Original Article
Fishery-dependent and -independent data lead to consistent estimations of essential habitats

Maria Grazia Pennino[1]*, David Conesa[2], Antonio López-Quílez[2], Facundo Muñoz[3], Angel Fernández[4], and Jose Maria Bellido[4]

[1]*Department of Ecology, Universidade Federal do Rio Grande do Norte—UFRN, Campus Universitário s/n, Lagoa Nova. CEP 59.098-970, Natal, RN, Brazil*
[2]*Departament d'Estadística i Investigació Operativa, Universitat de València, C/Dr. Moliner 50, Burjassot, Valencia 46100, Spain*
[3]*INRA, UR588, Amélioration, Génétique et Physiologie Forestières, Orléans F-45075, France*
[4]*Instituto Español de Oceanografía, Centro Oceanográfico de Murcia, C/Varadero 1, San Pedro del Pinatar, Murcia 30740, Spain*

*Corresponding author: e-mail: graziapennino@yahoo.it

Species mapping is an essential tool for conservation programmes as it provides clear pictures of the distribution of marine resources. However, in fishery ecology, the amount of objective scientific information is limited and data may not always be directly comparable. Information about the distribution of marine species can be derived from two main sources: fishery-independent data (scientific surveys at sea) and fishery-dependent data (collection and sampling by observers in commercial vessels). The aim of this paper is to compare whether these two different sources produce similar, complementary, or different results. We compare them in the specific context of identifying the Essential Fish Habitats of three elasmobranch species (*S. canicula*, *G. melastomus*, and *E. spinax*). Similarity and prediction statistics are used to compare the two different spatial patterns obtained by applying the same Bayesian spatio-temporal modelling approach in the two sources. Results showed that the spatial patterns obtained are similar, although differences are present. In particular, models based on fishery-dependent data are better able to identify temporal relationships between the probability of presence of the species and seasonal environmental variables. In contrast, fishery-independent data better discriminate spatial locations where a species is present or absent. Besides the spatial and temporal differences of the two datasets, the consistency of habitat results highlights the inclusion in each dataset of most of the environmental envelope of each species, both in time and space. Consequently, sampling data should be adapted to each species to reasonably cover their environmental envelope, and a combination of datasets will likely provide a better habitat estimation than using each dataset independently. These findings can be useful in helping fishery managers improve definition of survey design and analyses.

Keywords: Bayesian hierarchical models, cross validation, elasmobranchs, species distribution models, survey design.

Introduction

There is an increasing demand from marine ecologists for more effective protection of essential fish habitats (EFHs) to help restoration of fish stocks and sustainable exploitation of marine fish (e.g. Schmitten, 1999; Mumby *et al.*, 2004). EFHs have been defined as the habitats necessary for a species to spawn, breed, feed, and grow to maturity (US Sustainable Fisheries Act, 1996, Public Law 104–297), particularly for vulnerable species and fish size.

Mapping is an essential tool with which to identify EFHs, as it provides a clear picture of the distribution and extent of marine resources (Pennino *et al.*, 2014). However, in fishery ecology, while the importance of spatial management of marine resources is widely recognized, the amount of objective scientific information is limited and the data may not always be directly comparable due to habitat variations between sampling periods.

Information about the distribution of marine species can essentially be derived from two main sources, namely, fishery-independent data (scientific surveys at sea) and fishery-dependent data (collection and sampling by observers in commercial vessels). These two sources may provide different types of information with additional

details, and should be carefully selected according to the nature of the study and the objectives pursued.

Fishery-independent surveys mainly rely on expensive research programmes at sea carried out over relatively short periods of time. Survey data are considered to be of higher quality because sampling and collection are scientifically designed and standardized. In addition, both sampling statistics and the biological information on target species are taken into consideration during survey design (Board, 2000). However, they generally have a limited coverage in space and time (in terms of seasonality as well as the number of years of available data), which could lead to biased and imprecise estimations (Hilborn and Walters, 2013). According to Hilborn and Walters (2013), there are two main deficiencies in fishery-independent surveys: timing in relation to the seasonal cycle and location of the target species, and the fact that only a limited amount of data can be collected. Both deficiencies lead to unrepresentative sampling.

Fishery-dependent surveys can provide a long time-series, wide spatial coverage all year-round, and information on a large variety of target species, gear types, landing sites, and distribution channels (Lunn and Dearden, 2006). Data of this kind sometimes lack particular details such as the location of fishing grounds and species identity (catches are mostly identified to a higher taxonomic level), and also to be taken into consideration are issues of bias due to constraints imposed by management and the deliberate misreporting of catches. Some of the disadvantages of fishery-dependent surveys can be overcome by using on-board observers but inevitably only a small fraction of fishing activity can be covered in this way. Another issue arises because the sampled fishing locations are often repeated and samples are collected using preferential sampling because fishing fleets are commercially driven, unlike the systematic random selection of a scientific survey.

Within this context, our study makes use of two different sources of fishery data to test similar, complementary, or different spatial patterns of species EFHs in the two different datasets. We considered information gathered by on-board observers and data from the MEDIterranean Trawl Survey (MEDITS) of three of the main elasmobranch species (*Scyliorhinus canicula*, *Galeus melastomus*, and *Etmopterus spinax*) in the western Mediterranean Sea. Bayesian hierarchical spatio-temporal models were then used to identify EFHs and develop probabilistic spatial scenarios of the three species. Finally, similarity and prediction statistics were used to compare the resulting predictive spatial structures obtained with the two different fishery data sources.

This approach can help fishery managers to better define survey designs and analyses. The approach applied in this study also improves the understanding of species distribution and demonstrates the spatial needs for the management of any biologically important vulnerable species under threat.

Additionally, this research is useful for any study with different data sources and sampling schemes. Combining the different data sources into a robust and consistent statistical framework will provide a better insight into the study case, in this paper the EFH of three elasmobranch species in the western Mediterranean Sea.

