Ocean-scale prediction of whale shark distribution

Ana Sequeira[1]*, Camille Mellin[1,2], David Rowat[3], Mark G. Meekan[4] and Corey J. A. Bradshaw[1,5]

[1]*The Environment Institute and School of Earth and Environmental Sciences, University of Adelaide, Adelaide, SA 5005, Australia,* [2]*Australian Institute of Marine Science, PMB No. 3, Townsville MC, Townsville, Qld 4810, Australia,* [3]*Marine Conservation Society, Seychelles, PO Box 1299, Victoria, Mahe, Seychelles,* [4]*Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling Hwy, Crawley, WA 6009, Australia,* [5]*South Australian Research and Development Institute, PO Box 120, Henley Beach, SA 5022, Australia*

ABSTRACT

Aim Predicting distribution patterns of whale sharks (*Rhincodon typus*, Smith 1828) in the open ocean remains elusive owing to few pelagic records. We developed multivariate distribution models of seasonally variant whale shark distributions derived from tuna purse-seine fishery data. We tested the hypotheses that whale sharks use a narrow temperature range, are more abundant in productive waters and select sites closer to continents than the open ocean.

Location Indian Ocean.

Methods We compared a 17-year time series of observations of whale sharks associated with tuna purse-seine sets with chlorophyll *a* concentration and sea surface temperature data extracted from satellite images. Different sets of pseudo-absences based on random distributions, distance to shark locations and tuna catch were generated to account for spatiotemporal variation in sampling effort and probability of detection. We applied generalized linear, spatial mixed-effects and Maximum Entropy models to predict seasonal variation in habitat suitability and produced maps of distribution.

Results The saturated generalized linear models including bathymetric slope, depth, distance to shore, the quadratic of mean sea surface temperature, sea surface temperature variance and chlorophyll *a* had the highest relative statistical support, with the highest percent deviance explained when using random pseudo-absences with fixed effect-only models and the tuna pseudo-absences with mixed-effects models (e.g. 58% and 26% in autumn, respectively). Maximum Entropy results suggested that whale sharks responded mainly to variation in depth, chlorophyll *a* and temperature in all seasons. Bathymetric slope had only a minor influence on the presence.

Main conclusions Whale shark habitat suitability in the Indian Ocean is mainly correlated with spatial variation in sea surface temperature. The relative influence of this predictor provides a basis for predicting habitat suitability in the open ocean, possibly giving insights into the migratory behaviour of the world's largest fish. Our results also provide a baseline for temperature-dependent predictions of distributional changes in the future.

Keywords

Colour imagery, Indian Ocean, *Rhincodon typus*, satellite data, sea surface temperature, species distribution models, tuna purse-seine fisheries

*Correspondence: Ana Sequeira, The Environment Institute and School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia. E-mail: ana.martinssequeira@adelaide.edu.au

INTRODUCTION

Species distribution modelling has been applied widely in conservation ecology (Guisan & Zimmermann, 2000; Hirzel *et al.*, 2002; Guisan & Thuiller, 2005; Phillips *et al.*, 2009)

to understand the role of environmental conditions driving distribution and abundance patterns. Predictions arising from this approach are essential to determine the likely effects of habitat change on persistence probability or community structure (Araújo & Williams, 2000; Beaumont *et al.*, 2007).

Such models incorporate information on environmental conditions and combine these with the known distribution of a species or population to define its ecological niche (Hutchison, 1957; Kearney & Porter, 2009) and predict its probability of occurrence in a location where no biological information is currently available (Robertson et al., 2003).

The use of these models in terrestrial systems is commonplace (Lehmann et al., 2002); however, their application in non-terrestrial systems is rare with only a few marine examples such as seagrass beds (Kelly et al., 2001), bryophytes (Sérgio et al., 2007), coral communities (Garza-Pérez et al., 2004; Beger & Possingham, 2008; Tittensor et al., 2009; Mellin et al., 2010), shellfish in coastal ecosystems (Sequeira et al., 2008) and dolphins/whales in the Mediterranean Sea (Praca & Gannier, 2007) and in the eastern tropical Pacific Ocean (Oviedo & Solís, 2008). To our knowledge, they have not been applied to the distribution of widely dispersed pelagic fishes in open oceans. Data collection in marine environments is often 'very difficult, resource-intensive and expensive' (Richardson & Poloczanska, 2008), and the logistics of this task is much greater in the pelagic realm where work is typically based aboard large vessels. There is limited opportunity for synoptic sampling because few research programmes can afford multiple ships operating simultaneously, and data sets of distribution of fishes typically contain only reports of presence; therefore, confirming true absences can be difficult (Zaniewski et al., 2002).

Although some statistical techniques can cope with presence-only data – e.g. ENFA (Hirzel et al., 2002), BIOCLIM (Thuiller et al., 2009), Maximum Entropy (MaxEnt) (Phillips & Dudík, 2008) – and produce acceptable results (Zaniewski et al., 2002; Elith et al., 2006), regression methods such as boosted regression trees (De'ath, 2007; Elith et al., 2008), multivariate adaptive regression splines (Leathwick et al., 2005) or generalized linear or additive models (Nelder & Wedderburn, 1972; Hastie & Tibshirani, 1986) tend to provide more realistic projections when using reliable and accurate absence data (Zaniewski et al., 2002; Brotons et al., 2004). Where such data are unavailable to use in regression models, an alternative approach is to generate pseudo-absences that should, ideally, also account for any spatial bias in the sampling effort (Phillips et al., 2009).

The whale shark (Rhincodon typus, Smith 1828) is the largest of all fish and can reach more than 12 m in total length (Stevens, 2007). Although little is known about its habitat selection or migration patterns, the whale shark appears to be a highly mobile species. It is distributed in oceanic and coastal tropical waters (Last & Stevens, 1994; Compagno, 2001) from 30° N to 35° S, spending most of its time (often > 80%) at the surface (waters < 20 m deep), with regular excursions to depths of 300–500 m during the day when transiting the open ocean (Wilson et al., 2006; Brunnschweiler et al., 2009). Whale sharks are known to aggregate nearshore in a number of coastal locations (e.g. see Rowat, 2007). This behaviour makes them the subject of highly lucrative ecotourism industries around the world (Martin, 2007; Rowat & Engelhardt, 2007), although in the past, such aggregations have also been targeted

by commercial fishing (Joung et al., 1996; Pravin, 2000; Stevens, 2007; White & Cavanagh, 2007). While large-scale fisheries have closed, artisanal and small-scale fisheries for the species still exist in many parts of the tropics (Stewart & Wilson, 2005; White & Cavanagh, 2007; Riley et al., 2009). As the whale shark is considered a vulnerable species (IUCN, 2010), Bradshaw et al. (2008) stressed the importance of understanding their migratory behaviour because they can travel from regions where they are protected to those where they are still under threat.

At present, there is little information on the habitat requirements and pelagic distribution of any filter-feeding shark (Sims et al., 2003; Southall et al., 2006), although movements of whale sharks have been associated with climate and surface water conditions (Cárdenas-Palomo et al., 2010; Sleeman et al., 2010a), plankton blooms (Colman, 1997; Kumari & Raman, 2010) and other potential food sources (Wilson, 2002; Meekan et al., 2009). These fine-scale studies (100s m–100s km) and other hypotheses have never been formally tested at the ocean-basin scale (1000s km) (Southall et al., 2006; Cárdenas-Torres et al., 2007; Rowat, 2007).

Access to a 17-year data set of whale shark sightings generated by the tuna fishing industry in the Indian Ocean offered the opportunity to examine distribution patterns of this species in the open ocean. By combining these data with environmental time series measured via remote sensing, we can potentially infer seasonal patterns in habitat suitability for whale sharks at spatial scales likely to encompass the range of entire populations. Our main objectives were to: (1) determine the spatiotemporal patterns in habitat suitability for whale sharks over a broad (ocean basin) spatial scale and (2) identify the most important determinants of whale shark distribution; we tested the hypotheses that specific temperature ranges, productivity values and distance from shore functions drive seasonal spatial distribution of whale sharks.

METHODS

Data set collected by tuna fisheries (whale shark presences)

Whale sharks, mainly associated with net sets for tuna catch (hereafter referred to as whale shark sightings), were recorded by purse-seine fishers registered at the Indian Ocean Tuna Commission (IOTC; Pianet et al., 2009), and these data were made available by vessel owners through the Institut de Recherche pour le Développement (IRD; France). The data set consisted of a 17-year (1991–2007) time series of (1) whale shark sightings, date and location (longitude and latitude at 0.01° precision) for a total of 1185 whale sharks and (2) tuna catch (tonnes) and fishing effort (days) pooled for each 5° grid cell, per month and per year (6618 records total).

