Environmental factors and megafauna spatio-temporal co-occurrence with purse-seine fisheries

Lauriane Escalle, Maria Grazia Pennino, Daniel Gaertner, Pierre Chavance, Alicia Delgado de Molina, Hervé Demarcq, Evgeny Romanov, Bastien Merigot

Environmental factors and megafauna spatio-temporal co-occurrence with purse-seine fisheries

LAURIANE ESCALLE,[1,2]* MARIA GRAZIA PENNINO,[2,†] DANIEL GAERTNER,[2] PIERRE CHAVANCE,[2] ALICIA DELGADO DE MOLINA,[3] HERVÉ DEMARCQ,[2] EVGENY ROMANOV[4] AND BASTIEN MERIGOT[1]

[1]*Université Montpellier, UMR MARBEC, Station Ifremer, Av. Jean Monnet, BP 171, Sète 34203, France*
[2]*Institut de Recherche pour le Développement, UMR MARBEC, Station Ifremer, Av. Jean Monnet, BP 171, Sète 34203, France*
[3]*Instituto Español de Oceanografía, Apdo. Correos 1373, 38080 S/C Tenerife, Canary Island, Spain*
[4]*CAP RUN – ARDA, Magasin No 10, Port Ouest, Le Port, Île de la Réunion 97420, France*

ABSTRACT

Tropical tuna purse-seine fisheries spatially co-occur with various megafauna species, such as whale sharks, dolphins and baleen whales in all oceans of the world. Here, we analyzed a 10-year (2002–2011) dataset from logbooks of European tropical tuna purse-seine vessels operating in the tropical Eastern Atlantic and Western Indian Oceans, with the aim of identifying the principle environmental variables under which such co-occurrence appear. We applied a Delta-model approach using Generalized Additive Models (GAM) and Boosted Regression Trees (BRT) models, accounting for spatial autocorrelation using a contiguity matrix based on a residuals autocovariate (RAC) approach. The variables that contributed most in the models were chlorophyll-a concentration in the Atlantic Ocean, as well as depth and monsoon in the Indian Ocean. High co-occurrence between whale sharks, baleen whales and tuna purse-seine fisheries were mostly observed in productive areas during particular seasons. In light of the lack of a full coverage scientific observer on board program, the large, long-term dataset obtained from logbooks of tuna purse-seine vessels is highly important for identifying seasonal and spatial co-occurrence between the distribution of fisheries and megafauna, and the underlying environmental variables. This study can help to design conservation management measures for megafauna species within the framework of spatial fishery management strategies.

Key words: cetaceans, Eastern Atlantic Ocean, generalized additive models-boosted regression trees, marine conservation, purse-seine fishery, residual autocovariate, Western Indian Ocean, whale sharks

INTRODUCTION

In the open ocean, the tropical tuna purse-seine fishery co-occurs with several megafauna species, including sharks and cetaceans, which are often emblematic species very vulnerable to natural and anthropogenic impacts. Some of these species, such as whale sharks (*Rhincodon typus*), baleen whales or dolphins, may serve as sighting cues by purse-seine fishermen for the presence of tuna schools at the surface of the sea and/or may be caught accidentally when fishing for the targeted species (Romanov, 2002; Dagorn *et al.*, 2013; Capietto *et al.*, 2014). Considering the decline of their populations and the multiple threats they have to face, whale sharks and all cetacean species which co-occur with purse-seine fisheries have been listed by the International Union for Conservation of Nature (IUCN; http://www.redlist.org).

The whale shark is the world's largest chondrichthyan (Rowat and Brooks, 2012) and occurs across all warm temperate seas. Whale sharks spend most of their time near the surface to feed on pelagic invertebrates (e.g., krill, shrimp eggs and copepods) or small forage fish (Rowat and Brooks, 2012). While feeding near the surface, whale sharks typically swim slowly and often aggregate large numbers of juvenile tunas and other bony fishes (Gaertner and Medina-Gaertner, 1999). For this reason, whale sharks are considered as a 'living' Fish Aggregation Device (FAD) by fishers. In the Eastern Atlantic and the Western Indian Oceans, whale sharks co-occurrence with the

**Correspondence*. e-mail: lauriane.escalle@yahoo.fr
†Present address: Instituto Español de Oceanografía, Centro Oceanográfico de Murcia, C/Varadero 1, 30740 San Pedro del Pinatar, Spain.

purse-seine fishery is relatively low (1.5% of the total number of fishing sets are associated to whale sharks) and takes place in specific areas and periods: the waters off Gabon in the Atlantic Ocean from April to September and in the Mozambique Channel in the Indian Ocean between April and May (Sequeira *et al.*, 2012; Capietto *et al.*, 2014).

In the case of dolphins, schools of large yellowfin tuna (*Thunnus albacares*) are known to associate with the pantropical spotted dolphin (*Stenella attenuata*) and the spinner dolphin (*Stenella longirostris*) in the Eastern tropical Pacific Ocean, probably to reduce the risk of predation (Perrin, 1968; Hall, 1998; Scott *et al.*, 2012). In this ocean, purse-seine vessels chased then encircled dolphin groups to catch associated tuna schools (Hall, 1998). While the tuna-dolphin association has been observed in the Atlantic and Indian Oceans (Levenez *et al.*, 1979; Ballance and Pitman, 1998), few dolphin-associated sets have been detected (0.05%, Escalle *et al.*, 2015). Indeed, in these oceans, most of the co-occurrence between purse-seine fishing operations and cetaceans (3% of all fishing sets) involved baleen whales (Bryde's whale *Balaenoptera edeni*, fin whale *Balaenoptera physalus*, sei whale *Balaenoptera borealis* and humpback whales *Megaptera novaeangliae*) and occurred mostly east of the Seychelles from December to March and in the Mozambique Channel from December to May in the Indian Ocean, and off Gabon from April to September in the Atlantic Ocean (Escalle *et al.*, 2015). It is assumed that tunas and baleen whales may form foraging associations to feed on the same prey (Romanov, 2002), such as small pelagic fishes.

At the ocean basin scale, various environmental variables such as water temperature, primary production, currents and eddies may directly influence the distribution of large marine species or indirectly affect them through influences on the distribution of their prey (Ready *et al.*, 2010; Forney *et al.*, 2012; Sequeira *et al.*, 2012; Monsarrat *et al.*, 2015). For example, as filter feeders, whale sharks can be attracted to high productivity events which result in increased zooplankton concentration (Heyman *et al.*, 2001). Similarly, the large-scale spatial distribution of cetaceans is considered to be principally influenced by water temperature but also locally by the distribution of prey species (Ballance and Pitman, 1998; Mannocci *et al.*, 2014). In particular, baleen whales are assumed to require high densities of prey to fulfill their high metabolic requirements (Piatt and Methven, 1992).

In the Atlantic and Indian Oceans, the European tropical tuna purse-seine fleet covers a large area of the pelagic ocean. Cetaceans and whale sharks sightings are reported in logbooks as potential sightings cues for tuna schools. Here a long-term dataset of logbooks was used to investigate the co-occurrence of whale sharks or cetaceans and fishing operation in relation to local environmental conditions. However, sightings are dependent on the intensity of the fishing effort, which itself undergoes seasonal variation across the fishing areas. For this reason, an ecological study of the distribution of these megafauna species is beyond the scope of this study. The aim of this work was to identify environmental variables that may be linked to the spatio-temporal co-occurrence between the tuna purse-seine fishery and whale-shark or cetaceans across seasons in the tropical Eastern Atlantic and Western Indian Oceans. Deeper knowledge of the spatio-temporal patterns of fishery/megafauna co-occurrence could allow for the further development of spatial fishery management strategies. Detailed maps could provide an essential tool for identifying areas where the co-occurrence is high and could contribute to the conservation management of these species within an ecosystem-based fishery management approach.

MATERIAL AND METHODS

Studied regions and seasons

The regions analyzed represent the main fishing grounds of European (French and Spanish) purse-seine vessels operating in the Eastern tropical Atlantic Ocean and the Western tropical Indian Ocean (see Appendix A in Data S1, for a map of seasonal variations in sighting effort). These regions are influenced by different factors. The Eastern Atlantic Ocean circulation is influenced by the Benguela and Canary Currents that generate seasonal upwellings along the coast from South Africa to Gabon between July and September (Hardman-Mountford *et al.*, 2003) and from Mauritania to Senegal between November and May (Benazzouz *et al.*, 2014). The Congo River also has a major influence on environmental conditions in the area throughout the year and especially during the wet season (April to September), with significant input of freshwater and dissolved organic matter boosting primary production (Hardman-Mountford *et al.*, 2003). In the Western Indian Ocean, water circulation is driven by monsoonal atmospheric circulation and reflects complex interactions of the seasonally alternating Somali Current with the South Equatorial Counter Current and the South Equatorial Current (Schott *et al.*, 2009). Several areas with specific circulation can be identified: (i) the northern part of the region where monsoon-generated seasonal Somalian-Arabian upwelling drastically affects

productivity, (ii) the Seychelles-Chagos thermocline ridge (55°E–65°E; 5°S–12°S) features a productive open-ocean upwelling area during the North-East monsoon (Hermes and Reason, 2008), and (iii) the Mozambique Channel which has a complex circulation influenced by mesoscale eddies (Schott *et al.*, 2009).

In this context of a high difference in hydro-climatic conditions between both oceans, the year was divided into four seasons for each ocean differently. In the Atlantic Ocean, the seasons were defined as four quarters starting in January (labeled 1–4). In the Indian Ocean, the seasons reflected monsoon [North-East (NE) from December to March and South-West (SW) from June to September] and inter-monsoon [South-West (ISW) in April–May and North-East (INE) in October–November] periods (Capietto *et al.*, 2014; Escalle *et al.*, 2015).

Fishery logbooks

The sightings and effort data were obtained from logbook records from all European tropical tuna purse-seine vessels. In both oceans, the activities of the fleets have been monitored by the Institut de Recherche pour le Développement (IRD) and the Instituto Espagñol de Oceanografía (IEO), for French and Spanish vessels, respectively (average of 35 and 59 vessels per year over the 2002–2011 period).

We considered a 10-year time series (2002–2011) of purse-seine activities, to match the available environmental dataset (see below). An 'activity' was defined as a record reported by vessel captains, which is (i) made for each fishing set, or (ii) in cases when no fishing set occurred during 1 day the main activity of the day (e.g., search for tuna or transit between fishing areas) is recorded with geographic position at noon [~1.2–2.2° (180–240 km) traveled by vessel per day[1]]. In this way, most of the activities are fishing sets and more than one activity per day can be recorded. Each record of activity included time, geographic position (only at the beginning of a fishing set), information on associations between tuna schools and baleen whales, dolphins, whale sharks, a flock of birds, or floating objects (natural log or FAD). For each fishing set, estimated weight and catch composition of targeted tuna species were also reported. Whale shark or cetacean (baleen whales or dolphins) sightings (up to 5 nautical miles) are defined as the presence of individuals (the number of individuals was not recorded) during any vessel activity (i.e., fishing set, search for tuna or transit). Separate models for dolphins could not be fit considering the relatively low number of sightings (59 and 89). Here, we thus define megafauna and purse seine fishery co-occurrence as any sighting recorded in logbooks, as we assumed that captains mainly recorded sightings indicating the presence of tuna schools or made during a fishing set. In contrast, we define an association as the assemblage of a tuna school with a megafauna species (a group or one individual) that can be formed for ecological consideration (e.g. foraging associations). Activity and sighting data were then aggregated into 1° squares (~12 300 km^2 at the equator) in R 3.0.2. (R Development Core Team, 2014).

Environmental variables

With the aim of characterizing (i) the epipelagic environment, and tentatively (ii) the distribution of whale sharks and cetaceans' main prey (macroplankton, micronekton, fish), two physical variables: Sea Surface Temperature (SST) and Eddy Kinetic Energy (EKE), one biological variable: chlorophyll-a concentration (CHL), and three bathymetric variables: depth (Depth), slope (Slope) and distance to land (Land-Dist) were considered (Table 1). While we had access to different data sources [e.g., MODIS[2] (2002–2011) and AVHRR[3] (1981–2011) for SST, and MODIS (2002–2011) and SeaWIFS[4] (1997–2007) for CHL], only a single sensor time series was selected for each variable to avoid instrument-specific biases in the data, in our case MODIS for both SST and CHL. The appropriate environmental dataset was selected for each variable to cover the largest period possible and to provide an appropriate spatial resolution (at least 1° resolution) (see Table 1). All environmental data were aggregated (i.e., averaged) at a 1° × 1° resolution to reduce missing values induced by cloud cover and to provide the equivalent spatial resolution to that of the fishery activity data.

In addition to environmental variables, a seasonal effect was introduced in the models (see below) as a categorical variable, to account for the spatio-temporal

[1]Approximate maximum distance when considering a purse seiners cruising (transiting or searching for tuna school) speed of 10–13 knots (18–24 km h^{-1}), a crew watch of ~10 h a day (daylight), and 1° at the equator (111 km).

[2]MODIS – Moderate Resolution Imaging Spectroradiometer (http://oceancolor.gsfc.nasa.gov).

[3]AVHRR – Advanced Very High Resolution Radiometer (http://www.nodc.noaa.gov/sog/pathfinder4 km/).

[4]SeaWIFS – Sea-viewing Wide Field-of-view Sensor (http://oceancolor.gsfc.nasa.gov/).

Table 1. Predictor variables used for modelling whale shark and cetacean co-occurrence with purse-seine fisheries in the Atlantic and Indian Oceans.

Variables	Abbreviation	Resolution		Sensor	Information
		Spatial	Temporal		
Sea Surface Temperature (°C)	SST	4 km	8 days	MODIS-Aqua*	SST and CHL were extracted from satellite data, covering the period 2002–2011.
Chlorophyll-a concentration ($mg.m^{-3}$)	CHL	4 km	8 days	MODIS-Aqua*	
Eddy Kinetic Energy ($m^2.s^2$)	EKE	0.33°	7 days	w-aviso 3 (MSLA†)	EKE corresponds to the Sea Level Anomaly intensity, for the 2002–2011 period. $EKE=1/2x(U^2+V^2)$
Distance to land (km)	LandDist	0.5°	–	AquaMaps dataset (Kaschner *et al.*, 2008)	Distance to the nearest shoreline.
Depth (m)	Depth	0.02°	–	NOAA-NGDC ETOPO1 Global Relief Model (Amante and Eakins, 2009)	Extracted from this 1 arc-minute global relief model of Earth's surface, which integrates land topography and ocean bathymetry, for each exact location of an activity.
Slope (%)	Slope		–		Calculated from the depth with the slope function of the 'raster' package (Hijmans *et al.*, 2014) in R 3.0.2.

*MODIS - Moderate Resolution Imaging Spectroradiometer http://oceancolor.gsfc.nasa.gov/.
†MSLA - Maps of Sea Level Anomalies & geostrophic velocity anomalies http://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/msla.html.

co-occurrence between fisheries and species. Seasons are known to highly influence the spatial distribution of whale sharks and cetaceans (Robineau, 1991; Sequeira *et al.*, 2012; Capietto *et al.*, 2014; Escalle *et al.*, 2015) and a preliminary analysis (Kruskal–Wallis tests) revealed a non-random distribution of whale shark and cetacean sightings between the different seasons. Activities, sightings and environmental variables were thus averaged per 1° square and season (quarters or monsoon periods, over the whole 2002–2011 period). While an exploratory analysis also found annual variability in the time-series, our interest was to describe average trends in oceanographic features and produce seasonally predictive maps. Therefore, we focused our analysis on the seasonal effect, which also increased computational efficiency. Thus, the candidate explanatory variables were SST, EKE, CHL, Depth, Slope, LandDist and Seasons, and the response variable was the presence/absence per 1° square for the binomial model and the number of sightings per 1° square for the count model (see next section). Correlation between environmental explanatory variables was assessed using a draftsman's plot and the *r* Pearson's correlation index. Variables were not highly correlated ($r < 0.6$), and were thus all considered in further analyzes.

Model approach

In both oceans, our study aimed to develop two separate models of the count of whale sharks or cetaceans co-occurrence with the purse-seine fishery in relation to environmental variables and seasons. Whale shark and cetacean sightings are rare events compared to the number of purse-seine fishing activities (Table 2). As a consequence, exploratory analysis highlighted some specific features inherent in the dataset which, together, directed our modelling strategy: (i) a large proportion of zeros, (ii) a non-linear relationship between the response variable and explanatory variables, (iii) spatial autocorrelation and (iv) high variability in the intensity of fishing effort with areas and seasons. Various statistical approaches can handle one or more of these data features [e.g., Zuur *et al.*, 2012 for zero-inflated data; Lin and Zhang, 1999 for Generalized Additive Mixed Model (GAMM)], but none can do so simultaneously and/or provided poor accuracy when applied to our dataset.

As a result, we used a Delta-model approach (Lo *et al.*, 1992) to deal with high numbers of zero sightings, which includes two stages: (i) modelling presence/absence in order to obtain the envelope of the predicted probability of presence of the species studied

Table 2. Main statistics of French and Spanish logbook datasets used in the binomial and count models of whale sharks and cetaceans (baleen whales and dolphins) in the Atlantic and Indian Oceans.

	Atlantic Ocean	Indian Ocean
2002–2011 Period		
Number of fishing activities* (fishing sets and search/transit)	116 386 (60 520 and 55 866)	159 091 (92 272 and 66 819)
Number of whale shark sightings	674	455
Number of cetacean sightings	1128 (1067 and 61)†	1275 (1183 and 92)†
Binomial model		
Number of fishing activities*,‡	112 867	158 828
Number of whale shark sightings	674	454
Number of grids with whale sharks present§	163	146
Number of cetacean sightings	1113 (1056 and 57)†	1275 (1183 and 92)†
Number of grids with cetaceans present§	347 (322 and 39)†	427 (397 and 81)†
Whale shark count model		
Number of fishing activities*	57 689	124 443
Number of whale shark sightings	611	440
Cetacean count model		
Number of fishing activities*	96 442	129 205
Number of cetacean sightings	1113 (1056 and 57)†	1196 (1122 and 74)†

*Activities include all fishing sets, as well as the search for tuna school or transit between fishing areas.
†The numbers in brackets are the details for baleen whales and dolphins, respectively.
‡Reduced dataset following the removal of missing values in environmental data due to cloud coverage, implemented in the models.
§One or more sightings per 1°× 1° grid cell.

(see Monsarrat *et al.*, 2015 as an example) and (ii) modelling the number of sightings (i.e., count data) of the studied species by the purse-seine fishery, only in areas where species were predicted to be present. For both stages, the candidate explanatory variables included all environmental variables, the season, and all possible interaction terms.

