SCIENTIFIC REPORTS

OPEN

Assessing cetacean surveys throughout the Mediterranean Sea: a gap analysis in environmental space

Laura Mannocci [1,2], Jason J. Roberts[1], Patrick N. Halpin [1], Matthieu Authier [3], Oliver Boisseau [4,5], Mohamed Nejmeddine Bradai[6], Ana Cañadas [7], Carla Chicote[8], Léa David [9], Nathalie Di-Méglio[9], Caterina M. Fortuna [10], Alexandros Frantzis [11], Manel Gazo [8], Tilen Genov [12,13,14], Philip S. Hammond [14], Draško Holcer [15,16], Kristin Kaschner [17], Dani Kerem [18], Giancarlo Lauriano[10], Tim Lewis[4,5,19], Giuseppe Notarbartolo di Sciara [20], Simone Panigada[20], Juan Antonio Raga[21], Aviad Scheinin[18,22], Vincent Ridoux[23], Adriana Vella [24,25] & Joseph Vella[25,26]

Heterogeneous data collection in the marine environment has led to large gaps in our knowledge of marine species distributions. To fill these gaps, models calibrated on existing data may be used to predict species distributions in unsampled areas, given that available data are sufficiently representative. Our objective was to evaluate the feasibility of mapping cetacean densities across the entire Mediterranean Sea using models calibrated on available survey data and various environmental covariates. We aggregated 302,481 km of line transect survey effort conducted in the Mediterranean Sea within the past 20 years by many organisations. Survey coverage was highly heterogeneous geographically and seasonally: large data gaps were present in the eastern and southern Mediterranean

[1]Marine Geospatial Ecology Laboratory, Nicholas School of the Environment, Duke University, Durham, NC, 27708, USA. [2]UMR MARBEC (IRD, Ifremer, Université de Montpellier, CNRS), Institut Français de Recherche pour l'Exploitation de la Mer, Avenue Jean Monnet, CS 30171, 34203, Sète, France. [3]Observatoire PELAGIS UMS 3462 Université de La Rochelle/CNRS, 5 allées de l'Océan, 17 000, La Rochelle, France. [4]Marine Conservation Research (MCR), 94 High Street, Kelvedon, CO5 9AA, UK. [5]Song of the Whale research team, International Fund for Animal Welfare (IFAW), 87-90 Albert Embankment, London, SE1 7UD, UK. [6]Institut National des Sciences et Technologies de la Mer (INSTM), Centre de Sfax, B.P. 1035, Sfax, 3018, Tunisia. [7]Alnilam Research and Conservation, Pradillos 29, 28491, Navacerrada, Madrid, Spain. [8]SUBMON - Marine Environmental Services, Rabassa, 49, 08024, Barcelona, Spain. [9]EcoOcéan Institut, 18 rue des Hospices, 34090, Montpellier, France. [10]Italian National Institute for Environmental Protection and Research (ISPRA), via Vitaliano Brancati 60, 00144, Rome, Italy. [11]Pelagos Cetacean Research Institute, Terpsichoris 21, 16671, Vouliagmeni, Greece. [12]Morigenos - Slovenian Marine Mammal Society, Kidričevo nabrežje 4, 6330, Piran, Slovenia. [13]Department of Biodiversity, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, 6000, Koper, Slovenia. [14]Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, KY16 8LB, Scotland, UK. [15]Blue World Institute of Marine Research and Conservation, Kaštel 24, HR-51551, Veli Lošinj, Croatia. [16]Croatian Natural History Museum, Demetrova 1, 10000, Zagreb, Croatia. [17]Department of Biometry and Environmental Systems Analysis, Albert-Ludwigs University Freiburg, Tennenbacher Straße 4, 79106, Freiburg i. Br., Germany. [18]Israel Marine Mammal Research & Assistance Center, Institute of Maritime Studies, School of Marine Sciences, The University of Haifa, Mt Carmel, 31095, Haifa, Israel. [19]North Atlantic & Mediterranean Sperm Whale Catalogue (NAMSC), London, United Kingdom. [20]Tethys Research Institute, Acquario Civico, Viale G.B. Gadio 2, 20121, Milano, Italy. [21]Unidad de Zoología Marina, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, University of Valencia, Aptdo 22085, 46071, Valencia, Spain. [22]The Morris Kahn Marine Research Centre, The University of Haifa, Haifa, Israel. [23]Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 Université de La Rochelle/CNRS, 2 avenue Olympe de Gouges, 17000, La Rochelle, France. [24]Conservation Biology Research Group, Department of Biology, University of Malta, Msida, MSD2080, Malta. [25]The Biological Conservation Research Foundation, BICREF, PO BOX 30, Hamrun, Malta. [26]Department of Computer Information Systems, University of Malta, Msida, MSD2080, Malta. Correspondence and requests for materials should be addressed to L.M. (email: laura.mannocci@duke.edu)

1

and in non-summer months. We mapped the extent of interpolation *versus* extrapolation and the proportion of data nearby in environmental space when models calibrated on existing survey data were used for prediction across the entire Mediterranean Sea. Using model predictions to map cetacean densities in the eastern and southern Mediterranean, characterised by warmer, less productive waters, and more intense eddy activity, would lead to potentially unreliable extrapolations. We stress the need for systematic surveys of cetaceans in these environmentally unique Mediterranean waters, particularly in non-summer months.

A gap analysis—defined here as the process of assembling various datasets over a desired study area to identify where or when knowledge is lacking—is the first important step towards the development of large-scale species distribution models (SDMs)[1–5]. In the marine realm, data collection is particularly difficult, resource intensive and expensive[6]. As a result, our knowledge of mobile marine species distributions and densities is far from complete, underlining the need for gap analyses across large ocean basins. Gap analyses of mobile marine species observation data have led to identification of geographic areas and seasons with important knowledge gaps[7,8]. For example, a global gap analysis of line transect surveys used to derive abundance estimates and habitat-based density models of cetaceans has revealed large geographic gaps in the Southern Hemisphere and seasonal gaps in non-summer months[8].

Assessing the geographic and seasonal coverage of species observation datasets is informative but not sufficient to assess the feasibility of SDMs, which typically rely on inferred species-environment relationships to derive predictions. It is critical to assess the coverage of datasets in environmental space to evaluate the extent of extrapolation when a SDM calibrated on existing data is used for prediction across a study region. Extrapolation in environmental space can lead to highly uncertain predictions because species-environment relationships are unknown in unsampled environments[9,10]. The extent of extrapolation has traditionally been visualized with environmental envelopes based on the ranges of individual covariates spanned by the data[10–12]. However, approaches based on such univariate envelopes fail to detect combinations of covariates that are within the univariate environmental space but outside the multivariate environmental space[13]. An analysis of datasets' coverage in multivariate environmental space can flag potentially unreliable extrapolations resulting from these novel combinations[10,14,15].

The Mediterranean Sea (Fig. 1) is unique among large sea basins because it constitutes a miniature ocean with contrasted physical, climatic and biological characteristics[16], and supports a highly diverse marine fauna including large mobile animals such as cetaceans[17,18]. The Mediterranean Sea biodiversity is undergoing profound alterations as high levels of anthropogenic pressures synergistically interact with the effects of climate change[19,20]. The manmade Suez Canal, which exposes the Mediterranean Sea to the distinct fauna of the Red Sea, caused the former to become the globally most invaded marine ecosystem[21], augmenting and hastening biodiversity shifts. Despite these marked anthropogenic pressures, the collection of systematic data to assess marine animal abundances and responses to these stressors has been heterogeneous throughout the Mediterranean Sea, reflecting the uneven distribution of funding for population monitoring[22]. Cetacean population monitoring represents no exception: line transect survey programs to estimate cetacean abundances have been implemented mostly by European countries in the northwestern and central Mediterranean[23–26]. The wide range of some cetaceans across the Mediterranean Sea (e.g., fin whale and sperm whale[27–29]), combined with their vulnerability to the multiple anthropogenic pressures[30–33], stress the need to develop predictive models to map their densities throughout the Mediterranean Sea.

As part of a regional-scale collaboration, we assembled for the first time line transect survey data collected across the Mediterranean Sea to identify gaps in the geographic, temporal, and environmental coverage of survey effort. Our objective was to evaluate the feasibility of mapping cetacean densities in the entire Mediterranean Sea by using models calibrated on available survey data and various environmental covariates. Our approach gives novel insights on traditional gap analyses solely based on spatiotemporal coverage, helps prioritise future survey efforts in the Mediterranean Sea, and is widely applicable to other marine regions and taxa.

Results

Spatiotemporal coverage of surveys.   We aggregated line transect surveys conducted by 12 organisations, including universities, consultancy companies and non-governmental organisations, resulting in a total of 302,481 km of effort (Table 1, Fig. 2). Aerial surveys represented 149,225 km of effort and shipboard surveys represented 153,256 km of effort.

Survey effort was concentrated in the northwestern Mediterranean and comparatively patchy in the eastern and southern Mediterranean (Fig. 3a). The Algero-Provençal basin was the Mediterranean subregion that received the largest amount of survey effort, followed by the Tyrrhenian Sea/eastern Ligurian Sea and the Alborán Sea/Strait of Gibraltar (Table 2). In comparison, the Levantine Sea and Aegean Sea subregions received only 2.1% of the effort each.

Aerial surveys were conducted off Spain west of the Balearic Islands, in the Pelagos Sanctuary (a marine protected area for marine mammals in the Ligurian Sea established by international treaty[34]), in the Tyrrhenian Sea, in the Strait of Sicily (especially around the Maltese Islands), in the northern Ionian Sea, and in the Adriatic Sea (Fig. 3b). Shipboard surveys were conducted primarily in the northern Alborán Sea, in the western Ligurian Sea, in the eastern Ionian Sea, and around the Maltase islands. Shipboard survey effort was very patchy in the rest of the Mediterranean Sea (Fig. 3c).