Material and methods
The selected species
Because of their high level of vulnerability, cartilaginous fish are considered a priority in fishery research and management. In 1999, the FAO developed an International Plan of Action for the Conservation and Management of Sharks (IPOA-Shark); in 2003 the United Nations Environment Programme (UNEP) launched the Mediterranean action plan for cartilaginous fish, and in February 2009, the European Commission adopted the first EU Action Plan for the Conservation and Management of Sharks. The aim of the European Commission plan is to ensure that effective steps are taken to help rebuild threatened stocks, and to set down guidelines for the sustainable management of the fisheries concerned. The plan also includes measures to improve scientific knowledge about their stocks and fisheries (Mendoza *et al.*, 2014).

For these reasons, for this study we selected the three elasmobranch species with the highest catches, and that are a significant representation of the elasmobranch community in the studied area (Pennino *et al.*, 2013), namely, *Scyliorhinus canicula*, *Galeus melastomus*, and *Etmopterus spinax*.

Study area
The study area was Southeast Spain (western Mediterranean Sea), extending from the area of Cartagena and the Gulf of Alicante, between 37°15.6′ and 39°18.6′N, and 1°39.0′W and 1°52.2′E (Figure 1).

Fishery-dependent data
Fishery-dependent data were extracted from the Spanish Oceanographic Institute (Instituto Español de Oceanografía, IEO) observers' database, from a sampling fleet of 27 trawler boats, for the period 2006–2011. This sampling usually involves two to three expert observers, comprising a total of 400 hauls for the 6-year study period. From this database, we used the geographical location (latitude and longitude) and occurrence of the elasmobranch species for each sampled haul. An average point between the starting and finishing point of each haul was used as the geographical location.

The entire bottom trawl fleet operating in the area of study comprises 169 vessels landing an average of 8000 t per year. Bottom trawling of the seabed usually takes place on the shelf, harvesting a multi-specific catch with European hake (*Merluccius merluccius*) as the main target species. Because elasmobranchs are not target species of this commercial fishery, their occurrences are unbiased indicators of their presence/absence patterns. This also means that the selection of the sampling locations does not depend on the values of the spatial variable and so they are stochastically independent of the field process. Indeed, in studies aiming to model the distribution of the target species, the sampling process that determines the data locations and the species observations is not independent (Diggle *et al.*, 2010). Data from commercial fisheries are a clear example of preferential sampling since fishing fleets are likely to fish in areas where they are also likely to find target species. In all these cases, a preferential-sampling model is required to model the species distribution.

Fishery-independent data
Fishery-independent data were collected during the MEDITS (EU-funded MEDIterranean Trawl Survey) carried out from spring to early summer (April to June) across the whole Mediterranean Sea since 2000 (Bertrand *et al.*, 2002). The MEDITS survey uses a stratified sampling design based on depth (5 bathymetric strata: 10–50, 51–100, 101–200, 201–500, and 501–700 m). Sampling stations were placed randomly within each stratum at the beginning of the project. In all subsequent years, sampling was performed in similar locations. Specifically, 99 trawl hauls were extracted from the MEDITS database for the entire time series.
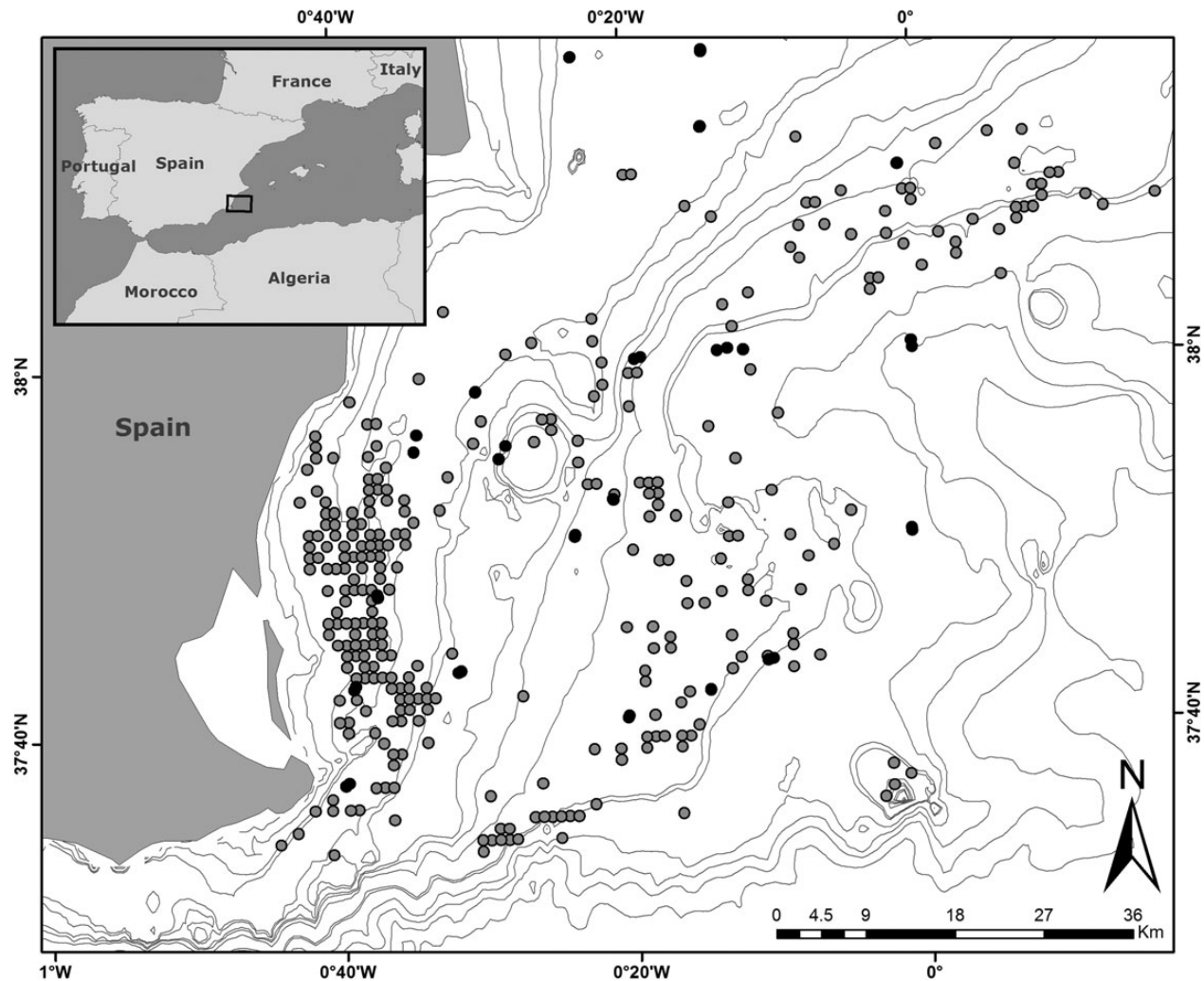
Figure 1. Map of the study area with sampling locations. In grey dots, fishery-dependent data and in black dots fishery-independent data.