The data set covered 172,800 km^2 of the Indian Ocean between 30° N–30° S and 35–100° E (Fig. 1a). When all years were combined, a total of 811 whale sharks were recorded in autumn (April–June; 68% of records), 68 in winter (July–September;
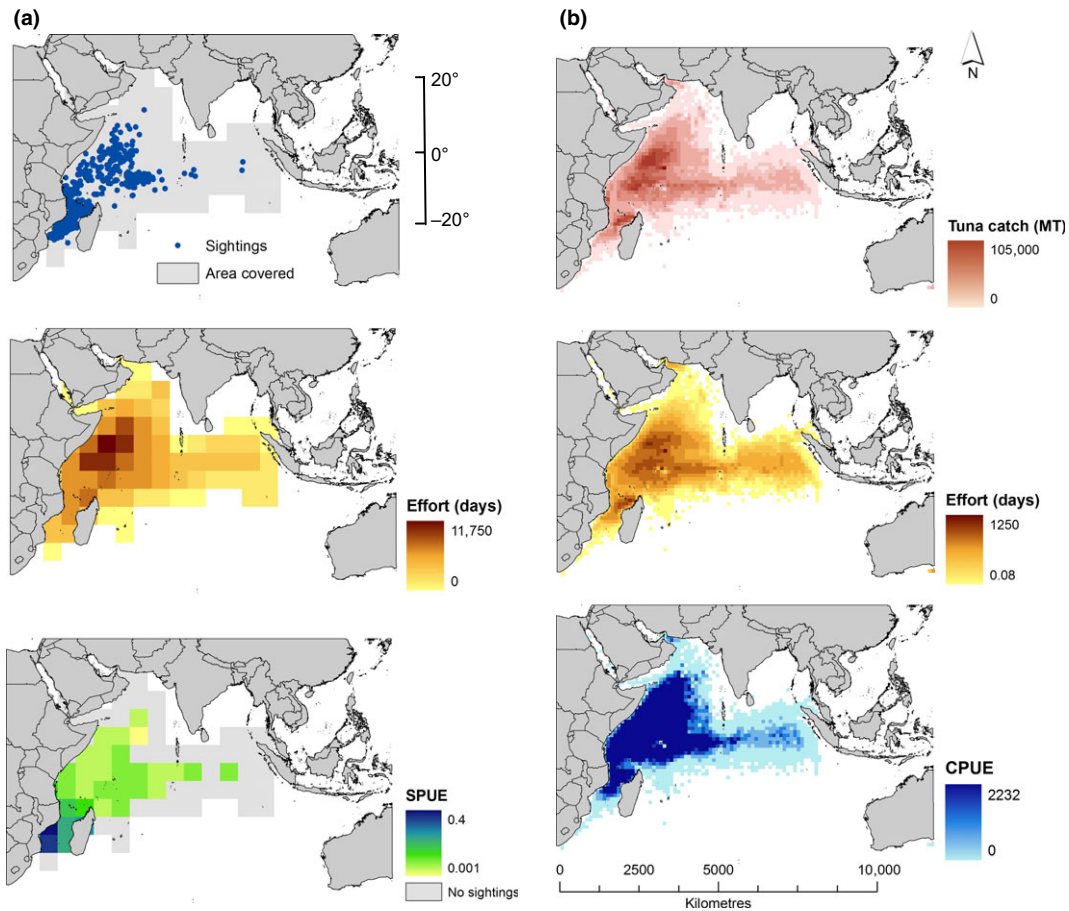
Figure 1 Indian Ocean Tuna Commission (IOTC) data sets collected from 1991 to 2007 on tuna purse-seine fisheries in the Indian Ocean. (a) Tuna-associated fisheries per 5° grid cell including whale shark (*Rhincodon typus*, Smith 1828) associated net sets (hereafter referred to as sightings): top panel – area covered by the fisheries (grey) and 1185 whale shark sightings (blue dots); middle panel – associated effort in days spent fishing within the area sampled; bottom panel – whale shark sightings per unit effort (SPUE) of associated fisheries (in sightings per fishing day). Grey areas represent sampled area where no whale sharks were sighted. (b) Tuna total fisheries per 1° grid cell: top panel – total tuna catch; middle panel – total fishing effort in days; bottom panel – total tuna catch per unit effort (tons per fishing day).

6%), 191 in spring (October–December; 16%) and 115 in summer (January–March; 10%) – austral seasons. We divided the number of sightings by the fishing effort of associated catches to calculate a sightings-per-unit-effort index (SPUE – with unit effort being fishing days), for each month across the study area (Fig. 1a). We applied a Kruskal–Wallis analysis to the whale shark locations per season to test whether the distributions of whale shark SPUE varied among seasons.

We also obtained a second data set of tuna catches in the Indian Ocean derived from the same raw data as the whale shark-associated sets, covering the same spatial extent but aggregated at a finer scale (1° resolution) (Pianet *et al.*, 2009). This data set consisted of: total tuna catch (tonnes) and fishing effort (days) pooled for each 1° grid cell per month (Fig. 1b). As for the 5° resolution data set, we estimated the normalized catch-per-unit-effort (CPUE) by dividing the total tuna catch by the fishing effort in each grid cell for each month across the study area.

To reduce bias associated with temporal autocorrelation and test the hypothesis that specific environmental parameters (e.g. SST, or Chl *a* – used here as a proxy for food availability)

affect whale shark distributions in a specific season, we split the data into four groups each of them corresponding to quarters of the year and defined here as summer (January–March), autumn (April–June), winter (July–September) and spring (October–December).

Pseudo-absence data set generation

Whale shark data set was presence-only, as true absences were unknown. We therefore generated pseudo-absences for use in generalized linear models (GLM) and spatial generalized linear mixed-effects models (GLMM). The pseudo-absence locations were also included as background data in additional MaxEnt runs, to obtain more comparable results between the two modelling procedures.

To assess the influence of the method used to generate pseudo-absences on model outputs, we generated the same number of pseudo-absences as presence records for each season following three different methods: (1) random selection of non-presence grid cells within the 9-km grid over the entire area

accounting for uneven sampling bias by weighting pseudo-absences by the fishing effort in the same area (random), (2) selection of non-presence grid cells with a probability that was weighted by the inverse distance to the whale shark presences (IDW), assuming that detection probability was higher near recorded presences and to account for the same spatial bias and (3) selection weighted by the total tuna catch from the 1° data set (tuna), assuming that whale sharks and tuna would have similar distribution patterns owing to their association – it is a common practice among tuna fishers to use whale sharks as indicator species when targeting tuna (Matsunaga *et al.*, 2003), making whale shark absences more likely to occur in areas of high tuna catch (and high fishing effort), but where no whale shark sightings were recorded.

We generated each set of pseudo-absences 100 times using the *srswor* function (simple random sampling without replacement) from the *Sampling* package in R (R Development Core Team, 2010) prior to their use in GLM and GLMM.

Environmental variables

We collated a environmental data set at 9-km resolution over the study area that included bathymetry (including mean depth in meter and slope in degree), distance to shore (km), distance to shelf (km), seasonal mean and standard deviation of chlorophyll *a* concentration (Chl *a* in mg m^{-3}) and sea surface temperature (SST in °C). We initially collated bathymetry data (see Fig. S1 in Supporting Information) across the study area at approximately 1.7-km resolution using the one-min grid for the General Bathymetry Chart of the Oceans (GEBCO, 2003). Mean depth and slope were then calculated for each 9-km grid cell. For each whale shark sighting, we calculated the shortest distances to the coast and to the continental shelf in ArcGIS 9.2 (ESRI Inc., 2008; Redlands, California, USA) with the *Near* tool using a World Equidistant Cylindrical coordinate system.

We obtained Chl *a* and SST data at a 9-km resolution from relayed image composites from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS - http://oceancolor.gsfc.nasa.gov/Sea WiFS/) and Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua - http://oceancolor.gsfc.nasa.gov/DOCS/MOD ISA_processing.html), respectively (see Figs S2 and S3 in Supporting Information). We averaged weekly daytime measures available since September 1997 for SeaWiFS and July 2002 for MODIS-Aqua. For each season, we calculated the resulting Chl *a* and SST mean and standard deviation in R (R Development Core Team, 2010). Owing to non-Gaussian data, we investigated monotonic relationships among environmental predictors using Spearman's rank correlation coefficient (ρ).

Models

Generalized linear and mixed-effects models

We set GLM with a binomial error distribution and a logit link function to compare, for each season, the predictive ability of all possible combinations of environmental predictors. We included a quadratic term for SST to account for the possibility of a selected temperature range given that they are poikilotherms and external temperatures likely affect their metabolic processes (Bullock, 1955).We compared models based on two bias-corrected indices of parsimony (Burnham & Anderson, 2004): the information-theoretic Akaike's information criterion corrected for small sample sizes (AIC$_c$) and the Bayesian information criterion (BIC). AIC$_c$ favours more complex models (with higher predictive capacity) when tapering effects exist and sample sizes are large, whereas BIC tends to identify the main drivers of complex relationships (Link & Barker, 2006). We assessed each model's strength of evidence relative to the entire model set by calculating relative model weights (*w*AIC$_c$ and *w*BIC). We used the percentage of deviance explained (DE) to quantify each model's goodness-of-fit.

