Global Distribution Models for Whale Sharks

(Assessing Occurrence Trends of Highly Migratory Marine Species)

Ana Micaela Martins Sequeira

(M.Sc. Modelling the Marine Environment)

A thesis submitted to the

THE UNIVERSITY
of ADELAIDE

SCHOOL OF EARTH AND ENVIRONMENTAL SCIENCES

Australia

in fulfilment of the requirements for the degree of

Doctor of Philosophy

February 2013
(Quando eu era pequenina...)

- A minha mãe ensinou-me como explicar sempre o que eu queria.
- O meu pai ensinou-me que não podemos ter sempre o que queremos no momento em que queremos.

Completar o meu doutoramento teria sido impossível se eu não tivesse aprendido estas duas lições.

Esta tese é dedicada aos meus pais.

Translation:

(When I was a little girl...)

- My mum taught me how to always explain what I wanted.
- My father taught me that we cannot always have what we want, when we want it.

Completing my Ph.D. would never have been possible without these two lessons

I dedicate my Ph.D. to my parents.
TABLE OF CONTENTS

TABLE OF FIGURES .................................................................................................................. x

TABLE OF SUPPLEMENTARY FIGURES ................................................................................. xiii

APPENDIX B ......................................................................................................................... xiii

APPENDIX C ........................................................................................................................... xiv

APPENDIX D ........................................................................................................................... xiv

TABLE OF TABLES ............................................................................................................... xv

TABLE OF SUPPLEMENTARY TABLES ............................................................................... xvii

APPENDIX A ......................................................................................................................... xvii

APPENDIX B ......................................................................................................................... xvii

APPENDIX C ........................................................................................................................... xviii

APPENDIX D ........................................................................................................................... xviii

SUMMARY ................................................................................................................................ xix

STATEMENT OF ORIGINALITY ............................................................................................... xxi

ACKNOWLEDGMENTS .............................................................................................................. xxiii

PREFACE ................................................................................................................................. xxvii

CHAPTER I. GENERAL INTRODUCTION ............................................................................. 30

  Background ....................................................................................................................... 30

  Whale shark biology and ecology .................................................................................. 37

  Outline of the thesis ....................................................................................................... 44

CHAPTER II. Inferred global connectivity of whale shark populations ................................. 47

  Statement of authorship ............................................................................................... 49
# Table of Contents

- Overview of my research ................................................................. 196
  - A worldwide perspective .......................................................... 199
  - Better management ...................................................................... 201
  - Future improvements .................................................................. 203
  - Future directions ......................................................................... 209
- AFTERWORD .................................................................................. 216
- APPENDIX A – Inferred global connectivity of whale shark populations .... 217
- APPENDIX B – Ocean-scale prediction of whale shark distribution .......... 219
  - Receiving operator characteristic curve / area under the cure (ROC/AUC) .... 227
  - Boyce index comparison between MaxEnt and GLM/GLMM models ........ 228
  - BIOMOD, GBM and BRT models for the Indian Ocean ...................... 229
  - BIOMOD .................................................................................. 229
  - BRT / GBM ............................................................................. 233
- APPENDIX C – Spatial and temporal predictions of decadal trends in Indian Ocean whale sharks ......................................................... 236
  - Model predictors (Step 1) ............................................................. 236
  - Prevalence test (down-weighting pseudo-absences) ......................... 242
- APPENDIX D – Predicting current and future global distributions of whale sharks ................................................................. 243
  - BIOMOD – Global model ............................................................. 244
- ADDENDUM – Thesis amendments .................................................. 247
  - Outcome of thesis examination .................................................. 247
  - Letter from Examiner 1 ............................................................... 247
TABLE OF FIGURES

Figure 1: Diagram showing the components and main decision steps of species distribution modelling. Adapted from Franklin (2010) ........................................ 33

Figure 2: Whale shark (Rhincodon typus) (Smith, 1828) observed at Ningaloo Reef, Western Australia. ............................................................... 38

Figure 3: Whale shark distribution range and sighting reports known as of 2009. 43

Figure 4: Graphical illustration of the number of representative whale shark studies (since 1997). .............................................................. 58

Figure 5: Compilation of worldwide whale shark occurrences .......................... 60

Figure 6: Global sequence of whale shark monthly appearance timings (3 years presented) in different locations worldwide. .............................. 63

Figure 7: Global overview of published whale shark tracks .............................. 69

Figure 8: Whale shark migration patterns ...................................................... 76

Figure 9: Indian Ocean Tuna Commission (IOTC) datasets collected from 1991 to 2007 on tuna purse-seine fisheries in the Indian Ocean .......... 91

Figure 10: 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 ......................... 97

Figure 11: Habitat suitability of whale sharks (Rhincodon typus, Smith 1828) in the Indian Ocean during autumn ........................................... 100

Figure 12: Seasonal ensemble result for whale sharks (Rhincodon typus, Smith 1828) habitat suitability ......................................................... 107
Figure 13: Temporal variation in whale sharks sighted by the tuna purse-seine fisheries and comparison with climatic indices.................................................................122

Figure 14: Example of variation in fishing effort (days – top row) and number of whale sharks sighted (bottom row).....................................................................................123

Figure 15: Results of the fishing effort analysis across the western Indian Ocean where the purse-seine fisheries operated in more than one year.........................124

Figure 16: Partial effect of time on whale shark presence..................................................132

Figure 17: Partial effects of NINO4 and Time (after accounting for climatic contributions) in the probability of whale shark presence.................................133

Figure 18: Whale shark (Rhincodon typus, Smith 1828) datasets recorded by tuna purse-seiners .......................................................................................................................146

Figure 19: Predicted habitat suitability of whale sharks (Rhincodon typus, Smith 1828) in the Atlantic (left) and Pacific (right) Oceans during the months of April to June.................................................................151