In terms of inter-annual survey coverage, two peaks were apparent: a small peak in the early 2000s and a larger peak between 2009 and 2013 (Fig. 4a) corresponding to the implementation of major aerial surveys. Survey effort

2
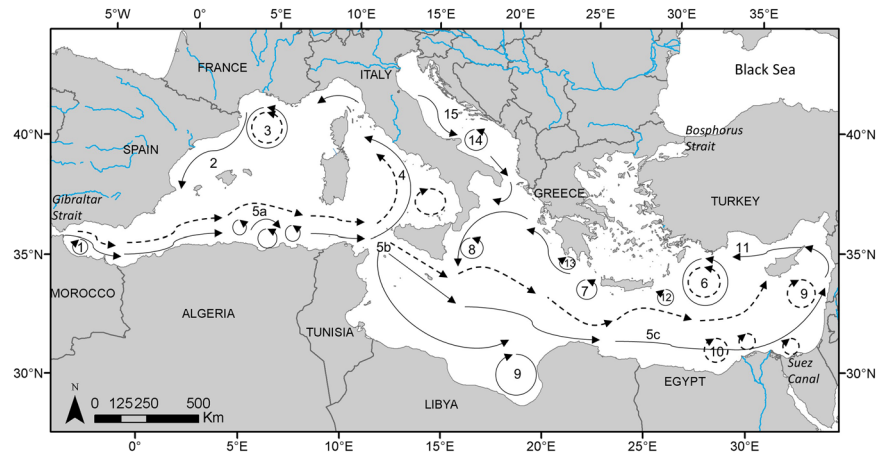
Figure 1. Map of main surface currents and gyres in the Mediterranean Sea. Dashed arrows represent summer circulation, plain arrows represent winter circulation. (1) Western Alborán gyre; (2) Ligurian-Provençal current; (3) Lions Gyre; (4) Thyrrhenian cyclonic circulation with summer weakening and eastern anticyclone; (5a) Algerian current and eddies, (5b) Atlantic Ionian stream and (5c) mid-Mediterranean jet; (6) Rhodes gyres; (7) Western Cretan gyre; (8) Western Ionian gyre; (9) Gulf of Sirte anticyclone; (10) Shikmona and Mers a-Matruh gyres; (11) Cicilian and Asia Minor current; (12) Iera-Petra gyre; (13) Pelops gyre; (14) Southern Adriatic gyre; (15) Western Adriatic coastal current. Figure adapted from Pinardi and Masseti (2000[52]. The map was generated with ArcGIS (http://desktop.arcgis.com/en/) (version 10.2.2).

Surveying entities	Platform	Surveyed years	Surveyed subregion[1]	Total effort (km)	References
Alnitak – ALNILAM	Ship	1997–2011	Alborán Sea/Strait of Gibraltar	43,283	[33,66,67]
BWI – ISPRA	Aircraft	2010, 2013	Adriatic Sea	16,796	[68]
EcoOcéan Institut and partners[2]	Ship	1997–2002; 2005–2015	Algero-Provençal basin	52,608	[31]
IFAW – MCR	Ship	2003, 2004, 2005, 2007, 2013	Basin-wide	17,824	[45,69,70]
IMMRAC	Ship	2005	Levantine Sea	1,458	[46]
INSTM	Ship	2001, 2003, 2005	Strait of Sicily/Tunisian Plateau/Gulf of Sirte and Tyrrhenian Sea/eastern Ligurian Sea	2,352	[71]
PELAGIS Observatory	Aircraft	2011, 2012	Algero-Provençal basin and Tyrrhenian Sea/eastern Ligurian Sea	32,592	[26]
Pelagos Cetacean Research Institute	Ship	2001–2014	Ionian Sea/Central Mediterranean and Aegean Sea	16,742	[72]
SUBMON Marine Environmental Services	Ship	2010 2011 2015	Algero-Provençal basin	2,951	[73–75]
TETHYS – ISPRA	Aircraft and ship	2008–2011; 2013, 2014, 2016	Algero-Provençal basin and Tyrrhenian Sea/eastern Ligurian Sea	64,795	[23,25,76–78]
CBRG, University of Malta[3]	Aircraft and ship	1997–2015	Strait of Sicily/Tunisian Plateau/Gulf of Sirte	24,704	[24,79,80]
University of Valencia	Aircraft	2000–2003; 2010, 2011, 2013	Algero-Provençal basin	26,376	[81,82]

Table 1. Details of Mediterranean line transect surveys incorporated in this gap analysis. Surveying entities: BWI = Blue World Institute of Marine Research and Conservation; CBRG = Conservation Biology Research Group; IFAW = International Fund for Animal Welfare; IMMRAC = Israel Marine Mammal Research and Assistance Center; INSTM = Institut National des Sciences et Technologies de la Mer; ISPRA = Italian National Institute for Environmental Protection and Research; MCR = Marine Conservation Research. [1]Mediterranean subregions following previous studies[18,56]. [2]Partners: École Pratique des Hautes Études, WWF-France, Swiss Cetacean Society, Cybelle Planète, Participe Futur and Fondation Nicolas Hulot. [3]A selection of the aerial and shipboard survey data collected by CBRG around the Maltese Islands was used in this analysis. Thus, the reported 24,704 km of effort represents part of the actual aerial and shipboard survey effort.

was exceptionally low in 2015. The northern Alborán Sea, the northern Gulf of Lions and Ligurian Sea, the area around Maltese Islands, and the Hellenic Trench were covered by surveys in most years, reflecting long-term monitoring efforts (Supplementary Fig. S1).

3
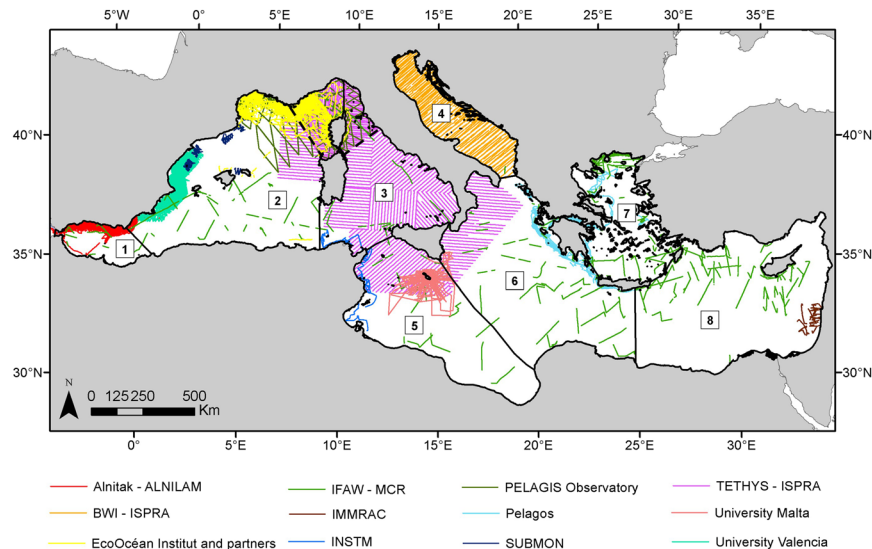
Figure 2. Line transect surveys in the Mediterranean Sea. Colours represent entities responsible for these surveys. Mediterranean subregions following Notarbartolo di Sciara (2016) and UNEP-MAP-RAC/SPA (2010)[18,56]: (1) Alborán Sea/Strait of Gibraltar, (2) Algero-Provençal Basin, (3) Tyrrhenian Sea/eastern Ligurian Sea, (4) Adriatic Sea, (5) Strait of Sicily/Tunisian Plateau/Gulf of Sirte, (6) Ionian Sea/Central Mediterranean, (7) Aegean Sea, (8) Levantine Sea. The location of the Pelagos Sanctuary[34] is indicated with black dashed lines. Surveying entities: BWI = Blue World Institute of Marine Research and Conservation; ISPRA = Italian National Institute for Environmental Protection and Research; IMMRAC = Israel Marine Mammal Research and Assistance Center; INSTM = Institut National des Sciences et Technologies de la Mer; IFAW = International Fund for Animal Welfare; MCR = Marine Conservation Research. The map was generated with ArcGIS (http://desktop.arcgis.com/en/) (version 10.2.2).

In terms of intra-annual survey coverage, the amount of survey effort was largest in July and August and lowest in February and April (Fig. 4b). Most surveys were conducted from May to September. Survey effort was spread across the Mediterranean Sea from May to September, but restricted to the northwestern and central Mediterranean the rest of the year (Supplementary Fig. S2).

Environmental coverage of surveys. *Univariate environmental coverage.* The surveys spanned broad ranges of values for all considered static covariates (depth, slope, distance to seamounts, and distance to canyons), resulting in limited univariate extrapolation throughout the Mediterranean Sea (Table 3) (Supplementary Figs S3 and S4).

The surveys spanned narrower ranges of values for the considered dynamic covariates (sea surface temperature, primary productivity, and eddy kinetic energy), resulting in a larger univariate extrapolation (especially in non-summer months) (Table 3).

The extent of univariate extrapolation would be largest with the model including sea surface temperature only. In winter and spring, extrapolation would be needed to predict cetacean densities in warmer waters of the southeastern Mediterranean and in colder waters of the northern Mediterranean (Gulf of Lions, Adriatic Sea, and northern Aegean Sea) (Fig. 5a, Supplementary Figs S5 and S6). In summer and autumn, extrapolation would occur in warmer waters of the Levantine Sea and of the Tunisian Plateau (but with a reduced extrapolation compared to winter and spring).