To deal with non-linear and non-monotonic relationships between response and explanatory variables, we used two statistical methods in both previous steps: Generalized Additive Models (GAM; Hastie and Tibshirani, 1986), and Boosted Regression Trees (BRT, Elith *et al.*, 2008). GAMs have previously been used on complex species distribution patterns (Guisan *et al.*, 2002), e.g., whale sharks, cetaceans (Afonso *et al.*, 2014; Mannocci *et al.*, 2014) using the 'mgcv' package in R (Wood, 2013). In GAM models, we used the default thin plate regression splines as the smoothing function (Wood, 2003), "ti" parameter was used for interactions terms, and we limited the smoothing to 4 degrees of freedom for each spline to avoid overfitting. BRT combines regression trees and boosting methods to fit complex non-linear relationships between predictors and the response variable, automatically handling interactions between predictors (Elith *et al.*, 2008). The 'train' function from the R package 'caret (Kuhn, 2008) uses a resampling method to evaluate the effect of model tuning parameters on performance and to select the 'optimal' model. To keep the model simpler and easier to interpret, we chose to build the model with a tree complexity of 2 and a learning rate of 0.01; the optimal number of boosting trees was assessed with the 'gbm.step' function from the 'dismo' package (Hijmans *et al.*, 2014). To compare with the GAM, results from BRT were expressed in percent (using the total percentage of deviance explained by the model) as the relative influence of each variable to the model.

To account for the spatial autocorrelation in the data, we implemented the residuals autocovariate (RAC) approach (Crase *et al.*, 2012) during both stages of the Delta-model, for both GAM and BRT. Here the spatial autocorrelation was included by adding another term to the model (the autocovariate), which represents the influence of neighbor observations on the response variable at a particular location (1° square). For each model, the RAC approach was implemented as follows: first, the model (GAM or BRT in our case) was computed with a variables backward stepwise selection procedure based on the information theory criteria ('AIC') and this procedure and the principle of parsimony were used to determine the number of interaction terms. Second, residuals from the selected model were calculated for each grid cell

and were used to compute the autocovariate by a focal calculation. This allowed cells from a selected neighbor to have a weight of 1 and all other cells a weight of 0. Finally, the residuals autocovariate was considered as an explanatory variable in the previous model (Crase *et al.*, 2012). Spatial autocorrelation was tested for each model by calculation of Moran's index and the Moran statistical test (R package 'spdep', see Bivan, 2010), which indicates a correlation between observations depending on the distance between them.

The GAM and BRT binomial RAC models based on the presence/absence of whale sharks and cetaceans were performed separately for each ocean. A binomial distribution (logit function) was applied to deal with binary data and to eliminate the possibility of negative fitted values (Zuur, 2009). Effort (number of activities) could not be accounted for in the binomial models, but only in the models of count data (see below), because its addition in the form of an offset term was not appropriate, and including it in the weight term strongly decreased the model performance and prediction. Validation of the binomial models was conducted through an internal 5-fold cross-validation in which the relationship between occurrence data and the environmental variables was modeled using a training dataset (created by a random selection of 75% of the data) and the quality of predictions was then assessed using a validation dataset (created by a random selection of 25% of the data), as advised by Fielding and Bell (1997). This calibration-validation procedure was repeated five times for each method and averaged the resulting measures of model performance. During the model validation process, a confusion matrix that records the number of true/false positive/negative cases predicted by the model was generated. From this matrix, all the performance statistics were derived (e.g., area under the curve, specificity and sensitivity (Fielding and Bell, 1997), see Appendices B–E in Data S1 for definition and value derived from the models). Model validation was performed using the 'PresenceAbsence' package (Freeman, 2012). With the aim of calculating the envelope of the presence of whale shark or cetaceans, the conversion of the predicted probability of presence into a binary presence/absence response was done for a value above a threshold derived from the cross-validation (see Appendices B–E in Data S1).

Once the envelope of presence was obtained, different distributions were tested to model the number of sightings in areas where the species/group was predicted to be present (negative-binomial, quasi Poisson and Poisson). To account for the non-constant sighting effort between 1° square, the number of vessel activities was implemented as an offset in the count model (Kotze *et al.*, 2012). For whale shark and cetaceans, separately, the best model between classic GAM and BRT, and GAM and BRT RAC models was then selected based on the above-mentioned prediction performance of binomial models, as well as residuals analysis, a plot of predicted versus observed values, and deviance explained for both binomial and count models (see Appendices B–E in Data S1). The quasi-Poisson GAM and the Poisson BRT were the distributions that provided the most accurate results for our datasets. Note that an averaging ensemble approach (Araújo and New, 2007), based on the combined predictions of the GAM and BRT, was also tested. However, in each case, only one model (i.e., GAM or BRT) provided the best performances, whereas averaging the GAM and BRT predictions weakened them. For this reason, the averaging approach was not adopted. Predictions by ocean per season from GAM and BRT models were projected onto a 1° × 1° grid, limited to areas with reported fishing activities by the European purse-seine fleet.

RESULTS

Between 2002 and 2011, totals of 112 867 and 158 8281 fishery activities (110 452 and 132 682 fishing days) were recorded in the Atlantic and Indian Oceans, respectively[5] (see Table 2 for original statistics of logbook data), and over 1093 and 1194 1° square grid cells were sampled. From this selected datasets, 674 and 454 whale shark sightings; and 1113 and 1275 cetacean sightings were recorded in the Atlantic and Indian Oceans, respectively (Table 2). For whale sharks, the best binomial models were BRT RAC in the Atlantic Ocean with all explanatory variables retained, and GAM RAC in the Indian Ocean with only Depth, EKE, CHL, Season and the interaction between CHL and EKE (Table 3, Appendices B and C in Data S1). For cetaceans, the best binomial models were GAM RAC in the Atlantic, with all explanatory variables significant, and BRT RAC in the Indian Ocean with all explanatory variables and the interaction between Depth and Slope retained (Table 3, Appendices D and E in Data S1).

For whale sharks and for cetaceans, the best model to fit the number of sightings data was the BRT poisson RAC model (see Appendices B–E in Data S1 for details). The percentage of deviance explained, for

[5]After the removal of 3.02% and 0.17% of activities as a result of missing values in the environmental dataset due to the cloud cover.

Table 3. Binomial and count models used for whale sharks and cetaceans in the Atlantic (AO) and Indian Oceans (IO). Numbers in brackets indicate the relative contribution of each explanatory variable (for BRT only).

	Explanatory variables	% De	R^2	Pearson r	Moran's I	Moran P-value
Whale shark AO						
BRT binomial RAC	CHL (5.4), SST (4.4), LandDist (2.4), Depth (1.6), EKE (1.4), Season (1.0), Slope (0.2), Autocovariate (19.0)	35.5 (16.5)		0.51	0.06	$<2.2e^{-16}$
BRT poisson RAC	CHL (5.5), LandDist (5.1), Slope (5.1), SST (3.5), Depth (3.5), Season (0.4), EKE (0.1), Autocovariate (59.7)	83.0 (23.3)		0.94	0.02	0.06
Whale shark IO						
GAM binomial RAC	Depth, EKE, CHL, Season, Autocovariate, ti (CHL, EKE)	12.3	0.06	0.26	0.12	$<2.2e^{-16}$
BRT poisson RAC	Season (3.2), Depth (3.0), EKE (2.0), SST (1.1), CHL (0.5), Slope (0.1), Autocovariate (60.2)	70.1 (9.9)		0.54	0.03	$0.10e^{-03}$
Cetacean AO						
GAM binomial RAC	Depth, SST, CHL, EKE, Slope, LandDist, Season, Autocovariate, ti(Depth, Slope)	21.1	0.16	0.41	0.15	$1.2e^{-10}$
BRT poisson RAC	SST (7.6), CHL (6.6), Slope (3.7), Depth (1.9), Season (0.6), EKE (0.6), LandDist (0.5), Autocovariate (50.6)	72.1 (21.5)		0.92	0.02	$0.10e^{-02}$
Cetacean IO						
BRT binomial RAC	CHL (5.5), SST (3.2), EKE (2.1), LandDist (3.8), Depth (1.8), Slope (0.5), Season (0.2), Autocovariate (15.0)	30.2 (15.2)		0.53	0.15	$2.0e^{-16}$
BRT poisson RAC	Depth (21.2), Season (4.3), Slope (3.0), SST (3.0), CHL (0.8), EKE (0.5), LandDist (0.3), Autocovariate (35.7)	68.2 (32.5)		0.77	0.04	0.28

% De = total percentage of the deviance explained by each model (in brackets are the percentages of deviance explained by environmental and season variables); R^2 = pseudo coefficient of determination; r = Pearson's correlation index between observed and predicted values; I = Moran's index maximum absolute value for each model and associated P-value.

each group of megafauna, ranged from 9.9% to 32.5% (Table 3, Appendices B–E in Data S1) when the season and environmental variables were considered (68.2–83.0% when the deviance from the autocovariate term is also considered). Pearson's correlation index between observed and predicted values was always higher than 0.54 (Table 3, Appendices B–E in Data S1). An absence (whale shark in the Atlantic Ocean and cetaceans in the Indian Ocean) or weak (for whale shark in the Indian Ocean and cetaceans in the Atlantic Ocean; P-value <0.05 but Moran's index of autocorrelation <0.02; see Appendices C and D in Data S1) spatial autocorrelation was detected, highlighting that it was accurately handled by the RAC method.

For whale sharks in the Atlantic Ocean, the model explained 23.3% of the total deviance (83.0% with the autocovariate term), and the most significant variables were CHL, LandDist and Slope (5.5%, 5.1% and 5.1% of relative influence, respectively; Table 3). In the Indian Ocean, 9.9% of deviance was explained by environmental and season variables (70.1% with the autocovariate term, Table 3), with the most significant variables being Season, Depth, EKE and SST (3.2%, 3.0%, 2.0% and 1.1% of the total deviance, respectively; Table 3).

For cetaceans in the Atlantic Ocean, the total deviance explained was 21.5% (a total of 72.1% with the autocovariate term). SST (7.6%), CHL (6.6%), Slope (3.7%) and Depth (1.9%) were the most significant variables (Table 3). In the Indian Ocean, 32.5% of the deviance of the model was explained by season and environmental variables (68.2% due to the autocovariate term). In this model, depth accounted for the vast majority of the relative influence of each variable (21.2%). Nonetheless, season (4.3%), slope (3.1%) and SST (3.0%) still represented a significant influence in the model.

Results showed several hotspots of whale shark and purse-seine fishery co-occurrence (Fig. 1a), which highly varied with seasons (Fig. 2): (i) in the Atlantic Ocean, in the eastern part of the Gulf of Guinea, especially along the coast from Gabon to Angola during July–September, (ii) in the Indian Ocean, around the Seychelles during the NE monsoon, (iii) in the Mozambique Channel during ISW monsoon, and (iv) in the area north of the Seychelles and off Somalia EEZ (around 5°N–55°E) during the SW monsoon period.

For cetaceans, the predicted number of sightings showed wider dispersal, especially in the Indian Ocean (Fig. 1b), and high variations after seasonal and monsoonal patterns were also detected (Fig. 3). In the Atlantic Ocean, cetacean/fishery co-occurrence varied from (i) the north of the Gulf of Guinea, above the equator, from 20°W to 5°E during season 1 and 4, (ii) north-east of the Gulf of Guinea, especially in the waters off Gabon during season 2 and 3, and (iii) in the waters off Mauritania to Senegal during season 2 (Fig. 3). In the Indian Ocean, prediction of cetacean sightings by the purse seine fishery ranged from (i) around the Seychelles (10°S–0°N and 50°E–75°E) during the NE monsoon, (ii) the north of the Mozambique Channel during the ISW monsoon and (iii) an area around and to the north of the Seychelles (10°S–8°N and 40°E–65°E) during the SW and the INE monsoon periods (Fig. 3).

DISCUSSION

Using a long-term tropical tuna purse-seine logbook dataset, we investigated links between the co-occurrence of purse-seine fisheries with whale sharks and with cetaceans in relation to environmental variables in the tropical Eastern Atlantic and Western Indian Oceans. By implementing quantitative models, which included environmental variables, our study is complementary to previous works that roughly identified spatio-temporal variations of co-occurrence of fisheries and megafauna (Capietto *et al.*, 2014; Escalle *et al.*, 2015). By incorporating a spatial correlation contiguity matrix as an explanatory variable in GAM and BRT using the residuals autocovariate approach, we accounted for the autocorrelated nature of the data. This methodology, which was recently developed by Crase *et al.*, 2012, showed good predictive results for both megafauna groups studied. In particular, it allows the spatial correlation, present in our dataset, to be

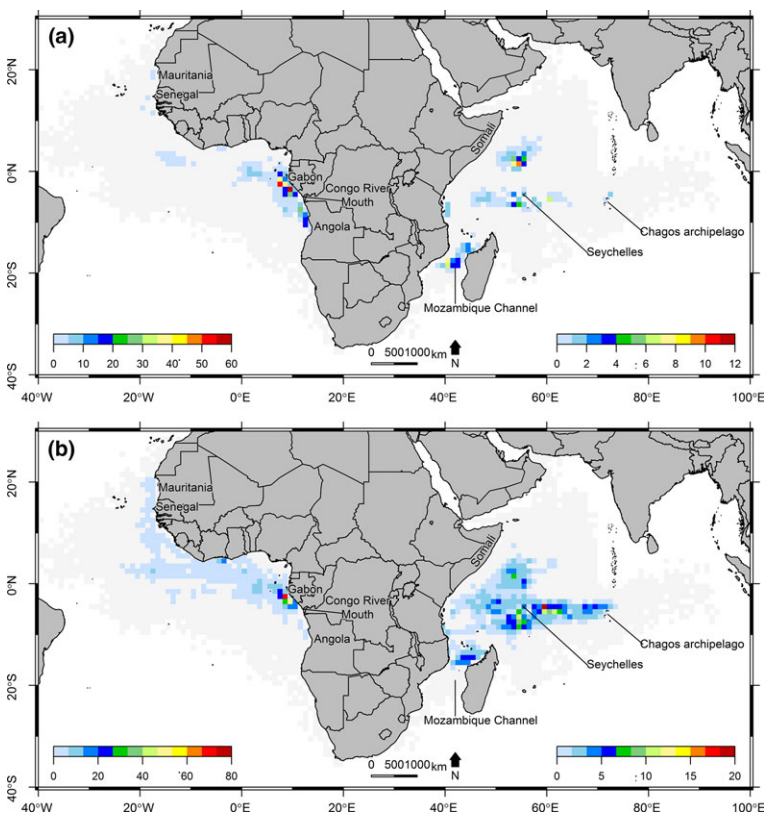
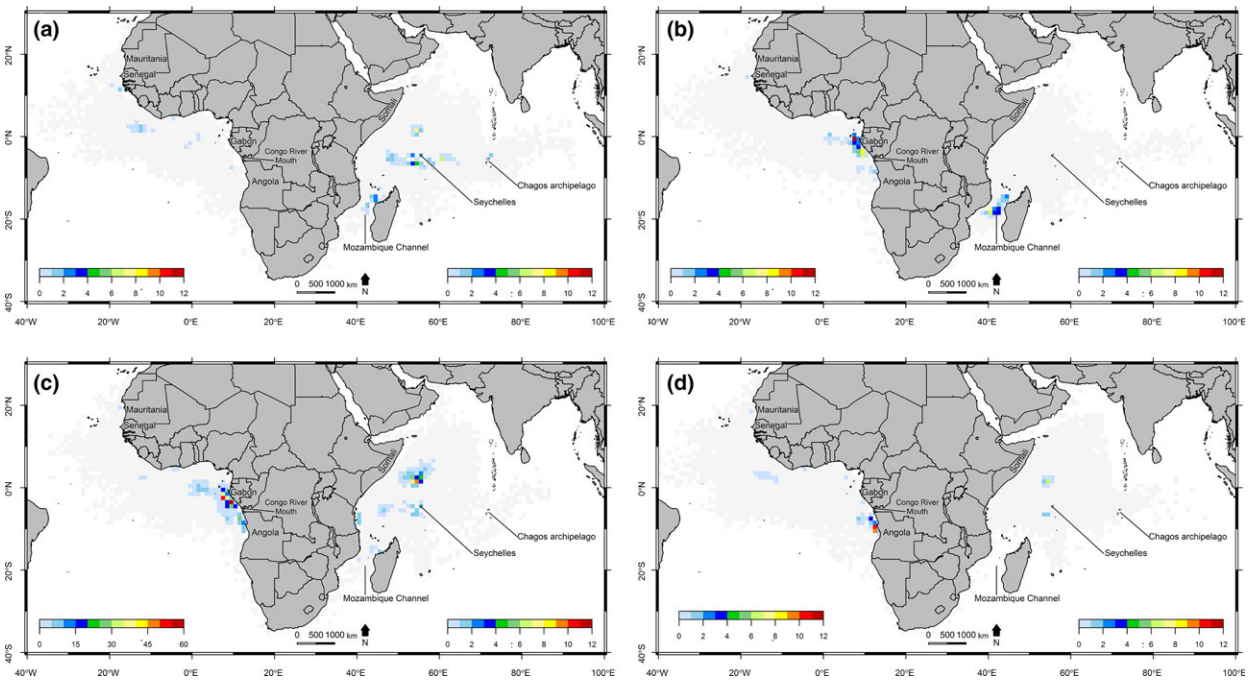


Figure 1. Prediction of co-occurrence between (a) whale sharks and (b) cetaceans and purse-seine fisheries in the tropical Eastern Atlantic and Western Indian Oceans, derived from Boosted Regression Trees (BRT) Poisson residuals autocovariate (RAC) models. Maximum values of (a) 60 and 12 whale sharks, and (b) 80 and 20 cetaceans, co-occurrence with the purse seine fishery per 1° square, in the Atlantic and Indian Oceans, respectively.

Figure 2. Seasonal distribution of the predicted co-occurrence between whale sharks and purse-seine fisheries in the tropical Eastern Atlantic and Western Indian Oceans, derived from Boosted Regression Trees (BRT) Poisson residuals autocovariate (RAC) models. (a) season 1 (Atlantic Ocean) and NE monsoon (Indian Ocean), (b) season 2 and ISW monsoon, (c) season 3 and SW monsoon, and (d) season 4 and INE monsoon. Maximum values for the legend scale vary depending on the ocean and the season (12, 75 or 20 whale sharks/purse-seine fisheries co-occurrence per 1° square).



