

Occupancy estimation of marine species: dealing with imperfect detectability

Yiannis Issaris^{1,2,*}, Stelios Katsanevakis¹, Maria Salomidi³, Kostas Tsiamis³,
Nikos Katsiaras³, George Verriopoulos²

¹Institute of Marine Biological Resources, Hellenic Centre for Marine Research, Anavyssos 19013, Greece

²Department of Zoology–Marine Biology, School of Biology, University of Athens, Athens 15784, Greece

³Institute of Oceanography, Hellenic Centre for Marine Research, Anavyssos 19013, Greece

ABSTRACT: Underwater visual surveys are frequently used in monitoring programmes of marine populations. Species occupancy, defined as the probability of presence in a sampling unit, is a commonly used state variable. Imperfect detectability is a serious issue in such studies and, if ignored, may lead to incorrect inferences and erroneous management decisions. In this paper, we propose a methodology and field protocol for underwater visual surveys implemented by multiple observers. This approach can be applied for an unbiased occupancy estimation of marine species by explicitly incorporating imperfect detection into the modelling process. Based on a case study carried out in a Greek coastal area (Saronikos Gulf), the benefits of the proposed approach were demonstrated. Using a sufficient number of observers, the probability of recording false absences (i.e. the probability that the target species was present in a site but not detected) was minimized and occupancy estimation was greatly improved. For the whelk *Stramonita haemastoma* in the case study area, single-observer occupancy estimates were negatively biased and varied significantly (between 0.64 and 0.89) depending on the observer, while with the proposed methodology, using 5 observers, the obtained occupancy estimate had the value of 0.93. The probability of false absence was high in the single-observer case (between 0.10 and 0.30), and rather low with any combination of 3 observers (<0.025), while it dropped to practically 0 with 5 observers. As demonstrated in the case of the alien green alga *Codium fragile fragile*, occupancy models provide a flexible framework for relating occupancy to spatial and environmental covariates, testing ecological hypotheses and producing predictive distributional maps. Overall, the presented methodology and its potential extensions could prove extremely useful in a variety of applications in the marine environment.

KEY WORDS: Occupancy · Underwater visual survey · Detectability · Monitoring

—Resale or republication not permitted without written consent of the publisher—

INTRODUCTION

Monitoring programmes are a key component of ecological research and adaptive management. To discriminate among competing hypotheses about the state and dynamics of populations and communities, researchers need to collect relevant field data. In the context of marine conservation and sustainable exploitation of marine resources, there is a need for

estimates of the system state and its trends in order to decide on appropriate management actions. Adequate monitoring of the population(s) responses to management actions is the cornerstone of adaptive maritime spatial planning and successful marine resource management (Day 2008, Douvère & Ehler 2011).

Monitoring programmes for animal or plant populations are typically based on obtaining estimates of a

*Email: yiannis@yissaris.com

state variable, appropriately replicated in space and time. The most commonly used state variables are population density, abundance and percent cover. However, such estimations are often costly and require substantial effort (e.g. Lancia et al. 1994, Pollock et al. 2002, MacKenzie et al. 2006) or may be unfeasible for various reasons, e.g. in the case of rare or elusive species (Thompson 2004). Alternatively, species occupancy, defined as the proportion of area, patches or sampling units occupied (or as the probability of presence in a sampling unit) may be seen as a low-cost surrogate (MacKenzie et al. 2006). Moreover, in some cases occupancy is the appropriate state variable and would be chosen in the first place, e.g. in studies of distribution and range (Scott et al. 2002), alien invasions (Hanspach et al. 2008), metapopulation studies (Moilanen 2002), community studies (Martinez-Solano et al. 2003, Weber et al. 2004) and large-scale monitoring (Manley et al. 2004).

The estimation of occupancy is based on presence-absence data. It is often confounded by detection probability, i.e. the probability of correctly recording the presence of an individual or species within the area of interest. The detection of a species in a sampling unit confirms its presence with certainty, assuming that there is no misidentification (Royle & Link 2006). However, failure to detect the species may be due either to the actual absence of the species or to the failure to detect the species that is actually present at the site during the survey (a 'false absence'; MacKenzie et al. 2006). This could be due to e.g. the inability of the observer to identify the target species, the cryptic behaviour of the species that may be hidden in crevices or burrowing or camouflaged, and its high mobility. Hence, the true absence of a species from a sampling unit is often impossible to be inferred solely from presence-absence data. Failure to properly account for imperfect detectability may lead to incorrect inferences about the system and erroneous management decisions; occupancy is underestimated, colonization rates are biased, and habitat relationships may well be misleading, particularly if detectability also changes among different habitats (Moilanen 2002, Tyre et al. 2003, Gu & Swihart 2004, MacKenzie et al. 2006).

A set of methods that permit inference about occupancy based on presence-absence data have been developed, taking into account the imperfect detection of the target species (Nichols & Karanth 2002, MacKenzie et al. 2006 and references therein, Dorazio et al. 2006). All methods involve multiple visits to each site, where the target species is either detected,

with probability p , or not detected, with probability $1 - p$. The goal is to estimate the proportion of sites that are occupied, ψ , accepting that the target species is not always detected.

MacKenzie et al. (2006) provided a synthesis of the literature on estimating occupancy-type metrics, while explicitly accounting for detection probability. The methods described by MacKenzie et al. (2006) have been widely applied in the terrestrial and freshwater environments for various taxa, such as amphibians (Campbell Grant et al. 2009, Sewell et al. 2010), birds (Dorazio & Royle 2005, Ferraz et al. 2007), mammals (O'Connell et al. 2006, Weller 2008, Moritz et al. 2008) and insects (MacKenzie et al. 2006, Pellet 2008).

A variety of techniques accounting for imperfect detectability, mostly based on distance sampling (e.g. Katsanevakis 2009, Ronconi & Burger 2009) and mark-recapture (e.g. D'Anna & Pipitone 2000, Bradshaw et al. 2007), have been used in the marine environment. However, methods for estimating occupancy of marine species from presence-absence surveys that take into account imperfect detectability have not been applied yet. Studies on the distributional range of marine species, marine alien invasions, species richness or large-scale monitoring have generally ignored detectability issues. Presence-absence underwater surveys are being widely used to estimate occupancy, but they routinely assume that detectability is perfect (i.e. equal to 1) and thus commonly underestimate occupancy and species richness of demersal communities (MacNeil et al. 2008a,b, Canary et al. 2010). In an assessment of the structure of coral reef fish communities in Tanzania, detection probabilities of 47 reef fish families varied substantially among fish family groups and among sites, and ranged between 0.05 and 0.54 (MacNeil et al. 2008a). That study also found that schooling behaviour, reef fish functional group and fish length can affect reef fish detectability. The very low detectability figures reported by MacNeil et al. (2008a) highlight the need to account for detectability in presence-absence underwater surveys, to make unbiased estimations of species occurrence.