Extrapolation in more productive coastal waters would occur year-round on the Tunisian Plateau and in non-summer months in the northern Adriatic Sea and off Egypt. In spring and autumn, extrapolation would occur in the less productive offshore waters of the southeastern Mediterranean (Fig. 5b, Supplementary Figs S7 and S8).

The extent of univariate extrapolation would be lowest with the model including eddy kinetic energy only. Extrapolation would occur in waters characterised by more intense eddy activity in the Alborán Sea and Levantine Sea, mostly in winter and spring (Fig. 5c, Supplementary Figs S9 and S10).

Multivariate environmental coverage. As expected, the extent of extrapolation was larger with combinations of covariates than with individual covariates (Table 3). For example, the extrapolation extent with the combination of sea surface temperature, primary productivity, and eddy kinetic energy was larger than the extrapolation extents with these three covariates individually.

For the model including all four static covariates, extrapolation would be limited to 3.7% of the Mediterranean Sea (Table 3). Extrapolation would mostly occur in offshore waters of the central Mediterranean and off Egypt (Fig. 6a). The proportions of data nearby in the multivariate space defined by static covariates were lowest on the Tunisian Plateau, in the northern Adriatic Sea, and in offshore waters throughout the Mediterranean Sea (Fig. 6b). Differences between extrapolation metrics (the extent of extrapolation and the proportion of data nearby) are presented in the Methods and discussed further below.
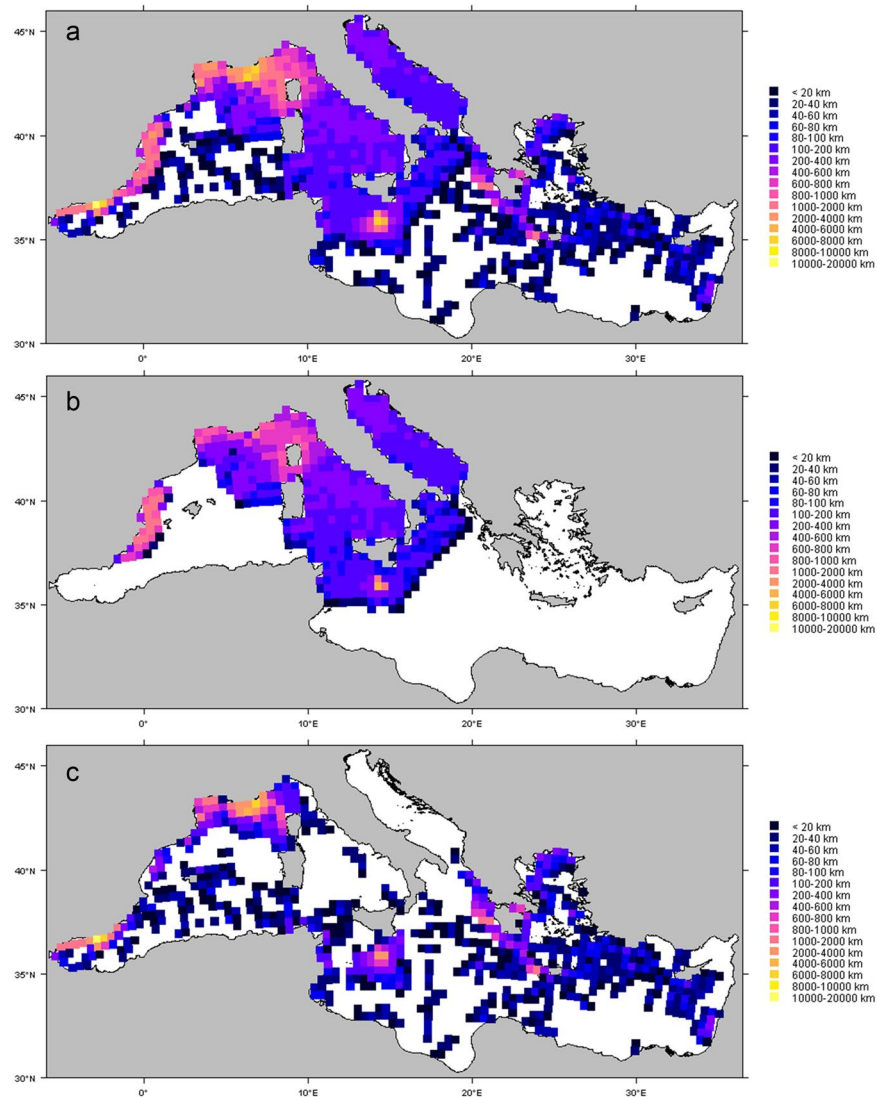
Figure 3. Geographic coverage of effort for: (**a**) all surveys, (**b**) aerial surveys only and (**c**) shipboard surveys only. Effort was aggregated on a 20×20 km grid for visualization (10×10 km cells used for the analysis were too small to be visible on a map of the entire Mediterranean Sea). The colour scale represents effort in km per 20×20 km grid cell and is the same for all three maps. Blank cells represent zero effort. The maps were generated with R (https://www.r-project.org/) (version 3.1.1).

Mediterranean subregion	Area (km²)	Area (%)	Effort (km)	Effort (%)
Alborán Sea/Strait of Gibraltar	62,134	2.5	38,415	13.2
Algero-Provençal basin	515,739	20.5	131,571	45.3
Tyrrhenian Sea/eastern Ligurian Sea	267,808	10.7	39,296	13.5
Adriatic Sea	133,364	5.3	16,204	5.6
Strait of Sicily/Tunisian Plateau/Gulf of Sirte	346,705	13.8	29,879	10.3
Ionian Sea/Central Mediterranean	497,523	19.8	22,482	7.7
Aegean Sea	187,984	7.5	6,214	2.1
Levantine Sea	501,476	20.0	6,127	2.1

Table 2. Overall survey effort per Mediterranean subregion (defined following previous studies[18,56]).

For the model including all three dynamic covariates, extrapolation would occur in 49.9% of the Mediterranean Sea on average (minimum: 8.1% in September; maximum: 55.1% in February) (Table 3). Extrapolation would be widespread in the eastern Mediterranean and in parts of the western Mediterranean (e.g., the Alborán Sea and the Ligurian Sea), especially in winter and spring (Fig. 7a, Supplementary Fig. S11).
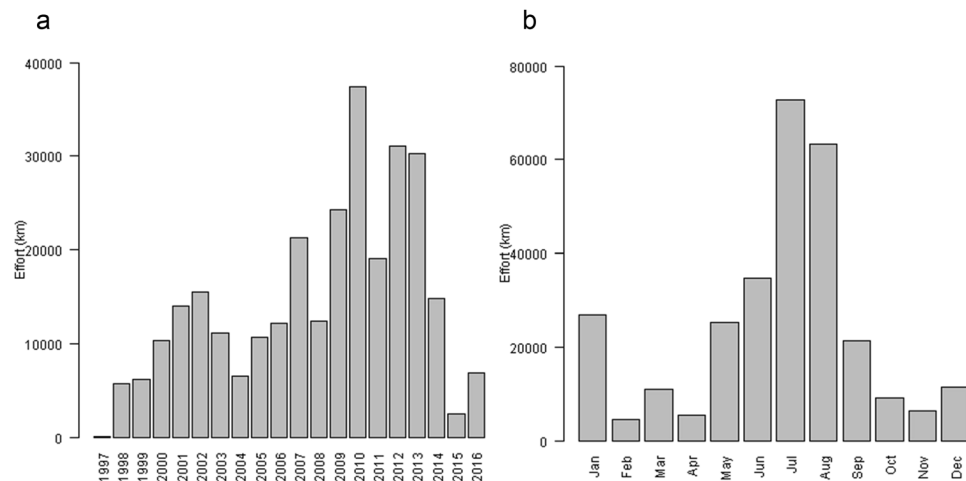
Figure 4. Overall survey effort (**a**) per year and (**b**) per month in the entire Mediterranean Sea for the study period (October 1997-April 2016). Note that years 1997 and 2016 did not include all months.

Univariate extrapolation	
Depth	0.1%
Slope	0.0%
Distance to seamounts	0.0%
Distance to canyons	0.0%
SST	40.8% (0.5–40.8%)
PP	0.6% (0.0–21.0%)
EKE	0.1% (0–17.1%)
Multivariate extrapolation	
All static covariates (depth, slope, distance to seamounts, distance to canyons)	3.7%
All dynamic covariates (SST, PP, EKE)	49.9% (8.1–55.1%)
All static and dynamic covariates (depth, slope, distance to seamounts, distance to canyons, SST, PP, EKE)	80.1% (55.5–96.5%)

Table 3. Spatial extent of extrapolation (i.e., the percentage of cells of the study area where extrapolation occurred) with single covariates and combinations of covariates. For dynamic covariates, the mean extent of extrapolation averaged over the 12 month period is provided, followed by the minimum and maximum monthly extents in parentheses. SST: sea surface temperature; PP: primary productivity; EKE: eddy kinetic energy.

In summer, extrapolation would be mostly limited to the Levantine Sea. The proportions of data nearby in the multivariate space defined by the three dynamic covariates were low in most of the eastern Mediterranean and in the northern and southern parts of the western Mediterranean (Fig. 7b, Supplementary Fig. S12). In autumn, higher proportions of data nearby appeared in a mid-latitude band. Some areas such as the Alborán Sea and the northern Adriatic Sea had low proportions of data nearby year-round.