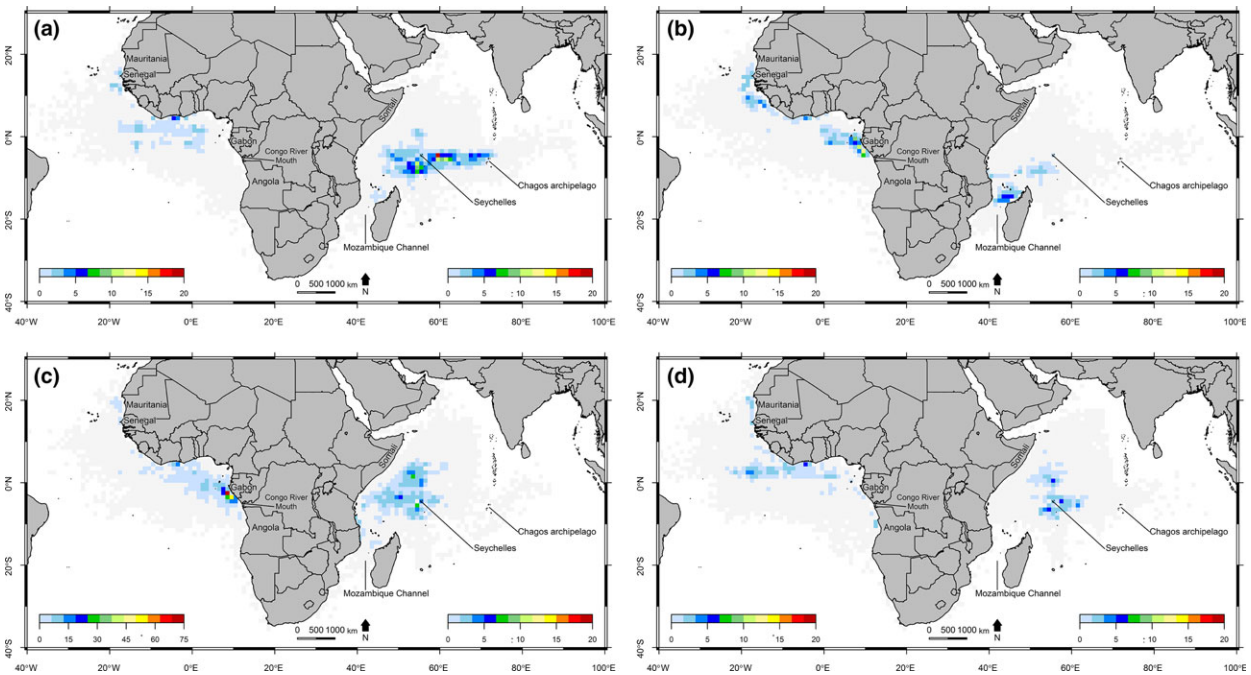
dealt with and can be applied to rare species from fisheries-dependent data (i.e., containing large proportions of zeros with varying intensity in the observational effort), which cover large spatial scales. It should be stressed that the relatively low contribution of each environmental explanatory variable to the total deviance of the model (<1% to 21.2%) is similar or higher than those reported in other modeling studies on the distribution of marine megafauna (e.g., Forney *et al.*, 2012; Mannocci *et al.*, 2014 for cetaceans, and Afonso *et al.*, 2014 for whale sharks). It should be mentioned that the logbook dataset used here contains inherent limitations, such as underreporting or missed sightings by captains (e.g., when megafauna species are not associated with a tuna school). While captains recorded sightings of either baleen whales or dolphins, we used the same model for a broad group of cetacean species (both baleen whales and dolphins, which was not discussed further as a result of the low number of dolphin sightings), each with a different ecology and biology. Furthermore, scientific observer data were not used in complement to logbook data as (i) many sightings would have been counted twice, (ii) it was not considered appropriate to have data from two different sources covering the same fishing trips (i.e., when observers are onboard, captains still fill logbooks) and (iii) scientific observers data have relatively low and irregular coverage rate (in time and space). Also, while in the Pacific Ocean it has been shown that the tuna-dolphin association is promoted by a shallow mixed layer and hypoxic oxygen minimum zone (Scott *et al.*, 2012), these variables were not included here as the tuna–dolphin association is encountered less frequently.

Whale sharks

The chlorophyll-a concentration explained a relevant portion of the deviance of the whale shark model in the Atlantic Ocean. Areas with high whale shark and fishery co-occurrence (i.e., the waters off Gabon to Angola, Figures 1a and 2) matched areas with high chlorophyll-a concentrations[6] (Fig. 4a, Appendix F in Data S1), especially from April to September. This might reflect a larger regional phenomenon occurring

[6]Very high chlorophyll-a concentration is also detected year-round in the latitudinal band from 24–27°S (Fig. 4a). This represents the Benguela Current area, with high primary production, but very low overlap with purse-seine fishing activities.

Figure 3. Seasonal distribution of the predicted co-occurrence between cetaceans and purse-seine fisheries in the tropical Eastern Atlantic and Western Indian Oceans, derived from Boosted Regression Trees (BRT) Poisson residuals autocovariate (RAC) models. (a) season 1 (Atlantic Ocean) and NE monsoon (Indian Ocean), (b) season 2 and ISW monsoon, (c) season 3 and SW monsoon (maximum value of 60 cetaceans/purse-seine fisheries co-occurrence per 1° square in the Atlantic Ocean), and (d) season 4 and INE monsoon.



in the Gulf of Guinea during this period linked to the principal rainy season in this area. A major discharge of fresh water, dissolved organic matter and floating objects from rivers into the oceanic system (specifically the Congo River) occurs during this period and peaks from April to June, with the emergence of a coastal upwelling system (Van Bennekom and Berger, 1984). Such a productive area may explain the presence of whale sharks and tunas, by supporting high densities of prey. In the Western Indian Ocean, an opposite result was obtained. The concentration of chlorophyll-a was not a significant environmental variable in the model (Table 3). One explanation for this result could be that, in this ocean, values of chlorophyll-a concentration are lower than in the Atlantic (average of 0.5 ± 1.4 and 0.2 ± 0.4 mg m^{-3} in the Atlantic and Indian Oceans, respectively; Appendix F in Data S1). This might be because of very high concentrations of chlorophyll-a recorded in the Atlantic Ocean in areas influenced by the Congo River discharge. While seasonal peaks in productivity in the Indian Ocean, such as in the Somalia-Arabian upwelling, are higher than permanent upwellings in the Eastern Atlantic Ocean (McCreary *et al.*, 2013), they do not feature in our model, as few sightings were recorded directly in this area. However, in the Indian Ocean two small regions showing high whale sharks and fishery co-occurrence were also in very productive waters, as a result of the development of a seasonal upwelling (i) in the Seychelles–Chagos thermocline ridge during the NE monsoon (Hermes and Reason, 2008), and (ii) on the periphery of the Somalian–Arabian upwelling during the NE monsoon. Thus, it can be concluded that in both oceans, whale sharks occurred in highly productive environments. Such spatial co-occurrence was expected considering that the whale shark is a filter feeder specifically adapted to feed on prey in high densities (Rowat and Brooks, 2012).

Eddy kinetic energy showed the exact inverse of chlorophyll-a, explaining a significant part of the model deviance in the Indian Ocean but not in the Atlantic Ocean. This may be a result of the high values of this parameter in the Mozambique Channel (Appendix F in Data S1), related to the presence of mesoscale eddies (Tew-Kai and Marsac, 2010). It is known that the edge of eddies provides high biological production that supports large aggregations of micronekton, which in turn attracts top predators, such as seabirds and tunas (Tew-Kai and Eand Marsac, 2010).

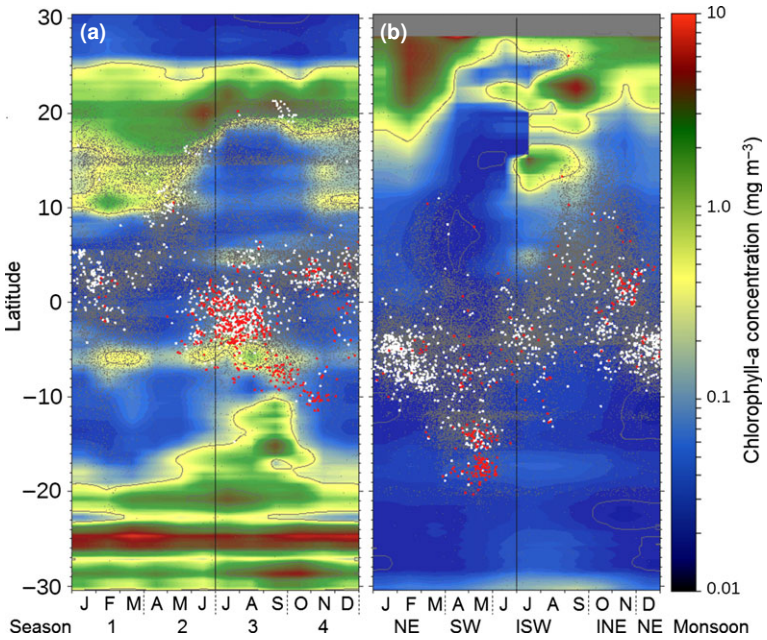


Figure 4. Chlorophyll-a concentration and latitudinal and seasonal locations of whale shark (red dots) and cetacean (white dots) sightings from 2002 to 2011, for (a) the Eastern tropical Atlantic Ocean and (b) the Western tropical Indian Ocean. Grey dots represent purse-seiner activities. Background colors represent the seasonality of chlorophyll-a concentration averaged over the envelope of longitudes of purse-seine activities (see Fig. 1).

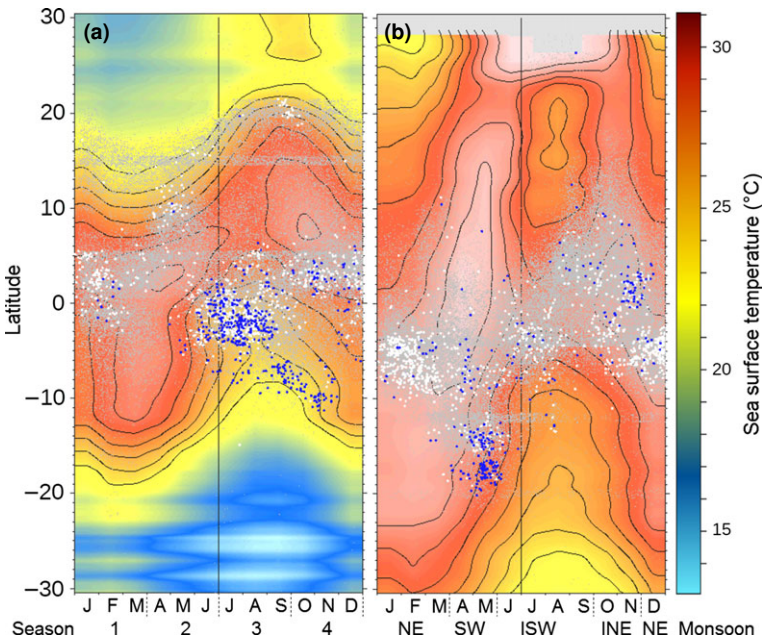


Figure 5. Sea Surface Temperature and latitudinal and seasonal locations of whale shark (blue dots) and cetacean (white dots) sightings from 2002 to 2011, for (a) the Eastern tropical Atlantic Ocean and (b) the Western tropical Indian Ocean. Grey dots represent purse-seiner activities. The background colors represent the seasonality of Sea Surface Temperature averaged over the envelope of longitudes of purse-seine activities (see Fig. 1).

While, in both oceans SST also played a significant role in the models, the range of observed temperature varied in each ocean. In the Indian Ocean, whale shark sightings mostly occurred between 27.5 and 29.5°C (Fig. 5b) as previously described by Sequeira *et al.* (2012), whereas in the Atlantic Ocean, most whale shark sightings were recorded between 22.6 and 26.5°C (Fig. 5a). In the waters off Gabon to Angola, SST was the lowest during season 2 and 3, linked to the Congo River discharge (Van Bennekom and Berger, 1984) and corresponding to the peak in the presence of whale sharks. In general, whale sharks are considered to occur mostly in warm temperatures (Rowat and Brooks, 2012; Sequeira *et al.*, 2012), but here their highest co-occurrence with the purse-seine fishery was observed during periods when the SST values were lower than their reported preferred temperature range. This pattern may indicate that individual

whale sharks could have a feeding advantage by being present in very productive areas rather than restricting their movements to warmer but less productive areas.

Cetaceans

In both oceans, cetacean sightings recorded by purse-seine captains were mainly of large baleen whales (94%), which include four species: Bryde's whale, fin whale, sei whale and humpback whale (based on observation from scientific observers). The number of dolphin sightings was too low to be further discussed.

In the Indian Ocean, depth explained a large part of the deviance in the model (21.2%). For instance, areas with higher predicted co-occurrence (e.g., around the Seychelles or the north of the Mozambique Channel) were part of the continental margin, in a depth range between 500 and 3000 m. In the Atlantic Ocean, cetaceans and fishery co-occurrence were also slightly linked to depth (e.g., the waters off Gabon to Angola and Mauritania to Senegal). Consequently, such environments may provide conditions that are suitable for both cetaceans and tuna, as was observed for whale sharks. Oceanic waters may be the preferred habitat for both species groups, and shallower areas may have the advantage of receiving terrestrial inputs, which boost productivity. Consequently, areas close to continental shelves, where coastal and oceanic habitats converge, may contain high densities of prey for both species groups.

In areas and seasons that showed high cetacean co-occurrence with tuna purse-seiners in the Atlantic Ocean, chlorophyll-a concentrations were also very high (Figures 3 and 4a; Appendix F in Data S1). Similarly to whale sharks, this variable was not significant at the scale of the entire Indian Ocean (Fig. 4b), but areas with high fishery/cetacean co-occurrence also arise in productive areas. For SST, results were again similar to the whale sharks models. In the Atlantic Ocean the co-occurrence of cetaceans and the purse-seine fishery primarily predicted across a range of moderate SST values (22.0–27.0°C), whereas in the Indian Ocean, it was by SSTs between 28.0–29.0°C (Figures 3 and 5; Appendix F in Data S1).

Similarly to the whale shark models, the season was the second most important variable in the Indian Ocean, whereas it was of marginal relevance in the Atlantic Ocean. Cetaceans/fishery co-occurrence was highly prevalent during two specific seasons in the Atlantic, whereas in the Indian Ocean, patterns of co-occurrence (Fig. 3) showed a closer reflection of the seasonal changes in purse-seine fishing grounds (Appendix A in Data S1) throughout the year. The inter-ocean difference in seasonal influence may thus be explained by different spatio-temporal fishing patterns. It is worth mentioning that cetaceans probably have a larger spatial distribution across the Atlantic and Indian Oceans than represented here, but our study was based on the co-occurrence with the surface tuna fisheries, and results are thus limited by their spatial extent.

Previous studies performed in waters along the coastline from Gabon to Angola found that Bryde's, humpback (the two most abundant cetacean species), fin and sei whales were present in this area from April to September (Weir, 2007; de Boer, 2010). Bryde's whales inhabit Gabonese waters from May to July, before migrating to South Africa (Weir, 2007; de Boer, 2010). Also, its foraging behavior on schools of small pelagic fish, such as *Sardinella* spp., close to the surface has been observed in Gabon (de Boer, 2010). Humpback whales are also present from May to October, but primarily in shallow waters (<200 m) as they use the area as calving and mating grounds during winter, after feeding in the Southern Ocean (Weir, 2007; de Boer, 2010). The two other species of baleen whales, fin and sei whales, are also believed to undertake similar seasonal migrations between winter breeding grounds and a summer feeding area in the Antarctic (Weir, 2007). However, sei whale was considered the only species performing these seasonal migration and also feeding in winter in Gabonese waters (Budker and Collignon, 1952). All four baleen whale species also occur in the Indian Ocean and have been reported in the Seychelles and the Mozambique Channel (Robineau, 1991; Mannocci *et al.*, 2014).

Baleen whales and surface tuna schools may form foraging associations to feed on the same prey species (Romanov, 2002), a behaviour sometimes used by purse-seine fishers to improve their fishing success. This strategy may thus explain the high co-occurrence between baleen whales and purse-seine fisheries observed in areas that support high densities of prey common to both baleen whales and tunas. Tunas in surface aggregations feed primarily on epipelagic fishes, crustaceans and small cephalopods, such as the oceanic light fish (*Vinciguerria nimbaria*) or the bigeye cigarfish (*Cubiceps pauciradiatus*) (Romanov, 2002; Potier *et al.*, 2004) and Bryde's whale, as mentioned above, also forage on pelagic schooling fish (de Boer, 2010). Areas of high abundance of bigeye cigarfish in the Western Indian Ocean shown by Potier *et al.* (2008) clearly overlap baleen whales hotspots identified in the present study. Additionally, similar distributions between whale-associated sets and bigeye cigarfish abundance estimated during trawl surveys have been

found during the NE and IWS monsoon periods along the Seychelles–Chagos thermocline ridge (Romanov, unpublished data). This confirms the high co-occurrence of baleen whales and the tuna fishery in this productive environment and further reinforces previous assumptions of baleen whales and tuna feeding associations on shared prey species.

CONCLUSION

As a result of the immensity of the open ocean environment, conducting large-scale scientific observation programs on all fishing vessels (i.e., 100% coverage) dedicated to the study of the co-occurrence between fisheries and megafauna species remains difficult. Fisheries-related data have inherent limitations, such as non-reporting or underreporting of interactions, and imprecise species identification. The use of a long-term dataset obtained from commercial fisheries, such as the logbooks from tropical tuna purse-seine vessels, is however of major interest for identifying seasons and regions of co-occurrence between fisheries and megafauna, and to explore the environmental conditions linked to this co-occurrence.

In summary, by implementing quantitative models including environmental variables, our study highlighted that high co-occurrence of both whale sharks and cetaceans with purse-seine fisheries were in productive areas during particular seasons. Other environmental variables, reflecting the continental shelves (depth, slope and distance to land) or eddy kinetic energy, also highlighted the link with productive environments. Whale sharks and cetaceans co-occurrence with the fishery were also associated with SST, with both groups detected in lower SST ranges in the Atlantic Ocean than in the Indian Ocean. Finally, our results suggest that seasonal variability could have different effects in the Atlantic or the Indian Oceans on the distribution of megafauna/fishery co-occurrence, with greater seasonal effects in our models in the Indian Ocean. The identification of areas and seasons linked to a particular environmental condition with high co-occurrence of purse-seine fisheries and both whale sharks and cetaceans, described here, could facilitate conservation management measures for these species, such as the ban of intentional encirclements in these specific areas and seasons.