In this study, we propose a methodology and field protocol for occupancy estimation of marine species during underwater visual surveys by explicitly incorporating imperfect detection into the modelling process. The proposed approach is an adaptation of the single-season occupancy models of MacKenzie et al. (2006). It is applicable to diver-based (with SCUBA or snorkelling) underwater visual surveys (UVS), but can easily be extended to video-based

surveys with the use of towed cameras, remotely operated vehicles or autonomous operated vehicles. The proposed approach was applied in a snorkelling UVS in the Saronikos Gulf (Aegean Sea, Eastern Mediterranean).

MATERIALS AND METHODS

Scheme of occupancy surveys

An area of interest can be considered as a population of sampling units for which a value of 1 or 0 can be assigned, denoting the presence or absence, respectively, of the target species within their boundaries. Sampling units may be spatially defined in the study area in a systematic way (e.g. grid cells of specific size, transects of specific length) or chosen from discreetly occurring natural formations and habitats (e.g. colonies of sponges on a sandy bottom, patches of seagrass, underwater caves, island coastlines of an archipelago). The appropriate sampling design is generally driven by the hypothesis. However, the final selection can sometimes be just a matter of scale. A number of s sites are selected from the population of all sampling units. Inference about the study area is made from the collection and analysis of presence–absence data at the s sites. Such inference can be based on either a design-based approach (e.g. random sampling or stratified random sampling) or a model-based approach, where occupancy status is modelled in relation to environmental or spatial covariates. These models are used to predict occupancy in the entire study area based on the distributional patterns of covariates.

UVS methods are used to detect the presence of at least 1 individual of the target species at each site. It is assumed that there is no false detection because of misidentification or other reasons. The sites are also assumed to be closed to changes in occupancy during the survey, meaning that no new sites are being occupied and no sites are abandoned during the survey; this period of population closure is defined as ‘a season’. Further modifications for highly mobile species that violate this assumption are considered in the ‘Discussion’. A species may or may not be detected when present at a site, and its detection is assumed to be independent of the detection of the species by another observer or the detection of the species at other sites.

In the general scheme of a single-season occupancy survey (MacKenzie et al. 2006), each of the s sites is visited K times by the same or different

observers, i.e. K surveys of each site are conducted. In this paper, we propose that each marine site is surveyed once by K observers, simultaneously but independently searching for the target species. The detections of each observer would represent a ‘survey’ sensu MacKenzie et al. (2006); a single visit to each site would suffice to collect the necessary presence–absence data for estimating both occupancy and probability of detection.

The resulting presence–absence data for each site surveyed by all observers is a sequence of detections (1) and non-detections (0) that form a detection history H_i of the target species at each site. The set of these detection histories is used to estimate the proportion of sites occupied by the species, based on the general approach described by MacKenzie et al. (2002, 2006). Data gathered from independent observers were considered as independent multiple surveys of the same site.

Modelling occupancy and detectability

Occupancy ψ is jointly modelled with probability of detection p under a model-based approach, with models being based on straightforward probabilistic arguments describing the underlying processes that may have caused a given detection history to be observed (MacKenzie 2005). On the basis of the conceptual model, there are 2 stochastic processes occurring that affect the detection of the target species at a site (MacKenzie et al. 2006). A site might be either occupied (with probability ψ) or unoccupied (with probability $1 - \psi$) by the target species. If the site is unoccupied, the target species will not be detected. If the site is occupied, at each survey j the target species will either be detected (with probability p_j) or pass undetected (with probability $q_j = 1 - p_j$).

For example, consider the detection history $H_i = 101$ (denoting that site i was surveyed by 3 observers, with the target species being detected by the first and third observers). The probability of this history would be $\Pr(H_i = 101) = \psi p_1 q_2 p_3$. For sites where the target species is never detected there are 2 possibilities: the species is either present but never detected, or genuinely absent. Thus, $\Pr(H_i = 000) = \psi q_1 q_2 q_3 + (1 - \psi)$. By deriving such expressions for each of the s observed detection histories, assuming independent observations, the likelihood of the data will be

$$L(\psi, \mathbf{p} | H_1, H_2, \dots, H_s) = \prod_{i=1}^s \Pr(H_i) \quad (1)$$

where \mathbf{p} is the vector of detection probabilities.

Incorporating covariates

The potential relationships between model parameters (occupancy and detection probabilities) and the characteristics of the sites (e.g. habitat type, depth, slope) or environmental (e.g. annual average surface temperature, current intensity, water turbidity) and geographical variables (longitude, latitude) can be investigated. In addition, different observers can have quite dissimilar capabilities as regards the detection of the target species (MacNeil et al. 2008b), and the observer effect should normally be considered. Moreover, the probability of detecting the target species in a specific location is often a direct function of the species' local abundance. Covariates (related to all of the above) are incorporated by using the logistic model

$$\theta_i = \exp(Y_i\beta) \cdot (1 + \exp(Y_i\beta))^{-1} \quad (2)$$

where θ_i is the probability of interest (occupancy or detection probability), Y_i are the covariates to be modelled, and β denotes the vector of the covariate coefficients to be estimated (MacKenzie et al. 2006). Standard maximum likelihood techniques are applied to obtain estimates of the model parameters.

Model selection and multi-model inference

An information theory approach may be followed in order to infer occupancy patterns. According to this approach, data analysis is assumed to be the integrated process of an *a priori* specification of a set

of candidate models (based on the science of the problem), model selection based on the principle of parsimony according to the Akaike information criterion (AIC; Akaike 1973) and the estimation of parameters and their precision (Burnham & Anderson 2002). The principle of parsimony implies the selection of a model with the smallest possible number of parameters for adequate representation of the data, a bias versus variance trade-off. Furthermore, rather than estimating parameters using only the 'best' model, parameters (i.e. occupancy and detection probabilities) can be estimated using several or even all of the models considered. This procedure is termed multi-model inference and has several theoretical and practical advantages (Burnham & Anderson 2002, Katsanevakis 2006).