We also applied a 10-fold cross-validation using 1000 iterations to assess the mean prediction error of the model that maximized DE for 10% of the observations that were randomly selected and left out of the training data set. We assessed the predictive power of the models according to Cohen's Kappa statistics κ that measures agreement/accuracy (Cohen, 1960) and varies from ≤ 0 for no agreement between observed data and hypothetical probability to 1 when perfect agreement occurs. Following Woodby *et al.* (2009), we considered that a model had a poor performance when κ < 0.4, good when 0.4 < κ < 0.75 and excellent when κ > 0.75; these ranges were based on the (κ) agreement scale proposed by Landis & Koch (1977). For each index of model performance, we calculated the median across the results obtained for each of the 100 replicates of each pseudo-absence data set.

We assessed potential spatial autocorrelation in both observations and GLM residuals as a function of distance between sites based on Moran's I (Diggle & Ribeiro, 2007) after a Bonferroni correction (Legendre & Legendre, 1998). We used the spatial correlation structure that gave the best fit to the null model to define the error covariance matrix in a spatial GLMM, coding the 1° grid cell as a random effect within each GLM fitted previously (e.g. Mellin *et al.*, 2010). We fit GLMM using the penalized quasi-likelihood (Venables & Ripley, 2002) and derived predictions for the entire area sampled in each season. We then built prediction maps in ArcGIS 9.2.

Following Araújo & New (2006), we used an ensemble approach to combine the full range of results obtained by the different techniques used to account for pseudo-absences. We weighted the contribution of each model according to its percentage of deviance explained to build an ensemble prediction for seasonal distribution of whale sharks in the Indian Ocean.

Maximum Entropy

We compared GLM results to predictions obtained using a presence-only modelling technique, MaxEnt (Phillips *et al.*, 2004, 2006, 2009; Phillips & Dudík, 2008), using the software

provided by the authors (MaxEnt version 3.3.3e November 2010, AT&T Labs Research, Princeton, NJ, USA).

Maximum Entropy is a tool for generating species distribution models from presence-only data. This modelling tool uses covariate data from species presence locations and background sampling to estimate habitat suitability for the species occurrence (for a detailed statistical explanation of MaxEnt, see Elith *et al.*, 2011).

To make MaxEnt models more comparable to the GLM/GLMM, we used the same data sets for presences and pseudo-absences to generate similar models per season with restricted settings (i.e. we fitted only linear and quadratic features) in MaxEnt (for details on MaxEnt features options see Elith *et al.*, 2011). Owing to the functionality of MaxEnt to use presence-only data, we ran additional models making use of the entire background where environmental data were available and allowing the model to use all the features available in the console: *auto features* (linear, quadratic, product, threshold and hinge).

We projected results to the entire area sampled within each season, using the area under the curve (AUC) to measure model performance. For models using the same presence and pseudo-absence data sets, we calculated the Kappa statistic to compare the results directly to the GLM performance. We selected the jack-knife test option in all model sets to infer the relative importance of each variable.

RESULTS

Owing to the uneven fishing effort in each 5° cell within the area covered by the tuna purse-seine fisheries (e.g. ≤ 100 days

in the eastern side and ≥ 11,000 days in the western side of the Indian Ocean; Fig. 1a), the aggregated normalized whale shark data were used instead of the whale shark sightings alone. SPUE varied from 2.10×10^{-3} to a maximum of 444×10^{-3} in each grid cell during the season when most whale sharks were sighted (autumn), followed by SPUE between 2.45×10^{-3} and 200×10^{-3} in summer, 0.86×10^{-3}– 160×10^{-3} in winter and 0.46×10^{-3}– 37.9×10^{-3} in spring.

Sighting (SPUE) patterns of whale sharks shifted between seasons (Fig. 2); there was a non-random (Kruskal–Wallis $H_3 = 33.702$; $P < 0.001$) clockwise shift in relative occurrence from the Mozambique Channel in autumn, through the areas around the Equator in the western Indian Ocean in winter, spreading east in spring and returning to the Mozambique Channel in summer. We found a strong positive correlation (Spearman's $\rho = 0.89$) between whale shark SPUE and tuna CPUE recorded by the same tuna purse fleets. We used the tuna catch in the 1° resolution data set to weight the probability of selecting a pseudo-absence in the following analyses, as these data were available at a much finer spatial resolution.

Seasonal standard deviation of Chl *a* and its mean value were highly correlated in all seasons ($-0.92 <$ Spearman's $\rho < 0.87$; $P < 0.001$), as well as the distance to the continental shelf and to the shore ($\rho = 0.9115$; $P < 0.001$), and depth and distance to the shelf ($\rho = 0.6146$; $P < 0.001$). Even though distance to shelf is potentially more informative regarding whale shark distribution than distance to shore, we omitted the standard deviation of Chl *a* and distance to the shelf (instead of standard deviation of Chl *a* and both distance to shore and depth) from the list of candidate predictors to construct the model set with the lowest number of uncorrelated variables.
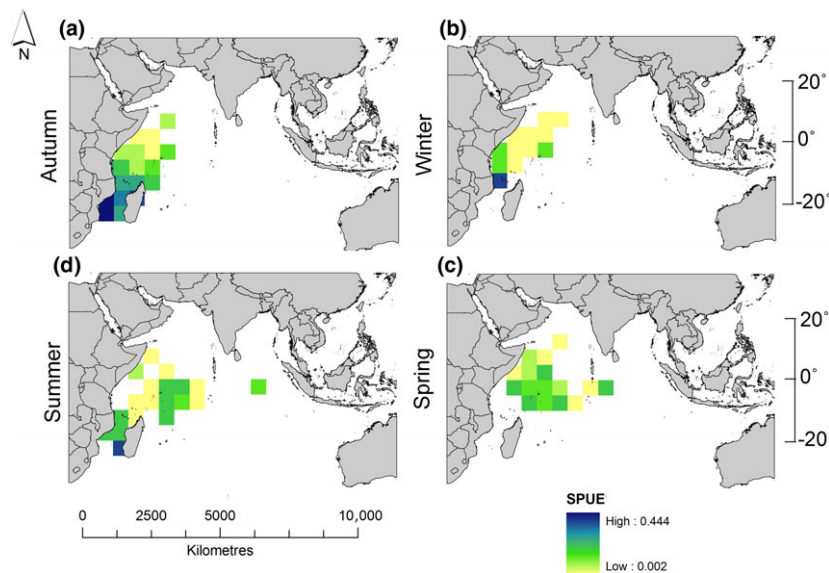

Figure 2 Seasonal variation in aggregated whale shark (*Rhincodon typus*, Smith 1828) sightings per unit effort (SPUE, where effort corresponds to the number of fishing days) in the area of the Indian Ocean sampled by the Indian Ocean Tuna Commission (IOTC) from 1991 to 2007: (a) autumn (April–June): 811 whale sharks recorded (68% of sightings); (b) winter (July–September): 68 whale sharks sighted (6% of sightings); (c) spring (October–December): 191 whale sharks sighted (16% of sightings); (d) summer (January–March): 115 whale sharks sighted (10% of total sightings).

Generalized linear and mixed-effects models

The percentage of deviance explained was highest for the generalized linear model including all environmental variables in all seasons, irrespective of the technique used for generating pseudo-absences (Table 1). The highest values were obtained with random pseudo-absence (57% in autumn and 20% in summer) and lowest with IDW pseudo-absences (< 15% in all seasons).

Statistical support (wAIC$_c$) was greatest for the model including all environmental variables as well, but only when using random and the tuna pseudo-absences in all seasons, except in spring when removing Chl a increased support using the tuna data set (Table 1). The top-ranked model according to wBIC only matched the one that also maximized wAIC$_c$ in autumn for the three pseudo-absence data sets (Fig. S4 in Supporting Information), and in winter and summer when using random pseudo-absences (Table 1).

Both observations and GLM residuals were spatially auto-correlated ($P < 0.001$). The spatial correlation structure that gave the best fit to the null model varied between the exponential and spherical, with a shape parameter of 5 or 10, depending on the method used for generating pseudo-absences (an example of the resulting Moran's I plots is shown in Fig. S6 in Supporting Information).

The pseudo-absence data set that provided the best results also differed between modelling techniques; for example, in autumn, the highest deviance was explained with random when using GLM, and with tuna when using GLMM. Percent deviance explained was generally higher for all GLM in all seasons. After accounting for the spatial autocorrelation (Table 3), GLMM using tuna pseudo-absences explained the highest deviance (25.5% in autumn, 23.7% in summer, 11.1% in spring and 5.3% in winter).

For each season, predictive maps derived from different pseudo-absence data sets resulted in similar patterns (Fig. 3), while greater differences occurred among seasons (Fig. 4). During autumn, highly suitable areas were concentrated near the Mozambique Channel and close to shore in the south-eastern side of the African continent. A shift in habitat suitability occurred in winter towards the north and central western Indian Ocean, spreading towards the east in a 'C' shape (surrounding the Equator) in spring and stretching from east to west south of the Equator in summer.