Data of the occurrence of the elasmobranch species were extracted from the same time series as the fishery-dependent data (2006–2011) and spatial area (Gulf of Alicante) (Figure 1).

Environmental data

We considered three oceanographic variables: sea surface temperature (SST), chlorophyll a concentration (Chl a), and salinity, and four bathymetric features: bathymetry, type of substratum, slope, and distance to shore. These variables were selected as they are known to impact on the habitat selection of elasmobranch species across the Mediterranean Sea and in particular in the studied area (Pennino et al., 2013; Lauria et al., 2015).

The oceanographic satellite data (SST, Chl a, and salinity) were extracted as the monthly mean from the aqua-MODIS sensor (http://oceandata.sci.gsfc.nasa.gov/).

Bathymetry and type of substratum data were obtained from the IEO geoportal, accessible from the website of the Spanish Institute of Oceanography (www.ieo.es). The bathymetric map of the study area was interpolated using GRASS GIS (www.grass.fbk.eu) to obtain the depth value at any precise location. Specifically, first we rasterized contours with a resolution of 500 m then extracted specific values using the function r.surf.contour (www.grass.osgeo.org/wiki/Contourlines_to_DEM).

The slope and distance to shore information were derived from the bathymetric map using Slope Spatial Analyst and the Near (World Equidistant Cylindrical coordinate system) tools of ArcGis 9.2 (ESRI Inc., 2008; Redlands, CA, USA).

The original type of substratum polygon shapefile used in the model includes a classification of 10 levels of sediment granulometry categories. These categories covered mixed substrata that are mid-way between gravel and rocky seabeds (i.e. coarse gravel); they were aggregated in one group labelled hard substrata. Finally, to reduce variability in the analyses, and as we do not have observations for all 10 categories, we combined them into three groups: sand, mud, and hard substrata (Figure 2).

All environmental data were aggregated at a resolution of $0.05° \times 0.05°$ ($\sim$5500 m) using the "raster" package (Hijmans 2014) in R (R Development Core Team, 2015).

Collinearity between explanatory environmental variables was tested using a draftsman's plot and the Pearson correlation index. The variables were not highly correlated ($r < 0.5$), so all variables were considered in further analyses.

Modelling species occurrence

Species distribution models (SDMs) allow researchers to predict quantities of interest at unsampled locations based on measured
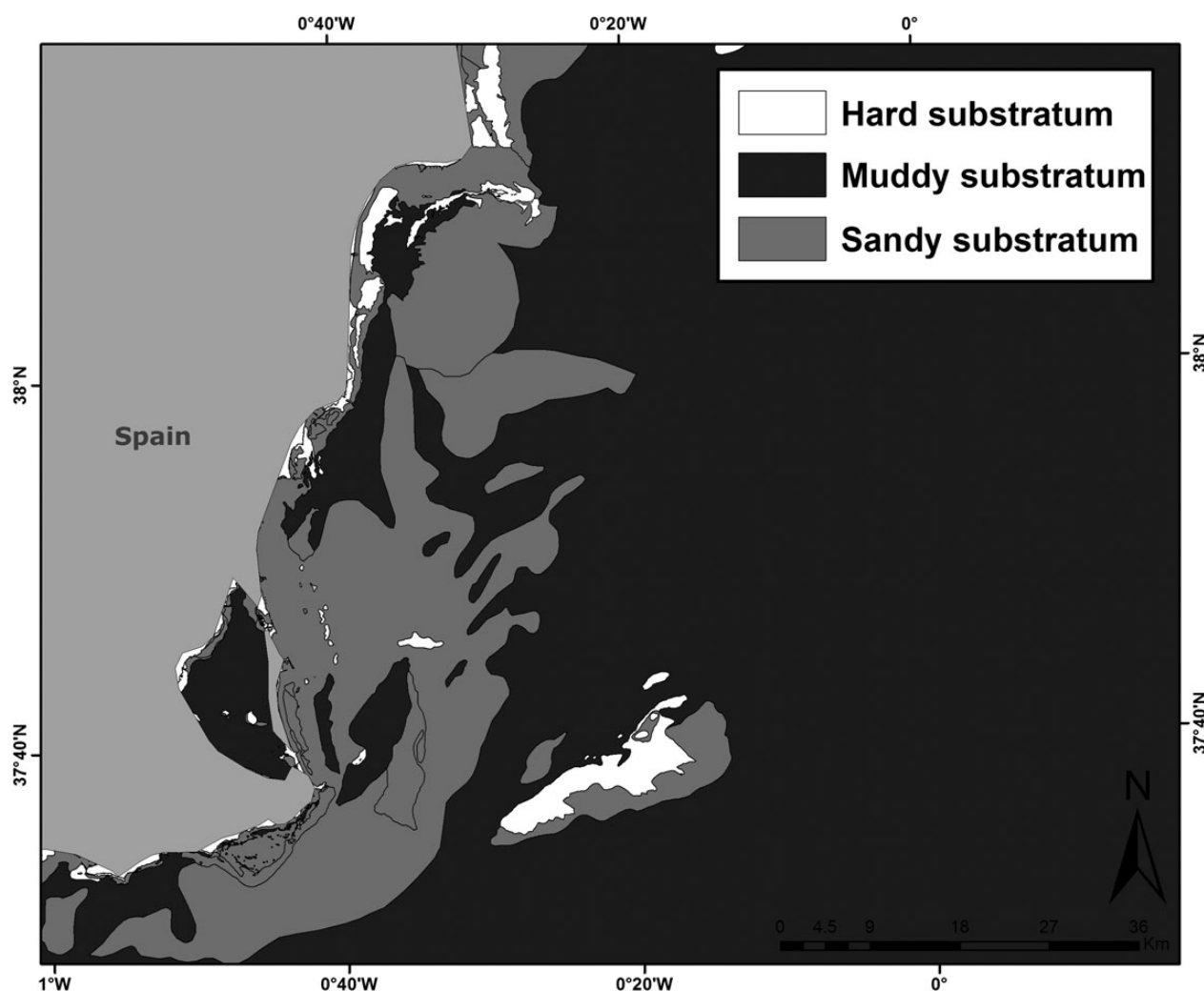
Figure 2. Map of the seabed type of the study area with three different categories: (a) hard substratum, (b) muddy substratum, and (c) sandy substratum.