For the model including all static and dynamic covariates, extrapolation would increase to 80.1% of the Mediterranean Sea on average (minimum: 55.5% in July; maximum: 96.5% in February) (Table 3). Extrapolation would be widespread in all seasons except in the western Mediterranean in summer (Fig. 8a, Supplementary Fig. S13). Overall, the proportions of data nearby in the multivariate space defined by all static and dynamic covariates were lower compared to the proportions of data nearby in the multivariate space defined by dynamic covariates only (as seen by the comparison of colour scales between Figs 8b and 7b). In winter and spring, the proportions of data nearby were low throughout the southeastern Mediterranean and in offshore parts of the western basin (Fig. 8b, Supplementary Fig. S14). In summer, the proportions of data nearby were overall higher in coastal areas, particularly to the north. In autumn, the proportions of data nearby were low in most of the eastern Mediterranean and in the northwestern Mediterranean.

Discussion

We aggregated over 300,000 km of cetacean line transect surveys suitable for estimating abundance that were conducted in the Mediterranean Sea within the past 20 years by numerous organisations. Survey coverage was heterogeneous geographically and seasonally: large data gaps were present in the eastern and southern Mediterranean in all seasons and elsewhere in non-summer months. Survey coverage was also heterogeneous in environmental space. Surveys covered a number of static bathymetric features (e.g., coastal areas, continental slopes, submarine canyons and seamounts), but not the full range of dynamic oceanographic conditions found within the
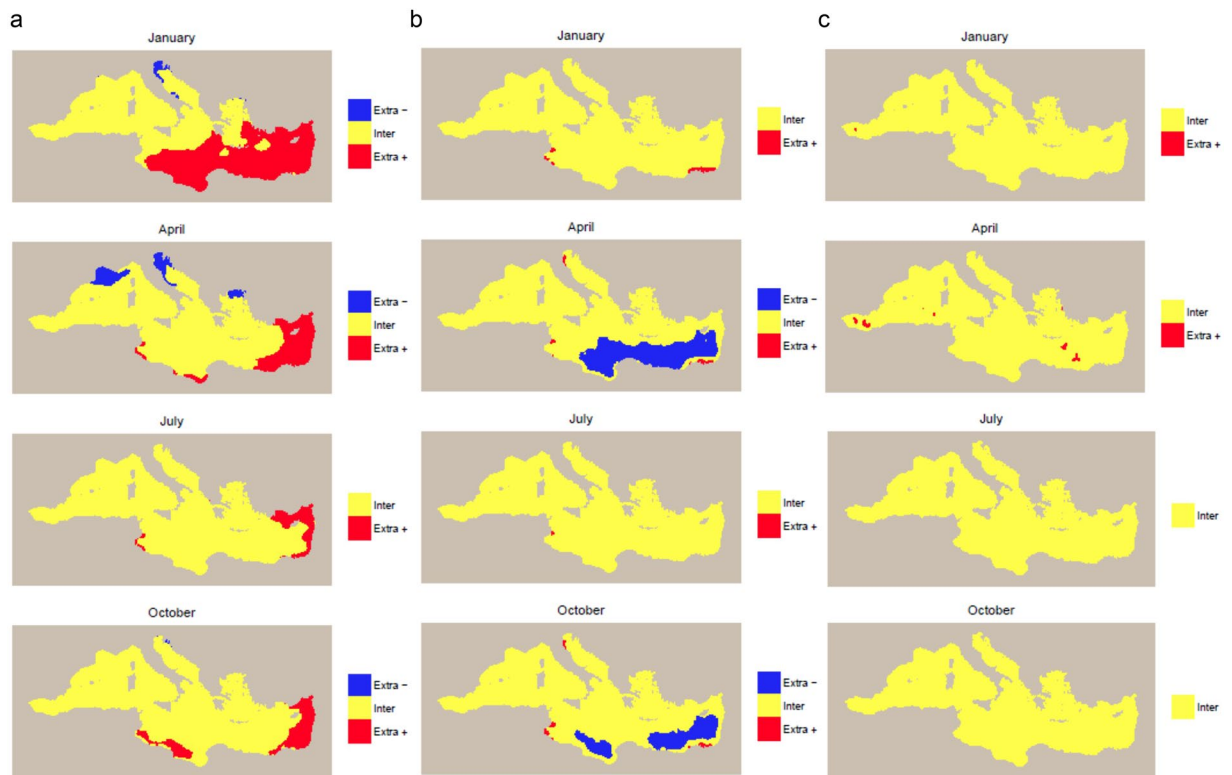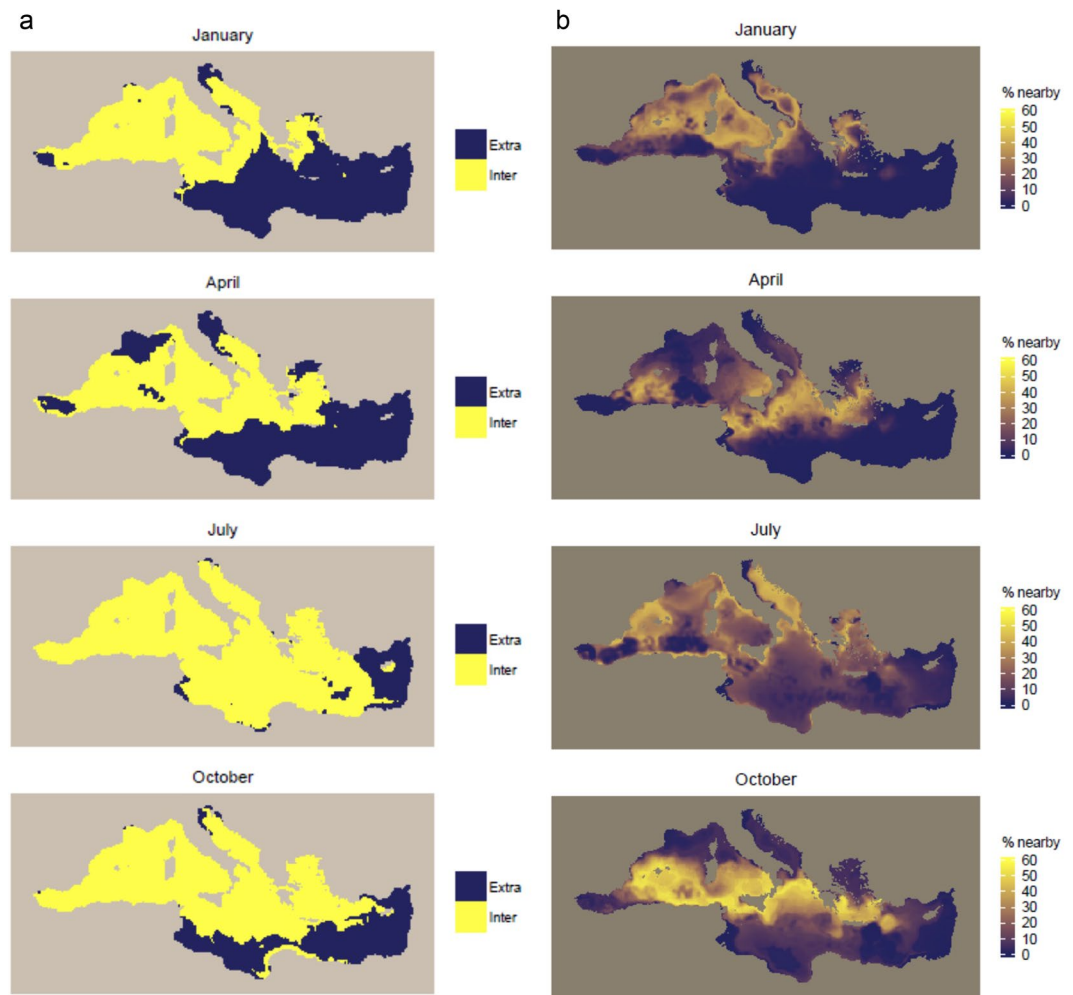
Figure 5. Extent of extrapolation *versus* interpolation if models calibrated on the available survey data were used for prediction across the Mediterranean Sea. (**a**) Model including sea surface temperature only; (**b**) model including primary productivity only; (**c**) model including eddy kinetic energy only. Cells where extrapolation to lower/higher values would occur are indicated in blue/red. Cells where interpolation would occur are indicated in yellow. Results for January, April, July, and October, corresponding to the middle month of solar seasons, are shown. Results for all months are shown in Supplementary Figs S6, S8 and S10. The maps were generated with R (https://www.r-project.org/) (version 3.1.1).
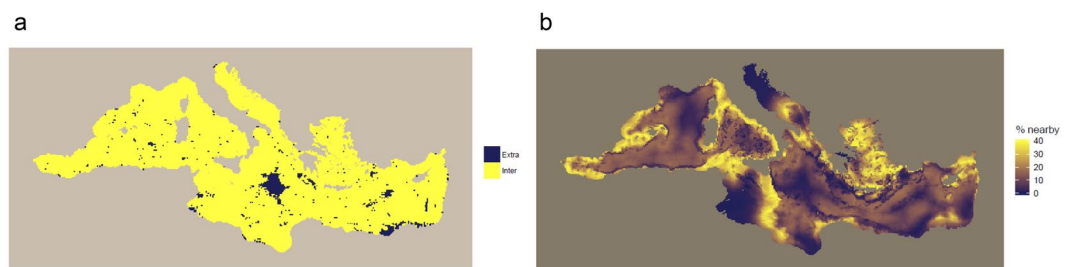
Figure 6. (**a**) Extent of extrapolation (dark blue) *versus* interpolation (yellow), and (**b**) proportion of prediction points near available data points in the multivariate environmental space defined by all considered static covariates if a model including all static covariates calibrated on the available survey data was used for prediction across the Mediterranean Sea. In (**b**), dark blue/yellow represents areas where predictions would potentially be unreliable/reliable. The definition of neighborhood in multivariate environmental space is provided in the Methods. The maps were generated with R (https://www.r-project.org/) (version 3.1.1).