Maximum Entropy

The pseudo-absence data sets derived from the three different techniques resulted in similar maps in MaxEnt in each season (Fig. 1). MaxEnt prediction maps were generally consistent with that predicted by the GLM and GLMM (Fig. 3), except for spring – where MaxEnt presented the central equatorial area of the Indian Ocean as the most suitable habitat. Within

Table 1 Summary of generalized linear models relating probability of whale shark (*Rhincodon typus*, Smith 1828) occurrence to ocean properties: slope, depth, distance to shore (shore), mean sea surface temperature (SST mean) and its quadratic term (SST mean^2), SST standard deviation (SST SD) and chlorophyll a (Chl a).

Season	Autumn			Winter			Spring			Summer		
Model	wAIC$_c$	wBIC	%DE	wAIC$_c$	wBIC	%DE	wAIC$_c$	wBIC	%DE	wAIC$_c$	wBIC	%DE
Random												
Phys + SST var	–	–	–	0.25	0.65	34.04	*0.01*	*0.04*	18.00	–	–	–
Saturated	1.00	1.00	57.05	0.99	0.99	34.97	0.75	0.35	34.72	0.99	0.96	19.88
IDW												
Slope	–	–	–	*0.05*	*0.07*	0.15	–	–	–	*0.01*	*0.04*	0.20
Depth	–	–	–	*0.08*	*0.09*	0.55	–	–	–	*0.04*	0.11	1.08
Shore	–	–	–	*0.05*	*0.06*	0.11	–	–	–	*0.03*	*0.07*	0.70
SST var	–	–	–	0.21	*0.02*	4.51	0.20	0.89	11.80	0.25	*0.03*	3.98
Chl a	–	–	–	*0.07*	*0.07*	0.32	–	–	–	*0.04*	*0.09*	0.97
Phys + SST var	–	–	–	*0.07*	*0.00*	7.47	0.51	*0.01*	13.93	0.14	*0.07*	6.34
Saturated	0.99	0.99	11.88	*0.09*	–	8.91	0.24	0.00	14.16	0.14	0.00	7.14
Null	–	–	–	0.13	0.60	*0.00*	–	–	–	*0.02*	0.34	0.00
Tuna												
SST var	–	–	–	*0.06*	0.90	17.92	*0.00*	0.37	22.64	–	–	–
Phys + SST var	–	–	–	0.26	*0.00*	20.60	0.66	0.55	27.37	0.45	0.32	18.19
Saturated	1.00	1.00	42.61	0.86	*0.08*	26.72	0.32	*0.05*	27.77	0.54	*0.10*	19.10

Shown for each model are the bias-corrected model evidence based on weights of Akaike's information criterion corrected for small sample sizes (wAIC$_c$), weights of Bayesian information criterion (wBIC) and the percentage of deviance explained (%DE). Three different methods were used for generating pseudo-absences: Random, inversely distant to whale shark sighting locations (IDW) and based on total tuna catch (Tuna). Results shown only for cases where wAIC$_c > 0.001$, and values ≤ 0.1 are shown in italic. Note: slope, depth, distance to shore referred to together as physical variables (Phys); mean sea surface temperature (SST mean), its quadratic term (SST mean^2) and SST standard deviation (SST SD) referred to together as SST var (SST var).
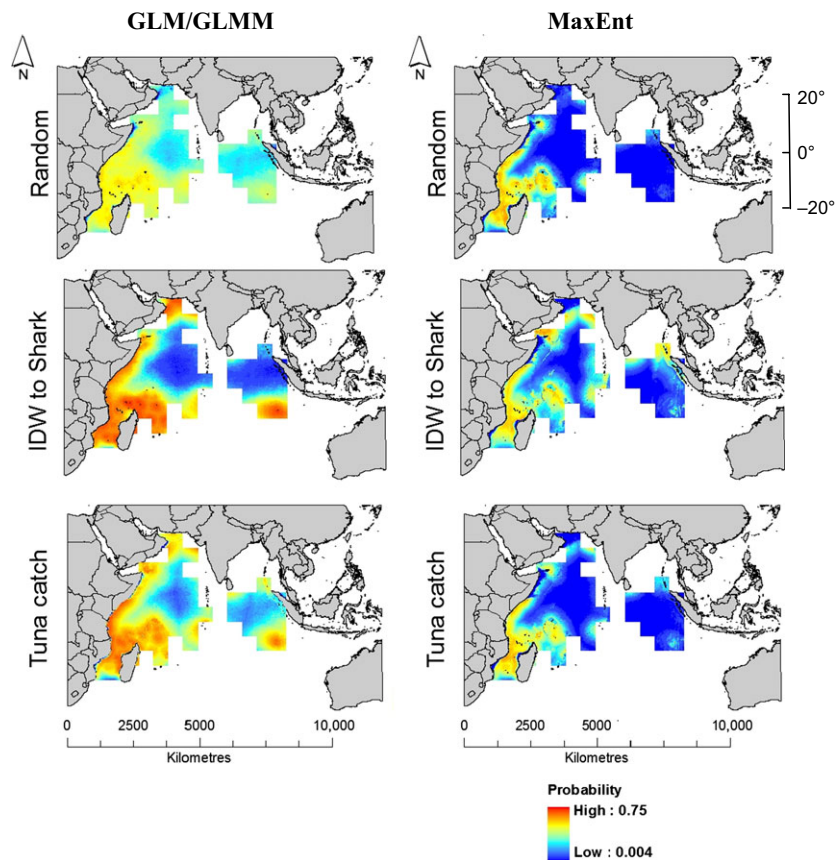
Figure 3 Habitat suitability of whale sharks (*Rhincodon typus*, Smith 1828) in the Indian Ocean during autumn. The prediction maps in the left panel are derived from generalized linear mixed-effects models, and those in the right panel are derived from Maximum Entropy when using three different techniques for generating pseudo-absences: (i) randomly (Random), (ii) based on probability weighted by the inverse distance to whale shark presence locations (IDW to shark) and (iii) by a probability directly weighted by total tuna catch (Tuna catch), respectively, per row.

the total area sampled, the region likely to be more suitable for whale shark occurrence in autumn was the Mozambique Channel, followed by the western equatorial Indian Ocean in winter, the central area of the Indian Ocean in spring, and more dispersed, but already including the Mozambique Channel again, in summer.

The MaxEnt variable importance ranking differed both among seasons and the pseudo-absence data set used in the model, while the percent of variable contribution varied mainly within seasons (Tables 2 and S1 in Supporting Information). In autumn, the more important variable was the quadratic term of SST for all techniques, while the variable with the highest percentage of contribution for the model results was distance to shore. For both winter and spring, the highest-ranked variables varied mainly between Chl *a* (mean), SST (mean), SST (SD) and the quadratic SST term, while for summer, depth was an important variable with the highest percentage contribution.

The jack-knife test identified Chl *a* (in winter and spring) and physical variables (in both autumn and summer) as those with important individual effects in all model sets (Table 2). The most important single variables were mainly producing

the poorest model results when excluded from the set of predictors.

Area under the curve obtained with the MaxEnt models was generally low, varying from 0.574 to 0.721 (Table 2). In all seasons, the highest AUC values were obtained for models with the random, followed by tuna pseudo-absences data sets (around 0.7), while the lower scores were obtained when the IDW data set was used (below 0.63). k Obtained for MaxEnt were nearly always less than those obtained from GLM, and generally higher than those from the GLMM accounting for spatial autocorrelation (Table 3).

Maximum Entropy models using the full background data available but with restricted settings, i.e. keeping the features restricted to linear and quadratic, produced similar results to the models using random and tuna pseudo-absences in terms of variables percentage of contribution and permutation importance (Tables 2 and S1 in Supporting Information). When the same model was allowed to use auto features, i.e. all the feature types available, results were somewhat different and mostly only Chl *a* (mean) came out as an important predictor (Table 2). Despite having the highest AUC (> 0.92), the predicted suitable area seemed to be more restricted with this model set.
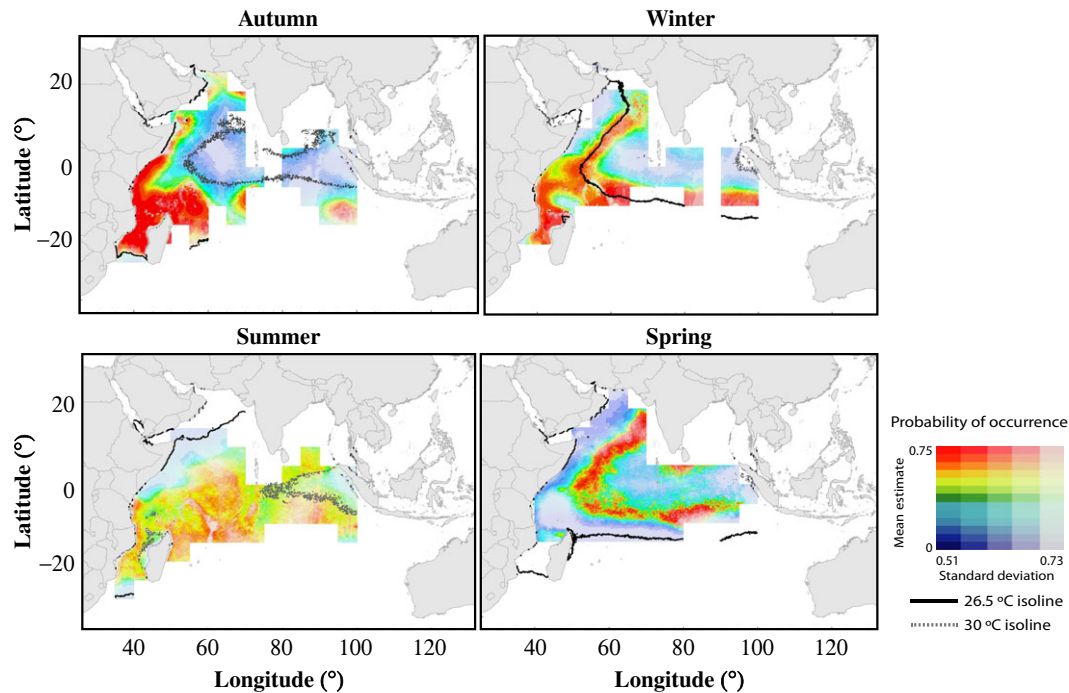
Figure 4 Seasonal ensemble result for whale sharks (*Rhincodon typus*, Smith 1828) habitat suitability overlayed by sea surface temperature isolines of 26.5 and 30 °C. Standard deviation of the estimated probability of habitat suitability is shown with fading colours as per figure legend.