values at nearby sampled locations, within the range of sampled environments, and within the same general time frame as the sampled environmental variables (Elith and Leathwick, 2009). Many approaches have been proposed to predict the spatial distribution of species; however, these approaches do not always provide accurate results if they are run without taking into account the large amount of variability in the measurements of response and environmental variables. Our approach to solve this issue is to use the Bayesian statistical methodology.

Bayesian methods have several advantages over traditional ones and are increasingly being used in fishery research (i.e. Punt and Hilborn, 1997; Millar and Meyer, 2000; Colloca *et al.*, 2009; Vermard *et al.*, 2010; Paradinas *et al.*, 2015). They provide a more realistic and accurate estimation of uncertainty because they can consider both the observed data and the model parameters as random variables (Banerjee *et al.*, 2004).

But the main reason for using the Bayesian approach is the way it deals with this kind of data, namely, geostatistical or point-referenced data. Bayesian modelling provides a way to incorporate the spatial component to model the residual spatial autocorrelation (Gelfand *et al.*, 2006). Moreover, it also allows us to deal with typical extensions of SDM models in which estimation and prediction can

be computationally difficult. Note that these models can also be considered to be a spatial extension of generalized linear models because the modelling process describes the variability in the response variable as a function of the explanatory variables with the addition of a stochastic spatial effect. Finally, hierarchical Bayesian structures can also be used to identify additional covariates that can improve the model fit or the existence of area effects that can affect species occurrence (Paradinas *et al.*, 2015).

We used a hierarchical Bayesian spatio-temporal approach to estimate the probability of occurrence of the three elasmobranch species (*S. canicula*, *G. melastomus*, and *E. spinax*). The response variable Z_{ij} is a binary variable that represents the presence (1) or absence (0) of the species in each fishing location identified in the fishing landings. Consequently, the conditional distribution of the data is $Z_{ij} \sim \mathrm{Ber}(\pi_{ij})$, where π_{ij} is the probability of occurrence at location i ($i = 1, \ldots n$) and year j ($j = 1, \ldots 6$), assuming that observations are conditionally independent given π_{ij}. These probabilities were modelled using the following hierarchical model:

$$Z_{ij} \sim \mathrm{Ber}(\pi_{ij}),$$
$$\mathrm{logit}(\pi_{ij}) = X_{ij}\boldsymbol{\beta} + Y_j + W_i,$$

where β is the vector of regression parameters, X_{ij} is the matrix of the explanatory covariates at location i and year j, Y_j is the component of the temporal random effect at year j and W_i represents the spatially structured random effect at location i. The latter effect represents unaccounted local environmental factors which introduce a positive spatial autocorrelation in the latent probabilities of occurrence.

A vague zero-mean Gaussian prior distribution with a variance of 100 was used for all of the parameters involved in the fixed effects, while for the spatial effect a zero-mean prior Gaussian distribution with a Matérn covariance structure was assumed (see Muñoz et al., 2013 for more detailed information about how to express the prior knowledge on spatial effects). Finally, a vague Gamma prior distribution with shape and scale parameters of 1 and 5e − 05, respectively, was assumed on the precision parameter λ_y of the temporal component.

Estimation and prediction were performed using the Integrated Nested Laplace Approximation methodology (Rue et al., 2009) and the package (www.r-inla.org) implemented in R software.

Variables were selected beginning with all possible interaction terms based on the deviance information criterion (DIC) (Spiegelhalter et al., 2002) and on the cross-validated logarithmic score (LCPO) measure (Roos and Held, 2011). Specifically, DIC was used as a measure of goodness of fit, while LCPO was used as a measure of the predictive quality of the models. DIC and LCPO are inversely related to the compromise between fit, parsimony, and predictive quality.

Model validation
Model validation was performed through 10 times cross validation based on randomly selected training and test datasets created by a random selection of 80 and 20% of each dataset, respectively (Fielding and Bell, 1997), using the "PresenceAbsence" package in R (Freeman and Moisen, 2008). The model performance was assessed using the area under the receiver-operating characteristic curve (AUC) (Fielding and Bell, 1997) and the true skill statistic (TSS) (Allouche et al., 2006).

AUC has been widely used in the species distribution modelling literature (Elith et al., 2006). It measures the model's ability to discriminate between sites in which the species is present and those in which it is absent. The values for AUC range from 0 to 1, where 0.5 indicates a performance no better than random, values between 0.7 and 0.9 are useful to indicate results of presence/absence different from random, and values >0.9 are excellent to ensure that the results are different from random.