Mediterranean Sea. In particular, Mediterranean waters characterised by comparatively warmer temperatures, lower productivity and higher eddy activity were poorly surveyed for cetaceans. This raised the prospect that cetacean density models fitted to environmental covariates would have to be extrapolated in order to provide predictions for the entire Mediterranean Sea.

To assess extrapolation in environmental space, we mapped two metrics: the extent of extrapolation *versus* interpolation—a binary metric indicating whether a point is within the convex hull enclosing the sampled environmental space—and the proportion of data nearby in multivariate environmental space—a continuous metric derived from Gower's distance. The first metric distinguishes predictions in unsampled geographic space that are within sampled environmental space (interpolation) from predictions in unsampled geographic *and*
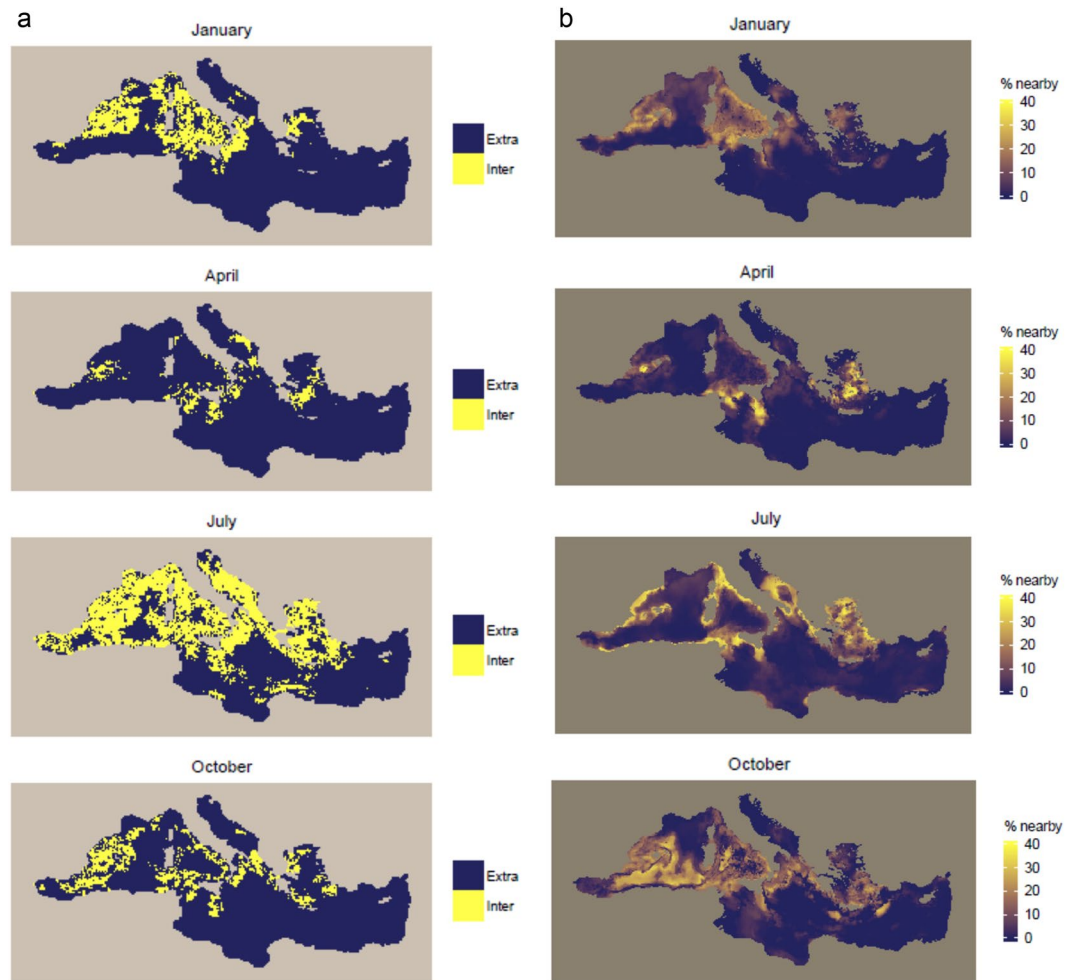
Figure 7. (**a**) Extent of extrapolation (dark blue) *versus* interpolation (yellow), and (**b**) proportion of prediction points near available data points in the multivariate environmental space defined by all considered dynamic covariates if a model including all dynamic covariates calibrated on the available survey data was used for prediction across the Mediterranean Sea. In (**b**), dark blue/yellow represents areas where predictions would potentially be unreliable/reliable. Results for January, April, July, and October, corresponding to the middle month of solar seasons, are shown. Results for all months are shown in Supplementary Figs S11 and S12. The definition of neighborhood in multivariate environmental space is provided in the Methods. The maps were generated with R (https://www.r-project.org/) (version 3.1.1).

environmental space (extrapolation). It can be used to restrict predictions to the sampled environmental space by overlaying a mask on the prediction map or by flagging extrapolations in unsampled environmental space[13,35]; the standard errors of such extrapolations can then potentially be inflated until new data are acquired. The second metric provides a quantitative measure of the reliability of extrapolation in multivariate environmental space. It can be used as a measure of prediction uncertainty when maps of species distributions are needed for an entire region (e.g., to support management decisions). This metric is useful for differentiating predictions that are just outside the sampled environmental space from those that are far outside (and thus less reliable). It also identifies predictions that occur in sparsely sampled regions of environmental space despite being classified as interpolations by falling within sampled environmental space. For example, this situation was observed in the northern Adriatic Sea where low proportions of data nearby would occur although interpolation is predominant (see Fig. 6). Using the binary metric that distinguishes interpolation from extrapolation as the only measure of the reliability of predictions would lead to an overly high degree of confidence in predictions in the northern Adriatic Sea.

Mesgaran *et al.*[15] proposed a modification of the Mahalanobis distance to assess deviations between the prediction and the calibration dataset using the mean and linear correlations between covariates[15]. We did not follow their approach because the covariates we selected *a priori* were not highly correlated (see Methods). Conn *et al.*[36] assessed extrapolation using a model-based convex hull derived from a generalization of Cook's independent variable hull. Because our study objective was to conduct a gap analysis of survey effort in environmental space (rather than to fit models), we adopted a model-independent approach.

Figure 8. (**a**) Extent of extrapolation (dark blue) *versus* interpolation (yellow), and (**b**) proportion of prediction points near available data points in the multivariate environmental space defined by all considered static and dynamic covariates if a model including all static and dynamic covariates calibrated on the available survey data was used for prediction across the Mediterranean Sea. In (**b**), dark blue/yellow represents areas where predictions would potentially be unreliable/reliable. Results for January, April, July, and October, corresponding to the middle month of solar seasons, are shown. Results for all months are shown in Supplementary Figs S13 and S14. The definition of neighborhood in multivariate environmental space is provided in the Methods. The maps were generated with R (https://www.r-project.org/) (version 3.1.1).

We found that the extent of extrapolation was larger for combinations of covariates than for individual covariates. This is a common issue when predicting SDMs beyond the scope of the data: as more covariates are included in a model, new individual covariate values and new combinations of covariates appear, rapidly increasing the extent of extrapolation[13]. For example, in a study of small pelagic fish in the Gulf of Lions, models that included single covariates yielded less than 5% extrapolation while models that included five covariates exceeded 50% extrapolation[37]. For the Mediterranean cetacean surveys, a model that included all seven covariates resulted in 80% extrapolation; this model would be overly complex (in terms of having too many parameters) for reliably predicting cetacean densities at the regional scale. To predict species distributions or densities beyond the scope of the calibration data using SDMs, we recommend choosing models with a small number of covariates to limit the extent of extrapolation in environmental space.

Our analysis showed that the extent of extrapolation would be lower with static covariates than with dynamic covariates. Advantages of static covariates for modelling cetacean densities include generally low measurement errors and easy acquisition and processing leading to prompt model fitting. However, static covariates are rather indirectly related to cetacean densities. For example, steep seafloor slope may aggregate cetaceans via oceanographic processes that enhance prey availability and accessibility, assuming cetaceans are involved in foraging activities. Indirect ecological relationships are known to transfer poorly to new geographic regions because ecological processes underlying cetacean distributions may differ between surveyed and unsurveyed regions[38,39].

By design, a gap analysis does not require the observation data (from which the response variable is derived) because it aims at assessing how representative sampling effort is of the environmental space where predictions are to be made. However, for the gap analysis to be useful, the environmental covariates to be tested should be

the ones that are most likely to be selected when fitting the models. This can be done *a priori* (as in this study) by using sets of covariates that are ecologically-relevant and likely to be selected in future cetacean density models. This could also be done *a posteriori* for a given case study by using the set of covariates that has been actually selected in the model selection process. In this latter case, the response variable would be needed, not for the gap analysis itself, but to help identify the set of most relevant environmental covariates to be tested. When it comes to model development, it is important to keep in mind that (1) sample size of species observations affects model accuracy and precision[40] and (2) the selected model strongly influences predictions, particularly when predicting outside the sampled range of the covariate data[41,42].