DISCUSSION

Niche-based models provide an alternative means for generating information about species distributions when conventional sampling methods are expensive, logistically difficult and produce unreliable results (e.g. when sampling for rare species - Edwards *et al.*, 2005; Guisan *et al.*, 2006). In the past, occurrence data for whale sharks have been collected at spatial scales that constitute only a small part of the animal's range, usually within coastal waters where nearshore aggregations form (Beckley *et al.*, 1997; Meekan *et al.*, 2006; Graham & Roberts, 2007; Jonahson & Harding, 2007; Rowat *et al.*, 2009). This analysis of sightings collected by fisheries in the open ocean provides the first opportunity to predict the pelagic distribution of this wide-ranging species, even if collected at a coarse spatial resolution (see Barbosa *et al.*, 2010 on downscaled projections derived from low resolution data). The predictive maps produced by our models revealed a seasonal shift in whale shark habitat suitability following a clockwise direction from the south-west Indian Ocean in autumn, to the central (north and south of the equator) Indian Ocean in winter and spring, and then back to the southern Indian Ocean in summer (Fig. 4). Given that our analysis accounted as much as possible for seasonal differences in sampling effort, this clockwise shift likely results from seasonal changes in environmental conditions, such as variation in temperature, that seem to be driving whale shark distribution patterns within the Indian Ocean. Indeed, surface water properties were used before to explain variation in the temporal distribution

patterns of whale sharks (Wilson *et al.*, 2001; Wilson, 2002; Sleeman *et al.*, 2007; Cárdenas-Palomo *et al.*, 2010), although our study is the first to test these hypotheses spatially and by season at the scale of almost an entire ocean basin.

Sea surface temperature was the main variable affecting the relative occurrence of whale sharks, with the resulting predictive maps reflecting the 'C' shape of the SST patterns (Fig. S3 in Supporting Information). Despite average temperatures ranging between 23 and 34 °C (Fig. S3 in Supporting Information), around 65% of the whale shark sightings occurred between just 27.5 and 29 °C, and 90% occurred between 26.5 and 30 °C (PathFinder AVHRR – temperatures recorded during the same weeks whale sharks were spotted). It seems therefore that whale sharks use only a narrow SST range, which is in accordance with our hypothesis that a restricted temperature regime exists for this species, thus justifying the inclusion of the quadratic term in our models. Whale sharks appear to avoid high temperatures that might elevate metabolic rates and food requirements, and excessively low temperatures that limit metabolic function. However, even though whale sharks used only a small band of averaged temperatures, these are not exclusive (e.g. Turnbull & Randell, 2006), and they might move outside this envelope for other reasons such as foraging. Other species have previously been described as predominantly occurring in a small range of temperatures, e.g. leatherback turtle (15–33 °C; McMahon & Hays, 2006), salmon sharks (2–8 °C in winter; Weng *et al.*, 2005) and white sharks (10–14 °C; Boustany *et al.*, 2002).Our results are consistent with much of what is known about the

Table 2 Summary of the Maximum Entropy (MaxEnt) models relating probability of whale shark (*Rhincodon typus*, Smith 1828) occurrence to individual ocean properties: slope, depth, distance to shore (shore), mean sea surface temperature (SST mean) and its quadratic term (SST mean^2), SST standard deviation (SST SD) and chlorophyll *a* (Chl *a*).

Season	Autumn	Winter	Spring	Summer
Random – linear and quadratic features				
Top ranked	SST mean^2	Chl *a*	SST mean	Depth
Best/worst	Shore	Chl *a*	Chl *a*/SST SD	Depth
AUC	0.721	0.701	0.691	0.668
IDW – linear and quadratic features				
Top ranked	SST mean^2	Depth	Chl *a*	Depth
Best/worst	Depth	Chl *a*	Chl *a*	Depth
AUC	0.625	0.574	0.627	0.607
Tuna – linear and quadratic features				
Top ranked	SST mean^2	SST mean^2	SST mean	Depth
Best/worst	Shore	Chl *a*	Chl *a*/SST SD	Depth
AUC	0.710	0.673	0.672	0.676
Background – linear and quadratic features				
Top ranked	SST mean^2	SST mean^2	SST mean	SST mean
Best/worst	Depth	Chl *a*	SST mean^2/shore	Depth
AUC	0.928	0.837	0.858	0.843
Background – auto features				
Top ranked	Chl *a*	Chl *a*	SST mean^2	Chl *a*
Best/worst	Chl *a*	Chl *a*	SST mean/Chl *a*	Chl *a*
AUC	0.961	0.920	0.956	0.928

Three different methods were used for generating pseudo-absences: Random, inversely distant to whale shark sighting locations (IDW) and based on total tuna catch (Tuna), and results are shown in rows 1–3 when only linear and quadratic features were used. The two last rows show results when MaxEnt model was given the full background (Background) with covariate data available and varying the feature type used in the model. The jack-knife test results showing which environmental variable had the highest gain when used in isolation (best) and which environmental variables decreased the gain the most when omitted (worst) are also shown together with the value obtained for the area under the curve (AUC) test. Only top-ranked variables according to permutation importance are shown here – the complete table with scores per variable can be seen in the supplementary material (Table S1 in Supporting Information).

occurrence of these sharks from finer-scale (km–100s km) studies of nearshore aggregations in the Indian Ocean (Fig. S7 in Supporting Information). According to Rowat (2007), whale sharks aggregate off South Africa, Mozambique and Kenya mainly in summer where our models predicted a relatively high probability of occurrence. South African and Mozambique aggregations persist into autumn, which is consistent with our predictions. A higher density of whale sharks between January and May were also detected by Beckley *et al.* (1997) and Cliff *et al.* (2007) in the far north of the South African coast. During autumn, there are peaks in whale shark abundance off Gujarat (India) and Thailand, another observation that is in accordance with our model predictions (Theberge & Dearden, 2006 recorded an increase in whale shark sightings starting in October and peaking in May). Peak aggregations off Tanzania, Kenya and Seychelles occur in winter, when our 0.4 probability isoline covers these areas. In the Seychelles, whale sharks peak in abundance during the spring and winter (Rowat & Engelhardt, 2007), while in the Maldives, abundances peak in spring, as our models predict (Fig. S7 – the 0.4 probability isoline covers the referred locations in the correspondent seasons). These aggregations are both consistent with our model outcomes. However, peaks in abundance of sharks observed off the coasts of Bangladesh and

in the Mozambique Channel in spring could not be compared to our model outputs because the available data did not extend to these regions at these times.

Our prediction of the occurrence of whale sharks along the Madagascar coastline throughout the remainder of the year is corroborated by opportunistic observations (Jonahson & Harding, 2007). The Mozambique Channel also has suitable conditions for whale sharks almost year-round (Pierce *et al.*, 2010). The presence of sharks in the Channel might be influenced by series of rotating gyres that spin off sequentially southwards down into the Channel (DiMarco *et al.*, 2002); these are thought to entrain tuna and might do the same for whale sharks. In our models, the high likelihood of whale shark occurrence in this area was driven mainly by the suitable SST and/or productivity ranges observed there (we did not have access to geostrophic current data). However, we found no strong evidence to explain why the Channel is an important whale shark habitat, except that Chl *a* never dropped below 0.1 mg m^{-3} there (Fig. S2 in Supporting Information).

Generally, we found that SST was a better predictor of whale shark distributions than Chl *a*. The latter was used as a proxy for food availability (zooplankton) for whale sharks, although trophic links between phytoplankton and zooplankton are not necessarily direct, strong or immediate. There are likely at least

Table 3 Comparison of results for the Maximum Entropy (MaxEnt) and generalized linear (GLM) and mixed-effects (GLMM) models – where spatial autocorrelation was included for the three pseudo-absence techniques used (Random, IDW and Tuna).