Figure 20: Yearly variation in (a) number of whale sharks sighted, (b) relative whale shark occurrences, and (c) days spent fishing..................................................154

Figure 21: Partial effect of time on the probability of whale shark presence during the months of April to June.................................................................156

Figure 22: Whale shark sightings per unit effort during the months of April to June in the Atlantic (top row; 1981 - 2010) and Pacific (bottom row; 2000 - 2010)...157

Figure 23: Fisheries data derived from tuna purse-seine fishery logbooks in the Atlantic, Indian and western Pacific Oceans..................................................179

Figure 24: Boxplots for major environmental predictors within the area covered by the fisheries in each ocean. Upper and lower limits indicate the maximum and minimum values in each dataset.................................................................185
Figure 25: Predicted whale shark habitat suitability within the area covered by tuna purse-seine fisheries. ................................................................. 186

Figure 26: Global predictions of current seasonal habitat suitability for whale sharks. ........................................................................................................... 188

Figure 27: Predicted shift in global of whale sharks habitat suitability for 2070 under a no-climate-policy reference (no greenhouse gas emissions) scenario. ....... 189

Figure 28: Map of world organisations managing fisheries. ................................................. 210

Figure 29: Conditional modes of the random year intercept term for models including time as fixed effect (quadratic term; left), and without the fixed temporal trend (right). ................................................................. 293
TABLE OF SUPPLEMENTARY FIGURES

APPENDIX B

Figure B1: Bathymetry of the Indian Ocean as per the General Bathymetric Chart of
the Oceans (1 minute grid ~ 1.8 km). ...........................................................222

Figure B2: Seasonal chlorophyll a concentration (Chl a) obtained from averaged of
weekly SeaWiFS satellite composites at a 9 km spatial resolution from 1997
to the end of 2007. .......................................................................................222

Figure B3: 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. ...........................................................................................223

Figure B4: Probabilities (top panel) and resulting pseudo-absences locations (bottom
panel) generated by three different techniques. ...........................................223

Figure B5: Moran's I plots showing the reduction in spatial autocorrelation in the GLM
residuals. ......................................................................................................224

Figure B6: 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.......................................................................225

Figure B7: Seasonal habitat suitability of whale sharks (Rhincodon typus, Smith 1828)
in the Indian Ocean. .....................................................................................226

Figure B8: Whale shark autumn distribution in the Indian Ocean as a result of the
ensemble of models in BIOMOD.................................................................232

Figure B9: 3D interpretation of the two top ranked pairwise interactions found with
BRT.............................................................................................................234
Figure B10: Map with probability of whale shark occurrence in the Indian Ocean obtained by BRT. .................................................................235

APPENDIX C

Figure C1: Purse-seine fishing effort (in days) during autumn for each year from 1991 to 2007...........................................................................................................238

Figure C2: Habitat suitability of whale sharks (Rhincodon typus Smith 1828) in the Indian Ocean during autumn.................................................................239

Figure C3: Correlation between climatic predictor variables. ..............................240

Figure C4: Partial effect of the Indian Ocean Dipole (IOD) on the probability of whale shark sightings.................................................................241

Figure C5: Partial effect of time on whale shark presence, using a presence:pseudo-absence ratio of 0.5 (prevalence). .................................................................242

APPENDIX D

Figure D1: Global sea surface temperature averages from April to June..........243

Figure D2: Comparison of GLMM and BIOMOD ensemble prediction for the area covered by the tuna fisheries (mean probability). .................................246
TABLE OF TABLES

Table 1: List of published whale shark tracks from the Indian Ocean.........................64
Table 2: List of published whale shark tracks from the Pacific Ocean and Indo-Pacific area...............................................................................................................66
Table 3: List of published whale shark tracks from the Atlantic Ocean...................67
Table 4: Summary of generalised linear models relating probability of whale shark (Rhincodon typus, Smith 1828) occurrence to ocean properties.................99
Table 5: Comparison of results for the MaxEnt and generalised linear (GLM) and mixed effects (GLMM) models. ..............................................................103
Table 6: Summary of the MaxEnt models relating probability of whale shark (Rhincodon typus, Smith 1828) occurrence to individual ocean properties. 104
Table 7: Summary of the generalised linear mixed-effects models relating probability of whale shark occurrence to spatial and temporal predictors. ..............130
Table 8: Summary of spatial generalised linear models (step 1) relating probability of whale shark (Rhincodon typus, Smith 1828) occurrence to ocean properties. .....................................................................................................................152
Table 9: Summary of the temporal generalised linear mixed-effects models (step 2). .....................................................................................................................155
Table 10: Estimated weight of evidence for each temporal predictor used in the generalised linear mixed-effects models: Hsuit - spatial predictor derived from the spatial distribution models, effort - temporal variation in fishing effort and a linear and quadratic term for time (years). ...............................................155
Table 11: Number of whale shark sightings in each ocean per decade as recorded by tuna purse-seine fisheries. .................................................................177
Table 12: Summary of generalised linear mixed-effects models relating probability of whale shark occurrence to ocean properties in the Atlantic, Indian and western Pacific during the season of peak occurrences for each ocean. 181
TABLE OF SUPPLEMENTARY TABLES

APPENDIX A

Table A1: Summary of representative whale shark studies (since 1996) included in Figure 4 ........................................................................................................217

APPENDIX B

Table B1: Summary of the MaxEnt models relating probability of whale shark
(Rhincodon typus, Smith 1828) occurrence to individual ocean properties. .219

Table B2: Comparison of the Area Under the Curve (AUC) and Kappa results for the two modelling approaches used in chapter III: MAxEnt and GLM/GLMM for each of the pseudo-absence dataset generated method (P/A) and for each season: Autumn (Aut), Winter (Win), Spring (Spr) and Summer (Sum). ......227

Table B3: Boyce’s index for the three pseudo-absence generation techniques (P/A) for each modelling technique (MaxEnt and GLM) for autumn. Standard deviation was always < 0.001 ........................................................................................................228

Table B4: BIOMOD modeling summary .................................................................229

Table B5: Best model selection in BIOMOD .........................................................230

Table B6: Ranking of variable importance for each model considered within the BIOMOD platform. .................................................................231

Table B7: Analysis of pairwise interactions between variables using BRT ..........233
APPENDIX C

Table C1: Treatment given to each explanatory variable (standardization) used to obtained the spatial predictor (SpatialP) (Sequeira et al., 2012) for the first step of our temporal models; SST – sea surface temperature; Chl a – chlorophyll a.