Our study showed that if cetacean density models were calibrated on existing line transect surveys and predicted across the Mediterranean Sea, widespread extrapolations in environmental space would occur, leading to potentially unreliable predictions. However, it should be noted that the surveys available for this exercise were limited to those compatible with distance sampling, which is used to estimate absolute abundance. For certain management applications, relative abundance predictions will suffice. There are a number of additional surveys in the Mediterranean Sea that were not aimed at estimating absolute abundance with distance sampling but that followed similar methodology e.g.,[43,44] and that could be included in models of relative abundance. Including such surveys would likely reduce some of the gaps in environmental space.

Conclusion and Perspectives

Systematic survey efforts for cetaceans have previously been recognized as heterogeneous across the Mediterranean Sea[18,45,46], but this heterogeneity was never quantified. By aggregating surveys conducted throughout the Mediterranean Sea, we were able to identify gaps in the spatiotemporal and environmental coverage of cetacean surveys for the first time.

We strongly recommend that future surveys be conducted in the eastern and southern Mediterranean where waters are characterised by warmer sea surface temperature, lower productivity, and higher eddy kinetic energy. These environmental conditions are found nowhere else in surveyed regions; they are unique to the eastern and southern Mediterranean and were under-represented in our dataset. Conducting surveys there would be particularly valuable in non-summer months because oceanographic conditions differ the most from those of other areas surveyed in those months. We also recommend additional survey effort in the northern Mediterranean in non-summer months, for example in the Gulf of Lions in April, which is characterised by colder waters than other areas surveyed in that month.

A Mediterranean-wide survey is scheduled to take place in the summer 2018 thanks to the long-standing initiative of ACCOBAMS (Agreement on the Conservation of Cetaceans in the Black Sea, Mediterranean Sea and Contiguous Atlantic Area). This regional survey will be extremely helpful to fill existing data gaps by collecting data synoptically with a consistent methodology. Unfortunately, this survey will not be sufficient to fill data gaps in non-summer months in which environmental conditions (and by extension cetacean distributions) are widely different. Achieving a representative sample of environmental space is critical for surveys intended for habitat-based density modelling of cetaceans. An analysis of environmental space early in the survey planning process—for example assessing the environmental representativeness of a set of potential survey designs—can help ensure model-based predictions of cetacean densities are derived with limited extrapolation for the study region.

The natural next step of this study is to incorporate cetacean observations obtained from the aggregated surveys to build models and predict cetacean densities across the Mediterranean Sea. The resulting density maps could be related to maps of human activities such as ship traffic and underwater noise, which are particularly harmful to cetaceans in the Mediterranean Sea[30,47], in order to identify unsurveyed areas where populations may be more at risk. Predicted density maps could also be used for delineating areas of potentially high cetacean densities in poorly surveyed regions of the Mediterranean Sea, and thus be utilised as part of the Important Marine Mammal Area (IMMA) process of International Union for Conservation of Nature.

Methods

Study area.   The Mediterranean Sea is a semi-enclosed water body connected to the Atlantic Ocean by the Strait of Gibraltar, to the Black Sea by the Bosphorus, and since 1869 to the Red Sea by the Suez Canal. It is divided into a western and an eastern basin by a central ridge between Sicily and the Tunisian-Libyan coast. The Mediterranean Sea is mainly characterised by narrow continental shelves, steep slopes and extensive abyssal plains. It includes a variety of submarine canyons, mostly located along the continental slopes in the north. It also includes approximately one hundred seamounts (Supplementary Fig. S15), known to affect the distribution of pelagic species, including cetaceans[48–50]. The Mediterranean is an oligotrophic sea characterised by salty and nutrient-poor waters[51].

Circulation in the Mediterranean Sea is mainly driven by water flow through the Strait of Gibraltar, freshwater inputs from the main rivers (Nile, and to a lesser extent, Po, Rhone and Ebro), wind stress, and thermohaline and topographic features[52]. Atlantic Surface Water flows into the Mediterranean Sea through the Strait of Gibraltar and circulates in a cyclonic (counterclockwise) direction (Fig. 1). Water flow along the southern coasts generates short-lived mesoscale anticyclonic eddies (e.g., the eddy field off Algeria). To the north, water flow creates persistent cyclonic gyres (e.g., the Lions gyre) associated with upwelling of nutrient-rich waters that result in enhanced primary productivity[52]. As it moves eastward, surface water evaporates and becomes saltier, warmer and poorer in nutrients, resulting in a gradual decline in phytoplankton biomass and productivity from west to east[16,53]. As it becomes saltier and denser, the Atlantic Surface Water sinks in the Levantine Sea, returning westward as Levantine Intermediate Water before exiting into the Atlantic through the Strait of Gibraltar. During winter, water sinks in the Aegean, Adriatic and Ligurian seas and goes to the very bottom, creating the Mediterranean Deep Water[52].

Phytoplankton biomass and primary production have marked seasonal cycles in the Mediterranean Sea[54,55]. Phytoplankton blooms are primarily initiated in winter and spring by wind stress, causing mixing and nutrient uplift to surface layers. Uplift of nutrients also occurs at cyclonic eddies. While blooms are markedly seasonal and intense in the northwestern basin (e.g., in the Gulf of Lions), they are often sporadic and subject to significant inter-annual variability in the eastern basin. Stratification occurs in summer, resulting in a lower and more homogeneous phytoplankton biomass across the Mediterranean Sea.

The spatial extent for this study corresponded to the entire Mediterranean Sea, excluding enclosed lagoons and estuaries (surface area: 2.5 million km^2). The study area was projected to a custom Lambert Azimuthal Equal Area projection in ArcGIS 10.2.2 and gridded into 10×10 km cells for analysis, representing a compromise between the various resolutions of environmental covariates (see Supplementary Tables S1 and S2). The Mediterranean study area was divided into 8 distinct ecological subregions following previous studies[18,56] (Fig. 2).

Line transect surveys. We aggregated data from visual shipboard and aerial line transect surveys that followed distance sampling methodology, i.e., allowed the calculation of perpendicular distances to observed groups of cetaceans and thus the estimation of absolute abundances[57]. We considered both occasional large-scale surveys and recurrent small-scale surveys aimed at abundance estimation (homogeneous coverage probability was not a requirement for this analysis). We included only line transect surveys conducted after October 1997 to ensure the full set of environmental covariates was available (primary productivity data was available from October 1997; see Supplementary Table S2), and until April 2016.

Spatiotemporal coverage of surveys. For each survey, we created in ArcGIS a shapefile of on-effort tracklines with associated latitude, longitude and date. We merged shapefiles of all survey tracklines and added them to a file geodatabase. We then intersected and spatially joined the tracklines to the 10×10 km grid of the study area. To examine the spatiotemporal coverage of surveys, we summed survey effort per grid cell in the entire study area and created subsets of surveys per month, year, and platform (ship or aircraft).

Environmental covariates. We examined the environmental coverage of surveys with respect to four static covariates (depth, slope, distance to canyons, and distance to seamounts) and three dynamic covariates (sea surface temperature, primary productivity, and eddy kinetic energy) known to be important for cetaceans and likely to be included in future habitat-based density models (Supplementary Tables S1 and S2).

We created rasters of static and dynamic covariates (data sources detailed in Supplementary Tables S1 and S2) and resampled them to the 10×10 km grid of the study area (using nearest neighbour interpolation implemented in the Marine Geospatial Ecology Tools software[58]) in ArcGIS.

We used monthly contemporaneous and climatological resolutions for dynamic covariates[59]. For surveyed grid cells (i.e., representing the "available" dataset; see below), we used the month and year when the survey took place (contemporaneous resolution). For all grid cells of the study area (i.e., representing the "prediction" dataset; see below), we used a monthly climatological average describing long-term oceanographic conditions in the Mediterranean Sea. Doing so allowed us to compare surveyed oceanographic conditions with long-term oceanographic averages (rather than oceanographic conditions of a particular year). To obtain monthly climatological rasters, we averaged monthly rasters per grid cell from 1997–2016. These climatological means smooth out the inter-annual variability of oceanographic processes while maintaining the strong intra-annual variability which characterizes the Mediterranean Sea[52,54,55].

Environmental coverage of surveys. To analyse the environmental coverage of surveys, we examined two multivariate environmental metrics: the extent of extrapolation *versus* interpolation and the proportion of prediction points near available data points.

To evaluate the extent of extrapolation *versus* interpolation if models calibrated on the available survey data were used to predict cetacean densities in the entire Mediterranean Sea, we used the convex hull approach[37,60,61]. The convex hull of a set of points is defined as the smallest convex set that contains these points. Predictions inside the convex hull are interpolations while predictions outside the convex hull are extrapolations. The convex hull of a single covariate is the interval between the minimum and maximum data points (i.e., a univariate environmental envelope). The convex hull of two covariates is a polygon with vertices at the extreme points of the data points (i.e., a bivariate environmental envelope). A convex hull can be defined for any number of covariates, although visualization becomes difficult for more than two covariates and computational power starts to limit calculations for more than ten covariates[61]. As stressed by Authier *et al.* (2017), the assessment of the convex hull does not require model fitting (only effort and environmental covariate data are required)[37].

Here, environmental characteristics of the surveyed grid cells during the month and year of each survey represented the "available" dataset and monthly climatological environmental conditions of all grid cells of the study area represented the "prediction" dataset. We considered the seven individual covariates enumerated above and the three following combinations of covariates:

- all static covariates: depth, slope, distance to seamounts, and distance to canyons;
- all dynamic covariates: sea surface temperature, primary productivity, and eddy kinetic energy;
- all static and dynamic covariates: depth, slope, distance to seamounts, distance to canyons, sea surface temperature, primary productivity, and eddy kinetic energy.