Season	Autumn			Winter			Spring			Summer		
Model	κ	SD	%DE	κ	SD	%DE	κ	SD	%DE	κ	SD	%DE
Random												
MaxEnt	0.54	0.02	–	0.57	0.08	–	0.50	0.05	–	0.49	0.06	–
GLM	0.79	0.09	58.0	0.51	0.09	29.5	0.59	0.05	32.8	0.51	0.06	17.7
GLMM	0.36	0.02	23.9	0.04	0.07	1.3	0.00	0.01	0.0	0.22	0.07	6.3
IDW												
MaxEnt	0.34	0.03	–	0.10	0.09	–	0.25	0.05	–	0.31	0.06	–
GLM	0.35	0.02	10.9	0.00	0.00	0.0	0.33	0.06	13.3	0.00	0.00	0.0
GLMM	0.21	0.02	4.9	0.17	0.09	10.5	0.26	0.06	2.6	0.07	0.06	8.5
Tuna												
MaxEnt	0.64	0.02	–	0.37	0.08	–	0.48	0.05	–	0.43	0.06	–
GLM	0.68	0.02	42.3	0.47	0.08	30.9	0.51	0.05	24.4	0.56	0.56	22.5
GLMM	0.59	0.02	25.5	0.56	0.07	5.3	0.17	0.03	11.1	0.43	0.06	23.7

Shown for each model are the Cohen's Kappa statistic (κ) and its standard deviation (SD). The percentage of deviance explained (%DE) for each technique (shown for the average model predictions – weighted by $w\text{AIC}_c$) is also shown for the GLM and GLMM.

to be time-lags (and therefore spatial) between peaks of Chl *a* and zooplankton (Runge, 1998; Sleeman *et al.*, 2010b). The filtering effects of zooplankton on algae (sensu Runge, 1998) suggest that direct measurements of zooplankton abundance would provide a much better predictor of whale shark distribution than Chl *a* per se, which appears to be the case for filter-feeding basking sharks (Sims *et al.*, 2005). However, such data are only available over relatively restricted spatial scales; remote sensing provides the only means by which estimates of food availability can be obtained at the spatial scales relevant to oceanic patterns.

Both genetic (Castro *et al.*, 2007; Schmidt *et al.*, 2009) and satellite tracking data (Wilson *et al.*, 2007; Sleeman *et al.*, 2010b) indicate a capacity for long-distance dispersal in whale sharks. Despite no observations of individually identified (Speed *et al.*, 2007) whale sharks traversing the Indian Ocean basin, the basin-scale shift in distributions predicted by our models suggests at least some migratory behaviour at the individual level, implying that broadscale movements are possible such that individuals could visit several known aggregation sites as they follow suitable environmental conditions among seasons. We identified SST as a key determinant of whale shark distribution, thus forewarning that current aggregation locations might shift with a changing climate.

Our models of seasonal distribution were able to predict habitat suitability for whale sharks over a more extensive area than that covered by the sightings/tagging data alone. They can be used to assess interannual variability in sightings at an ocean scale. Fluctuating conditions measured at interannual scales using remote sensing can be used to infer interannual differences in the probability of occurrence over time and space outside of our study area. Such results would assist in predicting how seasonal aggregations might shift over space and time.

All three pseudo-absence techniques (random, IDW and tuna) resulted in similar predictive maps for each season. However, we found major differences among seasons within each technique, mainly in terms of the regression models' deviance explained. It should be noted here that because pseudo-absence locations vary within each model, the resultant deviance explained and AUC are not directly comparable. Nevertheless, they are useful to determine how absence locations influence the models' results, and we have used them to weight the predicted probability of occurrence derived from each model when generating the final ensemble prediction.

Generalized linear models explained the highest deviance for almost all models, but this approach does not account for spatial autocorrelation that can lead to an inflated explanatory power in models of species abundance (Lichstein *et al.*, 2002; Wintle & Bardos, 2006). The inclusion of a spatial correlation structure is usually necessary because it likely arises from the ecological processes that drive population dynamics (Mellin *et al.*, 2010). Our GLMM approach resulted in lower kappa statistics and deviance explained by the models. This change in results is not easy to interpret (Dormann *et al.*, 2007); however, we contend that the GLMM approach that accounted for some of the potential spatial bias provides predictions of higher confidence, as GLM residuals are highly spatially autocorrelated, confirming that these models are biased.

Our results show that the MaxEnt model can produce similar prediction maps to those generated by GLM based on the same input data sets. Being a much easier tool to employ, MaxEnt is useful to develop species distribution models quickly that give results analogous to more robust regression models. It is noteworthy, however, that assessment results (made here by means of Kappa statistics) were slightly lower with MaxEnt when compared to GLM. When spatial autocorrelation is not accounted for (either with MaxEnt or GLMs),

the random method for pseudo-absence selection generally resulted in better performances. In this context, MaxEnt models can be used even more efficiently by using the full set of available covariate data and letting the model randomly select the points used as 'pseudo-absences' (i.e. background). Results obtained by the runs with the full data set and the linear and quadratic features gave similar maps and better AUC results. When the MaxEnt model used all features, the same sort of pattern in habitat suitability was apparent; however, it produced much steeper spatial gradients across the region (Fig. S7 in Supporting Information). This resulted in a reduced capacity to predict some known whale shark areas of occurrence relative to GLM and GLMM predictions (Fig. S7 in Supporting Information). Additionally, MaxEnt identified depth, Chl *a* and SST as the main predictors, which contrasts somewhat with the support for more complex models using GLM and GLMM.

The coarse resolution of the input data, the necessity of relying mainly on surface data from remote sensing (as opposed to data integrated over all depths exploited by this species) and the lack of true absences mean that predictions should be taken only as an index of relative probability of occurrence. The tuna purse-seine fisheries covered a large area (172,800 km^2), but not all of the Indian Ocean. For this reason, there is less uncertainty in predictions for the western than the eastern part of the ocean basin. Additionally, about 70% of the sightings occurred during autumn, making predictions for other seasons relatively less robust.

There is a growing demand by managers for ecologists to supply more accurate results on the area of occurrence and distribution of ecological niches of species. Such data are fundamental in generating appropriate protection rules for management strategies (Lehmann *et al.*, 2002; Beger & Possingham, 2008; Urbina-Cardona & Flores-Villela, 2009). Combining data collected by the Regional Fisheries Management Organisations with our modelling approach, the timing of whale shark appearances at specific sites (e.g. at sites where they are still currently fished) can also be predicted and subsequently used to examine the drivers of observed population trends (Bradshaw, 2007; Bradshaw *et al.*, 2008). Further, whale sharks are frequently seen with large wounds or scars clearly derived from collision with boats or ship propellers (Speed *et al.*, 2008). Predicted areas for whale shark occurrence could thus be used as input information for management of shipping routes. In general, understanding the distribution and migration patterns of whale sharks is an essential precursor to identify possible mating and breeding areas, and to understand the potential effects of fisheries and ecotourism on the probability of the long-term persistence of the species.

REFERENCES

Araújo, M.B. & New, M. (2006) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.

Barbosa, A.M., Real, R. & Vargas, J.M. (2010) Use of coarse-resolution models of species' distributions to guide local conservation inferences. *Conservation Biology*, **24**, 1378–1387.

Beaumont, L.J., Pitman, A.J., Poulsen, M. & Hughes, L. (2007) Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology*, **13**, 1368–1385.

Beckley, L.E., Cliff, G., Smale, M.J. & Compagno, L.J.V. (1997) Recent strandings and sightings of whale sharks in South Africa. *Environmental Biology of Fishes*, **50**, 343–348.

Beger, M. & Possingham, H.P. (2008) Environmental factors that influence the distribution of coral reef fishes: modeling occurrence data for broad-scale conservation and management. *Marine Ecology Progress Series*, **361**, 1–13.

Boustany, A.M., Davis, S.F., Pyle, P., Anderson, S.D., Le Boeuf, B.J. & Block, B.A. (2002) Expanded niche for white sharks. *Nature*, **415**, 35–36.

Bradshaw, C.J.A. (2007) Swimming in the deep end of the gene pool: global population structure of an oceanic giant. *Molecular Ecology*, **16**, 5111–5114.

Bradshaw, C.J.A., Fitzpatrick, B.M., Steinberg, C.C., Brook, B.W. & Meekan, M.G. (2008) Decline in whale shark size and abundance at Ningaloo Reef over the past decade: the world's largest fish is getting smaller. *Biological Conservation*, **141**, 1894–1905.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.

Brunnschweiler, J.M., Baensch, H., Pierce, S.J. & Sims, D.W. (2009) Deep-diving behaviour of a whale shark *Rhincodon typus* during long-distance movement in the western Indian Ocean. *Journal of Fish Biology*, **74**, 706–714.

Bullock, T.H. (1955) Compensation for temperature in the metabolism and activity of poikilotherms. *Biological Reviews*, **30**, 311–342.

Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research*, **33**, 261–304.