Case study

Study area and target species

The proposed approach was applied in a snorkelling UVS in the Saronikos Gulf (Greece, Eastern Mediterranean; Fig. 1). The aim of the survey was to estimate occupancy patterns for a range of benthic species of the shallow water rocky subtidal communities. In this paper, we present results for 2 species: the gastropod *Stramonita haemastoma* (Linnaeus 1767: *Buccinum*) and the alien green alga *Codium fragile* (Suringar) Hariot subsp. *fragile*.

Stramonita haemastoma is a predatory cosmopolitan whelk that may reach up to ca. 80 mm in shell length and is found in rocky littoral habitats in warm temperate waters. It is a keystone species in many locations, greatly determining the structure and organization of intertidal communities. Although in the Atlantic Sea the species may attain densities of dozens to hundreds of individuals per m², in the Mediterranean Sea its densities are much lower (Rilov et al. 2001).

The green alga *Codium fragile fragile* (hereafter *C. fragile*) is one of the most common alien macroalgae in the Mediterranean Sea. It originates from the North Pacific Ocean and has spread to both hemispheres, mainly due to shipping (Carlton & Scanlon 1985). Populations of the species are relatively restricted in Greece; they

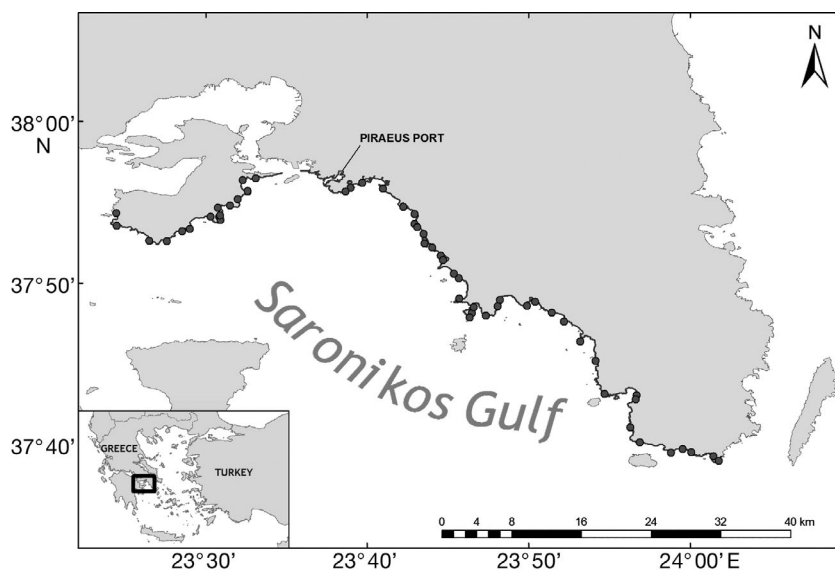


Fig. 1. Study area. Sampling sites are marked with grey dots

are only found at a small number of sites (mostly near ports and polluted environments) (Tsiamis et al. 2010). It was first observed in the Saronikos Gulf in 1998 (Tsiamis & Panayotidis 2007).

Design and field protocol

In total, 54 sites of hard substrate were surveyed during summer 2009 (Fig. 1). At each sampling site, a 200 m line was deployed in shallow waters <5 m deep, using a diving reel. Five observers (all marine scientists with various levels of experience in UVSs, trained for the identification of the target species) surveyed the transect independently (with a time lag of a few minutes) during fixed 20 min time intervals, searching for the target species. For *Codium fragile*, the observers recorded not only its presence but also the number of individuals detected; these data were classified into 2 relative abundance categories: (1) ≤ 5 detected individuals, (2) >5 detections.

Effect of multiple observers

In this study, we used the *Stramonita haemastoma* dataset to demonstrate the effect of the number of independent observers on the occupancy estimate. A model that assumed constant occupancy ψ within the study area and different detection probabilities p_j for each observer was fit to the presence-absence data. The model was fitted for all different combinations of 2, 3, 4 or 5 observers. The probability of a ‘false absence’, i.e. the probability that the species was present in a site, given that it was not detected, was estimated as

$$\psi_{\text{fa}} = \psi \prod_{j=1}^K (1 - p_j) \cdot \left[(1 - \psi) + \psi \prod_{j=1}^K (1 - p_j) \right]^{-1} \quad (3)$$

(MacKenzie et al. 2006). In the hypothetical case of a single-observer survey, occupancy was naively estimated as the ratio of the number of sites where the species was detected to the total number of surveyed sites.

Testing multiple hypotheses for occupancy patterns

The *Codium fragile* dataset was analysed following an information-theoretic approach and multi-model inference, i.e. estimation of occupancy and detection probabilities taking into account all plausible models and not just the best one. Our aim was to demon-

strate how occupancy patterns may be inferred by testing multiple hypotheses, while also taking into account detectability issues. Specifically, we tested the following hypotheses for the occupancy of *C. fragile* in the study area: (O1) occupancy was constant; (O2) occupancy varied with the distance from the highly industrialized and polluted port of Piraeus. Three assumptions for detectability were investigated: (D1) detectability was constant across sites and observer-independent; (D2) detectability was observer-dependent; (D3) detectability differed in relation to species abundance. Eight occupancy models were fitted for each target species by combining the 2 assumptions for occupancy with the 3 assumptions for detectability (Table 1). Seven covariates were used in these models: (1) *dist*, which was the standardized value of the distance of each site from the port of Piraeus, i.e. (value – mean)/SD; (2) *obs*, collectively representing $obs_1, obs_2, obs_3, obs_4, obs_5$, where $obs_i = 1$ if the survey was conducted by observer i , 0 otherwise; (3) *abund*, which was 0 if ≤ 5 individuals were detected in a site, 1 otherwise (Table 1). The software PRESENCE v. 3.1 (Hines 2006) was used to fit the candidate models.