ACKNOWLEDGEMENTS

We thank an anonymous reviewer for his constructive comments on an earlier version of the manuscript. We are grateful to the skippers and fishing companies involved in the logbook data collection (supported by IRD and IEO) and the teams of the 'Observatoire Thonier' (IRD) and the 'Centro Costero de Canarias' (IEO) for provided logbook data. The altimeter products were produced by Ssalto/Duacs and distributed by Aviso with support from CNES. L. Escalle is funded by a PhD grant from the University of Montpellier, and M.G. Pennino by the MORSE project (CEP&S 2011-Project ANR-11-CEPL-006, France). J.D. Filmalter has edited the English of the paper.

REFERENCES

Afonso, P., McGinty, N. and Machete, M. (2014) Dynamics of whale shark occurrence at their fringe oceanic habitat. *PLoS One* **9:**e102060.

Amante, C. and Eakins, B.W. (2009) *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis*. NOAA, National Geophysical Data Center, Marine Geology and Geophysics Division, Boulder, Colorado. http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/docs/ETOPO1.pdf [accessed 13 May 2016].

Araújo, M.B. and New, M. (2007) Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22:**42–47.

Ballance, L.T. and Pitman, R.L. (1998) Cetaceans of the western tropical Indian Ocean: distribution, relative abundance, and comparisons with cetacean communities of two other tropical ecosystems. *Mar. Mammal Sci.* **14:**429–459.

Benazzouz, A., Mordane, S., Orbi, A. *et al.* (2014) An improved coastal upwelling index from sea surface temperature using satellite-based approach – The case of the Canary Current upwelling system. *Cont. Shelf Res.* **81:**38–54.

Bivan, R. (2010) Package "spdep" – CRAN. Spatial Dependence: Weighting Schemes, Statistics and Models, Version 0.5-77. http://cran.r-project.org/web/packages/spdep [accessed 13 May 2016].

de Boer, M.N. (2010) Cetacean distribution and relative abundance in offshore Gabonese waters. *J. Mar. Biol. Assoc. U. K.* **90:**1613–1621.

Budker, P., Collignon, J. (1952) Trois campagnes baleinières au Gabon (1949-1950-1951). *Bulletin de l'Institut d'Etudes Centrafricaines* **3:**75–100.

Capietto, A., Escalle, L., Chavance, P. *et al.* (2014) Mortality of marine megafauna induced by fisheries: insights from the whale shark, the world's largest fish. *Biol. Conserv.* **174:**147–151.

Crase, B., Liedloff, A.C. and Wintle, B.A. (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* **35:**879–888.

Dagorn, L., Holland, K.N., Restrepo, V. and Moreno, G. (2013) Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? *Fish Fish.* **14:**391–415.

R Development Core Team. (2014) R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Elith, J., Leathwick, J.R. and Hastie, T. (2008) A working guide to boosted regression trees. *J. Anim. Ecol.* **77:**802–813.

Escalle, L., Capietto, A., Chavance, P. *et al.* (2015) Cetaceans and tuna purse seine fisheries in the Atlantic and Indian

Oceans: interactions but few mortalities. *Mar. Ecol. Prog. Ser.* **522:**255–268.

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

Forney, K.A., Ferguson, M.C., Becker, E.A. *et al.* (2012) Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endanger Species Res.* **16:**113–133.

Freeman, E. (2012) *Package "PresenceAbsence" - CRAN. Presence-Absence Model Evaluation*. Version 1.1.9. http://cran.r-project.org/web/packages/PresenceAbsence [accessed 13 May 2016].

Gaertner, D. and Medina-Gaertner, M. (1999) An overview of the tuna fishery in the southern Caribbean Sea. In: Proceedings of the International Workshop on Fishing For Tunas Associated with Floating Objects, La Jolla, CA, February 11–13, 1992. Inter–Am. Trop. Tuna Comm. Spec. Rep. **11:**66–86.

Guisan, A., Edwards, T.C. Jr and Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* **157:**89–100.

Hall, M.A. (1998) An ecological view of the tuna-dolphin problem: impacts and trade-offs. *Rev. Fish Biol. Fish* **8:**1–34.

Hardman-Mountford, N.J., Richardson, A.J., Agenbag, J.J. *et al.* (2003) Ocean climate of the South East Atlantic observed from satellite data and wind models. *Prog. Oceanogr.* **59:**181–221.

Hastie, T. and Tibshirani, R. (1986) Generalized additive models. *Stat. Sci.* **1:**297–310.

Hermes, J.C. and Reason, C.J.C. (2008) Annual cycle of the South Indian Ocean (Seychelles-Chagos) thermocline ridge in a regional ocean model. J. Geophys. Res. Oceans 1978–2012 113(C4). http://onlinelibrary.wiley.com/doi/10.1029/2007JC004363/full [accessed 13 May 2016].

Heyman, W.D., Graham, R.T., Kjerfve, B. and Johannes, R.E. (2001) Whale sharks *Rhincodon typus* aggregate to feed on fish spawn in Belize. *Mar. Ecol. Prog. Ser.* **215:**275–282.

Hijmans, R.J., Phillips, S.J., Leathwick, J.R. and Elith, J. (2014) *Package "dismo" – CRAN*. Species Distribution Modeling. Version 1.0-5. http://cran.r-project.org/web/packages/dismo/ [accessed 13 May 2016].

Kaschner, K., Ready, J.S., Agbayani, E. *et al.* (2008) *AquaMaps Environmental Dataset: Half-Degree Cells Authority File (HCAF)*. Version 07/2010. http://www.aquamaps.org/download/main.php [accessed 13 May 2016].

Kotze, D.J., O'Hara, R.B. and Lehvävirta, S. (2012) Dealing with varying detection probability, unequal sample sizes and clumped distributions in count data. *PLoS One* **7:**e40923.

Kuhn, M. (2008) Building predictive models in R using the caret package. *J. Stat. Softw.* **28:**1–26.

Levenez, J.-J., Fonteneau, A. and Regalado, R. (1979) Resultats d'une enquete sur l'importance des dauphins dans la pecherie thoniere FISM. *Collec. Vol. Sci. Papers ICCAT* **9:**176–179.

Lin, X.H. and Zhang, D.W. (1999) Inference in generalized additive mixed models by using smoothing splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61:**381–400.

Lo, N.C., Jacobson, L.D. and Squire, J.L. (1992) Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fish Aquat. Sci.* **49:**2515–2526.

Mannocci, L., Laran, S., Monestiez, P. *et al.* (2014) Predicting top predator habitats in the Southwest Indian Ocean. *Ecography* **37:**261–278.

McCreary, J.P., Yu, Z. and Hood, R.R. (2013) Dynamics of the Indian-Ocean oxygen minimum zones. *Prog. Oceanogr.* **112–113:**15–37.

Monsarrat, S., Pennino, M.G., Smith, T.D. *et al.* (2015) Historical summer distribution of the endangered North Atlantic right whale (*Eubalaena glacialis*): a hypothesis based on environmental preferences of a congeneric species. *Divers. Distrib.* **21:**925–937.

Perrin, W.F. (1968) The porpoise and the tuna. *Sea Front.* **14:**166–174.

Piatt, J.F. and Methven, D.A. (1992) Threshold foraging behavior of baleen whales. *Mar. Ecol. Prog. Ser. Oldendorf.* **84:**205–210.

Potier, M., Marsac, F., Lucas, V., Sabatié, R., Hallier, J.-P. and Ménard, F. (2004) Feeding partitioning among tuna taken in surface and mid-water layers: the case of yellowfin (*Thunnus albacares*) and bigeye (*T. obesus*) in the western tropical Indian Ocean. *West. Indian Ocean J. Mar. Sci.* **3:**51–62.

Potier, M., Romanov, E., Cherel, Y., Sabatié, R., Zamorov, V. and Ménard, F. (2008) Spatial distribution of *Cubiceps pauciradiatus* (Perciformes: Nomeidae) in the tropical Indian Ocean and its importance in the diet of large pelagic fishes. *Aquat. Living Resour.* **21:**123–134.

Ready, J., Kaschner, K., South, A.B. *et al.* (2010) Predicting the distributions of marine organisms at the global scale. *Ecol. Model.* **221:**467–478.

Robineau, D. 1991. Balaenopterid sightings in the western tropical Indian Ocean (Seychelles area), 1982–1986. In: The Indian Ocean Sanctuary Marine Mammal Technical Report No. 3. S. Teatherwood & G.P. Donovan (eds). Nairobi: UNEP, pp. 171–178.

Romanov, E.V. (2002) Bycatch in the tuna purse-seine fisheries of the western Indian Ocean. *Fish. Bull.* **100:**90–105.

Rowat, D. and Brooks, K.S. (2012) A review of the biology, fisheries and conservation of the whale shark *Rhincodon typus*. *J. Fish Biol.* **80:**1019–1056.

Schott, F.A., Xie, S.-P. and McCreary, J.P. (2009) Indian Ocean circulation and climate variability. *Rev. Geophys.* **47:**RG1002.

Scott, M., Chivers, S., Olson, R., Fiedler, P. and Holland, K. (2012) Pelagic predator associations: tuna and dolphins in the eastern tropical Pacific Ocean. *Mar. Ecol. Prog. Ser.* **458:**283–302.

Sequeira, A., Mellin, C., Rowat, D., Meekan, M.G. and Bradshaw, C.J.A. (2012) Ocean-scale prediction of whale shark distribution. *Divers. Distrib.* **18:**504–518.

Tew Kai, E. and Marsac, F. (2010) Influence of mesoscale eddies on spatial structuring of top predators' communities in the Mozambique Channel. *Prog. Oceanogr.* **86:**214–223.

Van Bennekom, A.J. and Berger, G.W. (1984) Hydrography and silica budget of the Angola Basin. *Neth. J. Sea Res.* **17:**149–200.

Weir, C.R. (2007) Occurrence and distribution of cetaceans off northern Angola, 2004/05. *J. Cetacean Res. Manag.* **9:**225–239.

Wood, S.N. (2003) Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65:**95–114.

Wood, S.N. (2013) *Package "mgcv" – CRAN. Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*, Version 1.7-27. http://cran.r-project.org/web/packages/mgcv [accessed 13 May 2016].

Zuur, A.F. (2009) *Mixed Effects Models and Extensions in Ecology With R*. New York: Springer-Verlag.http://books.google.fr/books?id=vQUNprFZKHsC.

Zuur, A.F., Savaliev, A.A. and Ieno, E.N. (2012) *Zero Inflated Models and Generalized Linear Mixed Models with R. Highland Statistics Ltd.* Highland Statistics Ltd.: United Kingdom, 336 pp.

SUPPORTING INFORMATION

Data S1. Seasonal distribution of purse-seine sighting effort (Appendix A), model calibration per megafauna group and ocean (Appendix B, C, D and E), and yearly and seasonal distribution of the environmental variables used in our study (Appendix F).

Appendix A. Sighting effort

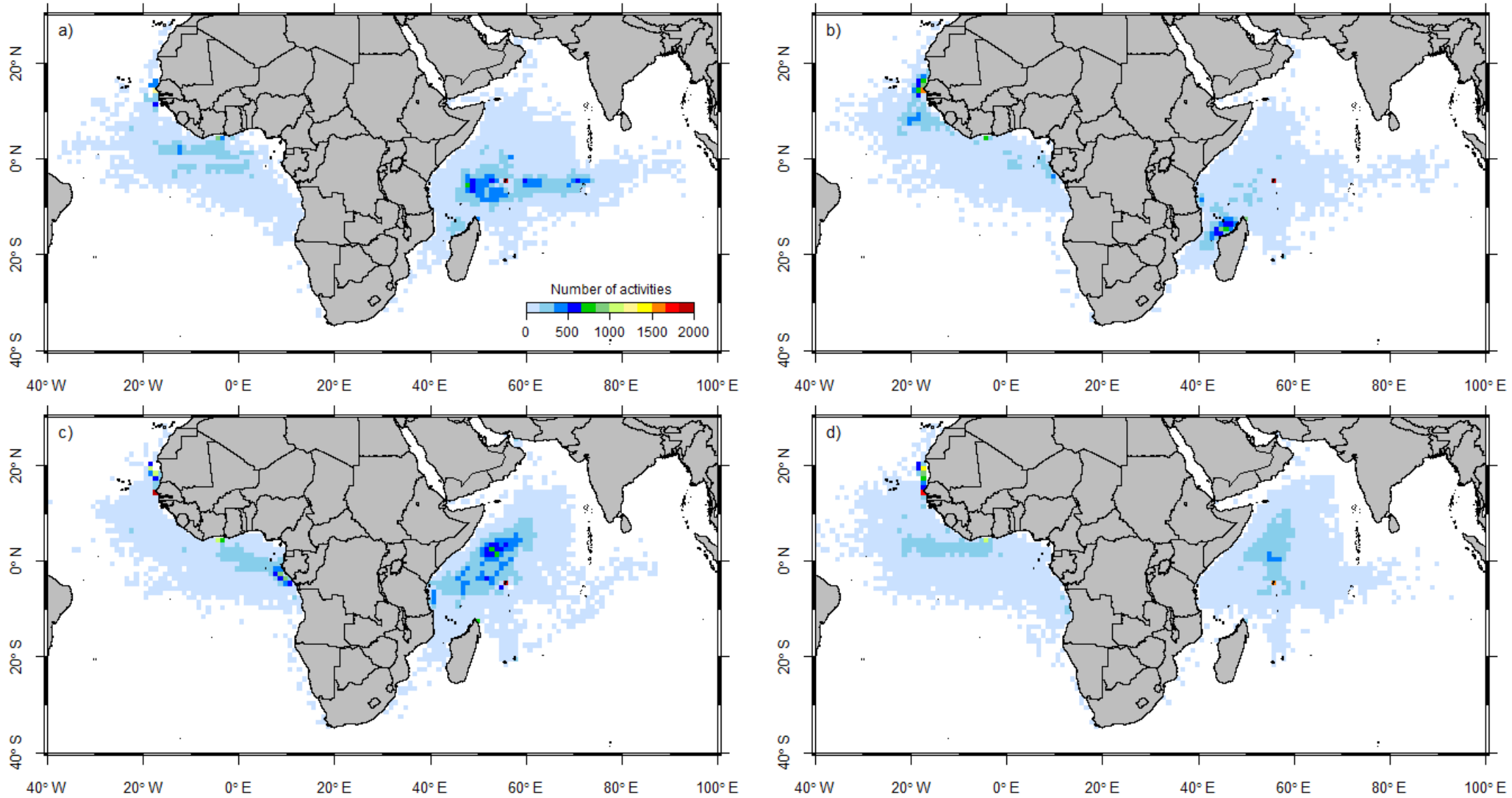


Figure A.1. Seasonal distribution of purse-seine sighting effort in the tropical Eastern Atlantic and Western Indian Oceans a) season 1 and NE monsoon, b) season 2 and ISW monsoon, c) season 3 and SW monsoon, and d) season 4 and INE monsoon.

Appendix B. Whale sharks in the Atlantic Ocean

Table B.1. GAM and BRT binomial models (presence/absence). Number in brackets indicate the percentage of the total deviance explained by each explanatory variables.

	Explanatory variables	% Deviance[1]	R2	Correlation [2]	Moran p value	Moran's I[3]
GAM binomial	SST, CHL, EKE, LandDist, Season, ti(EKE, LandDist)	22.4	0.14	0.38	$< 2.2e^{-16}$	0.14
GAM binomial RAC	SST, CHL, EKE, LandDist, Season, Autocovariate, ti(EKE, LandDist)	29.0	0.18	0.43	$< 2.2e^{-16}$	0.09
BRT binomial	CHL (8.4), SST (6.9), LandDist (5.0), Depth (2.7), EKE (2.7), Slope (2.6), Season (1.3)	29.3		0.49	$< 2.2e^{-16}$	0.12
BRT binomial RAC*	CHL (5.4), SST (4.4), LandDist (2.4), Depth (1.6), EKE (1.4), Season (1.0), Slope (0.2), Autocovariate (19.0)	35.5		0.51	$< 2.2e^{-16}$	0.06

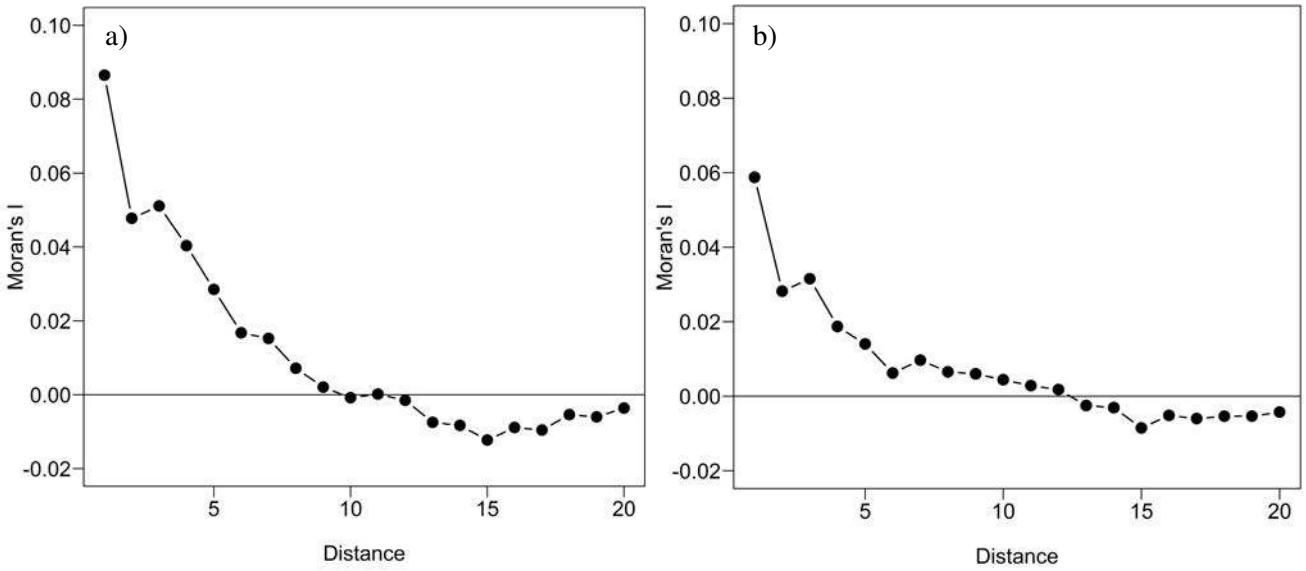


Figure B.1. Correlograms of a) GAM binomial RAC and b) BRT binomial RAC models.