An examination of Spearman's rank correlation coefficients[62] showed that these covariates were not highly correlated (Spearman's rho $< |0.6|$) in both available and prediction datasets.

To further quantify extrapolation in environmental space, we calculated Gower's distances (G^2) for the three combinations of covariates listed above. As for the convex hull, the calculation of Gower's distances does not require model fitting. The Gower's non-parametric distance between points i and j belonging to the available and prediction datasets, respectively, is defined as the average absolute distance between the values of these two points in each dimension, divided by the range of the data[61–63]. With K environmental dimensions, the Gower's distance formula is as follows:

$$G_{ij}^2 = \frac{1}{K}\sum_{k=1}^{K}\frac{|x_{ik} - x_{jk}|}{r_k}$$

where r_k is the difference between the largest and the smallest values of the available dataset for the k_{th} covariate. Because of range standardization, covariates have equal contribution to the Gower's distance. We considered a prediction point to be in the neighbourhood of available data points if it was situated within a radius of one geometric mean Gower's distance of all pairs of available data points[61]. We stress that this definition of neighbourhood relates to the multivariate environmental space defined by the considered environmental covariates, not geographic space. The larger the proportion of prediction points near available data points in this multivariate environmental space ("proportion of data nearby" for brevity), the more reliable the extrapolation.

We used the WhatIf package (version 1.5-6)[64] in R (version R-3.1.1)[65] to calculate convex hulls and Gower's distances. We projected the extent of extrapolation *versus* interpolation and the proportion of data nearby to geographic space for mapping.

Data availability statement. The aggregated survey dataset and R code to reproduce the gap analysis are made available via the Dryad repository (https://doi.org/10.5061/dryad.4pd33).

References

1. Thuiller, W. *et al*. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Glob. Change Biol.* **11**, 2234–2250 (2005).
2. Heikkinen, R. K., Luoto, M., Virkkala, R., Pearson, R. G. & Körber, J.-H. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Glob. Ecol. Biogeogr.* **16**, 754–763 (2007).
3. Tugores, M. P. *et al*. Habitat suitability modelling for sardine *Sardina pilchardus* in a highly diverse ecosystem: the Mediterranean Sea. *Mar. Ecol. Prog. Ser.* **443**, 181–205 (2011).
4. Druon, J.-N. *et al*. Habitat suitability of the Atlantic bluefin tuna by size class: An ecological niche approach. *Prog. Oceanogr*. (2016).
5. Roberts, J. J. *et al*. Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. *Sci. Rep*. **6**, 22615 (2016).
6. Richardson, A. J. & Poloczanska, E. S. Under-resourced, under threat. *Science* **320**, 1294–1295 (2008).
7. Kot, C. Y. *et al*. Spatio-temporal gap analysis of OBIS-SEAMAP project data: assessment and way forward. *PloS One* **5**(9), e12990 (2010).
8. Kaschner, K., Quick, N. J., Jewell, R., Williams, R. & Harris, C. M. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges. *PloS One* **7**, e44075 (2012).
9. Elith, J. & Leathwick, J. R. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).
10. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).
11. Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. Bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Divers. Distrib.* **20**, 1–9 (2014).
12. Mannocci, L., Monestiez, P., Spitz, J. & Ridoux, V. Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. *J. Biogeogr.* **42**, 1267–1280 (2015).
13. Zurell, D., Elith, J. & Schröder, B. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Divers. Distrib.* **18**, 628–634 (2012).
14. Dormann, C. F. *et al*. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
15. Mesgaran, M. B., Cousens, R. D. & Webber, B. L. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Divers. Distrib.* **20**, 1147–1159 (2014).
16. Bethoux, J. P. *et al*. The Mediterranean Sea: a miniature ocean for climatic and environmental studies and a key for the climatic functioning of the North Atlantic. *Prog. Oceanogr.* **44**, 131–146 (1999).
17. Coll, M. *et al*. The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. *PloS One* **5**, e11842 (2010).
18. Notarbartolo di Sciara, G. Chapter One - Marine Mammals in the Mediterranean Sea: An Overview. In *Advances in Marine Biology* Mediterranean Marine Mammal Ecology and Conservation (ed. Giuseppe Notarbartolo di Sciara, M. Podestà & B. E. Curry) **75**, 1–36 (Academic Press, 2016).
19. Lejeusne, C., Chevaldonné, P., Pergent-Martini, C., Boudouresque, C. F. & Pérez, T. Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. *Trends Ecol. Evol.* **25**, 250–260 (2010).
20. Micheli, F. *et al*. Cumulative Human Impacts on Mediterranean and Black Sea Marine Ecosystems: Assessing Current Pressures and Opportunities. *PLoS ONE* **8**, e79889 (2013).
21. Edelist, D., Rilov, G., Golani, D., Carlton, J. T. & Spanier, E. Restructuring the Sea: profound shifts in the world's most invaded marine ecosystem. *Divers. Distrib.* **19**, 69–77 (2013).
22. Bianchi, C. N. & Morri, C. Marine biodiversity of the Mediterranean Sea: situation, problems and prospects for future research. *Mar. Pollut. Bull.* **40**, 367–376 (2000).
23. Lauriano, G., Pierantonio, N., Donovan, G. & Panigada, S. Abundance and distribution of *Tursiops truncatus* in the Western Mediterranean Sea: an assessment towards the Marine Strategy Framework Directive requirements. *Mar. Environ. Res.* **100**, 86–93 (2014).
24. Vella, A. Aerial surveys to study cetaceans: implications for integrated conservation management and sustainable maritime development. in *Proceedings of the 29th European Cetacean Society Conference, March 2015* (Vella, A., Vella, N., & Mifsud, C. 2015).
25. Panigada, S. *et al*. Estimating cetacean density and abundance in the Central and Western Mediterranean Sea through aerial surveys: Implications for management. *Deep Sea Res. Part II* **141**, 41–58 (2017).
26. Laran, S. *et al*. Seasonal distribution and abundance of cetaceans within French waters- Part I: The North-Western Mediterranean, including the Pelagos sanctuary. *Deep Sea Res. Part II* **141**, 20–30 (2017).
27. Aïssi, M. *et al*. Large-scale seasonal distribution of fin whales (*Balaenoptera physalus*) in the central Mediterranean Sea. *J. Mar. Biol. Assoc. U. K.* **88**, 1253 (2008).
28. Frantzis, A., Airoldi, S., Notarbartolo-di-Sciara, G., Johnson, C. & Mazzariol, S. Inter-basin movements of Mediterranean sperm whales provide insight into their population structure and conservation. *Deep Sea Res. Part I* **58**, 454–459 (2011).