AIC and AIC differences, $\Delta AIC_i = AIC_i - AIC_{\min}$, were computed over all candidate models. To quantify the plausibility of each model, given the data and set of 6 models, the Akaike weight w of each model was calculated, where

$$w_i = \exp(-0.5\Delta_i) / \sum_j \exp(-0.5\Delta_j) \quad (4)$$

Table 1. Set of candidate models for occupancy and detectability of *Codium fragile* in the study area. The underlying assumptions are: (O1) occupancy was constant; (O2) occupancy varied with the distance from the port of Piraeus; (D1) detectability was constant across sites and observer-independent; (D2) detectability was observer-dependent; (D3) detectability differed in relation to species abundance. Constant occupancy or detection probability is denoted as $\psi(\cdot)$ or $p(\cdot)$, respectively. Dist.: Distance

Model notation	Covariates		Assumptions
	Occupancy	Detectability	
$\psi(\cdot) p(\cdot)$	–	–	O1, D1
$\psi(\cdot) p(obs)$	–	Observer	O1, D2
$\psi(\cdot) p(abund)$	–	Abundance	O1, D3
$\psi(\cdot) p(obs+abund)$	–	Observer & abundance	O1, D2 & D3
$\psi(dist) p(\cdot)$	Dist.	–	O2, D1
$\psi(dist) p(obs)$	Dist.	Observer	O2, D2
$\psi(dist) p(abund)$	Dist.	Abundance	O2, D3
$\psi(dist) p(obs+abund)$	Dist.	Observer & abundance	O2, D2 & D3

The Akaike weight is considered as the weight of evidence in favour of model i being the actual best model for the available set of models (Akaike 1983, Burnham & Anderson 2002). To assess absolute goodness of fit, a Pearson's chi-squared statistic was used based on the parametric bootstrap procedure described by MacKenzie & Bailey (2004). Estimates of the relative importance of predictor variables j were made by adding up the Akaike weights across all the models in the set where each variable occurred [$w_+(j)$]; the larger the sum of Akaike weights the more important that variable was, relative to the other variables (Burnham & Anderson 2002).

Model-averaged estimates of ψ and p were calculated by the formula

$$\hat{\theta} = \sum_i w_i \hat{\theta}_i \quad (5)$$

where $\hat{\theta}$ denotes a model averaged estimate of parameter θ (Burnham & Anderson 2002). The unconditional standard error of θ was estimated as

$$\text{se}(\hat{\theta}) = \sum_i w_i \left[\text{var}(\hat{\theta}_i | g_i) + (\hat{\theta}_i - \hat{\theta})^2 \right]^{1/2} \quad (6)$$

where $\text{var}(\hat{\theta}_i | g_i)$ is the variance of θ according to model g_i , conditional on the model (Burnham & Anderson 2002).

RESULTS

Effect of multiple observers on occupancy estimations — the *Stramonita haemastoma* example

In our case study, *S. haemastoma* was found in 50 of the 54 surveyed sites by at least 1 of the 5 observers. The occupancy of the species in the study area (using the data from all 5 observers) was estimated as $\psi = 0.93$ with a 95% CI of 0.82–0.97. However, detectability varied among observers, ranging between 0.70 and 0.96. Due to the varying competence of the 5 observers to detect the species, results from hypothetical single-observer presence–absence surveys greatly varied. Such single-observer occupancy estimates ranged between 0.64 and 0.89, depending on the observer (Fig. 2); in every case, occupancy was underestimated, as none of the observers managed to detect the species at all 50 sites.

When using multiple observers, despite differing detectability among observers, the variation in occupancy estimates was reduced. The higher the number of observers, the lower the variation in estimated values (Fig. 2), indicating that with multiple observers the effect of an inefficient observer on occupancy estimation becomes less important. The higher

the number of observers, the lower is the effect any single observer has on the estimated occupancy. It is also evident from Fig. 2 that the gain in accuracy of the occupancy estimates was large when the number of observers increased from 1 to 2, smaller when the number of observers increased from 2 to 3, and marginal at further increases in the number of observers.

The probability of a false absence ψ_{fa} was relatively high in the case of 1 observer (range: 0.10–0.30) but became practically null (0.0002) with 5 observers (Fig. 3). With 3 observers, ψ_{fa} was <0.025 for every combination of observers (average $\psi_{fa} = 0.011$), while with 4 observers, ψ_{fa} was <0.005 in every case (Fig. 3).

Multi-model inference of occupancy patterns — the *Codium fragile* example

Based on the full model, $\psi(\text{dist})p(\text{obs} \times \text{abund})$, there was no evidence of poor model fit (χ^2 test, $p = 0.60$). The best model ($w = 80.2\%$) was $\psi(\text{dist})p(\text{abund})$, which supported assumptions O2 and D3 (Table 2). Model $\psi(\text{dist})p(\text{obs} + \text{abund})$, which assumed O2, D2 and D3, was also somewhat supported by the data ($w = 18.6\%$). All other models had essentially no support (Table 2) and may thus be omitted from further consideration.

The expressions of the best [$\psi(\text{dist})p(\text{abund})$] and second best [$\psi(\text{dist})p(\text{obs} + \text{abund})$] models were, respectively:

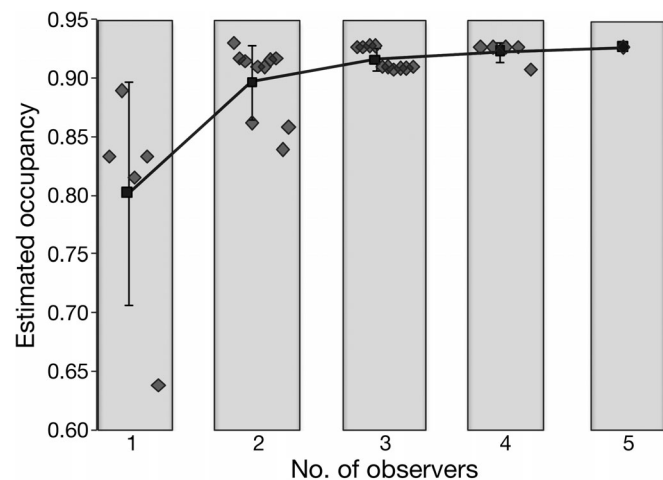


Fig. 2. *Stramonita haemastoma*. Estimated occupancy in the study area based on presence–absence data collected by 1, 2, 3, 4 or 5 observers. In each case of <5 observers, estimates for all possible combinations of observers (gray diamonds) are provided; the average estimated occupancy is also given (black rectangles, with error bars showing SD)

$$\begin{cases} \psi = \exp(-1.299-1.522dist) \times \\ [1+\exp(-1.299-1.522dist)]^{-1} \\ p = \exp(-0.234+2.78abund) \times \\ [1+\exp(-0.234+2.78abund)]^{-1} \end{cases}$$

$$\begin{cases} \psi = \exp(-1.373-1.492dist) \times \\ [1+\exp(-1.373-1.492dist)]^{-1} \\ p = \exp(A)[1+\exp(A)]^{-1} \\ A = -0.769obs_1 + 1.126obs_2 - 0.263obs_3 + \\ 1.126obs_4 - 0.753obs_5 + 2.73abund \end{cases}$$

With distances being standardized according to the formula $dist = (x - 40060) / 29041$, where x is the coastline distance from the port of Piraeus in metres.