APPENDIX D

Table D1: BIOMOD modelling summary for the runs with the global dataset.
Table D2: Model selection in BIOMOD.
Table D3: Ranking of variable importance for the global model considered within the BIOMOD platform.
SUMMARY

The processes driving distribution and abundance patterns of highly migratory marine species, such as filter-feeding sharks, remain largely unexplained. The whale shark (Rhincodon typus Smith 1828) is a filter-feeding chondrichthyan that can reach > 18 m in total length, making it the largest extant fish species. Its geographic range has been defined within all tropical and warm temperate waters around the globe. However, even though mitochondrial and microsatellite DNA studies have revealed low genetic differentiation among the three major ocean basins, most studies of the species are focussed on the scale of single aggregations. Our understanding of the species’ ecology is therefore based on only a small proportion of its life stages, such that we cannot yet adequately explain its biology and movement patterns (Chapter I). I present a worldwide conceptual model of possible whale shark migration routes, while suggesting a novel perspective for quantifying the species’ behaviour and ecology. This model can be used to trim the hypotheses related to whale shark movements and aggregation timings, thereby isolating possible mating and breeding areas that are currently unknown (Chapter II). In the next chapter, I quantify the seasonal suitable habitat availability in the Indian Ocean (ocean basin-scale study) by applying generalised linear, spatial mixed-effects and maximum entropy models to produce maps of whale shark habitat suitability (Chapter III). I then assess the inter-annual variation in known whale shark occurrences to unearth temporal trends in a large area of the Indian Ocean. The results from the Indian Ocean suggest both temporal and spatial variability in the whale sharks occurrence (Chapter IV). Therefore, I applied the same analysis to the Atlantic and Pacific Oceans using similar broad-scale datasets. While the results for the Pacific Ocean were inconclusive with respect to temporal
trends, in the Atlantic Ocean I found preliminary evidence for a cyclic regularity in whale shark occurrence (Chapter V). In Chapter VI, I build a model to predict global whale shark habitat suitability for the present, as well as within a climate change scenario for 2070. Finally, Chapter VII provides a general discussion of the work developed within this thesis and presents ideas for future research.
STATEMENT OF ORIGINALITY

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below*) resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University’s digital research repository, the Library catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Ana Micaela Martins Sequeira
Adelaide 27 February 2013


ACKNOWLEDGMENTS

Thanks to the Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) and European Social Funds (Ph.D. scholarship SFRH/BD/47465/2008), and to the University of Adelaide for funding my Ph.D. project. Also, thanks to the Australian Institute of Marine Science, Flying Sharks, Russel and Baudinette Travel Scholarship, and Society for Conservation Biology for additional funding / support.

I am thankful to my supervisors for the opportunity to undertake this project and for their support:

- Corey Bradshaw, who (in first place) believed in me and in my capacity to undertake this project, and fully supported me throughout (including overcoming my portuguenglish writing) always in a positive and encouraging way

- Mark Meekan, who made sure I had enough time to think, re-think and think again, and then finally be certain, that I wanted to do this Ph.D., and only then allowed me the opportunity to live unique “whale shark moments”

- Camille Mellin, who joined in as my Ph.D. supervisor after my arrival to Adelaide, and turned out to be my research “pillar” always with supportive and friendly advice.

Thanks to all the fishery-vessel owners (pole and line and purse-seiners) for access to log-book data used throughout my Ph.D. and to the Indian Ocean Tuna Commission and Institut de Recherche pour le Développement (France), the Azores Fisheries Observer Programme and the University of the Azores (Portugal), the Inter-American Tropical Tuna Commission and the Secretariat of the Pacific Community – in particular, A. Fonteneau, M. Herrera, R. Pianet, P. Afonso, M. A. Gaspar, M. Hall and P. Williams for extraction of the data. Also to D. Rowat for initial assistance with access to data.

Thanks to everybody involved in the whale shark field trips (2009 – 2012) for their help, partaking of knowledge and sharing of wonderful “whale shark moments”. Special thanks to: Iain Field, Michelle Thums, Kim Brooks and Terry Maxwell.
Thanks to the staff members from all the institutions involved, who in one way or another contributed and helped me throughout my Ph.D. In particular: Michelle Picard (for friendly support), Janelle Palmer, John Jennings, Louise Gillis, Maria Albanese and Lucy Zuzolo (from the University of Adelaide); Shona Pettigrew, Sharon Haswell, Susan Berry, Gheslaine Holmes and Mary Bell (from the Australian Institute of Marine Sciences), and Nuno Lima (from Fundação para a Ciência e Tecnologia).

Thanks to the Global Ecology Lab group for help and support. A special thanks to: Steven Delean (for helpful assistance with statistics), Salvador Herrando-Pérez (my mate in the lab), Lochran Trail, Nerissa Haby and Stephen Gregory.

While in Adelaide, my friends (Nadja Klafke, Suzanne McAllister, Kym Abrams, Esther Breed, Maria Marklund and Julian Felton), the King family (Gordon, Adrienne, Nicholas, Karen, Michael, Simone and Kyra – now part of my family as well) and the South Coast Wing Chun Kung Fu crew, kept the experience “human” ;-) and provided friendly and happy moments.