TSS corrects AUC for the dependence of prevalence on specificity (i.e. ability to correctly predict absences) and sensitivity (i.e. ability to correctly predict presence) (Allouche et al., 2006).

Comparison of model predictions
The spatial structure of predictions derived from the two different data sources were compared using the similarity statistics Schoener's and Warren's described by Warren et al. (2008). These statistics range from 0, which indicates no overlap between areas, to 1, which means that the distributions are identical. Both statistics assume probability distributions defined over geographic space, in which p_{Xi} (or p_{Yi}) denotes the probability assigned by the SDM for species X (or Y) to cell i. Specifically, Schoener's statistic for niche overlap is

$$D(p_X, p_Y) = 1 - \frac{1}{2}\sum_i |p_{Xi} - p_{Yi}|$$

while Warren's statistic is

$$I(p_X, p_Y) = 1 - \frac{1}{2}H_2(p_X - p_Y)$$

which is based on the Hellinger distance, defined as follows:

$$H(p_X, p_Y) = \sqrt{\sum_i (\sqrt{p_{Xi}} - \sqrt{p_{Yi}})^2}.$$

These analyses were carried out using the nicheOverlap function of the dismo package (Hijmans et al., 2013) in R software.

Results
From fishery-dependent data
A total of 400 trawl fishery operations were observed over a period of 6 years. The presence of S. canicula was recorded in 204 of these 400 hauls, while the frequency was 135 for G. melastomus and 54 for E. spinax. The main predictors of elasmobranch habitats in the western Mediterranean Sea were depth, slope, and type of substratum and a stochastic spatial component that accounted for the residual spatial autocorrelation (Table 1).

Specifically, results showed a positive relationship between bathymetry, slope, and the probability of occurrence of the three species (Table 1). The SST and Chl a concentration showed a negative relationship with the species occurrence but affected only the distribution of S. canicula. The final models with the best fits of G. melastomus and E. spinax did not include the SST and Chl a concentration as relevant variables.

No important yearly differences were found in this area for the occurrence of these species. All models with the temporal effect show higher DICs and LCPO than those without this effect.

Finally, distance to shore and salinity were not selected as relevant in any final model.

Figure 3 shows the median posterior probability of occurrence of the three species. In particular, S. canicula shows a greater probability of occurrence over high slope gradient, in deeper waters where Chl a and SST values are higher, and where there are sandy seabeds (Figure 3a).

For G. melastomus and E. spinax the probability of occurrence is higher in habitats associated with muddy substrata and sandy beds, mainly from deeper waters and high slope gradient (Figure 3b and c).

From fishery-independent data
From 2006 to 2011, 99 trawl hauls were sampled across the Gulf of Alicante during the MEDITS surveys. The presence of S. canicula was recorded in 75 of these hauls, G. melastomus in 21, and E. spinax in 11.

The main predictors of elasmobranch species were similar to the ones found with the fishery-dependent data, namely depth, slope, and type of substratum, plus a stochastic spatial component (Table 2).

Salinity, SST, Chl a concentration and distance to shore were not relevant to the occurrence of the studied species. In addition, as with the fishery-dependent data, no yearly differences were found in this area.

Results showed a positive relationship between bathymetry, slope, and the probability of occurrence of the three species (Table 2).

The type of substrata has a different influence on each elasmobranch species. The S. canicula species showed a higher probability of occurrence in sandy seabeds, while G. melastomus and E. spinax are more likely to be found in muddy substrata.

Table 1. Numerical summary of the posterior distributions of the fixed effects for the best model of the three studied species using the fishery-dependet data.

Species	Predictors	Mean	s.d.	$Q_{0.025}$	$Q_{0.5}$	$Q_{0.975}$
S. canicula	Intercept	0.23	2.14	−4.23	0.73	4.32
	Log Depth	1.06	0.63	−0.32	1.04	2.05
	Seabed(Mud)	−0.32	0.42	−1.16	−0.33	0.51
	Seabed(Hard)	−1.91	0.95	−3.87	−1.88	−0.12
	Slope	0.21	0.11	−0.51	0.22	0.74
	Chlorophyll *a*	−13.96	7.83	−29.96	−13.82	1.24
	SST	−0.52	0.27	−0.22	−0.51	1.05
G. melastomus	Intercept	−1.72	2.91	7.53	−1.89	4.94
	Log Depth	0.33	0.27	0.25	0.35	0.85
	Seabed(Mud)	0.41	0.50	0.40	0.42	1.29
	Seabed(Hard)	−0.72	1.28	−3.36	−0.68	1.69
	Slope	0.19	0.02	−0.06	0.20	0.33
E. spinax	Intercept	−2.35	12.76	−32.34	−2.72	28.65
	Log Depth	5.45	3.63	2.70	5.35	10.39
	Seabed(Mud)	0.08	0.85	−1.61	0.09	1.73
	Seabed(Hard)	−0.73	1.28	−3.36	−0.68	1.70
	Slope	0.09	0.03	−0.06	0.08	0.10