[1] % Deviance explained by each model.

[2] Pearson correlation index between observed and predicted values.

[3] Moran's index maximum absolute value for each model.

Cross validation

Binomial model performance was assessed using the area under the receiver-operating characteristic curve (AUC), and using specificity, sensitivity, and kappa derived from the confusion matrix. AUC allows occupied and unoccupied sites to be correctly distinguished. Specificity and sensitivity are the correctly classified proportion of true negatives and true positives respectively. They measure the model's ability to predict the presence or absence of a species according to the real occurrence of the species at a given location. Kappa measures the proportion of correctly classified locations, which is then compare to a threshold derived from the cross validation to convert these values into presence/absence predictions.

Table B.2. Cross Validation parameters from GAM and BRT binomial RAC models.

	GAM binomial RAC*	BRT binomial RAC
Threshold	0.14	0.04
AUC	0.53	0.53
Kappa	0.02	0.01
Sensitivity	0.56	0.71
Specificity	0.50	0.35

Table B.3. Statistics of observed and predicted values, and residuals of the binomial models (presence/ absence).

	Observed	GAM Binomial	GAM Binomial RAC	BRT Binomial	BRT Binomial RAC*
Observed or predicted values					
Mean	0.06	0.06	0.06	0.06	0.06
Median	0	0.02	0.02	0.03	0.02
Min	0	0	0	0	0
Max	1	1	1	0.65	0.83
Standard deviation	0.24	0.09	0.11	0.09	0.10
Residuals					
Mean		$1.82e^{-15}$	$-1.88e^{-15}$	$-5.54e^{-04}$	$-2.82e^{-04}$
Median		$-0.2e^{-02}$	-0.02	-0.02	-0.02
Standard deviation		0.22	0.21	0.21	0.20
Abs (residuals)					
Mean		0.10	0.09	0.09	0.09
Median		0.02	0.02	0.03	0.02
Standard deviation		0.20	0.19	0.19	0.18
(Residuals)^2					
Mean		0.05	0.05	0.04	0.04
Median		$5.80e^{-04}$	$4.06e^{-04}$	$8.35e^{-04}$	$4.41e^{-04}$
Standard deviation		0.17	0.16	0.19	0.15

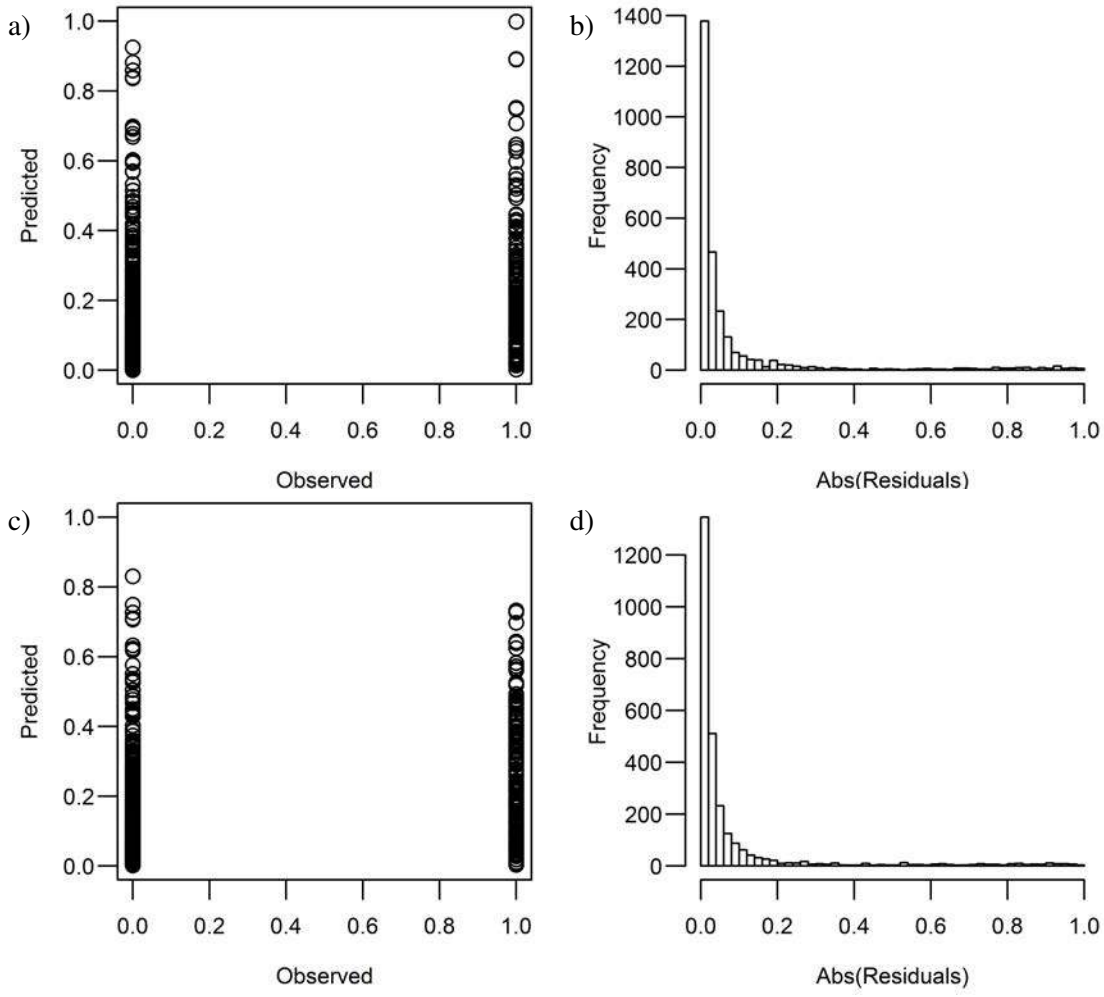


Figure B.2. Graph of the predicted and observed values and histogram the absolute value of the residuals for a) b) GAM binomial RAC and c) d) BRT binomial RAC models.

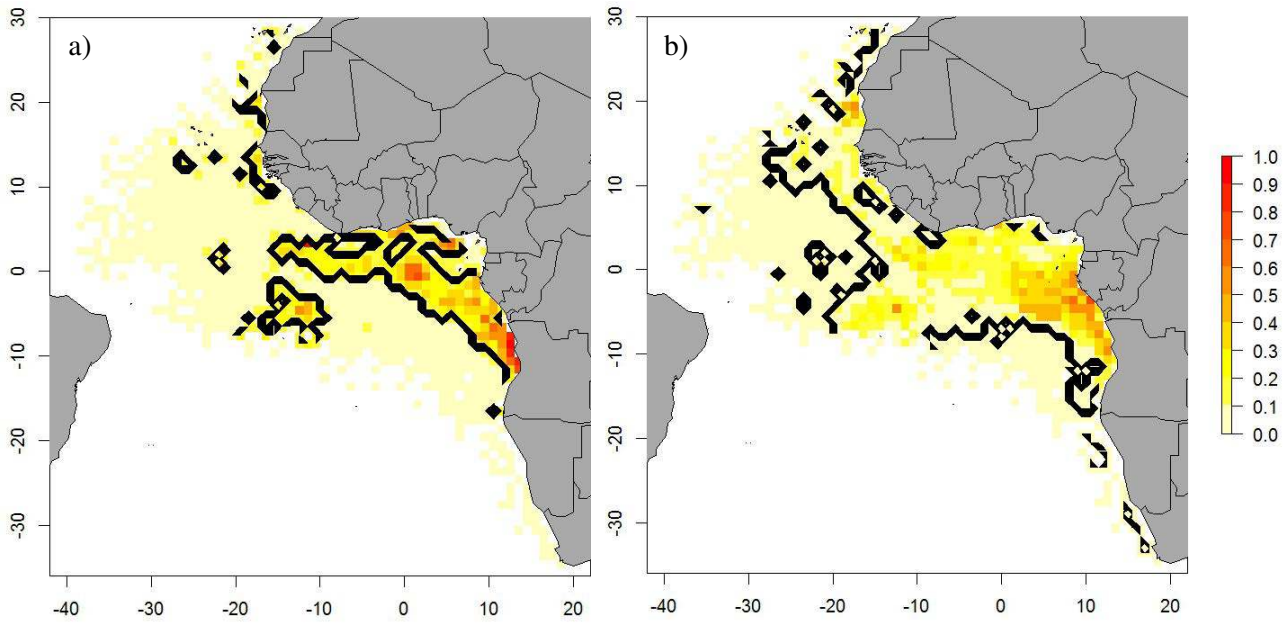


Figure B.3. Envelope of presence/absence from a) GAM binomial RAC and b) BRT binomial RAC models.

Table B.4. Count models (GAM quasi-poisson and BRT poisson). Number in brackets indicate the percentage of the total deviance explained by each explanatory variables.

	Explanatory variables	% Deviance	R2	Correlation	Moran p value	Moran's I
GAM quasi-poisson	SST, CHL, EKE, Depth, Slope, LandDist, Season	67.5	0.64	0.49	$< 2.2e^{-16}$	0.15
GAM quasi-poisson RAC	SST, CHL, EKE, Depth, Slope, LandDist, Season, Autocovariate	81.0	0.75	0.54	$7.5e^{-14}$	0.05
BRT poisson	SST (22.1), CHL (14.9), LandDist (11.2), Depth (10.6), Slope (9.3), Season (1.3), EKE (1.2)	70.6		0.87	$< 2.2e^{-16}$	0.15
BRT poisson RAC*	CHL (5.5), LandDist (5.1), Slope (5.1), SST (3.5), Depth (3.5), Season (0.4), EKE (0.1), Autocovariate (59.7%)	83.0		0.94	0.06	0.02

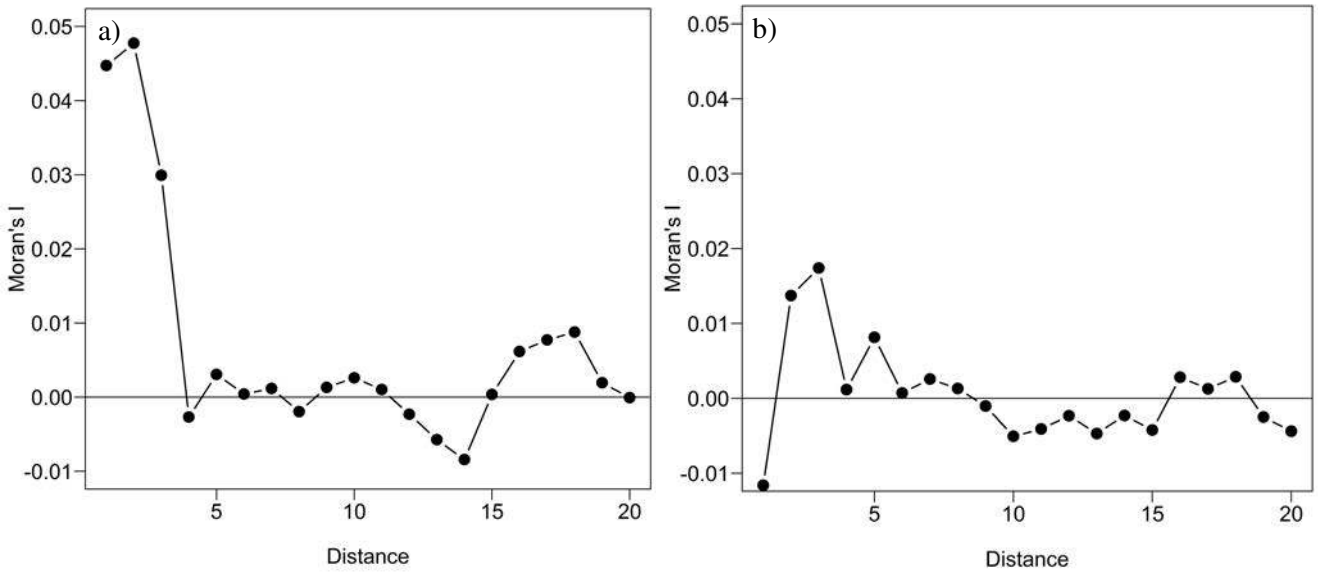


Figure B.4. Correlograms of a) GAM quasi-poisson RAC and b) BRT poisson RAC models.

Table B.5. Statistics of observed and predicted values, and residuals of the count models.

	Observed	GAM quasi-poisson	**GAM quasi-poisson RAC**	**BRT poisson**	**BRT poisson RAC***
Observed or predicted values					
Mean	0.36	0.81	0.89	0.38	0.39
Median	0	0.08	0.14	0.05	0.04
Min	0	0	0	0	0
Max	49	53	59.7	60.8	59.9
Standard deviation	2.58	3.03	3.13	2.41	2.67
Residuals					
Mean		-0.44	-0.53	-0.02	-0.02
Median		-0.06	-0.10	-0.04	-0.03
Standard deviation		2.86	2.78	1.28	0.89
Abs (residuals)					
Mean		0.82	0.86	0.34	0.25
Median		0.09	0.14	0.05	0.04
Standard deviation		2.78	2.70	1.23	0.85
(Residuals)^2					
Mean		8.38	8.02	1.63	0.79
Median		0.01	0.02	$0.3e^{-02}$	$0.1\ e^{-02}$
Standard deviation		88.3	84.7	15.6	8.32

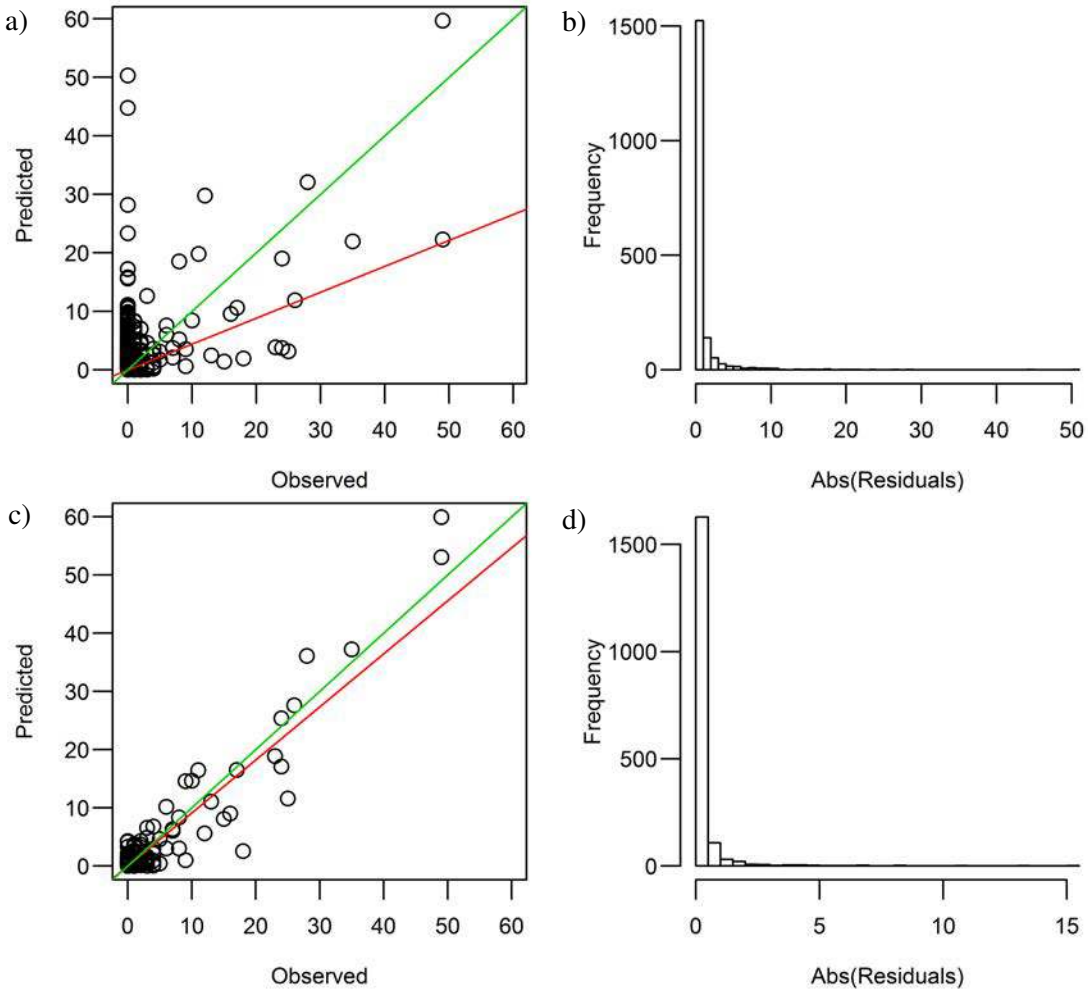


Figure B.5. Graph of the predicted and observed values and histogram the absolute value of the residuals for a) b) GAM quasi-poisson RAC c) d) BRT poisson RAC models.

Table C.1. GAM and BRT binomial models (presence/absence). Number in brackets indicate the percentage of the total deviance explained by each explanatory variables.