From the other side of the planet (where my reality makes sense...) there is not enough space to thank all: my family and beloved friends!

But I cannot avert mentioning here:

- My parents: Lurdes and Artur - who take care of me so well (even from afar)
- My sisters: Sónia and Suzete - their believe in me gives me strength
- My friends: Sofia, Célia and Inês - they keep me alive instead of just living

Thanks for being here, there... every time everywhere!

João Nuno (R.I.P.) thanks for dreaming the Australian dream with me - it became REAL!

Thanks for everything ♥Gary♥ (especially, for holding my hand).
As requested by the major funding organisation, here I include their logos:

Ph.D. scholarship SFRH/BD/47465/2008
PREFACE

“When you want something, the whole universe conspires to help you realise your desire” (Paulo Coelho)

My Ph.D. has a story! One which I cannot refrain from mentioning here - as I feel it belongs to this thesis as much as all the text, figures and tables.

It took me several years to ‘swap’ from ‘pure' chemistry (my background) to marine ecology, and ending up doing a Ph.D. on whale sharks. It was a winding road which involved working in several different environments: marine chemistry, oceanographic modelling, ecosystem modelling (linking hydrodynamics, to biogeochemistry and to species physiology), and finally, marine biology while working on connectivity of marine invasive species. During this process, my pursuit was focused on doing something about the global lack of knowledge on one of the most enigmatic (and impressive) species of the ocean. I wrote (and dropped) several drafts of possible research projects before the one I sent to Mark and Corey - which came back with positive feedback. After all my contacts, I finally got supervisors for my Ph.D. project. The hunt for funding followed...

Four years passed since the first email I sent to Mark until I finally got a scholarship which allowed me to start my Ph.D. in 2009. So no, the road was not easy. But in hindsight, I would not have been able to complete this Ph.D. if it had gone any other way. The experience I gained, the challenges involved in gathering new skills and working in different environments, the mind frame I got into before starting this Ph.D., and my perseverance were all vital to getting me here. They allowed the...
compilation of a comprehensive study on one species while still keeping a broader view on the wider applications of the study and tools developed. The ‘road’ would have felt longer and windier if I was alone - the credit for my perseverance does not belong just to me. I would like to formally thank all of those who supported my aspiration, listened to me *blabber* all these years, encouraged me whenever I needed it, and in the end, allowed me to believe that:

There is nothing more fulfilling than pursuing our dreams!

This thesis is the first outcome of my pursuit.

Note: This photo was taken during my first whale shark field trip in Ningaloo, Western Australia, when I wore for the first time the t-shirt I received as a present from my sister Sonia (after starting my Ph.D.), I include it here as an ode to all of those with whom I exchanged paths, and whom enriched me as a person.
Global Distribution Models for Whale Sharks

(Assessing Occurrence Trends of Highly Migratory Marine Species)
CHAPTER I. GENERAL INTRODUCTION

Background

The oceans cover 70% of the Earth’s surface, and hence play a vital role in the water cycle and climate regulation (Chahine, 1992). Global changes to this system are now affecting the hydrological cycle (Stocker & Raible, 2005), with major, if largely unpredictable, consequences for ocean life (Behrenfeld et al., 2006). Despite the plethora of ways in which humans exploit the marine environment, the marine realm is largely under-studied and its complexities remain therefore mysterious (Richardson & Poloczanska, 2008). This is mainly due to the physical and economic constraints of collecting data in the marine realm (Richardson & Poloczanska, 2008); for example, broad-scale biological data collection only became possible after the recent developments on satellite-sensor technology that allowed detection of changes in ocean colour associated with phytoplankton abundance (Sambrotto, 1999). However, this technology applies only to the sea surface, and so most of the ocean is still largely unsampled. The latter also applies to sampling marine species from higher trophic levels, which still relies on expedition surveys. In spatial ecology (Tilman & Kareiva, 1997), the distribution of species can be interpreted as a match between their ecological requirements and the environmental conditions available. Therefore, the difficulties of sampling the marine environment (through spatial surveys) hamper detection of the relevant spatial coverage to interpret these associations correctly. Modelling approaches are used to assist the understanding of the relationships between fundamental biological conditions (or processes) and species occurrence - as I discuss below.
The distribution of marine populations is also strongly associated with movement, whether through migration (e.g., herring; Dickey-Collas et al., 2009) or through larval dispersal (e.g., reef fishes; Jones et al., 2009). This leads many marine ecology studies to embrace and test metapopulation concepts and hypotheses (Hanski, 1998; Roughgarden et al., 1985), even if not specifically acknowledged. Metapopulation dynamics are the processes associated with population turnover, i.e., extinction versus colonisation (based on Levin’s model from 1969), between patchily distributed subpopulations with some degree of connectivity between them (Hanski & Gilpin, 1991). In terrestrial systems, this theory has been widely applied to a range of species including butterflies (Hanski, 1994), rabbits (Nielsen et al., 2008), rats (O’Brien et al., 2008) and bears (Hellgren et al., 2005). Indeed, fragmentation of terrestrial habitats makes the application of applied metapopulation theory highly relevant. In the marine environment, metapopulation concepts have been applied mostly to invertebrates and reef fishes (Sale & Kritzer, 2003); however, testing hypotheses at broader spatial scales might not be suitable or even practical for many marine species that are either highly migratory or widely dispersing (Kritzer & Sale, 2006).