Cárdenas-Palomo, N., Herrera-Silveira, J. & Reyes, Ó. (2010) Spatial and temporal distribution of physicochemical

features in the habitat of whale shark *Rhincodon typus* (Orectolobiformes: Rhincodontidae) in the north of Mexican Caribbean. *Revista de Biología Tropical*, **58**, 399–412.

Cárdenas-Torres, N., Enríquez-Andrade, R. & Rodríguez-Dowdell, N. (2007) Community-based management through ecotourism in Bahia de Los Angeles, Mexico. *Fisheries Research*, **84**, 114–118.

Castro, A.L.F., Stewart, B.S., Wilson, S.G., Hueter, R.E., Meekan, M.G., Motta, P.J., Bowen, B.W. & Karl, S.A. (2007) Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology*, **16**, 5183–5192.

Cliff, G., Anderson-Reade, M.D., Aitken, A.P., Charter, G.E. & Peddemors, V.M. (2007) Aerial census of whale sharks (*Rhincodon typus*) on the northern KwaZulu-Natal coast, South Africa. *Fisheries Research*, **84**, 41–46.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Colman, J.G. (1997) A review of the biology and ecology of the whale shark. *Journal of Fish Biology*, **51**, 1219–1234.

Compagno, L.J.V. (2001) *Sharks of the World: An Annotated and Illustrated Catalogue of Shark Species Known to Date, vol. 2: Bullhead, Mackerel and Carpet Sharks (Heterodontiformes, Lamniformes and Orectolobiformes)*. FAO, Species Catalogue for Fishery Purposes No.1; Rome.

De'ath, G. (2007) Boosted trees for ecological modelling and prediction. *Ecology*, **88**, 243–251.

Diggle, P. & Ribeiro, P.J. (2007) *Model-based geostatistics*. Springer, New York.

DiMarco, S.F., Chapman, P., Nowlin ., W.D. Jr, Hacker, P., Donohue, K., Luther, M., Johnson, G.C. & Toole, J. (2002) Volume transport and property distributions of the Mozambique Channel. *Deep-Sea Research II*, **49**, 1481–1511.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

Edwards, T.C. Jr, Cutler, D.R., Zimmermann, N.E., Geiser, L. & Alegria, J. (2005) Model-based stratifications for enhancing the detection of rare ecological events. *Ecology*, **86**, 1081–1090.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J., Leathwick, J.R. & T., H. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Garza-Pérez, J.R., Lehmann, A. & Arias-González, J.E. (2004) Spatial prediction of coral reef habitats: integrating ecology with spatial modeling and remote sensing. *Marine Ecology Progress Series*, **269**, 141–152.

GEBCO (2003) General bathymetric chart of the oceans – one minute grid, version 2.0. http://www.gebco.net.

Graham, R.T. & Roberts, C.M. (2007) Assessing the size, growth rate and structure of a seasonal population of whale sharks (*Rhincodon typus* Smith 1828) using conventional tagging and photo identification. *Fisheries Research*, **84**, 71–80.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

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

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E. (2006) Using niche-based models to improve the sampling of rare species. *Biological Conservation*, **20**, 501–511.

Hastie, T. & Tibshirani, R. (1986) Generalized additive models. *Statistical Science*, **1**, 297–318.

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.

Hutchison, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415–457.

IUCN (2010) *The IUCN red list of threatened species*. International Union for Conservation of Nature, www.iucnredlist.org.

Jonahson, M. & Harding, S. (2007) Occurrence of whale sharks (*Rhincodon typus*) in Madagascar. *Fisheries Research*, **84**, 132–135.

Joung, S.-J., Chen, C.-T., Clark, E., Uchida, S. & Huang, W.Y.P. (1996) The whale shark (*Rhincodon typus)* is a live-bearer: 300 embryos found in one 'megamamma' supreme. *Environmental Biology of Fishes*, **46**, 219–223.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species ranges. *Ecology Letters*, **12**, 334–350.

Kelly, N.M., Fonseca, M. & Whitfield, P. (2001) Predictive mapping for management and conservation of seagrass beds in North Carolina. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **11**, 437–451.

Kumari, B. & Raman, M. (2010) Whale shark habitat assessments in the northeastern Arabian Sea using satellite remote sensing. *International Journal of Remote Sensing*, **31**, 379–389.

Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Last, P.R. & Stevens, J.D. (1994) *Sharks and rays of Australia*. CSIRO, Australia.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, **50**, 2034–2052.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*, 2nd English edn. Elsevier, Amsterdam.

Lehmann, A., Overton, J.M. & Austin, M.P. (2002) Regression models for spatial prediction: their role for biodiversity and conservation. *Biodiversity and Conservation*, **11**, 2085–2092.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445–463.

Link, W.A. & Barker, R.J. (2006) Model weights and the foundations of multimodel inference. *Ecology*, **87**, 2626–2635.

Martin, R.A. (2007) A review of behavioural ecology of whale sharks (*Rhincodon typus*). *Fisheries Research*, **84**, 10–16.

Matsunaga, H., Nakano, H., Okamoto, H. & Suzuki, Z. (2003) Whale shark migration observed by pelagic tuna fishery near Japan. *16th Meeting of the Standing Committee on tuna and billfish* (ed J. Fisheries Research Agency), National Research Institute of Far Seas Fisheries, Shizuoka, **BBRG-12**, 1–7.

McMahon, C. & Hays, G.C. (2006) Thermal niche, large-scale movements and implications of climate change for a critically endangered marine vertebrate. *Global Change Biology*, **12**, 1330–1338.

Meekan, M.G., Bradshaw, C.J.A., Press, M., McLean, C., Richards, A., Quasnichka, S. & Taylor, J.G. (2006) Population size and structure of whale sharks *Rhincodon typus* at Ningaloo Reef, Western Australia. *Marine Ecology Progress Series*, **319**, 275–285.

Meekan, M.G., Jarman, S.N., McLean, C. & Schultz, M.B. (2009) DNA evidence of whale sharks (*Rhincodon typus*) feeding on red crab (*Gecarcoidea natalis*) larvae at Christmas Island, Australia. *Marine and Freshwater Research*, **60**, 607–609.

Mellin, C., Bradshaw, C.J.A., Meekan, M.G. & Caley, M.J. (2010) Environmental and spatial predictors of species richness and abundance in coral reef fishes. *Global Ecology and Biogeography*, **19**, 212–222.

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of Royal Statistics Society*, **135**, 370–384.

Oviedo, L. & Solís, M. (2008) Underwater topography determines critical breeding habitat for humpback whales near Osa Peninsula, Costa Rica: implications for marine protected areas. *Revista de Biologia Tropical* (Int. J. Trop. Biol.), **56**, 591–602.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning*. Banff, Canada, 655–662.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Pianet, R., Molina de, A.D., Doriso, J., Dewals, P., Norström, V. & Ariz, J. (2009) *Statistics of the main purse seine fleets fishing in the Indian Ocean (1981–2008)*. Indian Ocean Tuna Commission, Mombasa, Kenya.

Pierce, S.J., Méndez-Jiménez, A., Collins, K., Rosero-Caicedo, R. & Monadjem, A. (2010) Developing a code of conduct for whale shark interactions in Mozambique. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **20**, 782–788.

Praca, E. & Gannier, A. (2007) Ecological niche of three teuthophageous odontocetes in the northwestern Mediterranean Sea. *Ocean Science Discussions*, **4**, 785–815.

Pravin, P. (2000) Whale shark in the Indian coast – need for conservation. *Current science*, **79**, 310–315.

R Development Core Team (2010) R: A language and environment for statistical computing, reference index version 2.11.0. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Richardson, A.J. & Poloczanska, E.S. (2008) Under-resourced, under threat. *Science*, **320**, 1294–1295.

Riley, M.J., Harman, A. & Rees, R.G. (2009) Evidence of continued hunting of whale sharks *Rhincodon typus* in the Maldives. *Environmental Biology of Fishes*, **86**, 371–374.

Robertson, M.P., Peter, C.I., Villet, M.H. & Ripley, B.S. (2003) Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, **164**, 153–167.

Rowat, D. (2007) Occurrence of whale shark (*Rhincodon typus*) in the Indian Ocean: a case for regional conservation. *Fisheries Research*, **84**, 96–101.

Rowat, D. & Engelhardt, U. (2007) Seychelles: a case study of community involvement in the development of whale shark ecotourism and its socio-economic impact. *Fisheries Research*, **84**, 109–113.

Rowat, D., Gore, M., Meekan, M.G., Lawler, I.R. & Bradshaw, C.J.A. (2009) Aerial survey as a tool to estimate whale shark abundance trends. *Journal of Experimental Marine Biology and Ecology*, **368**, 1–8.

Runge, J.A. (1998) Should we expect a relationship between primary production and fisheries? The role of copepod dynamics as a filter of trophic variability. *Hydrobiologia*, **167/168**, 61–71.

Schmidt, J.V., Schmidt, C.L., Ozer, F., Ernst, R.E., Feldheim, K.A., Ashley, M.V. & Levine, M. (2009) Low genetic differentiation across three major ocean populations of the whale shark, *Rhincodon typus*. *PLoS One*, **4**, e4988.