This summary contains the mean, the standard deviation (s.d.), the median ($Q_{0.5}$), and a 95% credible interval, which is a central interval containing the 95% of the probability under the posterior distribution.
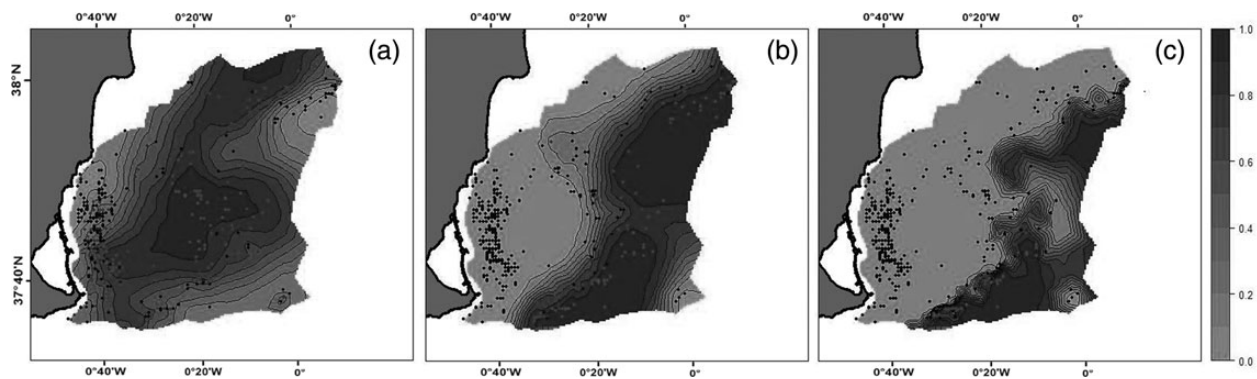
Figure 3. Median of the posterior probability of the presence of the studied elasmobranchs using fishery-dependent data: *S. canicula* (a), *G. melastomus* (b), and *E. spinax* (c). Sampling locations for the presence (in grey) and the absence (in black) were plotted.

Table 2. Numerical summary of the posterior distributions of the fixed effects for the best model of the three studied species using the fishery-independent data.

Species	Predictors	Mean	s.d.	$Q_{0.025}$	$Q_{0.5}$	$Q_{0.975}$
S. canicula	Intercept	0.55	0.65	0.12	0.47	1.13
	Log Depth	1.26	0.53	0.34	1.12	2.25
	Seabed(Mud)	−0.22	0.02	−1.54	−0.12	−0.01
	Seabed(Hard)	−1.34	0.58	−2.97	−1.14	−0.10
	Slope	0.34	0.05	0.03	0.15	0.98
G. melastomus	Intercept	1.23	0.98	0.78	1.12	1.89
	Log Depth	1.21	0.32	0.34	1.10	1.67
	Seabed(Mud)	0.87	0.50	0.23	0.76	1.15
	Seabed(Hard)	−0.97	0.57	−2.99	−0.87	−0.12
	Slope	0.67	0.02	0.24	0.59	1.01
E. spinax	Intercept	0.96	0.67	0.21	0.88	1.87
	Log Depth	4.67	0.78	1.89	4.46	6.78
	Seabed(Mud)	0.58	0.01	0.12	0.42	1.02
	Seabed(Hard)	−0.56	0.97	−2.98	−0.48	−0.02
	Slope	0.87	0.04	0.04	0.76	0.98

This summary contains the mean, the standard deviation (s.d.), the median ($Q_{0.5}$), and a 95% credible interval, which is a central interval containing the 95% of the probability under the posterior distribution.
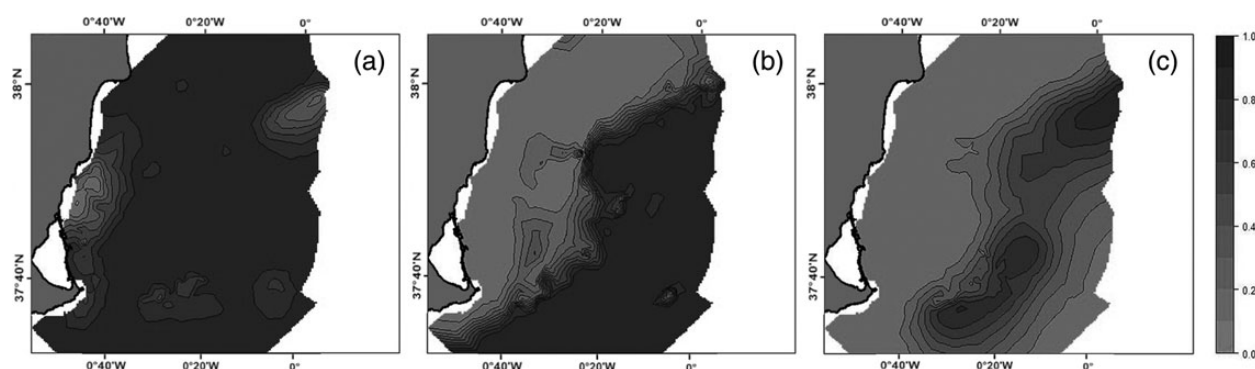
Figure 4. Median of the posterior probability of the presence of the studied elasmobranchs using fishery-independent data: *S. canicula* (a), *G. melastomus* (b), and *E. spinax* (c).

The posterior median probability of occurrence of the species is shown in Figure 4. Specifically, *G. melastomus* and *E. spinax* (Figure 4a and c) show a more marked probability of occurrence in deeper waters than *S. canicula* (Figure 3a).

Model prediction performance statistics

With both fishery-dependent and -independent data, all models achieved AUC values >0.80, which indicates excellent discrimination between locations where a species is present and those where it is absent (Table 3). Although the results are remarkably similar between the two data sources, AUC values for *G. melastomus* and *E. spinax* are slightly higher using the fishery-independent dataset. In contrast, fishery-dependent data for *S. canicula* showed higher AUC values (0.88 vs 0.85).

For all the models and both datasets, TSS values ranged between 0.60 and 0.80, also reflecting a good ability of the models to make true negative and true positive predictions (Table 3). Higher values of TSS are achieved with fishery-dependent data for both *S. canicula* and *G. melastomus*, while for *E. spinax* the opposite occurs.

Overall, the largest recorded difference is for the TSS values of the *G. melastomus* species with a difference of 0.26 between the two datasets.

Model comparison statistics

Overall, similarity statistics highlight a good level of overlapping of predictions from the two datasets. Indeed, for the three species Schoener's D ranged approximately from 0.7 to 0.9 while Warren's I was always >0.9.

Specifically, *S. canicula* was the species for which predictions from the two statistics were most similar ($I = 0.98$; $D = 0.87$). In contrast, the *G. melastomus* species had the lowest statistical values ($I = 0.91$; $D = 0.72$). Finally, the *I* value for *E. spinax* was 0.93, and for *D*, 0.74.