Despite our limited understanding of oceans relative to the terrestrial realm, global declines of commercially important fish stocks have been reported for years (e.g., Pauly et al., 2005), and climate- and fishing-driven shifts in the distribution of marine species are now being detected (Cheung et al., 2010; Dulvy et al., 2008; Edwards & Richardson, 2004; Hiddink & ter Hofstede, 2008; Perry et al., 2005). Together with evidence that marine biodiversity improves the ocean’s resilience to disturbance (Folke et al., 2004), it is imperative that studies are implemented to determine the underlying mechanisms driving these changes, as well as quantifying the magnitude of their impacts and identifying possible solutions.
To that end, more and more modelling tools are being developed and used to understand how the marine environment functions and how it might change in the near future. These range from hydrodynamic models that emulate ocean circulation or ocean-atmosphere interactions (Bush & Philander, 1998; Lazure & Dumas, 2008; Penven et al., 2001; Schmittner et al., 2002; Woods, 1985), to ecosystem models constructing marine trophic web (Baretta et al., 1995; deYoung et al., 2004), estimating coastal carrying capacity (Ferreira et al., 2008) and guiding sustainable management (Sequeira et al., 2008). Another class of model – ‘species distribution models’ (also known as ‘habitat suitability models’ or ‘resource selection functions’) – that has been applied mainly in terrestrial systems (Lehmann et al., 2002) are now increasingly being used in the marine environment (Elith & Leathwick, 2009; Robinson et al., 2011). These models combine information on species occurrence with concurrent environmental conditions to define a species’ ecological niche (Kearney & Porter, 2009), which can then be used to predict the probability of occurrence where no relevant biological information is available (Robertson et al., 2003). Species distribution models are therefore irreplaceable tools to assist resource management and conservation planning, such as in designing marine reserves (e.g., Beger & Possingham, 2008), assessing the consequences of climate change to biodiversity (Thuiller, 2007), or in determining species richness and abundance (e.g., Mellin et al., 2010b).

Franklin (2010) schematised the main requirements of species distribution models (Figure 1) – these include (i) the hypothesis to test (e.g., species occurrence depending on specific environmental conditions), (ii) data on species occurrence and assessment of their integrity, (iii) information on environmental variables within the study area at a relevant resolution, (iv) choice of a modelling method (dependent on
the previous components), (v) validation procedures and (vi) mapping tools (Elith & Leathwick, 2009; Franklin, 2010).

Figure 1: Diagram showing the components and main decision steps of species distribution modelling. Adapted from Franklin (2010).

The data on species occurrence first needs to be assessed for its integrity which involves determining the data type (occurrences and absences data), if there is any bias associated with the sampling, and the probability of detection of the species within the time frame of the sampling scheme. Biogeographical environmental data assist identifying the species ecological niche, and is needed as a digital map covering an area (i.e., a digital layer). These data are also useful at the stage of mapping the predicted occurrences. With both species occurrence and environmental data, the modelling procedure can take place. It typically involves the choice of modelling methods, variables to include, response variable and decision on any transformations needed in the data. Any bias derived from sampling procedures should be included in the models (e.g., through the use of ‘offsets’ or weighting the variables according to the known bias). Modelling validation can take different forms and be both quantitative (assess model fit and accuracy) and qualitative (compare the response results with real data – species data). An important step when the model’s objective is to explain the influence of environmental predictors in the occurrence of a species is to assess spatial autocorrelation. Nearby locations

NOTE:
This figure/table/image has been removed to comply with copyright regulations. It is included in the print copy of the thesis held by the University of Adelaide Library.
tend to have similar properties, and are likely to influence each other or be influenced by the same processes in one same pattern. This can affect the precision of the coefficients estimated and bias the variable selection by the model. Spatial correlograms of the residuals can help detecting spatial autocorrelation and some strategies can be used to reduce it (e.g., choosing observations that are within a pre-determined distance of others). When building the prediction map, an important step is the definition of a threshold to separate presence locations from absences. This map can be visually compared with real data or expert knowledge. Usually this map is built for the sampled area, although in some instances it might be useful to project outside the sampled area – extrapolation in space and time. This step should be careful, as when extrapolating to outside the range of the values for which the model was fit, the predictions can be erroneous (this step is indicated with a dashed line at the bottom of the scheme).

Addressing each of these model components can be challenging (Araújo & Guisan, 2006). For example, selecting the ‘right’ model depends on the nature of both the response variable (occurrence of species, such as whether presences, absences or both are available) and predictors (environmental variables) (Araújo & Guisan, 2006). Different model constructs are available depending on data types: box classifiers (e.g., BIOCLIM), machine learning (e.g., MaxEnt) (Phillips et al., 2004) and statistical/ regression methods (e.g., generalised linear or additive models) (Hastie & Tibshirani, 1986; Nelder & Wedderburn, 1972). Different methods also result in different performances (summarized in Franklin, 2010), but there is general agreement that regression methods using presence and absence data tend to provide more realistic projections (Brotons et al., 2004; Elith et al., 2006; Zaniewski et al., 2002). With the increasing pressures on biodiversity from global change, species distribution models have now been coupled to climate forecasting models to predict future shifts in species occurrence and habitat changes (Araújo et al., 2005; Cheung et al., 2010; Thuiller, 2004).

Shifts in the distribution of marine species associated with global changes (Hoegh-Guldberg & Bruno, 2010; Sumaila et al., 2011; Wernberg et al., 2011) have now been described for a range of taxa, from primary producers such as plankton (Beaugrand et al., 2009; Edwards, 2004; Southward et al., 1995), macroalgae
(Wernberg et al., 2011), and filter feeders (Berge et al., 2005), to grazers (Rose, 2005; Sabatés et al., 2006) and predators (e.g., cod; Beaugrand & Kirby, 2010; Dulvy et al., 2008; Perry et al., 2005). Moreover, modelling forecasts show that similar shifts are to be expected for other species (Cheung et al., 2010), along with possible extinctions (Caldeira & Wickett, 2003; Thuiller et al., 2004a).

