison, P.A., 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. Ecol. Mod. 154, 289–300.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.

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

Pranovi, F., Da Ponte, F., Raicevich, S., Giovanardi, O., 2004. A multidisciplinary study of the immediate effects of mechanical clam harvesting in the Venice Lagoon. ICES Mar. Sci. 61, 43–52.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.

Province of Venice, 2009. 2009. Piano per la gestione delle risorse alieutiche delle lagune di Venezia e Caorle. Dosson di Casier, Treviso (In Italian).

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, ISBN 3-900051-07-0, http://www.R-project.org.

Recknagel, F., 2001. Applications of machine learning to ecological modelling. Ecol. Mod. 146, 303–310.

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 Triennale (In Italian).

Santos, X., Brito, J.C., Sillero, N., Pleguezuelos, J.M., Llorente, G.A., Fahd, S., Parellada, X., 2006. Inferring habitat-suitability areas with ecological modelling techniques and GIS: a contribution to assess the conservation status of Vipera latastei. Biol. Conserv. 130, 416–425.

Seoane, J., Carrascal, L.M., Alonso, C.L., Palomino, D., 2005. Species-specific traits associated to prediction errors in bird habitat suitability modelling. Ecol. Mod. 185, 299–308.

Sfriso, A., Facca, C., Marcomini, A., 2005. Sedimentation rates and erosion processes in the lagoon of Venice. Environ. Int. 31, 983–992.

Solidoro, C., Cossarini, G., Pastres, R., 2002. Numerical analysis of the nutrient fluxes through the Venice Lagoon inlets. In: Campostrini, P. (Ed.), Scientific research and safeguarding of Venice, Corila Research Program 2001 results, Istituto Veneto di Scienze, Lettere ed Arti, Corila, Venice. , pp. 545–555.

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

Solidoro, C., Melaku Canu, D., Cucco, A., Umgiesser, G., 2004. A partition of the Venice Lagoon based on physical properties and analysis of general circulation. J. Mar. Syst. 51, 147–160.

Spillman, C.M., Hamilton, D.P., Hipsey, M.R., Imberger, J., 2008. A spatially resolved model of seasonal variations in phytoplankton and clam (Tapes philippinarum) biomass in Barbamarco Lagoon, Italy. Estuar. Coast Shelf S 79, 187–203.

Turolla, E., Zentilin, A., Pellizzato, M., Rossetti, E., 2008. La venericoltura in Italia a 25 anni dal suo Esordio. Il Pesce 3, 31–40 (In Italian).

Umgiesser, G., Melaku Canu, D.M., Cucco, A., Solidoro, C., 2004. A finite element model for the Venice Lagoon development, set up, calibration and validation. J. Mar. Syst. 51, 123–145.

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

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

Vincenzi, S., Caramori, G., Rossi, R., De Leo, G.A., 2007. 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). Mar. Pollut. Bull. 55, 579–590.