Discussion

Information on the mesoscale distribution patterns of species is lacking in fisheries research. This is particularly important to know the spatial and temporal mesoscale resolution for many commercial species and ecosystems (Mackinson, 2001). Hence, studies are required to develop spatially explicit predictive models needed for fishery management and to allow us to respond to change in fishery ecosystems (Mackinson, 2001). Despite recent awareness of the profound importance of properly identifying EFHs to more effective spatial fisheries management, much of our current

Table 3. Model prediction performance statistics for three species studied.

Species	AUC$_{FD}$	TSS$_{FD}$	AUC$_{FI}$	TSS$_{FI}$
S. canicula	0.88	0.68	0.85	0.66
G. melastomus	0.84	0.85	0.89	0.59
E. spinax	0.90	0.76	0.95	0.89

AUC, area under the receiver-operated characteristic curve; TSS, true skill statistic. AUC$_{FD}$ and TSS$_{FD}$ indicate the measures obtained using the fishery-dependent dataset, while AUC$_{FI}$ and TSS$_{FI}$ with the fishery-independent one.

understanding of species distribution and behaviour remains qualitative or highly uncertain (Moreno *et al.*, 2014).

Comparing and combining observations from different fishery data sources could provide a useful tool with which to bridge some gaps in our knowledge. Indeed, although the comparison of data sources is common in the context of stock assessment models, it has never been done for EFHs. In this study, we try to fill this gap by applying Bayesian spatial-temporal models to identify the EFHs of three elasmobranch species using data from two different fishery sources (fishery-independent and fishery-dependent surveys). We then compared the results using similarity and prediction statistics.

First, it should be noted that there are differences in the sampling design between these two sources that could influence the results. In particular, there is a considerable difference in sampling size between the two datasets: for the same period, fishery-dependent data have 400 observations compared with the 99 of the fishery-independent data. In addition, the observations from the two databases do not cover the same area. While fishery-independent data cover the whole study area, the fishery-dependent data cover only the southern part of the study area. Finally, the temporal coverage is not the same: fishery-dependent data are collected yearly while fishery-independent data are sampled only seasonally.

Despite these differences in sampling design, results show very similar spatial patterns between the two sources, although some discrepancies are present. In particular, fishery-dependent data revealed relationships with environmental variables, namely monthly SST and Chl *a*, that are not identified by fishery-independent data. This could be because the latter are always collected in a restricted period as we mentioned above, and thus can only capture a narrow picture of the distribution of the species. In support of this theory, fishery-independent

data are able to identify the relationship with bathymetric features and the occurrence of the species, possibly because the bathymetric features change less than the oceanographic variables. This could be useful information for researchers about which kind of dataset to use for studies that focus more on understanding the relationships between species and oceanographic variables.

However, although the observations of the MEDITS survey were less than one-quarter of those in the fishery-dependent dataset, the Bayesian spatial-temporal models yielded similar predictions. This could also be influenced by the Bayesian approach which allows researchers to achieve accurate estimations and predictions even with smaller observed datasets (Martínez-Abraín *et al.*, 2014).

From a biological point of view results are consistent with studies that explore the elasmobranchs distribution in similar and other habitats (Damalas *et al.*, 2010; Pennino *et al.*, 2013; Lauria *et al.* 2015), especially for bathymetric features.

Fishery-dependent data show greater ability to correctly predict temporal absences and presences than fishery-independent data. In contrast, fishery-independent data better discriminate locations where a species is present or absent, as shown by the AUC values. This could be an effect of the fact that a larger fishery-dependent dataset (400 observations vs 99 in fishery-independent data) could improve the overall predictive power of the model, while the more homogeneous sample design of MEDITS surveys could result in better assessments of the spatial distribution of the species.

It is worth stressing that we selected non-target species of commercial fishing, so sampled locations are an unbiased indicator of species occurrence as they are independently selected. In the opposite case where target species are selected, the sampling processes that determine the data locations and the species observations are stochastically dependent. Fishing vessels obviously go to areas where they are likely to find target species. To identify the EFHs of target species, fishery-independent data are the best option. In cases where only fishery-dependent data are available, a preferential-sampling model must be applied (Diggle *et al.*, 2010).

Maps of the predicted probability of occurrence for the three species show similar spatial patterns and this is also confirmed by similarity statistics. The similarity in species distributions between fishery-dependent data and fishery-independent data suggest that, overall, the commercial fleet sampled the study area in a similar fashion as did research vessels. As Fox and Starr (1996) point out, it therefore follows that fishery-dependent data can be used to evaluate or augment research data, and they may become the only cost-effective way to increase the amount of information available in specific areas.

In conclusion, these findings suggest that the two sources could have additional information and complementary results that can be jointly used for fishery management purposes, to better define the survey design and analyses as well to increase our knowledge about vulnerable species. For example, the restricted time coverage of fishery-independent data could be integrated using yearly fishery-dependent data, and the spatial uncertainty of fishery-dependent locations may be corrected using fishery-independent data.

Besides the spatial and temporal differences of the two datasets, the consistency of habitat results highlights the inclusion in each dataset of most of the environmental envelope of the species, both in time and space. Consequently, sampling data should be adapted to each species to reasonably cover their environmental envelope, and a combination of datasets will likely provide a better habitat estimation than using each dataset independently. The integration of multiple sources of information into science and management is a potentially invaluable tool that should not be overlooked. This research provides good results for data analysis of two different sources with a different sampling design. Sampling design usually restricts the statistical method to be used. Indeed, further analyses are required to develop a single model that can take into account all the variability that can arise from combining different sources of data collected with different sampling techniques and temporal and spatial coverage. The Bayesian approach may provide the answer to combine datasets from different sampling schemes and at the same time, maintain a sound statistical framework, which is especially useful in ecological studies.