Assessment of climate-induced shifts on the distribution of marine predators such as sharks has to date been hindered by our limited knowledge of their movement patterns and their susceptibility to environmental shifts. Nevertheless, many shark species are in decline, largely as a result of over-fishing (Baum et al., 2003; Field et al., 2009; Musick et al., 2000; Schindler et al., 2002). In some countries, shark landings have even surpassed those of most other commercially exploited species (e.g., Sri Lanka) (Barker & Schluessel, 2005). Due to their slow life histories (delayed maturity, long life spans, few offspring) sharks are particularly vulnerable to exploitation (Baum et al., 2003; Field et al., 2009; Garcia et al., 2007; Hutchings et al., 2012; Musick et al., 2000; Schindler et al., 2002). Consequently, many recent management strategies, increasingly reliant on good biological information and model projections, have been devised to maximise the long-term persistence of shark populations in the face of these pressures (Barker & Schluessel, 2005).

Appropriate management of shark populations is also a commercially lucrative ambition because of the high revenue some shark species bring to the tourism industry (Vianna et al., 2012). Sharks are often the subject of major tourism ventures, such as shark watching and diving (Anderson & Waheed, 2001; Brunnschweiler, 2010; Catlin & Jones, 2009), and some have argued that certain species are worth more alive than dead – a recurrent source of profit instead of a one-off income (Topelko & Dearden, 2009; Vianna et al., 2012). Moreover, sharks are also thought to play a key
role in controlling the food web structure and the dynamics of marine ecosystems (Frid et al., 2008; Myers et al., 2007; Scheffer et al., 2005). Despite the importance of this taxon, the basic biology and ecology of most sharks are still poorly quantified. For example, the economically important (tourism, fishing) whale shark – the largest fish species – is still largely unstudied. There is some evidence that whale sharks can travel across oceans, which raises the possibility of applying concepts and testing hypotheses associated with metapopulation theory. Depending on true connectivity (currently unquantified) of whale shark sub-populations worldwide, they might actually be part of a unique global meta-population, such that localised perturbations might have detrimental flow-on effects to other sub-populations in distant waters. Whale shark distribution is also strongly associated with temperature; they are generally found in warm and temperate seas (Last & Stevens, 2009) and spend most of their time at the surface (Gunn et al., 1999; Sleeman et al., 2010a; Sleeman et al., 2010b; Wilson et al., 2006) within a narrow range of temperatures (Sequeira et al., 2012). This limited temperature range suggests that their time in deeper waters is limited by thermoregulatory constraints (Thums et al., 2012) or minimises the time spent in deep, oxygen-poor water (Graham et al., 2006). Changes in water temperature associated with climate change (Hoegh-Guldberg & Bruno, 2010), are thus likely to change whale shark distribution or abundance as they re-equilibrate to new temperature regimes and adapt to changing prey distribution (Gutierrez et al., 2008).

Being filter feeders, whale sharks are also likely to be affected by the expected climate-driven changes in ocean productivity and food-web dynamics (Hoegh-Guldberg & Bruno, 2010). Plankton distribution is expected to shift and its abundance will likely change in response to the thermal properties of changing oceans (Reygondeau & Beaugrand, 2011). This logically suggest that such changes will
propagate through to higher trophic levels (Kirby & Beaugrand, 2009). This led Hayes et al. (2005) to refer to plankton monitoring programmes as “sentinels” to identify changes in marine ecosystems. Whale sharks have the capacity to travel large distances, so altered plankton distribution will repeatedly affect whale shark migration patterns (Chin et al., 2010). Moreover, whale sharks aggregations are usually associated with specific productivity events, such as coral spawn in Ningaloo, Western Australia (Taylor, 1996) that promotes an increase in zooplankton production. Changes in the timing of these events, as well as other impacts of climate change on coral reefs (e.g., habitat destruction within whale shark feeding areas), might therefore also impact whale shark occurrence (Chin et al., 2010).

Despite whale sharks spending the majority of their time at the surface (Gunn et al., 1999; Rowat et al., 2007; Thums et al., 2012), they can also dive deeply. This opens speculation to the potential for this species to use different depths (where water is cooler) as surface water temperatures increase. However, oxygen limitation might pose a problem for extended time spent in deeper waters (Pörtner, 2001; Prince & Goodyear, 2006). For most of the reasons given above, whale sharks have been referred to “as potentially the most vulnerable species [to climate change] in the pelagic ecological group” (Chin et al., 2010; in their assessment of the vulnerability of sharks and rays in the Great Barrier Reef), and present a challenging case study in the context of climate change and expected shifts in marine species distribution.

**Whale shark biology and ecology**

Whale sharks (*Rhincodon typus*) (Smith, 1828) are the largest of all fish (Figure 2), with records reporting animals as large as 18.8 and 20 m total length (Borrell et al.,
2011; Chen et al., 1997). However, due to difficulties associated with measuring such a large animal (Wintner, 2000), and lack of specification of the methods used in some studies reporting individual lengths, a commonly accepted maximum size is around 12 m (Stevens, 2007). The time it takes for a whale shark to reach such a size is still unknown. Some studies suggest that growth rates are slow (~20 to 34 cm year⁻¹) (Wintner, 2000), concluding that the species must be long-lived. In fact, a life span of 60 to 100 years has been suggested for this species (Pauly, 1997), and maturity is expected at lengths ≥ 9 m (Norman & Stevens, 2007).

A large body size, long life span and probable late maturity are characteristic traits of a K-selected ('slow') life history (Stearns, 1976). Species with such traits are also expected to produce only a few offspring. Despite the limited data on whale sharks reproduction, there is evidence this might not be the case for whale sharks. Most of what is known about whale shark reproduction derives from a single pregnant

NOTE:
This figure/table/image has been removed to comply with copyright regulations. It is included in the print copy of the thesis held by the University of Adelaide Library.