Between the 2 assumptions for occupancy, O2 had substantial support [$w_+(O3) = 99.1\%$] and prevailed against O1 [$w_+(O1) = 0.9\%$]. The coefficients of $dist$ were negative in both supported models, and thus it is concluded that the occupancy of *Codium fragile* declined with the distance from the port of Piraeus. Among the detectability assumptions, D3 had the highest support [$w_+(D3) = 99.7\%$], D2 had little support [$w_+(D2) = 18.9\%$], and D1 had essentially no support [$w_+(D1) = 0.2\%$]. Hence, detectability may not be considered constant but differing in relation to species abundance; the hypothesis that p differs depending on the observer may not be excluded. Based on the best model, p was 0.42 for the low-abundance sites and 0.93 for the sites of high abundance. The

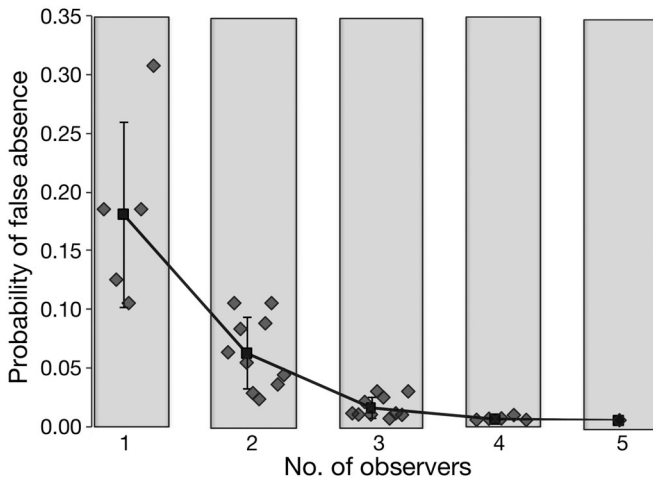


Fig. 3. *Stramonita haemastoma*. Probability of a ‘false absence’, ψ_{fa} , from a site in the study area (i.e. probability of presence given that the species was not detected), based on presence–absence data collected by 1, 2, 3, 4 or 5 observers. In each case of <5 observers, probabilities were estimated for all possible combinations of observers (gray diamonds); the average ψ_{fa} is also provided (black rectangles, with error bars showing SD)

probabilities of false absences were generally low ($<2.5\%$) for all sites.

Occupancy of *Codium fragile* in the study area based on the average model predictions (between the 2 supported models, weighted by w) is presented in a map (Fig. 4), where the decline in occupancy with increasing distance from the port of Piraeus is clearly evident.

DISCUSSION

Unbiased occupancy estimation by multiple-observer surveys

The proposed approach offers an efficient way to obtain unbiased occupancy estimates by allowing detection probabilities to be jointly estimated. Presence–absence visual surveys in the marine environment, as was evident in the cases of *Stramonita haemastoma* and *Codium fragile* suffer from both low and variable detection probabilities. For *S. haemastoma*, although quite abundant in the study area, the probability of a false absence reached as high as 30% for 1 observer. In the case of *S. haemastoma*, simple presence–absence surveys by single observers produced inaccurate results. Depending on the observer, occupancy estimates were negatively biased by as much as 31%. In the case of long-term monitoring studies or comparisons across environmental or spatial gradients, ignoring imperfect detectability might lead to misleading temporal or spatial population patterns.

Table 2. Summary of the model selection procedure examining factors predicting the probability of presence ψ of *Codium fragile* in 200 m standardized transects along the coastline of the Saronikos Gulf (during summer 2009), using simple logistic regression. Twice the negative log-likelihood ($-2l$), the number of parameters in the model (K), the relative difference in Akaike information criterion (ΔAIC) values compared to the top-ranked model (ΔAIC), and the AIC model weights (w) for all candidate models (as defined in Table 1) are reported. Models are ranked according to w

Model	$-2l$	K	ΔAIC	w (%)
$\psi(dist) p(abund)$	107.59	4	0.0	80.2
$\psi(dist) p(obs+abund)$	102.51	8	2.9	18.6
$\psi(\cdot) p(abund)$	118.99	3	9.4	0.7
$\psi(dist) p(\cdot)$	121.69	3	12.1	0.2
$\psi(\cdot) p(obs+abund)$	113.81	7	12.2	0.2
$\psi(dist) p(obs)$	117.03	7	15.4	0.0
$\psi(\cdot) p(\cdot)$	132.97	2	21.4	0.0
$\psi(\cdot) p(obs)$	128.31	6	24.7	0.0

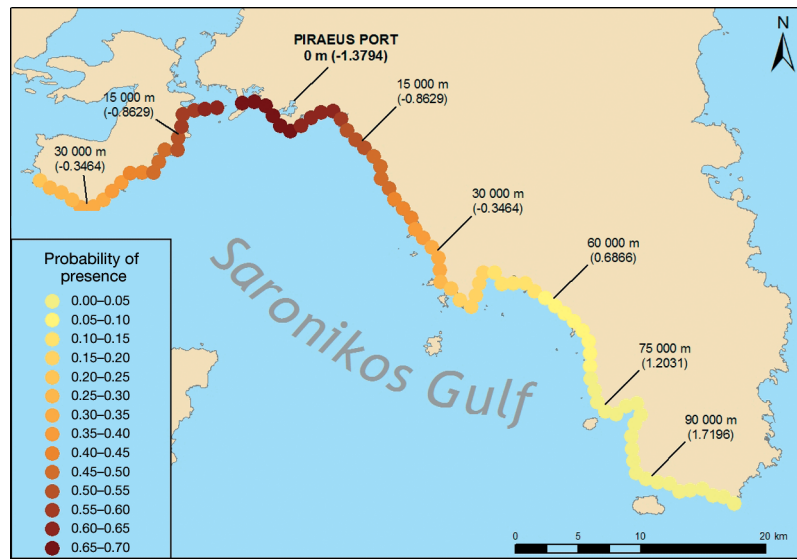


Fig. 4. *Codium fragile*. Occupancy along the coastline of the Saronikos Gulf. Representative values of the distance from the port of Piraeus are provided along the surveyed coastline (standardized within parentheses)

The observer effect is of paramount importance especially when spatial or temporal comparisons are made based on surveys conducted by different observers (e.g. Benedetti-Cecchi et al. 1996, MacKenzie et al. 2006, MacNeil et al. 2008b). The possibility of erroneously noting significant differences in occupancy due to observer-related differences in detectability is high when based on single-observer surveys. The increase in the variability of estimates due to unaccounted observer-related variation in detectability, which could easily mask temporal trends, is equally important. Using repetitive observations made at the sampling sites by multiple observers reduces the observer effect and offers solutions to all of these issues.