29. Panigada, S. *et al*. Satellite tagging of Mediterranean fin whales: working towards the identification of critical habitats and the focussing of mitigation measures. *Sci. Rep.* **7**, 3365 (2017).
30. Panigada, S. *et al*. Mediterranean fin whales at risk from fatal ship strikes. *Mar. Pollut. Bull.* **52**, 1287–1298 (2006).
31. David, L., Alleaume, S. & Guinet, C. High risk areas of collision between fin whales and ferries in the North-western Mediterranean sea. *J. Mar. Anim. Their Environ.* **4**, 17–28 (2011).
32. Castellote, M., Clark, C. W. & Lammers, M. O. Acoustic and behavioural changes by fin whales (*Balaenoptera physalus*) in response to shipping and airgun noise. *Biol. Conserv.* **147**, 115–122 (2012).
33. Cañadas, A. & Vázquez, J. A. Conserving Cuvier's beaked whales in the Alboran Sea (SW Mediterranean): Identification of high density areas to be avoided by intense man-made sound. *Biol. Conserv.* **178**, 155–162 (2014).
34. Notarbartolo di Sciara, G., Agardy, T., Hyrenbach, D., Scovazzi, T. & Van Klaveren, P. The Pelagos sanctuary for Mediterranean marine mammals. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **18**, 367–391 (2008).
35. Fitzpatrick, M. C. & Hargrove, W. W. The projection of species distribution models and the problem of non-analog climate. *Biodivers. Conserv.* **18**, 2255–2261 (2009).
36. Conn, P. B., Johnson, D. S. & Boveng, P. L. On Extrapolating Past the Range of Observed Data When Making Statistical Predictions in Ecology. *PloS ONE* **10**, e0141416 (2015).
37. Authier, M., Saraux, C. & Péron, C. Variable selection and accurate predictions in habitat modelling: a shrinkage approach. *Ecography* **39**, 001–012 (2017).
38. Randin, C. F. *et al*. Are niche-based species distribution models transferable in space? *J. Biogeogr.* **33**, 1689–1703 (2006).
39. Wenger, S. J. & Olden, J. D. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods Ecol. Evol.* **3**, 260–267 (2012).
40. Vaughan, I. P. & Ormerod, S. J. The continuing challenges of testing species distribution models. *J. Appl. Ecol.* **42**, 720–730 (2005).
41. Pearson, R. G. *et al*. Model-based uncertainty in species range prediction. *J. Biogeogr.* **33**, 1704–1711 (2006).
42. Mainali, K. P. *et al*. Projecting future expansion of invasive species: Comparing and improving methodologies for species distribution modeling. *Glob. Change Biol.* **21**, 4464–4480 (2015).
43. Genov, T., Kotnjek, P., Lesjak, J., Hace, A. & Fortuna, C. M. Bottlenose dolphins (*Tursiops truncatus*) in Slovenian and adjacent waters (northern Adriatic Sea). *Annales, Series Historia Naturalis* **18**, 227–244 (2008).
44. Pirotta, E., Matthiopoulos, J., MacKenzie, M., Scott-Hayward, L. & Rendell, L. Modelling sperm whale habitat preference: a novel approach combining transect and follow data. *Mar. Ecol. Prog. Ser.* **436**, 257–272 (2011).
45. Boisseau, O. *et al*. Encounter rates of cetaceans in the Mediterranean Sea and contiguous Atlantic area. *J. Mar. Biol. Assoc. U. K.* **90**, 1589–1599 (2010).
46. Kerem, D. *et al*. Update on the cetacean fauna of the Mediterranean Levantine basin. *Open Mar. Biol. J.* **6**, 6–27 (2012).
47. Rendell, L. E. & Gordon, J. C. D. Vocal response of long-finned pilot whales (*Globicephala melas*) to military sonar in the Ligurian sea. *Mar. Mammal Sci.* **15**, 198–204 (1999).
48. Mussi, B., Miragliuolo, A., Zucchini, A. & Pace, D. S. Occurrence and spatio-temporal distribution of sperm whale (*Physeter macrocephalus*) in the submarine canyon of Cuma (Tyrrhenian Sea, Italy). *Aquat. Conserv. Mar. Freshw. Ecosyst.* **24**, 59–70 (2014).
49. Tepsich, P., Rosso, M. & Halpin, P. N. & Moulins, A. Habitat preferences of two deep-diving cetacean species in the northern Ligurian Sea. *Mar. Ecol. Prog. Ser.* **508**, 247–260 (2014).
50. Fiori, C., Giancardo, L., Aïssi, M., Alessi, J. & Vassallo, P. Geostatistical modelling of spatial distribution of sperm whales in the Pelagos Sanctuary based on sparse count data and heterogeneous observations. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **24**, 41–49 (2014).
51. Longhurst, A. R. *Ecological geography of the sea*. (Academic Press, 2007).
52. Pinardi, N. & Masetti, E. Variability of the large scale general circulation of the Mediterranean Sea from observations and modelling: a review. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **158**, 153–173 (2000).
53. Bethoux, J. P. & Gentili, B. Functioning of the Mediterranean Sea: past and present changes related to freshwater input and climate changes. *J. Mar. Syst.* **20**, 33–47 (1999).
54. Bosc, E., Bricaud, A. & Antoine, D. Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations. *Glob. Biogeochem. Cycles* **18**, (2004).
55. D'Ortenzio, F. & Ribera d'Alcalà, M. On the trophic regimes of the Mediterranean Sea: a satellite analysis. *Biogeosciences* **6**, 139–148 (2009).
56. UNEP-MAP-RAC/SPA. Overview of scientific findings and criteria relevant to identifying SPAMIs in the Mediterranean open seas, including the deep sea. By Notarbartolo di Sciara, G. & Agardy, T. Edited by RAC/SPA, Tunis, Available at: http://bit.ly/1W8jste. (Accessed: 24th November 2015) (2010).
57. Buckland, S. T. *et al*. *Introduction to distance sampling: estimating abundance of biological populations*. (Oxford University Press, 2001).
58. Roberts, J. J., Best, B. D., Dunn, D. C., Treml, E. A. & Halpin, P. N. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environ. Model. Softw.* **25**, 1197–1207 (2010).
59. Mannocci, L. *et al*. Temporal resolutions in species distribution models of highly mobile marine animals: Recommendations for ecologists and managers. *Divers. Distrib.* **23**, 1098–1109 (2017).
60. Carpenter, G., Gillison, A. N. & Winter, J. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* **2**, 667–680 (1993).
61. King, G. & Zeng, L. When Can History Be Our Guide? The Pitfalls of Counterfactual Inference. *Int. Stud. Q.* **51**, 183–210 (2007).
62. Hollander, M. & Wolfe, D. *Non parametric statistical methods* (John Wiley & Sons, 1973).
63. Gower, J. C. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* **27**, 857–871 (1971).
64. Stoll, H., King, G. & Zeng, L. *WhatIf: Software for Evaluating Counterfactuals. R package version 1.5-6*, http://CRAN.R-project.org/package=WhatIf. (2014).
65. R Core Team. *R: A language and environment for statistical computing*. (Foundation for Statistical Computing, 2014).
66. Cañadas, A., Sagarminaga, R., De Stephanis, R., Urquiola, E. & Hammond, P. S. Habitat preference modelling as a conservation tool: proposals for marine protected areas for cetaceans in southern Spanish waters. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **15**, 495–521 (2005).
67. Cañadas, A. & Hammond, P. S. Model-based abundance estimates for bottlenose dolphins off southern Spain: implications for conservation and management. *J. Cetacean Res. Manag.* **8**, 13–27 (2006).
68. Fortuna, C. M., Holcer, D., Filidei, E., Dovovan, G. & Tunesi, L. First cetacean aerial survey in the Adriatic Sea: summer 2010. In *Seventh meeting of the ACCOBAMS Scientific Comittee* Document ACCOBAMS-SC7/2011/Doc06. Available at: https://bib.irb.hr/datoteka/560768.sc7_doc06.pdf (Accessed: 25th March 2016) (2011).
69. Lewis, T. *et al*. Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003. *J. Mar. Biol. Assoc. U. K.* **87**, 353–357 (2007).
70. Ryan, C. *et al*. A Visual and Acoustic Survey for Marine Mammals in the Eastern Mediterranean Sea during Summer 2013. Available at: http://www.marineconservationresearch.co.uk/ (Accessed: 24th November 2015) (2014).
71. Ben Naceur, L. *et al*. Recensement du grand dauphin *Tursiops truncatus* dans les eaux tunisiennes *Bull. Inst. Natn. Scien. Tech. Mer de Salammbô* **31**, 75–81 (2004).
72. Frantzis, A., Alexiadou, P. & Gkikopoulou, K. C. Sperm whale occurrence, site fidelity and population structure along the Hellenic Trench (Greece, Mediterranean Sea). *Aquat. Conserv. Mar. Freshw. Ecosyst.* **24**, 83–102 (2014).

73. Chicote, C., Parga, M. L., Castellote, M., Monna, A. & Gazo, M. Actuaciones complementarias al Life+ INDEMARES de inventario de tortugas marinas y cetáceos en las tres áreas marinas propuestas para la Red Natura 2000 en el mediterráneo Nor-occidental. Informe técnico-Fundación Biodiversidad. Memoria Technica. Available at: http://www.submon.org/ (Accessed: 15th April 2016) (2010).

74. Chicote, C., Cañadas, A. & Gazo, M. Actuaciones complementarias al Life INDEMARES de estudio, estima y directrices de conservación de las poblaciones de cetáceos en áreas marinas propuestas para Red Natura 2000 en el Mediterráneo nor-occidental. Informe técnico-Fundación Biodiversidad. Memoria Technica. Available at: http://www.submon.org/ (Accessed: 15th April 2016) (2012).

75. Chicote, C. *et al.* Proyecto Grampus: Estudio y seguimiento de la población de calderón gris asociada a cañones submarinos en la costa catalana. Informe técnico-Fundación Biodiversidad. Memoria Technica. Available at: http://www.submon.org/ (Accessed: 15th April 2016) (2015).

76. Lauriano, G., Panigada, S., Canneri, R., Zeichen, M. M. & Notarbartolo di Sciara, G. Abundance estimate of striped dolphins (*Stenella coeruleoalba*) in the Pelagos Sanctuary (NW Mediterranean Sea) by means of line transect survey. *J. Cetacean Res. Manag.* **11**, 279–283 (2010).

77. Lauriano, G., Panigada, S., Casale, P., Pierantonio, N. & Donovan, G. Aerial survey abundance estimates of the loggerhead sea turtle *Caretta caretta* in the Pelagos Sanctuary, northwestern Mediterranean Sea. *Mar. Ecol. Prog. Ser.* **437**, 291–302 (2011).

78. Panigada, S., Lauriano, G., Burt, L., Pierantonio, N. & Donovan, G. Monitoring Winter and Summer Abundance of Cetaceans in the Pelagos Sanctuary (Northwestern Mediterranean Sea) Through Aerial Surveys. *PLoS ONE* **6**, e22878 (2011).

79. Vella, A. Common dolphins (*delphinus delphis*) status in the central and southern mediterranean around the Maltese Islands. In *Proceedings of the workshop on common dolphins: current research, threats and issues* (Stockin, K., Vella, A. & Evans, P. G. H. 2005).

80. Vella, J. 2000–2014 cetacean sighting records in the central Mediterranean region from sea-users: An evaluation of long-term trends. in *Proceedings of the 29th European Cetacean Society Conference, March 2015* (Vella, A., Vella, N. & Mifsud, C. 2015).

81. Gomez de Segura, A., Crespo, E. A., Pedraza, S. N., Hammond, P. S. & Raga, J. A. Abundance of small cetaceans in waters of the central Spanish Mediterranean. *Mar. Biol.* **150**, 149–160 (2006).

82. Gomez de Segura, A., Hammond, P. S., Cañadas, A. & Raga, J. A. Comparing cetacean abundance estimates derived from spatial models and design-based line transect methods. *Mar. Ecol. Prog. Ser.* **329**, 289–299 (2007).

Acknowledgements

Author Contributions

L.M., J.J.R. and P.N.H. conceived the study. O.B., M.N.B., A.C., C.C., L.D., N.D.M., C.M.F., A.F., M.G., D.H., D.K., G.L., T.L., G.N.S., S.P, J.A.R., A.S., V.R., A.V., J.V. conducted the cetacean surveys. L.M., J.J.R and M.A. performed the analyses. L.M. wrote the manuscript and prepared the figures. T.G., P.S.H. and K.K. contributed to the manuscript. All authors reviewed and contributed to the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-19842-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Terms and Conditions