Figure 2: Whale shark (Rhincodon typus) (Smith, 1828) observed at Ningaloo Reef, Western Australia.
Photography by Wayne Osborne from the Australian Institute of Marine Science.
female caught in Taiwan back in 1995 (Joung et al., 1996). When this specimen was
examined, more than 300 embryos were found (1:1 female:male ratio), possibly all
fertilised by the same male (Schmidt et al., 2010). They were at different stages of
development with some in egg cases with yolk sac (~ 40-50 cm), and others ready to
be born, free from egg cases and yolk sacs (~ 60 cm). This discovery led to the
conclusion that the species is ovoviviparous, and possibly monoandric and with
capacity to store sperm (Schmidt et al., 2010). Monoandry is usually observed when
mating is only occasional, which might be the reason behind our still limited knowledge
on mating, reproduction, and early life stages. Observation of neonates (≤ 1 m) is also
rare, with only a few records (<20; Rowat & Brooks, 2012) mostly caught by fishers in
the open ocean (e.g., Wolfson, 1983). Despite the rarity of such occurrences, reports
of neonates span the three major oceans (Rowat et al., 2008; Wolfson, 1983).

Whale sharks are found in the Atlantic, Indian and Pacific Oceans. Being
poikilothermes, their body temperature depends on the external water temperature,
and their range of occurrence is considered to be largely restricted to tropical and
temperate waters, between latitudes of 30° north and south (Rowat & Brooks, 2012).
Despite their widespread occurrence, they are mostly observed in specific near-shore
locations where they form seasonal aggregations for reasons that remain unclear, e.g.,
in Ningaloo (Taylor, 1996), Gulf of Mexico (Hoffmayer et al., 2007) or Gulf of California
(Nelson & Eckert, 2007). Several studies have focused on monitoring movement
patterns starting at each of these locations (Eckert & Stewart, 2001; Hueter et al.,
2008; Wilson et al., 2006). However, only a few have captured movement beyond the
aggregation areas (Eckert & Stewart, 2001; Hueter et al., 2008; Rowat & Gore, 2007),
and even fewer have re-sighted the same shark in subsequent seasons (Graham et
al., 2006; Rowat et al., 2011). The few tracks showing movement beyond aggregations have revealed that whale sharks are possibly able to travel between ocean basins (Eckert & Stewart, 2001), but due to the paucity of tracking results obtained to date, data are still insufficient to confirm whether there is a general, recurrent movement of this species between major ocean masses. Although whale sharks spend most of their time at the surface (Gunn et al., 1999), they can frequently ‘dive’ deeply (> 1000 m) (Brunnschweiler & Sims, 2012; Rowat & Gore, 2007). The reasons behind their movement behaviour are yet to be understood, but some have suggested they might be associated with foraging and feeding (Brunnschweiler et al., 2009; Sleeman et al., 2010a).

Aggregations of whale sharks usually coincide with higher concentrations of plankton in surrounding waters (Heyman et al., 2001; Sleeman et al., 2007). Being filter feeders, they generally only consume small secondary grazers (e.g., krill or crab larvae; small fish) (Meekan et al., 2009; Wilson & Newbound, 2001), and therefore other species are commonly associated with them, including commercially valuable anchovy (Duffy, 2002) and herring (Wilson, 2002). This characteristic has been particularly helpful in oceanic tuna fisheries where whale sharks are used to spot tuna (Matsunaga et al., 2003) commonly aggregating under the shark (possibly foraging on similar prey) – in other words, whale sharks are used as fish aggregating devices.

Despite the evidence for their ability to embark on large migrations, some whale sharks are also commonly resighted within the same near-shore locations in different years (Norman, 2007), such as in Ningaloo (Holmberg et al., 2008; Taylor, 1996), the Seychelles (Rowat et al., 2009b), and Belize (Graham & Roberts, 2007). However, most of the sharks seen at these aggregations are immature males. Females are observed mostly in the Gulf of California (Eckert & Stewart, 2001;
Ramírez-Macías et al., 2007), and the Galapagos (Green, 2011), but they possibly also occur in South East Asia (Taiwan; Joung et al., 1996), and where a 46-cm newborn whale shark was recently seen (Philippines; Aca & Schmidt, 2011). Besides this possible sex segregation, separation by age or size has also been postulated (Ramírez-Macías et al., 2007; Rowat et al., 2011).

Due to their predictable aggregation behaviour, together with their harmless characteristics, whale sharks have become of economically important in several coastal locations around the world. Since 1993, when they first became the attraction of tourism in Ningaloo, Western Australia (Davis et al., 1997), this source of income, ranging from US$5 million (over a 14-week season in the Seychelles; Rowat & Engelhardt, 2007) to US$50 million (per year in Ningaloo; ABC News, 2005), has spread throughout aggregation areas. Whale shark-dedicated tourism now occurs in the Philippines (Quiros, 2005), Seychelles and Djibouti (Rowat et al., 2011), Mozambique (Pierce et al., 2010), Gulf of Mexico and Caribbean Sea (Heyman et al., 2010; Quiros, 2005), and Gulf of California (Cárdenas-Torres et al., 2007).

In the past, whale sharks have been targeted by fisheries mostly in southeast Asia where a single whale shark fin was valued at US$316 (White & Cavanagh, 2007). Due to declining landings (e.g., Chen & Phipps, 2002), commercial fisheries for whale sharks are now banned, with the last ban occurring in Taiwan as recently as 2008 (Council of Agriculture, 2007) (Articles 44 and 46 of the Fisheries Act, www.fa.gov.tw/en/LegalsActs/index.aspx). However, some of these fisheries continued their activity illegally (Riley et al., 2009). For a long time, whale sharks have also been used to assist the catch of commercially valuable fishes (Iwasaki, 1970). But while acting as fish aggregating devices, they are also caught (by-catch) in the tuna purse-seine nets (Matsunaga et al., 2003). Although they are commonly released from these
nets (Chassot et al., 2009), the extent of damage is unknown and survival rates have never been assessed.