There is undoubtedly a point beyond which it is more efficient to invest in spatial replication than to increase the number of observers. Increasing the number of surveyed sites would make it easier to detect significant effects (modelled covariates) in occupancy models. In the case of *Stramonita haemastoma*, it appears that 3 observers would suffice; the gain from further increasing the number of observers was small and not worth the extra cost and effort (however, this would depend on the accuracy required for the intended project/goal). As the ultimate goal of an occupancy survey is to achieve a desired level of precision in the estimation of occupancy with minimal effort, the number of required observers appears to be a key aspect in designing occupancy studies based on our proposed methodology. It has been proven that when detection probabilities are

high, it is more efficient to survey more sampling units (sites) rather than to increase the number of surveys per sampling unit (i.e. number of observers in our case), but as the detection probability decreases, more surveys per site are necessary (Tyre et al. 2003). MacKenzie & Royle (2005) and MacKenzie et al. (2006) provided specific advice on the number of repeated surveys per sampling unit considering the variance of the occupancy estimator under specific sampling schemes.

Modelling occupancy patterns

The modelling framework provided by MacKenzie et al. (2006) is based on 2 stochastic processes—for occupancy and detectability—that define whether the target species is detected at a site. By jointly modelling each of the 2 processes, this framework provides a direct means of investigating potential relationships with various environmental and spatial parameters, such as habitat type, distance from a pollution source, latitude, sea water temperature or salinity. Wildlife managers increasingly wish to extract more than an estimate of a state variable from their surveys and frequently want to relate animal occupancy to spatial and environmental variables. Inferred patterns from monitoring studies might provide useful insights about system behaviour and generate hypotheses that can be tested with manipulative experiments. In the case of the invasive alien species *Codium fragile*, the fact that the probability of occupancy was highest near the polluted port of Piraeus and declined with increased distance from the port generates 2 hypotheses that deserve further investigation: (1) the species has been introduced in the study area via shipping, initially near the port of Piraeus; (2) the species is resistant to marine pollution and might more easily get established in polluted areas due to a competitive advantage.

The information theory approach frees researchers from the limiting concept that the proper approximating model is somehow 'given'. When the data support evidence of more than 1 model, model averaging the predicted response variable across models is advantageous in reaching a robust inference that is not conditional on a single model (Burnham & Anderson 2002, Katsanevakis 2006).

Violation of model assumptions

Sites are assumed to be 'closed' to changes in occupancy during the survey period, i.e. the occupancy status remains constant. In the case of violation of the closure assumption, it is expected that the occupancy estimator will be unbiased if species move in and out of the sampling unit in a random way, although occupancy will now refer to proportions of sites 'used' by the target species (MacKenzie 2005). However, if movement into and out of the sampling unit is not random during the survey period (e.g. due to emigration, immigration or reproduction), occupancy will be biased. In particular, models relating occupancy to environmental and spatial covariates might give spurious results. For example, assume a survey is conducted along a coastline that is oriented across a latitudinal gradient, and (for reasons of cost efficiency) sites are surveyed consecutively along this gradient. If, during the survey, immigration of the target species into the study area causes an increase of overall occupancy, a false latitudinal gradient of occupancy will be inferred, because both true occupancy and the latitude of surveyed sites would be correlated with time. Any pattern of non-random animal movements even at much finer spatial or temporal scales might be a source of bias, unless properly modelled.

In the case of unmodelled occupancy heterogeneity (variation of occupancy probability among sites), the bias is relatively unknown compared to other model assumptions and more simulation studies are required. The estimated occupancy may appropriately reflect the average level of occupancy in the study area, but the variance is expected to be overestimated (MacKenzie et al. 2006). Unmodelled heterogeneity in detection probability p generally leads to negatively biased occupancy estimates. Low detection probabilities coupled with large variations (among sites or surveys) tend to increase the bias (Royle & Nichols 2003). Bias is further exaggerated in studies involving a small number of sites or a few repeated surveys at each site (MacKenzie et al. 2006).

Although the proposed approach can be applied to any type of marine species (benthic, demersal or pelagic), occupancy estimates of mobile species might be biased due to violation of the assumption of independent detectabilities among surveys. Mobile species, especially fish, might be affected substantially by the presence of observers, and can be classified into 4 major categories of behavioural response (Kulbicki 1998): (1) 'neutral', i.e. not affected; (2) 'shy', i.e. stay away from the observer; (3) 'curious',

i.e. attracted by the presence of an observer; and (4) 'secretive', i.e. tending to hide in holes or crevices as an observer approaches. Having a number of observers consecutively surveying the same site may introduce bias when targeting 'shy', 'secretive' or 'curious' species. Detectability of 'shy' and 'secretive' species might decline with successive surveys, as the species would tend to flee or hide after being disturbed by an observer, while the opposite could be true for 'curious' species. Part of the bias due to such behavioural responses of mobile species to observers could be avoided by defining appropriate survey-specific covariates when modelling detection probabilities, e.g. by defining a covariate $y_{ij} = 1$ if site i was surveyed by another observer before survey j , and 0 otherwise. Alternatively, the time interval between successive surveys carried out by different observers could be increased to allow observer-related effects to deflate in the meantime.

It is assumed that the observed presence of a species in a sampling unit is accurate. However, the possibility of misidentification during underwater surveys cannot always be excluded. Some species are difficult to identify *in situ* and can be easily confused with other species. In particular, when surveys are conducted by inexperienced observers or volunteers, the possibility of false positive records might be high. Even with low rates of false positive errors, the bias in occupancy rates can be extreme. Royle & Link (2006) developed a model for site occupancy that allows for false records of both absence and presence.