Acknowledgements

References

Allouche, O., Tsoar, A., and Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43: 1223–1232.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC, Boca Raton.

Bertrand, J. A., Gil De Sola, L., Papaconstantinou, C., Relini, G., and Souplet, A. 2002. The general specifications of the MEDITS surveys. Scientia Marina, 66(Suppl. 2): 9–17.

Board, O. S. 2000. Improving the Collection, Management, and Use of Marine Fisheries Data. National Academies Press, Washington, D.C.

Colloca, F., Bartolino, V., Lasinio, G. J., Maiorano, L., Sartor, P., and Ardizzone, G. 2009. Identifying fish nurseries using density and persistence measures. Marine Ecology Progress Series, 381: 287–296.

Damalas, D., Maravelias, C. D., Katsanevakis, S., Karageorgis, A. P., and Papaconstantinou, C. 2010. Seasonal abundance of non-commercial demersal fish in the eastern Mediterranean Sea in relation to hydrographic and sediment characteristics. Estuarine, Coastal and Shelf Science, 89: 107–118.

Diggle, P. J., Menezes, R., and Su, T. L. 2010. Geostatistical inference under preferential sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics), 59: 191–232.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., *et al.* 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29: 129–151.

Elith, J., and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40: 677.

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

Fox, D., and Starr, R. 1996. Comparison of commercial fishery and research catch data. Canadian Journal of Fisheries and Aquatic Sciences, 53: 2681–2694.

Freeman, E. A., and Moisen, G. 2008. PresenceAbsence: an R package for presence absence analysis. Journal of Statistical Software, 23: 1–31.

Gelfand, A. E., Silander, J. A., Wu, S., Latimer, A., Lewis, P. O., Rebelo, A. G., and Holder, M. 2006. Explaining species distribution patterns through hierarchical modeling. Bayesian Analysis, 1: 41–92.

Hilborn, R., and Walters, C. J. Eds. 2013. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media, Dordrecht.

Hijmans, R. J., Phillips, S. J., Leathwick, J. R., and Elith, J. 2013. Package "dismo" Version 0.9-3. http://www.ashrafawd.com/joomla/d/spss.pdf (accessed 25 March 2014).

Hijmans, R.J. 2014. Raster: Geographic data analysis and modeling. R package Version 2.2-12. https://cran.rproject.org/web/packages/raster/raster.pdf (accessed 25 March 2014).

Lauria, V., Gristina, M., Attrill, M. J., Fiorentino, F., and Garofalo, G. 2015. Predictive habitat suitability models to aid conservation of elasmobranch diversity in the central Mediterranean Sea. Scientific Reports, 5. doi:10.1038/srep13245.

Lunn, K. E., and Dearden, P. 2006. Monitoring small-scale marine fisheries: An example from Thailand's Ko Chang archipelago. Fisheries Research, 77: 60–71.

Mackinson, S. 2001. Integrating local and scientific knowledge: an example in fisheries science. Environmental Management, 27: 533–545.

Martínez-Abraín, A., Conesa, D., and Forte, A. 2014. Subjectivism as an unavoidable feature of ecological statistics. Animal Biodiversity and Conservation, 37: 141–143.

Mendoza, M., Garrido, D., and Bellido, J. M. 2014. Factors affecting the fishing impact on cartilaginous fishes in southeastern Spain (western Mediterranean Sea). Scientia Marina, 78(Suppl. 1): 67–76.

Millar, R. B., and Meyer, R. 2000. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. Canadian Journal of Fisheries and Aquatic Sciences, 57: 43–50.

Moreno, A., Lourenço, S., Pereira, J., Gaspar, M. B., Cabral, H. N., Pierce, G. J., and Santos, A. M. P. 2014. Essential habitats for pre-recruit *Octopus vulgaris* along the Portuguese coast. Fisheries Research, 152: 74–85.

Mumby, P. J., Edwards, A. J., Arias-González, J. E., Lindeman, K. C., Blackwell, P. G., Gall, A., Gorczynska, M., *et al.* 2004. Mangroves enhance the biomass of coral reef fish communities in the Caribbean. Nature, 427: 533–536.

Muñoz, F., Pennino, M. G., Conesa, D., López-Quílez, A., and Bellido, J. M. 2013. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. Stochastic Environmental Research and Risk Assessment, 27: 1171–1180.

Paradinas, I., Conesa, D., Pennino, M. G., Muñoz, F., Fernandez, A. M., López-Quílez, A., and Bellido, J. M. 2015. A Bayesian Spatio-temporal approach to identifying fish nurseries by validating persistence areas. Marine Ecology Progress Series, 528: 245–255.

Pennino, M. G., Muñoz, F., Conesa, D., López-Quílez, A., and Bellido, J. M. 2013. Modeling sensitive elasmobranch habitats. Journal of Sea Research, 83: 209–218.

Pennino, M. G., Muñoz, F., Conesa, D., López-Quílez, A., and Bellido, J. M. 2014. Bayesian spatio-temporal discard model in a demersal trawl fishery. Journal of Sea Research, 90: 44–53.

Punt, A. E., and Hilborn, R. A. Y. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries, 7: 35–63.

R Development Core Team. 2015. R: A language and environment for statistical computing. R Found Stat Comp, Vienna, Austria. http://www.R-project.org.

Roos, M., and Held, L. 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Analysis, 6: 259–278.

Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B, 71: 319–392.

Schmitten, R. A. 1999. Essential fish habitat: opportunities and challenges for the next millennium. In Fish Habitat: Essential Fish Habitat and Rehabilitation, pp. 3–10. Ed. by L. Benaka. American Fisheries Society, Bethesda, MD.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64: 583–639.

Vermard, Y., Rivot, E., Mahévas, S., Marchal, P., and Gascuel, D. 2010. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. Ecological Modelling, 221: 1757–1769.

Warren, D. L., Glor, R. E., and Turelli, M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution, 62: 2868–2883.

Handling editor: Ernesto Jardim