Sequeira, A., Ferreira, J.G., Hawkins, A.J.S., Nobre, A., Lourenço, P., Zhang, X.L., Yan, X. & Nickell, T. (2008) Trade-offs between shellfish aquaculture and benthic biodiversity: a modelling approach for sustainable management. *Aquaculture*, **274**, 313–328.

Sérgio, C., Figueira, R., Draper, D., Menezes, R. & Sousa, A.J. (2007) Modelling bryophyte distribution based on ecological

information for extent of occurrence assessment. *Biological Conservation*, **135**, 341–351.

Sims, D.W., Southall, E.J., Richardson, A.J., Reid, P.C. & Metcalfe, J.D. (2003) Seasonal movements and behaviour of basking sharks from archival tagging: no evidence of winter hibernation. *Marine Ecology Progress Series*, **248**, 187–196.

Sims, D.W., Southall, E.J., Tarling, G.A. & Metcalfe, J.D. (2005) Habitat-specific normal and reverse diel vertical migration in the plankton-feeding basking shark. *Journal of Animal Ecology*, **74**, 755–761.

Sleeman, J.C., Meekan, M.G., Wilson, S.G., Jenner, C.K.S., Jenner, M.N., Boggs, G.S., Steinberg, C.C. & Bradshaw, C.J.A. (2007) Biophysical correlates of relative abundances of marine megafauna at Ningaloo Reef, Western Australia. *Marine and Freshwater Research*, **58**, 608–623.

Sleeman, J.C., Meekan, M.G., Fitzpatrick, B.J., Steinberg, C.R., Ancel, R. & Bradshaw, C.J.A. (2010a) Oceanographic and atmospheric phenomena influence the abundance of whale sharks at Ningaloo Reef, Western Australia. *Journal of Experimental Marine Biology and Ecology*, **382**, 77–81.

Sleeman, J.C., Meekan, M.G., Wilson, S.G., Polovina, J.J., Stevens, J.D., Boggs, G.S. & Bradshaw, C.J.A. (2010b) To go or not to go with the flow: environmental influences on whale shark movement patterns. *Journal of Experimental Marine Biology and Ecology*, **390**, 84–98.

Southall, E.J., Sims, D.W., Witt, M.J. & Metcalfe, J.D. (2006) Seasonal space-use estimates of basking sharks in relation to protection and political-economic zones in the North-east Atlantic. *Biological conservation*, **132**, 33–39.

Speed, C.W., Meekan, M.G. & Bradshaw, C.J.A. (2007) Spot the match – wildlife photo-identification using information theory. *Frontiers in Zoology*, **4**, doi:10.1186/1742-9994-4-2.

Speed, C.W., Meekan, M.G., Rowat, D., Pierce, S.J., Marshall, A.D. & Bradshaw, C.J.A. (2008) Scarring patterns and relative mortality rates of Indian Ocean whale sharks. *Journal of Fish Biology*, **72**, 1488–1503.

Stevens, J.D. (2007) Whale shark (*Rhincodon typus*) biology and ecology: a review of the primary literature. *Fisheries Research*, **84**, 4–9.

Stewart, B.S. & Wilson, S.G. (2005) Threatened fishes of the world: *Rhincodon typus* (Smith 1828) (Rhincodontidae). *Environmental Biology of Fishes*, **74**, 184–185.

Theberge, M.M. & Dearden, P. (2006) Detecting a decline in whale shark *Rhincodon typus* sightings in the Andaman Sea, Thailand, using ecotourist operator-collected data. *Oryx*, **40**, 337–342.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD: a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 1–5.

Tittensor, D.P., Baco, A.R., Brewin, P.E., Clark, M.R., Consalvey, M., Hall-Spencer, J., Rowden, A.A., Schlacher, T., Stocks, K.I. & Rogers, A.D. (2009) Predicting global habitat suitability for stony corals on seamounts. *Journal of Biogeography*, **36**, 1111–1128.

Turnbull, S.D. & Randell, J.E. (2006) Rare occurrence of a *Rhincodon typus* (Whale shark) in the Bay of Fundy, Canada. *Northeastern Naturalist*, **13**, 57–58.

Urbina-Cardona, J.N. & Flores-Villela, O. (2009) Ecological-Niche modeling and prioritization of conservation-area networks for Mexican herpetofauna. *Conservation Biology*, **24**, 1031–1041.

Venables, W.N. & Ripley, B.D. (2002) *Modern applied statistics with S*, 4th edn. Springer, New york.

Weng, K.C., Castilho, P.C., Morrissette, J.M., Landeira-Fernandez, A.M., Holts, D.B., Schallert, R.J., Goldman, K.J. & Block, B.A. (2005) Satellite tagging and cardiac physiology reveal niche expansion in Salmon sharks. *Science*, **310**, 104.

White, W.T. & Cavanagh, R.D. (2007) Whale shark landings in Indonesian artisanal shark and ray fisheries. *Fisheries Research*, **84**, 128–131.

Wilson, S.G. (2002) A whale shark feeding in association with a school of giant herring at Ningaloo Reef, Western Australia. *Journal of the Royal Society of Western Australia*, **85**, 43–44.

Wilson, S.G., Taylor, J.G. & Pearce, A.F. (2001) The seasonal aggregation of whale sharks at Ningaloo Reef, Western Australia: currents, migrations and the El Niño/Southern Oscillation. *Environmental Biology of Fishes*, **61**, 1–11.

Wilson, S.G., Polovina, J.J., Stewart, B.S. & Meekan, M.G. (2006) Movements of whale sharks (*Rhincodon typus*) tagged at Ningaloo Reef, Western Australia. *Marine Biology*, **148**, 1157–1166.

Wilson, S.G., Stewart, B.S., Polovina, J.J., Meekan, M.G., Stevens, J.D. & Galuardi, B. (2007) Accuracy and precision of archival tag data: a multiple-tagging study conducted on a whale shark (*Rhincodon typus*) in the Indian Ocean. *Fisheries Oceanography*, **16**, 547–554.

Wintle, B.A. & Bardos, D.C. (2006) Modeling species-habitat relationships with spatially autocorrelated observation data. *Ecological Applications*, **16**, 1945–1958.

Woodby, D., Carlile, D. & Hulbert, L. (2009) Predictive modeling of coral distribution in the Central Aleutian Islands, USA. *Marine Ecology Progress Series*, **397**, 227–240.

Zaniewski, A.E., Lehmann, A. & Overton, J.M. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261–280.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Bathymetry of the Indian Ocean as per the General Bathymetric Chart of the Oceans (1-min grid approximately 1.8 km).

Figure S2 Seasonal chlorophyll *a* concentration (Chl *a*) obtained from averaged weekly SeaWiFS satellite composites at a 9-km spatial resolution from 1997 to the end of 2007.

Figure S3 Seasonal sea surface temperature (SST) obtained from averaged weekly MODIS-Aqua satellite composites at a 9-km spatial resolution from 2002 to the end of 2007.

Figure S4 Probabilities (top panel) and resulting pseudo-absences locations (bottom panel) generated by three different techniques (left: randomly, middle: inversely proportional to distance from shark locations and right: directly proportional to tuna catch).

Figure S5 Whale shark (*Rhincodon typus*, Smith 1828) habitat suitability for each season as predicted by MaxEnt when using the full background with only linear and quadratic features.

Figure S6 Moran's I plots showing the reduction in spatial autocorrelation in the GLM residuals when a random effect was included to the models.

Figure S7 Seasonal habitat suitability of whale sharks (*Rhincodon typus*, Smith 1828) in the Indian Ocean.

Table S1 Summary of the MaxEnt models relating probability of whale shark (*Rhincodon typus*, Smith 1828) occurrence to individual ocean properties: slope, depth, distance to shore (shore), mean sea surface temperature (SST mean) and its quadratic term (SST mean^2), SST standard deviation (SST SD) and chlorophyll *a* (Chl *a*).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Ana Sequeira is a PhD student at the University of Adelaide. Her main research interests are to develop models applied to the marine environment to describe key environmental processes, species distribution patterns and ecological interactions.

Camille Mellin is a postdoctoral research associate at the Australian Institute of Marine Science, and she has research interests in biogeography, marine ecology and the impact of global change on marine ecosystems.

David Rowat works primarily on whale shark ecology, distribution and migration; other research interests include turtle and cetacean ecology and the development of conservation measures for migratory sharks.

Mark Meekan is a Principal Research Scientist at the Australian Institute of Marine Science and has research interests in the ecology, demography and behaviour of tropical fishes.

Corey Bradshaw has research interests in population ecology (density feedback, sustainable harvest and extinction dynamics), climate change biology, behavioural ecology and invasive species.

Author contributions: D.R. provided data and assisted interpretation; A.S., C.M. and C.J.A.B. analysed the data and developed the models; C.B. and M.G.M. identified the relevant biological questions and provided interpretation; A.S., C.M. and C.J.A.B. led the writing.

Editor: Mark Burgman