Potential applications in the marine environment

Several methods have been developed to deal with detectability when estimating the abundance of marine populations (e.g. Buckland et al. 2001, 2004, Katsanevakis 2009). All of these approaches perform best when the numbers of individuals observed are relatively high. However, adequate sample sizes are difficult to obtain for rare and elusive species, which are often of management concern because of low population sizes and often limited geographical ranges (Thompson 2004). For such rare or elusive species (often endangered and/or under a protection status), occupancy is a good surrogate of abundance, as relatively small sample sizes are sufficient (MacKenzie et al. 2004). The presented approach is a good choice for revealing spatial distribution patterns and monitoring trends, and it is particularly suitable for endangered and protected species due to its non-destructive nature.

Marine biological invasions represent a recognized worldwide threat impacting biodiversity and local economies, and there is strong interest from the scientific community and international organizations to monitor invasive species (CBD 2000, EC 2008, Zenetos et al. 2010). In large-scale monitoring programmes that may span whole countries or wider regions (e.g. for the implementation of the Marine Strategy Framework Directive in Europe; EC 2008, 2010) the methodology proposed in this paper would be an appropriate, efficient and less biased (in relation to simple presence–absence surveys) approach to estimate the spatial distribution and range of invasive species. Multi-season occupancy surveys would additionally provide good estimates of invasion rates and help us understand the underlying dynamics of colonization. Such information would be valuable for managers and decision-makers in the process of assessing impacts of biological invasions, conducting risk analysis and taking mitigation measures.

Major long-term distributional shifts in the marine fauna as a response to climate change have been reported (e.g. Last et al. 2011, Philippart et al. 2011). Dramatic shifts in the distributional range of marine species have been projected for the following decades, leading to local extinctions, species invasions and substantial ecosystem disturbances (Cheung et al. 2009). There is a need for a concerted effort to collect and analyse distributional data for marine biota, especially for species dwelling in shallow coastal areas, which are more affected by the increase of sea water temperature. The proposed methodology for occupancy estimation and modelling offers a good, cost-effective solution for such long-term large-scale monitoring.

Novel tools and techniques from geographical information science, advanced machine-learning algorithms and presence–absence or presence-only data have been integrated to model complex species–environment relationships (e.g. Pittman & Brown 2011). Improving the quality of data by reducing the probability of false absences, especially when these probabilities correlate with the environmental variables used in the species–environment models, would improve the accuracy of species distribution predictions. Integration of the proposed multi-observer approach with such advanced distribution and habitat suitability models could increase their predictive performance.

In conclusion, the results of this study emphasize the importance of detectability in the context of presence–absence underwater visual surveys while proposing a cost-effective multiple observer ap-

proach that is best suited for the marine realm and successfully addresses the issue of false absences. This approach along with its potential extensions can prove to be a valuable tool in a variety of applications relating but not limited to the monitoring of marine populations.

Acknowledgements. We thank 3 anonymous reviewers for their constructive comments, which helped us improve the quality of the manuscript.

LITERATURE CITED

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaaki F (eds) Proc 2nd Int Symp Inf Theory, Akademiai Kiado, Budapest, p 267–281
- Akaike H (1983) Information measures and model selection. *Bull Int Stat Inst* 44:277–290
- Benedetti-Cecchi L, Airoldi L, Abbiati M, Cinelli F (1996) Estimating the abundance of benthic invertebrates: a comparison of procedures and variability between observers. *Mar Ecol Prog Ser* 138:93–101
- Bradshaw CJA, Mollet HF, Meekan MG (2007) Inferring population trends for the world's largest fish from mark–recapture estimates of survival. *J Anim Ecol* 76: 480–489
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, New York, NY
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2004) Advanced distance sampling: estimating abundance of biological populations. Oxford University Press, New York, NY
- Burnham KP, Anderson DR (2002) Model selection and multimodal inference, 2nd edn. Springer-Verlag, New York, NY
- Campbell Grant EH, Green LE, Lowe WH (2009) Salamander occupancy in headwater stream networks. *Freshw Biol* 54:1370–1378
- Carlton JT, Scanlon JA (1985) Progression and dispersal of an introduced alga: *Codium fragile* spp. *tomentosoides* (Chlorophyta) on the Atlantic coast of North America. *Bot Mar* 28:155–165
- CBD (Convention on Biological Diversity) (2000) Global strategy on invasive alien species. UNEP/CBD/SBSTTA/6/INF/9. Convention on Biological Diversity, Cape Town
- Cheung WWL, Lam VWY, Sarmiento JL, Kearney K, Watson R, Pauly D (2009) Projecting global marine biodiversity impacts under climate change scenarios. *Fish Fish* 10: 235–251
- D'Anna G, Pipitone C (2000) External tagging for underwater monitoring. In: Goñi R, Harmelin-Vivien M, Badalamenti F, Le Diréach L, Bernard G (eds) Introductory guide to methods for selected ecological studies in marine reserves. GIS Posidonie, Marseille, p 38–41
- Day J (2008) The need and practice of monitoring, evaluating and adapting marine planning and management—lessons from the Great Barrier Reef. *Mar Policy* 32: 823–831