	Explanatory variables	% Deviance[1]	R2	Correlation[2]	Moran p value	Moran's I[3]
GAM binomial	Depth, EKE, CHL, Season, ti(CHL, EKE)	10.4	0.05	0.24	$< 2.2e^{-16}$	0.12
GAM binomial RAC*	Depth, EKE, CHL, Season, Autocovariate, ti(CHL, EKE)	12.3	0.06	0.26	$< 2.2e^{-16}$	0.12
BRT binomial	EKE (5.8), CHL (5.2), Depth (3.1), Season (2.8), LandDist (2.5), SST (2.3), Slope (1.1)	22.9		0.20	$< 2.2e^{-16}$	0.01
BRT binomial RAC	EKE (5.0), CHL (4.7), Season (2.8), Depth (2.0), LandDist (2.0), SST (1.8), Slope (0.7), Autocovariate (5.2)	24.2		0.23	$< 2.2e^{-16}$	0.08

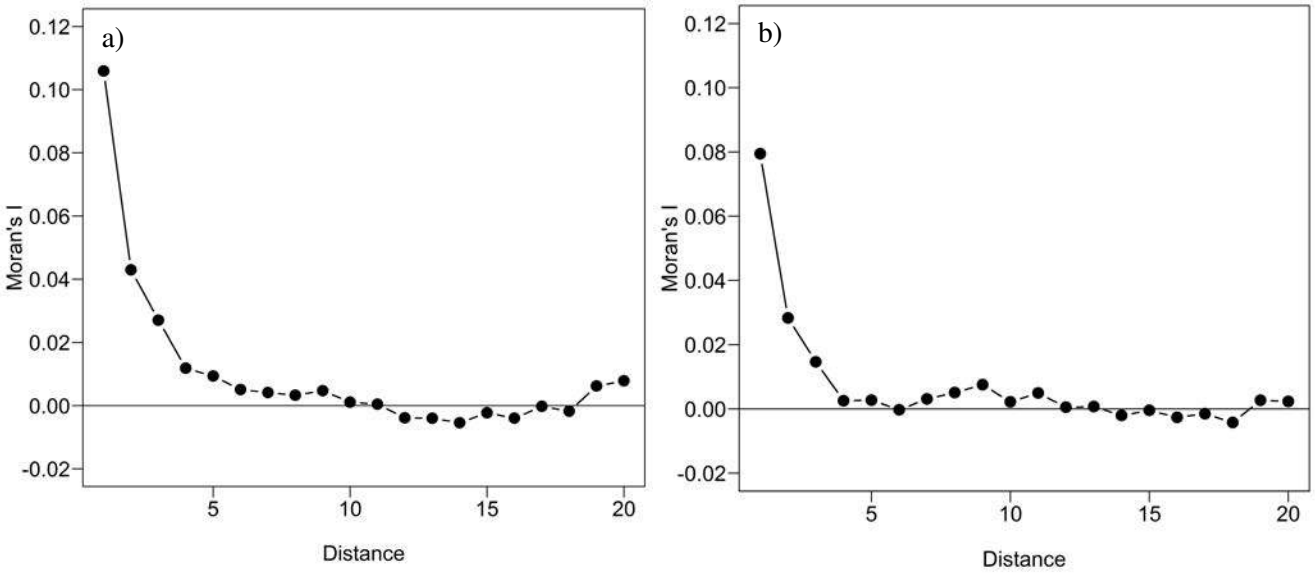


Figure C.1. Correlograms of a) GAM binomial RAC and b) BRT binomial RAC models.

[1] % Deviance explained by each model.

[2] Pearson correlation index between observed and predicted values.

[3] Moran's index maximum absolute value for each model.

Cross validation

Binomial model performance was assessed using the area under the receiver-operating characteristic curve (AUC), and using specificity, sensitivity, and kappa derived from the confusion matrix. AUC allows occupied and unoccupied sites to be correctly distinguished. Specificity and sensitivity are the correctly classified proportion of true negatives and true positives respectively. They measure the model's ability to predict the presence or absence of a species according to the real occurrence of the species at a given location. Kappa measures the proportion of correctly classified locations, which is then compare to a threshold derived from the cross validation to convert these values into presence/absence predictions.

Table. C.2 Cross Validation parameters from GAM and BRT binomial RAC models.

	GAM binomial RAC*	BRT binomial RAC
Threshold	0.05	0.13
AUC	0.54	0.03
Kappa	0.01	0.32
Sensitivity	0.60	0.32
Specificity	0.46	0.74

Table C.3. Statistics of observed and predicted values, and residuals of the binomial models (presence/ absence).

	Observed	GAM binomial	GAM binomial RAC*	BRT binomial	BRT binomial RAC
Observed or predicted values					
Mean	0.05	0.05	0.05	0.06	0.06
Median	0	0.03	0.03	0.03	0.03
Min	0	0	0	0	0
Max	1	0.34	0.57	0.53	0.65
Standard deviation	0.21	0.04	0.05	0.07	0.08
Residuals					
Mean		$1.26e^{-15}$	$3.74e^{-16}$	-0.01	-0.01
Median		-0.03	-0.03	-0.03	-0.03
Standard deviation		0.20	0.21	0.21	0.21
Abs (residuals)					
Mean		0.08	0.08	0.09	0.09
Median		0.03	0.03	0.03	0.03
Standard deviation		0.18	0.18	0.19	0.18
(Residuals)^2					
Mean		0.04	0.04	0.04	0.04
Median		$9.58e^{-04}$	$8.30e^{-04}$	$9.90e^{-04}$	$9.75e^{-04}$
Standard deviation		0.17	0.17	0.16	0.16

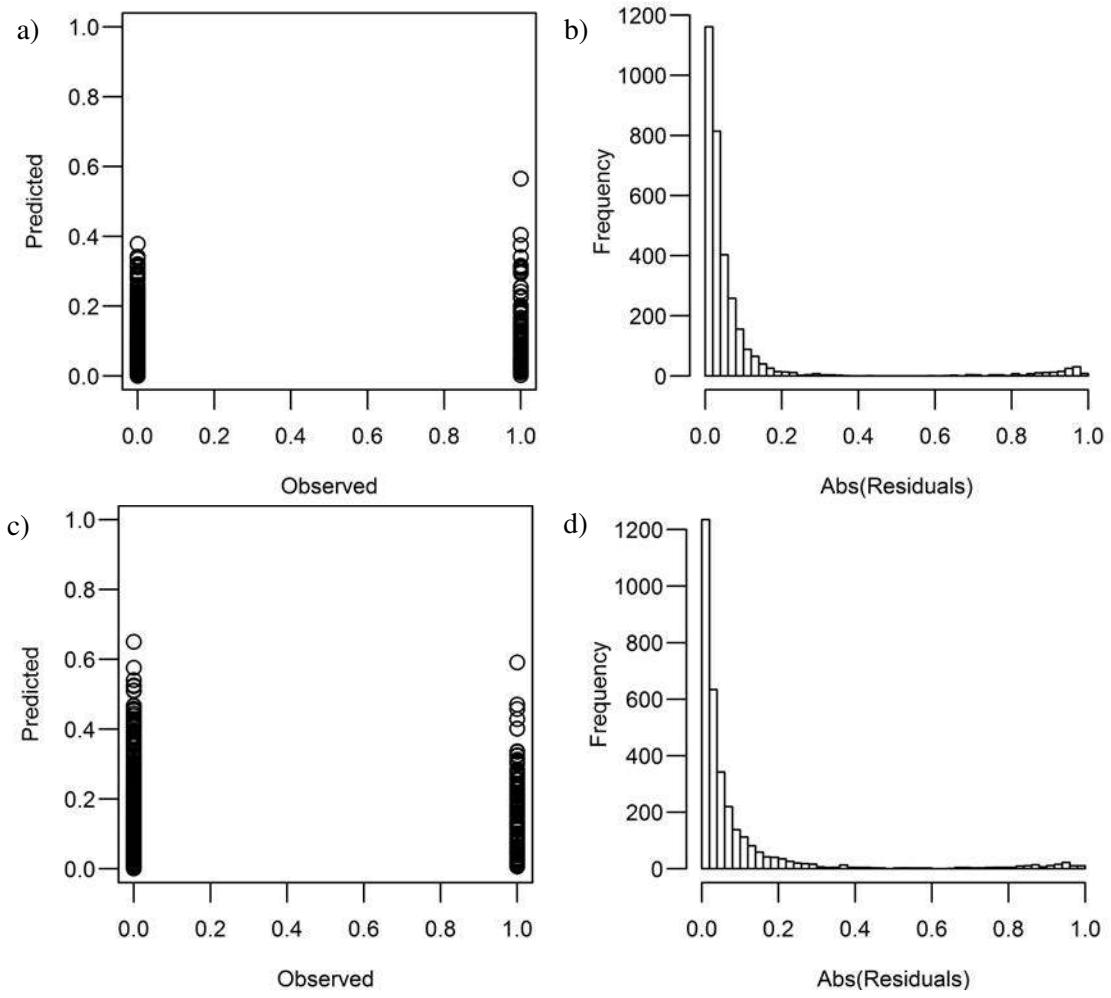


Figure C.2. Graph of the predicted and observed values and histogram the absolute value of the residuals for a) b) GAM binomial RAC c) d) BRT binomial RAC models.

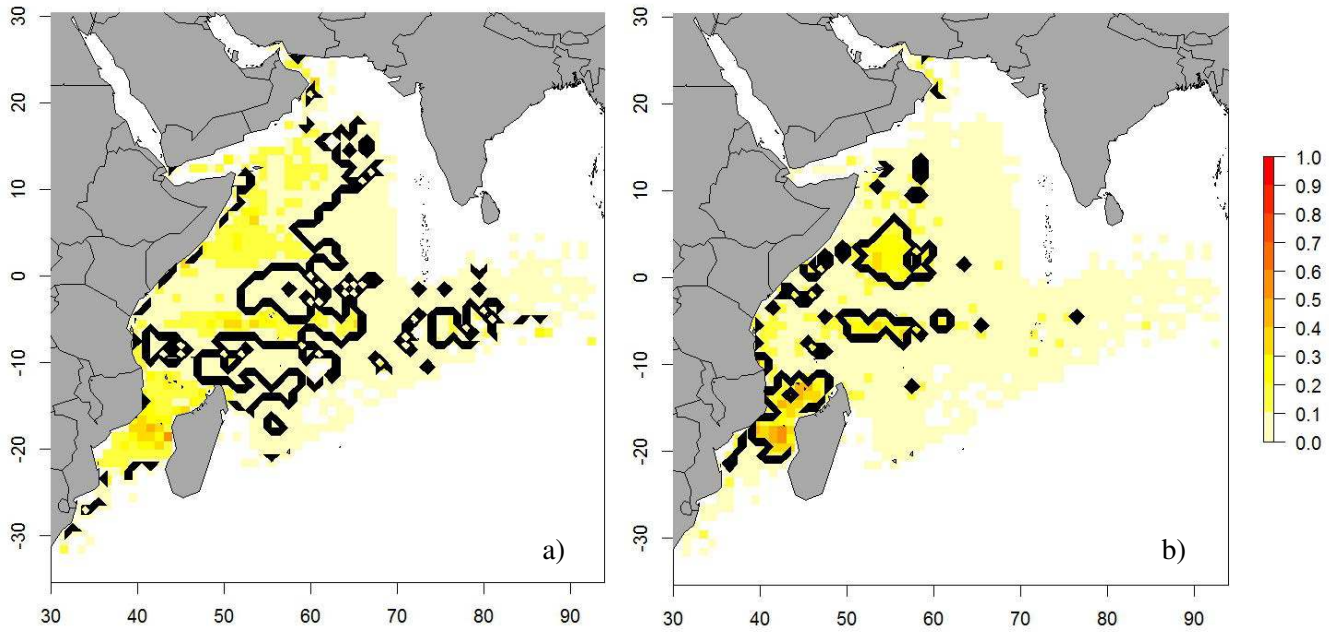


Figure C.3. Envelope of presence absence from a) GAM binomial RAC and b) BRT binomial RAC models.

Table C.4. Count models (GAM quasi-poisson and BRT poisson). Number in brackets indicate the percentage of the total deviance explained by each explanatory variables.

	Explanatory variables	% Deviance	R2	Correlation	Moran p value	Moran's I
GAM quasi-poisson	SST, CHL, EKE, Depth, Season	43.7	0.43	0.35	0.02	0.07
GAM quasi-poisson RAC	SST, CHL, EKE, Depth, Season, Autocovariate	59.7	0.68	0.56	0.14	0.02
BRT poisson	Season (22.8), Depth (18.6), EKE (11.7), SST (7.3), CHL (5.8), Slope (1.1)	67.4		0.31	$0.20e^{-03}$	0.05
BRT poisson RAC*	Season (3.2), Depth (3.0), EKE (2.0), SST (1.1), CHL (0.5), Slope (0.1), Autocovariate (60.2)	70.1		0.54	$0.01e^{-02}$	0.03

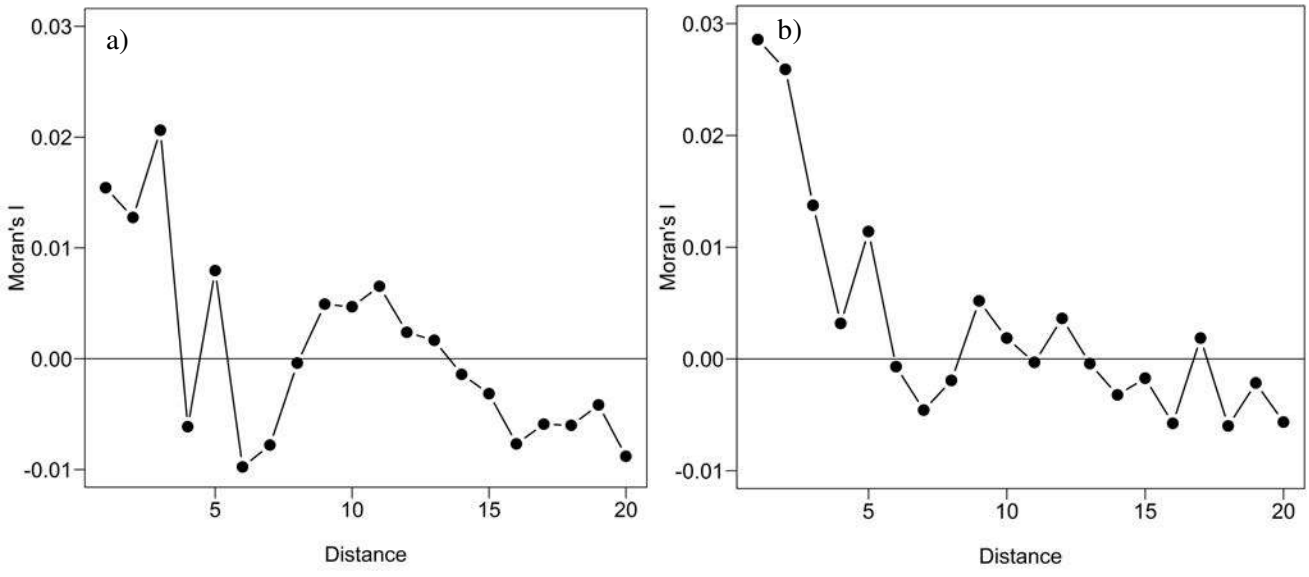


Figure C.4. Correlograms of a) GAM quasi-poisson RAC and b) BRT poisson RAC models.

Table C.5. Statistics of observed and predicted values, and residuals of the count models.

	Observed	GAM quasi-poisson	GAM quasi-poisson RAC	BRT poisson	BRT poisson RAC*
Observed or predicted values					
Mean	0.21	0.89	0.66	0.23	0.20
Median	0	0.37	0.28	0.05	0.04
Min	0	0	0	0	0
Max	36	23.0	31.2	3.9	10.2
Standard deviation	1.55	1.71	1.72	0.46	0.633
Residuals					
Mean		-0.64	-0.45	-0.02	0.01
Median		-0.34	-0.26	-0.03	-0.03
Standard deviation		1.86	1.55	1.47	1.32
Abs (residuals)					
Mean		0.87	0.66	0.36	0.30
Median		0.39	0.29	0.05	0.04
Standard deviation		1.76	1.47	1.43	1.28
(Residuals)^2					
Mean		3.87	2.59	2.17	1.74
Median		0.15	0.08	$0.20e^{-02}$	$0.10e^{-02}$
Standard deviation		24.90	27.99	32.70	24.30

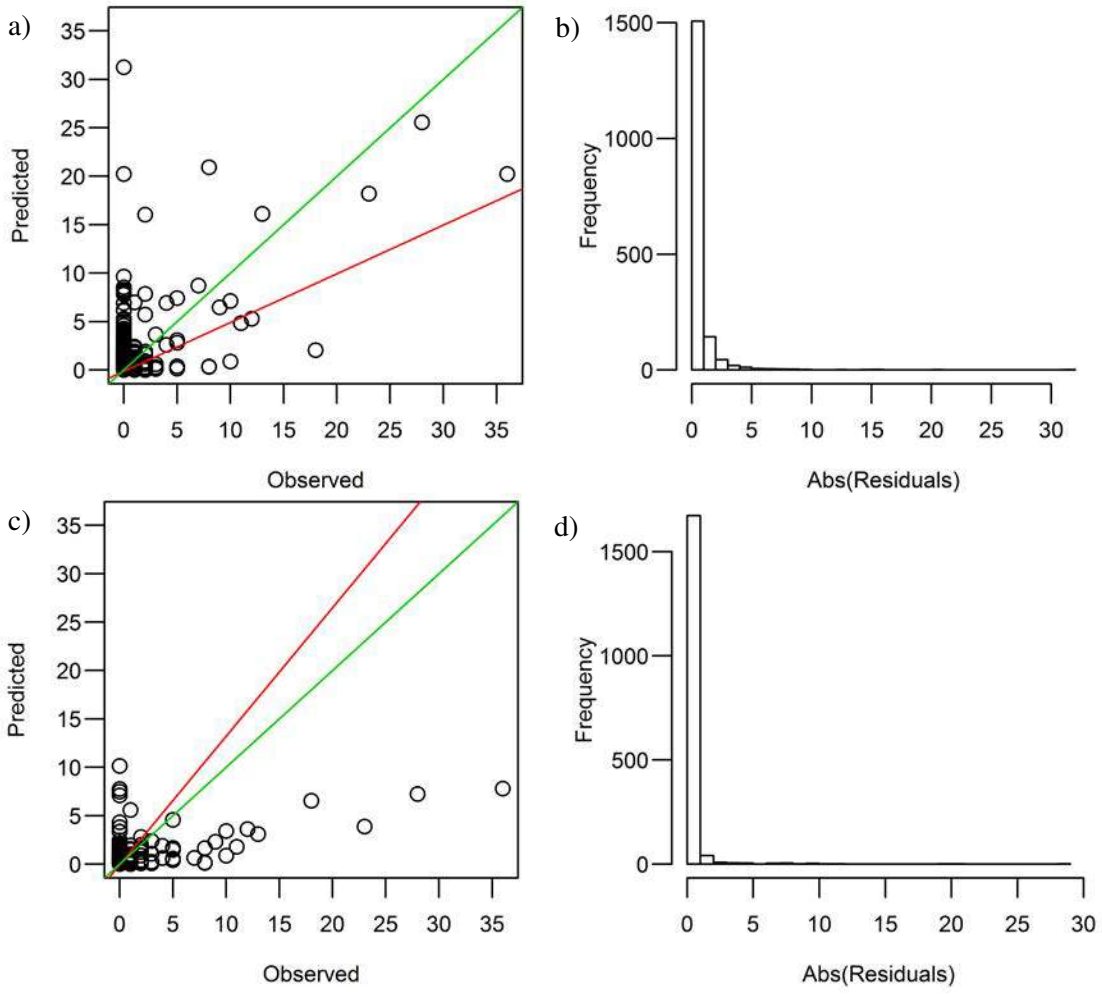


Figure C.5. Graph of the predicted and observed values and histogram the absolute value of the residuals for a) b) GAM quasi-poisson RAC c) d) BRT poisson RAC models.

Table D.1. GAM and BRT binomial models (presence/absence). Number in brackets indicate the percentage of the total deviance explained by each explanatory variables.

	Explanatory variables	% Deviance[1]	R2	Correlation[2]	Moran p value	Moran's I[3]
GAM binomial	Depth, SST, CHL, EKE, Slope, LandDist, Season, ti(Depth,Slope)	15.6	0.12	0.35	$< 2.2e^{-16}$	0.18
GAM binomial RAC*	Depth, SST, CHL, EKE, Slope, LandDist, Season, Autocovariate, ti(Depth,Slope)	21.1	0.16	0.41	$1.2e^{-10}$	0.15
BRT binomial	LandDist (6.3), EKE (4.6), Slope (4.0), Depth (3.0), SST (2.4), CHL (2.3), Season (0.3)	23.0		0.46	$< 2.2e^{-16}$	0.16
BRT binomial RAC	LandDist (4.7), CHL (3.2), EKE (3.1), SST (2.5), Slope (1.8), Depth (1.5), Season (0.1), Autocovariate (11.3)	28.1		0.50	$< 2.2e^{-16}$	0.12

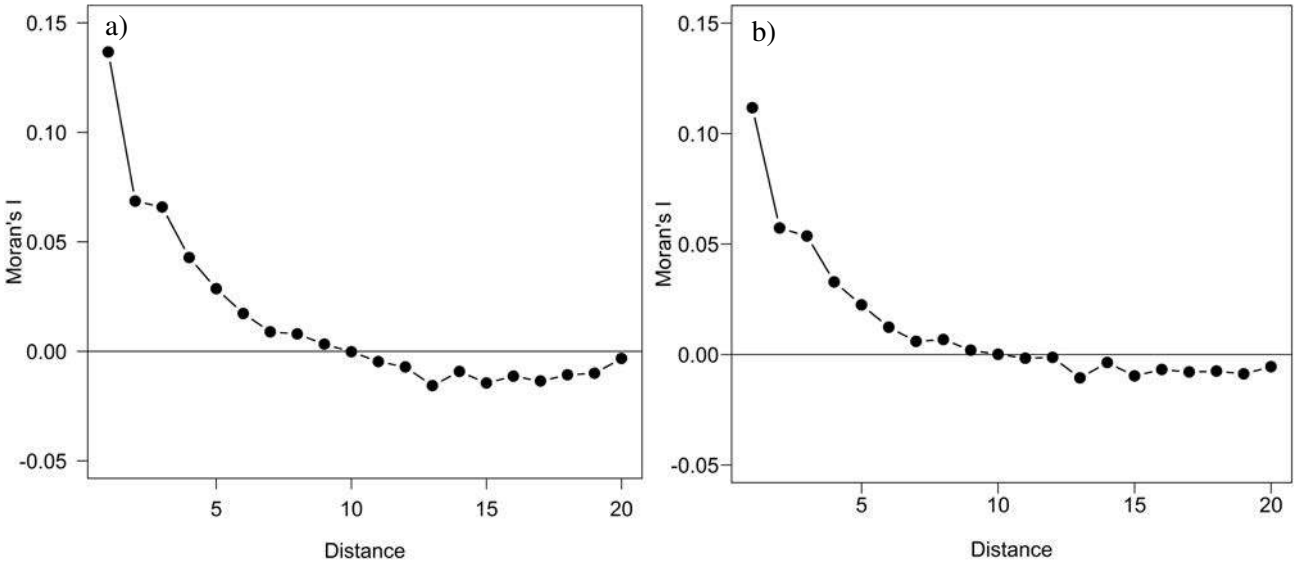


Figure D.1. Correlograms of a) GAM binomial RAC and b) BRT binomial RAC models.

[1] % Deviance explained by each model.

[2] Pearson correlation index between observed and predicted values.

[3] Moran's index maximum absolute value for each model.

Cross validation

Binomial model performance was assessed using the area under the receiver-operating characteristic curve (AUC), and using specificity, sensitivity, and kappa derived from the confusion matrix. AUC allows occupied and unoccupied sites to be correctly distinguished. Specificity and sensitivity are the correctly classified proportion of true negatives and true positives respectively. They measure the model's ability to predict the presence or absence of a species according to the real occurrence of the species at a given location. Kappa measures the proportion of correctly classified locations, which is then compare to a threshold derived from the cross validation to convert these values into presence/absence predictions.

Table. D.2 Cross Validation parameters from GAM and BRT binomial RAC models.

	GAM binomial RAC*	BRT binomial RAC
Threshold	0.10	0.15
AUC	0.54	0.54
Kappa	0.04	0.06
Sensitivity	0.65	0.58
Specificity	0.44	0.53

Table D.3. Statistics of observed and predicted values, and residuals of the binomial models (presence/ absence)

	Observed	GAM binomial	GAM binomial RAC*	BRT binomial	BRT binomial RAC
Observed or predicted values					
Mean	0.13	0.13	0.13	0.13	0.13
Median	0	0.09	0.07	0.08	0.07
Min	0	0	0	0	0
Max	1	0.85	0.92	0.79	0.86
Standard deviation	0.33	0.11	0.14	0.12	0.14
Residuals					
Mean		$2.21e^{-15}$	$-6.97e^{-16}$	$-6.79e^{-04}$	$-4.87e^{-04}$
Median		-0.01	-0.01	-0.06	-0.04
Standard deviation		0.31	0.30	0.30	0.29
Abs (residuals)					
Mean		0.19	0.18	0.18	0.17
Median		0.01	0.08	0.09	0.07
Standard deviation		0.21	0.24	0.23	0.23
$(Residuals)^2$					
Mean		0.10	0.09	0.09	0.08
Median		0.01	0.01	0.01	$0.40e^{-02}$
Standard deviation		0.21	0.20	0.19	0.19

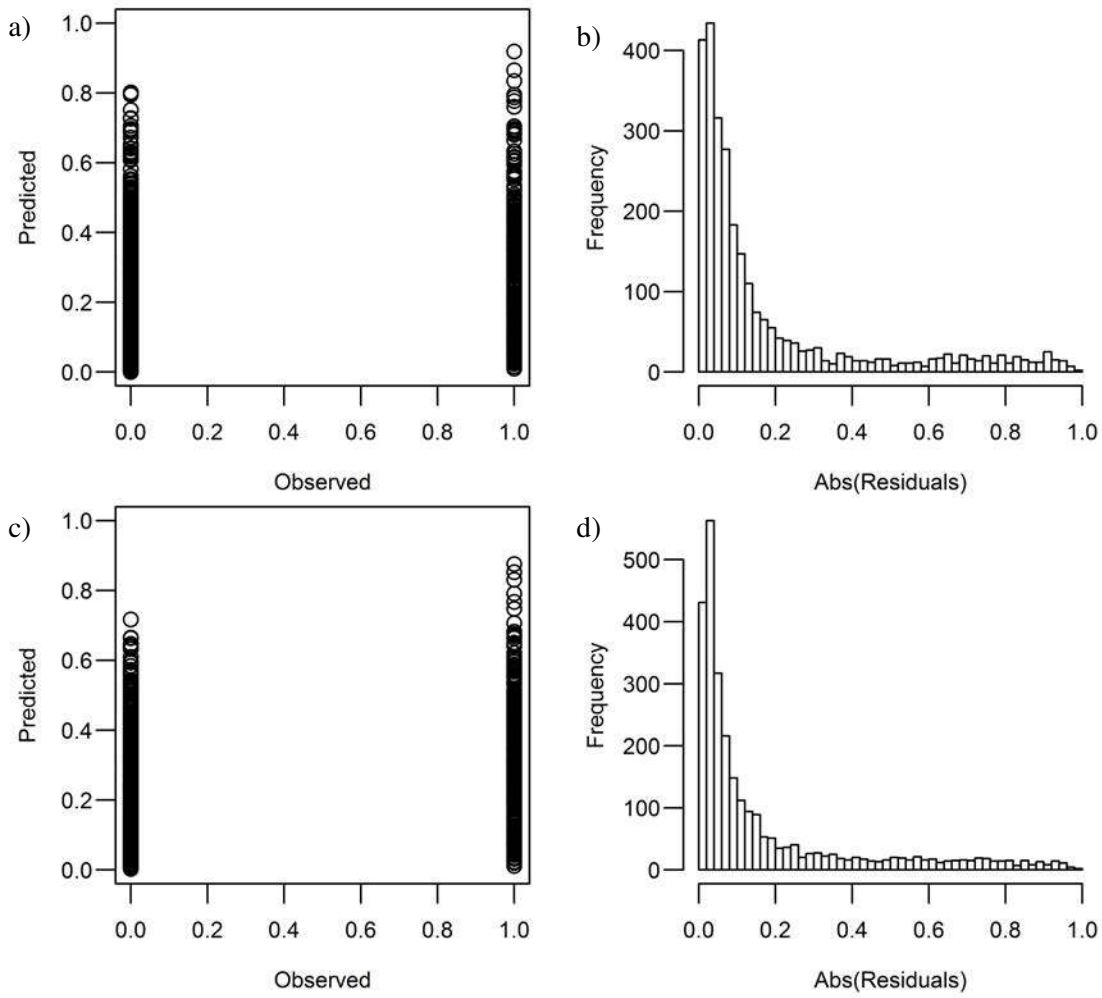


Figure D.2. Graph of the predicted and observed values and histogram the absolute value of the residuals for a) b) GAM binomial RAC and c) d) BRT binomial RAC models.

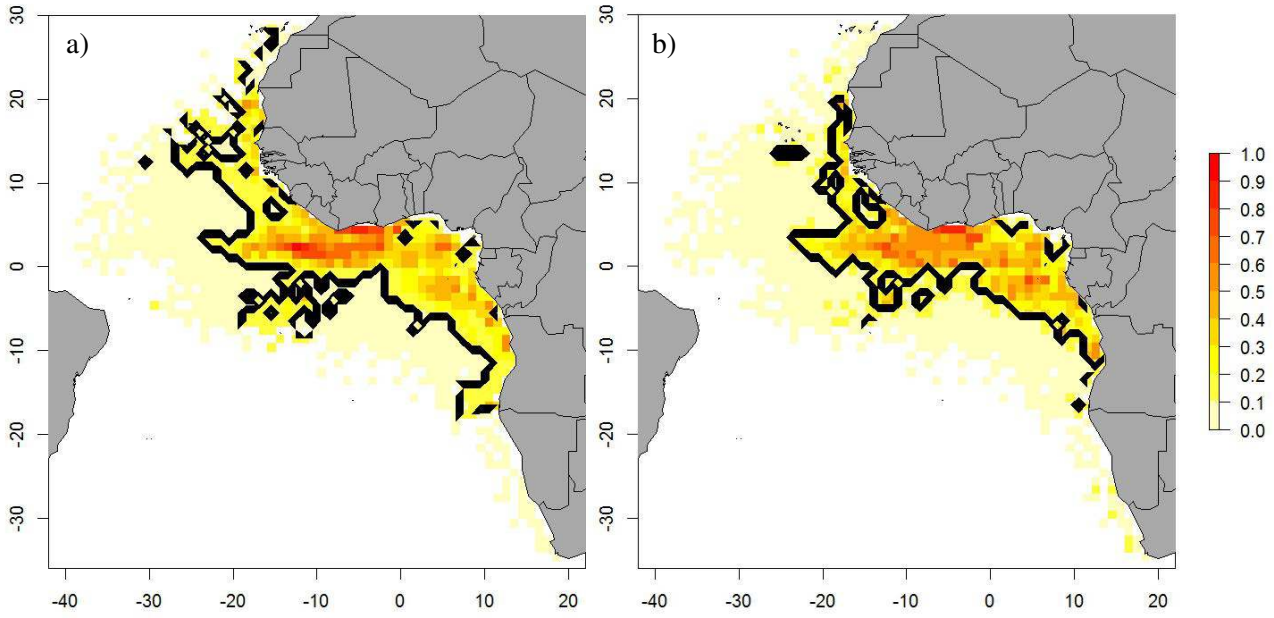


Figure D.3. Envelope of presence absence from a) GAM binomial RAC and b) BRT binomial RAC models.

Table D.4. Count models (GAM quasi-poisson and BRT poisson). Number in brackets indicate the percentage of the total deviance explained by each explanatory variables.

	Explanatory variables	% Deviance	R2	Correlation	Moran p value	Moran's I
GAM quasi-poisson	SST, CHL, EKE, Depth, Season, Slope	53.3	0.58	0.45	$<2.2e^{-16}$	0.12
GAM quasi-poisson RAC	SST, CHL, EKE, Depth, Season, Slope, Autocovariate	65.5	0.78	0.58	0.11	0.06
BRT poisson	SST (27.0), Slope (12.5), CHL (12.0), Depth (10.1), LandDist (2.8), Season (2.7), EKE (1.5)	68.7		0.91	$<2.2e^{-16}$	0.09
BRT poisson RAC*	SST (7.6), CHL (6.6), Slope (3.7), Depth(1.9), Season (0.6), EKE (0.6), LandDist (0.5), Autocovariate (50.6)	72.1		0.92	$0.10e^{-02}$	-0.02

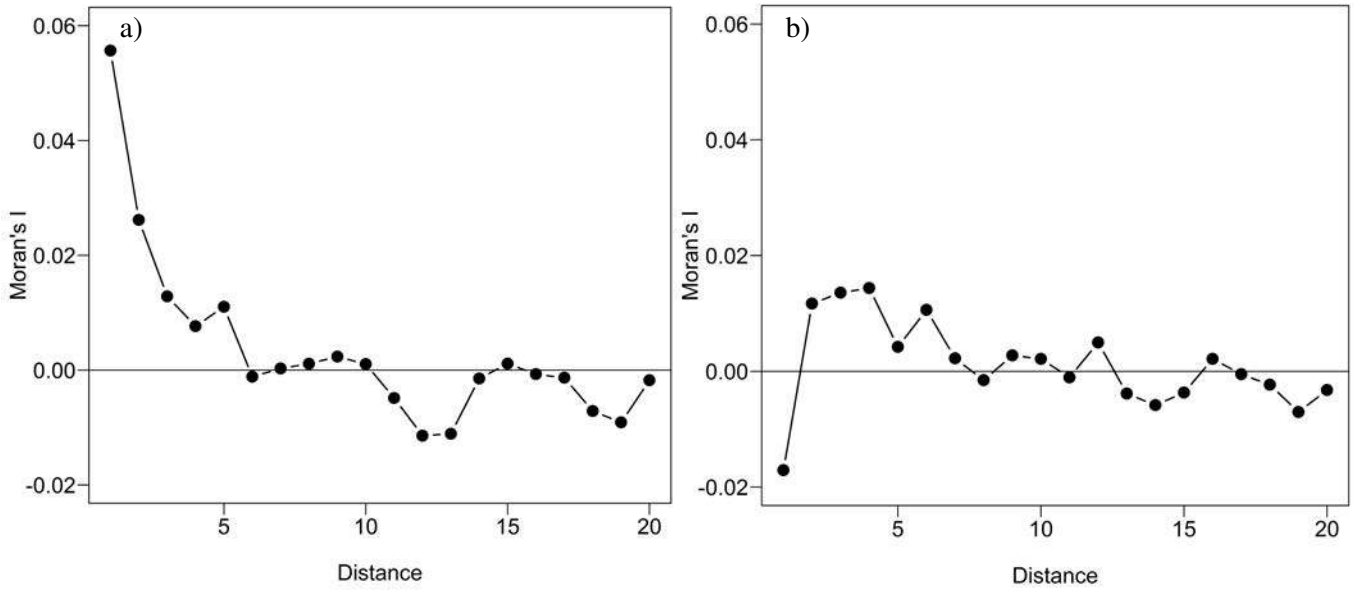


Figure D.4. Correlograms of a) GAM quasi-poisson RAC and b) BRT poisson RAC models.

Table D.5. Statistics of observed and predicted values, and residuals of the count models.

	Observed	GAM quasi-poisson	GAM quasi-poisson RAC	BRT poisson	BRT poisson RAC*
Observed or predicted values					
Mean	0.72	1.93	2.25	0.72	0.72
Median	0	0.82	1.15	0.22	0.21
Min	0	0.01	0.04	0	0
Max	81	49.8	66.0	71.1	72.9
Standard deviation	3.42	3.42	3.69	2.82	2.84
Residuals					
Mean		-1.20	-1.53	0.01	0.01
Median		-0.65	-0.92	-0.11	-0.12
Standard deviation		3.61	3.30	1.44	1.42
Abs (residuals)					
Mean		1.93	2.08	0.62	0.58
Median		0.85	1.13	0.22	0.21
Standard deviation		3.28	2.98	1.30	1.29
(Residuals)^2					
Mean		14.45	13.20	2.07	2.00
Median		0.72	1.27	0.05	0.04
Standard deviation		93.5	57.39	14.22	17.03

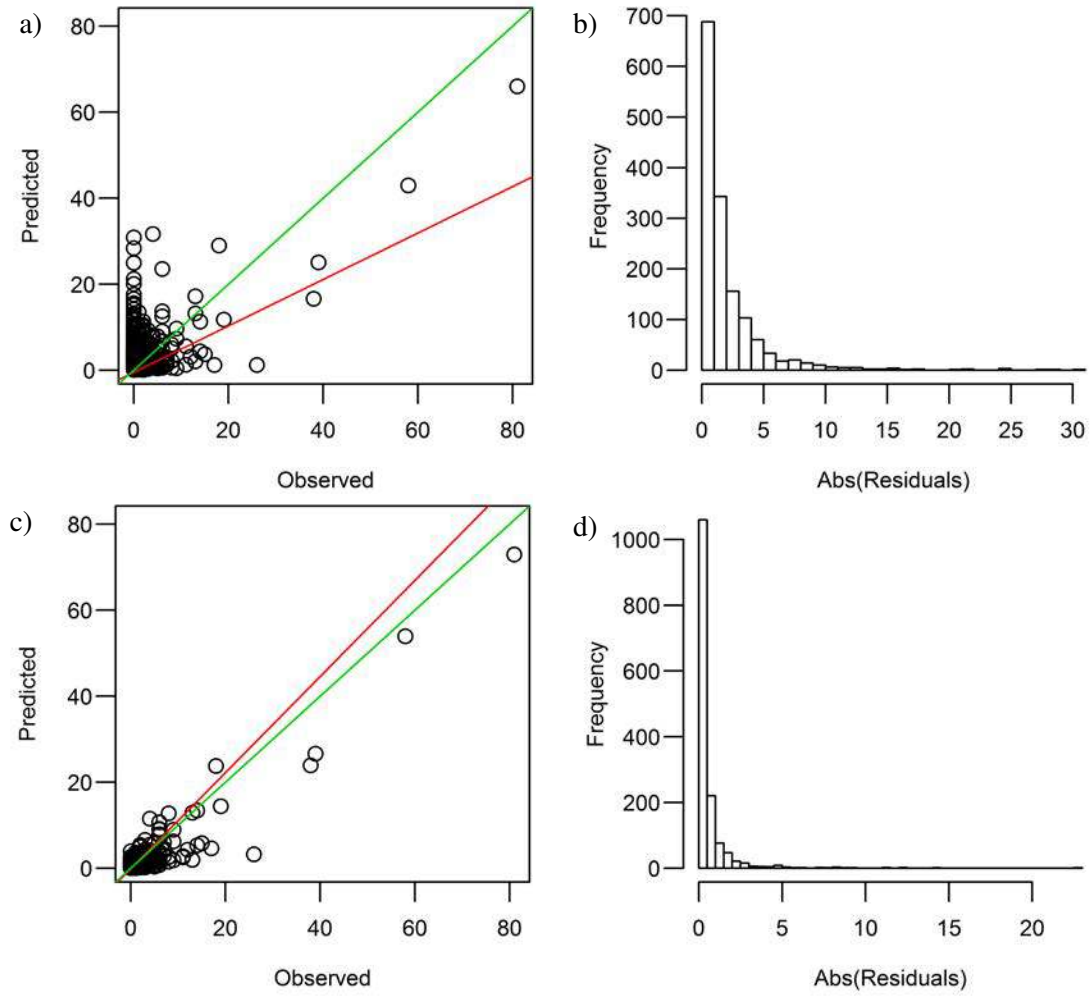


Figure D.5. Graph of the predicted and observed values and histogram the absolute value of the residuals for a) b) GAM quasi-poisson RAC c) d) BRT poisson RAC models.