- Dorazio RM, Royle JA (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *J Am Stat Assoc* 100:389–398
- Dorazio RM, Royle JA, Sonderstorm B, Glimskar A (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* 87: 842–854
- Douve F, Ehler CN (2011) The importance of monitoring and evaluation in adaptive maritime spatial planning. *J Coast Conserv* 15:305–311
- EC (European Commission) (2008) Directive of the European Parliament and the Council establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Directive 2008/56/EC, OJ L 164. European Commission, Brussels
- EC (2010) Decision on criteria and methodological standards on good environmental status of marine waters. Decision 2010/477/EU, OJ L 232. European Commission, Brussels
- Ferraz G, Nichols JD, Hines JE, Stouffer PC, Bierregaard RO Jr, Lovejoy TE (2007) A large-scale deforestation experiment: effects of patch area and isolation on Amazon birds. *Science* 315:238–241
- Gu W, Swihart RK (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biol Conserv* 116:195–203
- Hanspach J, Kühn I, Pyšek P, Boos E, Klotz S (2008) Correlates of naturalization and occupancy of introduced ornamentals in Germany. *Perspect Plant Ecol Evol Syst* 10: 241–250
- Hines JE (2006) PRESENCE2- Software to estimate patch occupancy and related parameters. Available at www.mbr-pwrc.usgs.gov/software/presence.html (accessed October 2010)
- Kanary L, Locke A, Watmough J (2010) Evaluating the effectiveness of SCUBA-based visual searches for an invasive tunicate, *Ciona intestinalis*, in a Prince Edward Island estuary. *Aquat Invasions* 5:41–47
- Katsanevakis S (2006) Modeling fish growth: model selection, multi-model inference and model selection uncertainty. *Fish Res* 81:229–235
- Katsanevakis S (2009) Estimating abundance of endangered marine benthic species using distance sampling through SCUBA diving: the *Pinna nobilis* (Mollusca: Bivalvia) example. In: Columbus AM, Kuznetsov L (eds) *Endangered species: new research*. Nova Science Publishers, Hauppauge, NY, p 81–115
- Kulbicki M (1998) How the acquired behaviour of commercial reef fishes may influence the results obtained from visual censuses. *J Exp Mar Biol Ecol* 222:11–30
- Lancia RA, Nichols JD, Pollock KH (1994) Estimating the number of animals in wildlife populations In: Bookhout TA (ed) *Research and management techniques for wildlife and habitats*. The Wildlife Society, Bethesda, MD, p 215–253
- Last PR, White WT, Gledhill DC, Hobday AJ, Brown R, Edgar GJ, Pecl G (2011) Long-term shifts in abundance and distribution of a temperate fish fauna: a response to climate change and fishing practices. *Glob Ecol Biogeogr* 20:58–72
- MacKenzie DI (2005) Was it there? Dealing with imperfect detection for species presence/absence data. *Aust N Z J Stat* 47:65–74
- MacKenzie DI, Bailey LL (2004) Assessing the fit of site occupancy models. *J Agric Biol Environ Stat* 9:300–318
- MacKenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. *J Appl Ecol* 42:1105–1114
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255
- MacKenzie DI, Royle JA, Brown JA, Nichols JD (2004) Occupancy estimation and modeling for rare and elusive populations. In: Thompson WL (ed) *Sampling rare or elusive species*. Island Press, Washington, DC, p 149–172
- MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2006) *Occupancy estimation and modeling*. Academic Press, San Diego, CA
- MacNeil MA, Tyler EHM, Fonnesebeck CJ, Rushton SP, Polunin NVC, Conroy MJ (2008a) Accounting for detectability in reef-fish biodiversity estimates. *Mar Ecol Prog Ser* 367:249–260
- MacNeil MA, Graham NAJ, Conroy MJ, Fonnesebeck CJ and others (2008b) Detection heterogeneity in underwater visual-census data. *J Fish Biol* 73:1748–1763
- Manley PN, Zielinski WJ, Schlesinger MD, Mori SR (2004) Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. *Ecol Appl* 14:296–310
- Martinez-Solano I, Bosch J, Garcia-Paris M (2003) Demographic trends and community stability in a montane amphibian assemblage. *Conserv Biol* 17:238–244
- Moilanen A (2002) Implications of empirical data quality to metapopulation model parameter estimation and application. *Oikos* 96:516–530
- Moritz C, Patton JL, Conroy CJ, Parra JL, White GC, Beissinger SR (2008) Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science* 322:261–264
- Nichols JD, Karanth KU (2002) Statistical concepts: assessing spatial distributions. In: Karanth KU, Nichols JD (eds) *Monitoring tigers and their prey: a manual for wildlife managers, researchers, and conservationists*. Centre for Wildlife Studies, Bangalore, p 29–38
- O’Connell AF Jr, Talancy NW, Bailey LL, Sauer JR, Cook R, Gilbert AT (2006) Estimating site occupancy and detection probability parameters for meso- and large mammals in a coastal ecosystem. *J Wildl Manag* 70:1625–1633
- Pellet J (2008) Seasonal variation in detectability of butterflies surveyed with Pollard walks. *J Insect Conserv* 12: 155–162
- Philippart CJM, Anadón R, Danovaro R, Dippner JW and others (2011) Impacts of climate change on European marine ecosystems: observations, expectations and indicators. *J Exp Mar Biol Ecol* 400:52–69
- Pittman SJ, Brown KA (2011) Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PLoS ONE* 6:e20583 doi:10.1371/journal.pone.0020583
- Pollock KH, Nichols JD, Simons TR, Farnsworth GL, Bailey LL, Sauer JR (2002) Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:105–119
- Rilov G, Benayahu Y, Gasith A (2001) Low abundance and skewed population structure of the whelk *Stramonita haemastoma* along the Israeli Mediterranean coast. *Mar Ecol Prog Ser* 218:189–202
- Ronconi RA, Burger AE (2009) Estimating seabird densities from vessels transects: distance sampling and implications for strip transects. *Aquat Biol* 4:297–309

- Royle JA, Link WA (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87:835–841
- Royle JA, Nichols JD (2003) Estimating abundance from repeated presence absence data or point counts. *Ecology* 84:777–790
- Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (2002) Predicting species occurrences. Island Press, Washington, DC
- Sewell D, Beebee TJC, Griffiths RA (2010) Optimising biodiversity assessments by volunteers: the application of occupancy modelling to large-scale amphibian surveys. *Biol Conserv* 143:2102–2110
- Thompson WL (2004) Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters. Island Press, Washington, DC
- Tsiamis K, Panayotidis P (2007) Occurrence of *Codium fragile* subsp. *tomentosoides* (van Goor) P.C. Silva (Chlorophyta: Bryopsidophyceae: Bryopsidales: Codiaceae) in Greece. *Aquat Invasions* 2:74–76
- Tsiamis K, Montesanto B, Panayotidis P, Katsaros C, Verlaque M (2010) Updated records and range expansion of alien marine macrophytes in Greece (2009). *Mediterr Mar Sci* 11:61–79
- Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP (2003) Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecol Appl* 13:1790–1801
- Weber D, Hinterman U, Zangger A (2004) Scale and trends in species richness: considerations for monitoring biological diversity for political purposes. *Glob Ecol Biogeogr* 13:97–104
- Weller TJ (2008) Using occupancy estimation to assess the effectiveness of a regional multiple-species conservation plan: bats in the Pacific Northwest. *Biol Conserv* 141:2279–2289
- Zenetos A, Gofas S, Verlaque M, Çinar ME and others (2010) Alien species in the Mediterranean Sea by 2010. A contribution to the application of European Union's Marine Strategy Framework Directive (MSFD). Part I. Spatial distribution. *Mediterr Mar Sci* 11:381–493

Editorial responsibility: Lisandro Benedetti-Cecchi,
Pisa, Italy

Submitted: May 20, 2011; Accepted: February 8, 2012
Proofs received from author(s): April 27, 2012