Appendix F. Environmental factors

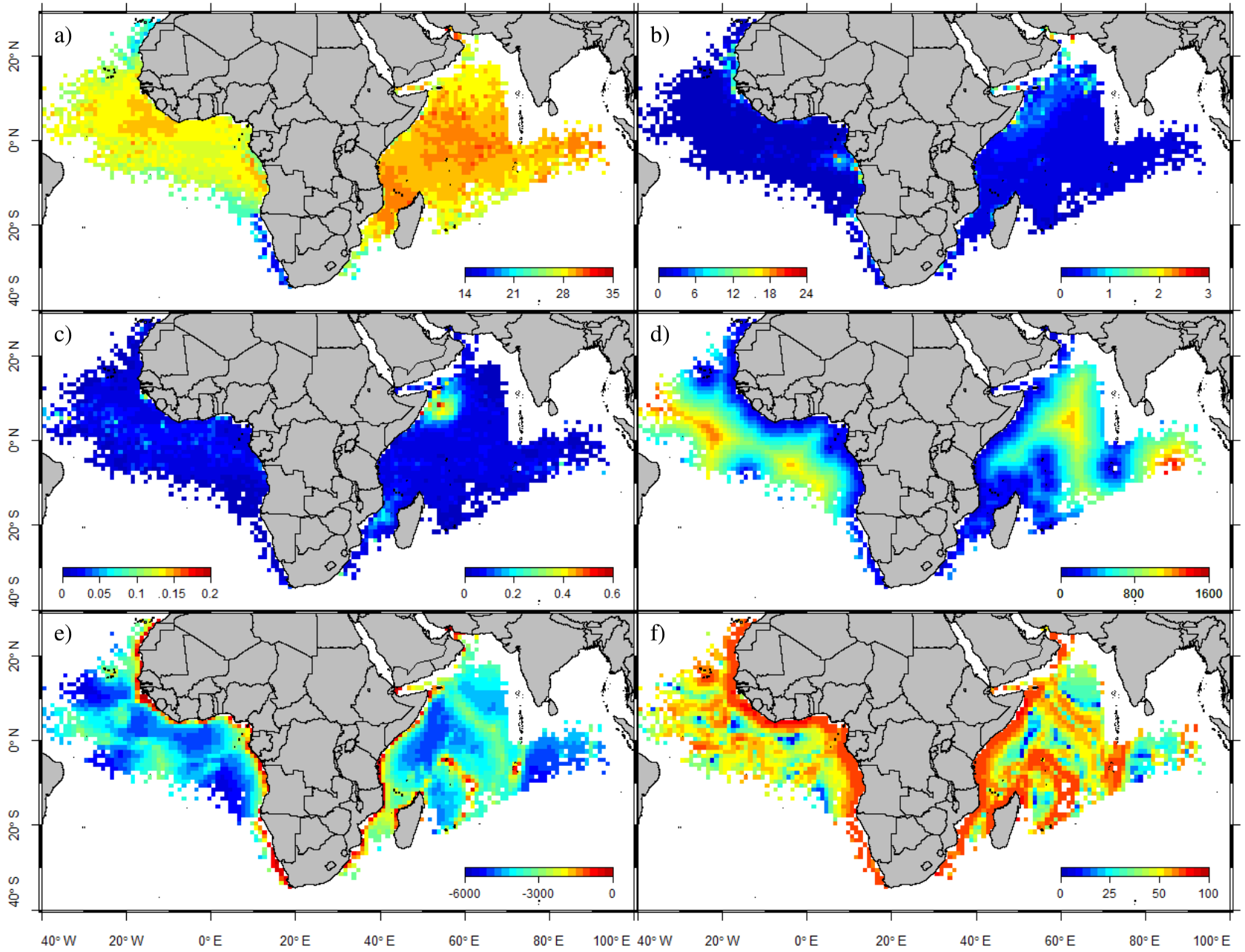


Figure F.1. a) Sea Surface Temperature (°C), b) Chlorophyll-a concentration (mg.m^{-3}), c) Eddy Kinetic Energy (m^2.s^2), d) Distance to land (km), e) Depth (m), and f) slope (%).

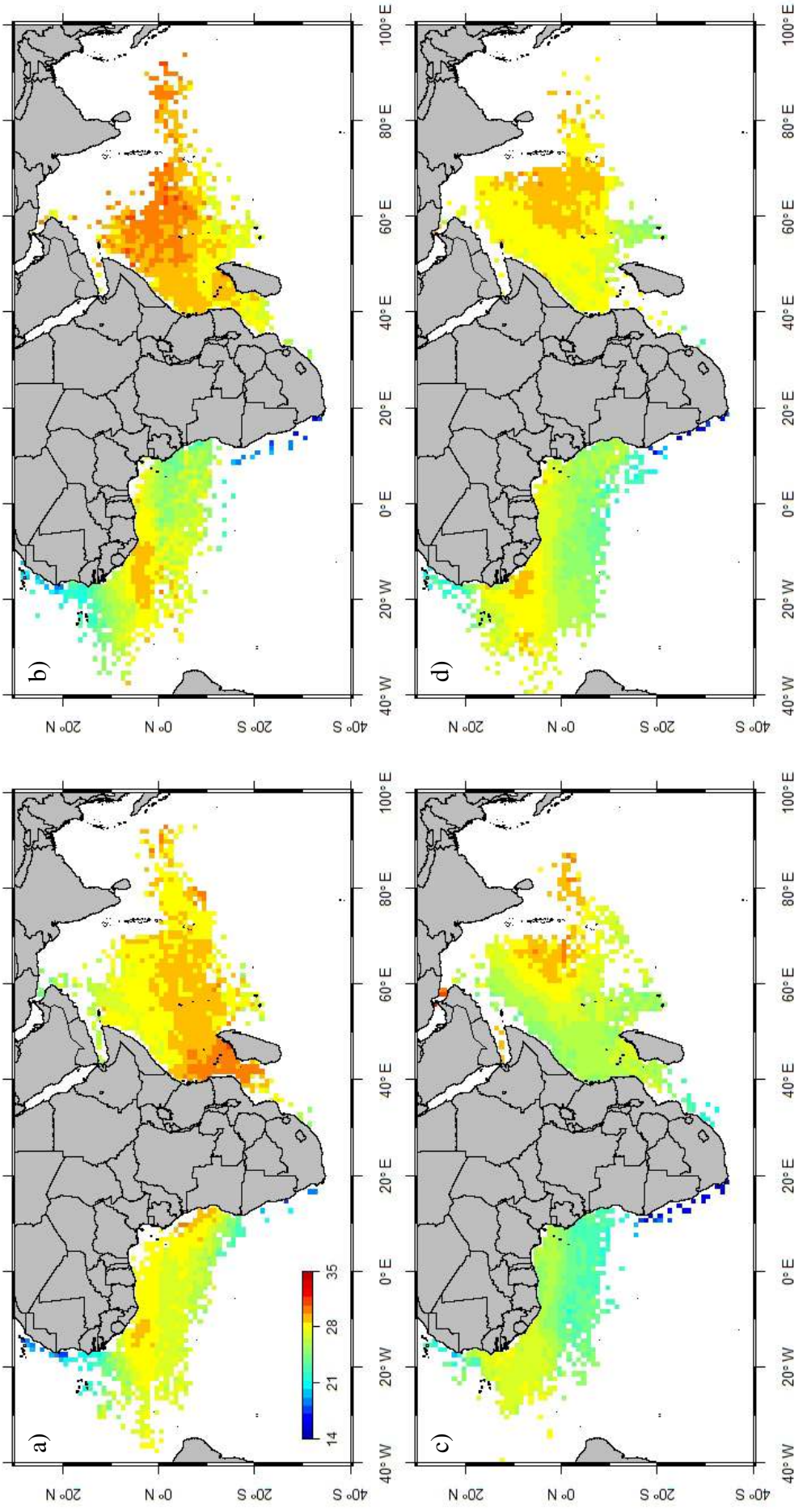


Figure F.2. Sea Surface Temperature (°C) seasonal variability in the Eastern Atlantic and Western Indian Oceans: a) Season 1 and North-East (NE) monsoon, b) Season 2 and intermediate South-West (ISW) monsoon, c) Season 3 and South-West (SW) monsoon, and d) Season 4 and intermediate North-East (INE) monsoon.

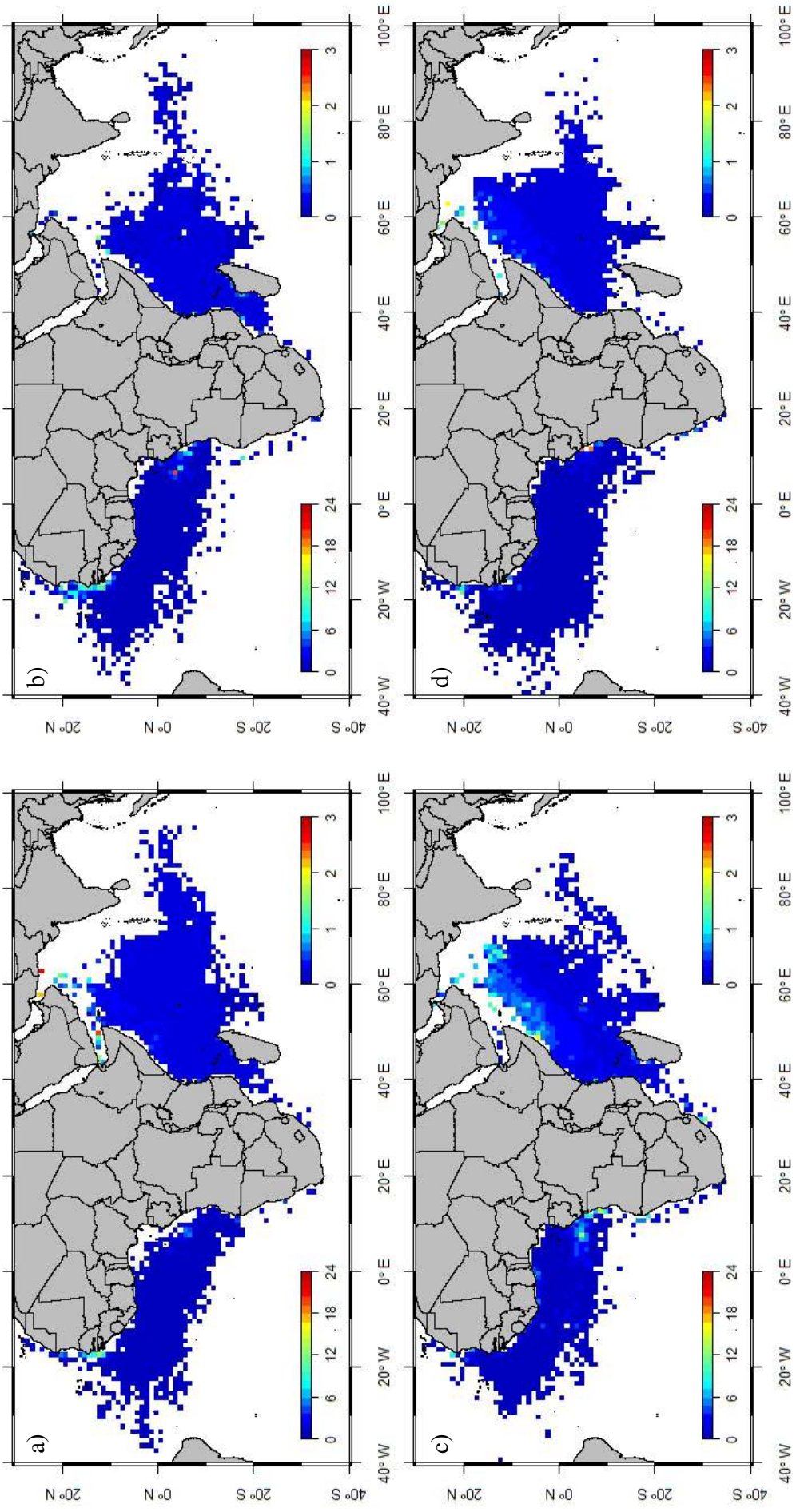


Figure F.3. Chlorophyll-a concentration ($mg.m^{-3}$) seasonal variability in the Eastern Atlantic and Western Indian Oceans: a) Season 1 and NE monsoon, b) Season 2 and ISW monsoon, c) Season 3 and SW monsoon, and d) Season 4 and INE monsoon.

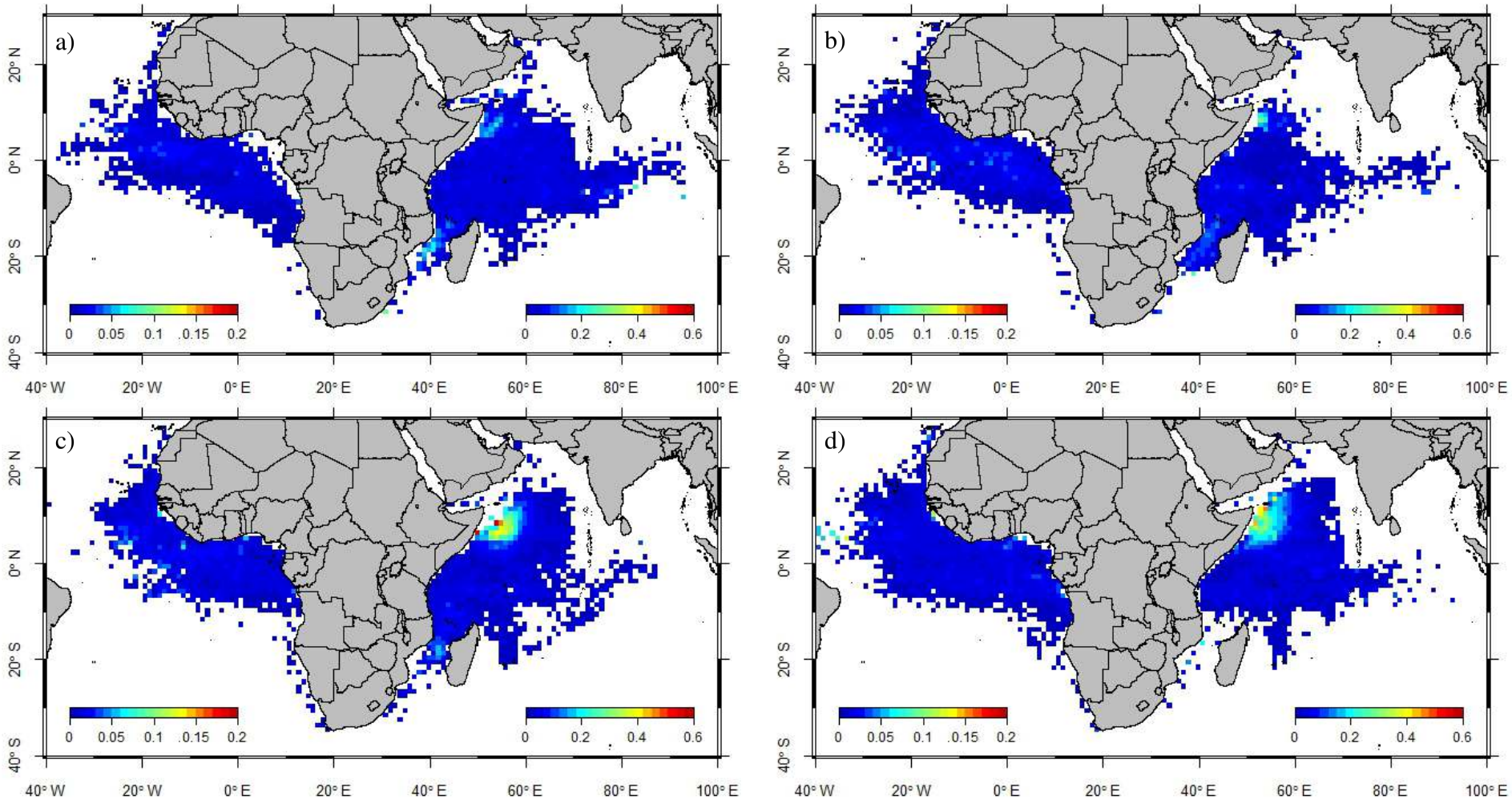


Figure F.4. Eddy Kinetic Energy ($m^2.s^2$) seasonal variability in the Eastern Atlantic and Western Indian Oceans: a) Season 1 and NE monsoon, b) Season 2 and ISW monsoon, c) Season 3 and SW monsoon, and d) Season 4 and INE monsoon.