Since 1990, whale sharks have been listed on the IUCN Red List of Threatened Species (IUCN, 2010), and since 2000 are classified as Vulnerable (i.e., at risk of extinction in the wild). Almost 200 years after the species was first described, basic knowledge of these sharks is still lacking, as is our understanding of the consequences of human activities on their long-term persistence. While impacts from targeted fisheries soon became evident through declining landings, the effects of current fishing (by-catch and illegal) are difficult to measure. The expansion of tourism at both new and previously known aggregations, might also have detrimental effects (Heyman et al., 2010), especially if management is poor.

Changes observed in whale shark populations from near-shore locations have been associated with a general decline (Bradshaw et al., 2008). With the broad-range of occurrence of whale sharks, pressure to develop large-scale conservation measures has increased (Rowat, 2007), however the currently unknown connectivity among aggregations (Figure 3) makes the assessment of the regional conservation status of whale sharks challenging. Recent genetic evidence demonstrated high connectivity between whale sharks from populations among different oceans (Castro et al., 2007; Schmidt et al., 2009); hence, one can hypothesise that whale sharks form a single, global population at least over generational scales. Most whale shark studies to date have focused on single aggregation locations, and are therefore strongly biased towards specific size, gender and habitat classes. Consequently, they have failed to capture global patterns of whale shark population ecology and connectivity.
Figure 3: Whale shark distribution range and sighting reports known as of 2009.

Black dots represent locations where whale sharks seasonally aggregate. N: free swimming neonates < 1m (Rowat et al., 2008; Wolfson, 1983); E: large female with 300 embryos (Joung et al., 1996); J: group of juveniles (~ 1 m) swimming with an adult whale shark (Pillai, 1998).
Outline of the thesis

My thesis describes and quantifies the global patterns of whale shark population distribution and connectivity for the first time, contextualising them within the theme of global change. Besides presenting a challenging case study in the context of climate change and likely shifts in marine species distribution, the elusive whale shark also represents a good model to understand biotic responses to oceanographic patterns; being a filter feeder, they should track primary production more closely than higher-order predators (Grémillet et al., 2008).

The main component of my thesis begins with formulating a hypothesis on the possible global connectivity of whale sharks. This is presented as a review (Chapter II) where I examined from a global perspective all of the whale shark literature and distributional data available to date, and then summarised this in a conceptual model of global whale shark connectivity. In this review I pointed out the major gaps in knowledge, and suggested some ways to fill these gaps. Later, I validated or explored these hypothetical links in subsequent chapters (Chapters III, IV, V and VI).

With most publications and data available for the Indian Ocean, I began here for my global analysis. In my first data chapter (Chapter III), I predicted the spatial and seasonal patterns in whale shark distribution within the Indian Ocean based on fisheries-collected observational data. Following this spatial analysis, and the current concerns about declining abundance (Bradshaw, 2007; Bradshaw et al., 2008), I assessed temporal trends in the Indian Ocean whale shark population, and compared the variation in whale shark sightings with the variation in climatic signals (Chapter IV). To test if my hypotheses regarding global connectivity are biologically realistic on a temporal scale, I analysed both the spatial and temporal patterns of whale shark occurrence also in the Atlantic and Pacific Oceans, again based on fisheries-collected
observational data (Chapter V). My goal here was to look for asynchronies in the probability of whale shark occurrence in the different oceans. This is to be expected if whale sharks are travelling between oceans. In the last data chapter (Chapter VI), I built a global distribution model to assess current global patterns of whale shark habitat suitability. Here I also projected whale shark distribution under a climate change scenario to provide the first quantitative prediction of how increasing sea surface temperatures might affect the distribution of this species. Finally, I revisit the general context in which this thesis is inserted to reveal the novelties of my approach. I point out how to overcome problems associated with incomplete, and often biased, data on highly migratory marine species, to generate appropriate statistical approaches which allow assessment of the possible climate change impacts on these species.

All my data chapters were written as ‘stand-alone’ manuscripts, and therefore some of the material presented in the Introduction and Methods sections are repeated in each chapter. These manuscripts are now published (Chapter III), in press (Chapter II and IV), submitted (Chapter VI) or in preparation for submission (Chapter V).

The main objectives of my Ph.D. were to:

1. provide a modelling framework to allow assessment of possible climate change impacts on wide-ranging elusive marine species (using the whale shark as a case study);

2. analyse the current knowledge about whale shark at a global scale and show how they might be connected within an hypothesised global meta-population;
3. assess the influence of environmental conditions on whale shark spatial
distribution at a regional and global scale;

4. determine temporal trends in whale shark occurrence in the three major
oceans and assess asynchrony in occurrences (as expected for a global
migrating population) or evidence of declining numbers; and

5. derive realistic predictions of this species current geographical range and
provide a modelling tool for exploring whale shark distribution under current
and future climate conditions.
CHAPTER II. Inferred global connectivity of whale shark populations

A. Sequeira\textsuperscript{a}, C. Mellin\textsuperscript{a,b}, M. Meekan\textsuperscript{c}, D. Sims\textsuperscript{d}, C. Bradshaw\textsuperscript{a,e}

\textsuperscript{a}The Environment Institute and School of Earth and Environmental Sciences, The University of Adelaide, South Australia 5005, Australia

\textsuperscript{b}Australian Institute of Marine Science, PMB No. 3, Townsville MC, Townsville, Queensland 4810, Australia

\textsuperscript{c}Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling Hwy, Crawley, Western Australia 6009, Australia.

\textsuperscript{d}Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

\textsuperscript{e}South Australian Research and Development Institute P.O. Box 120, Henley Beach, South Australia 5022, Australia

\textit{Journal of Fish Biology 2013; 82:2, 595 – 617}
INFERRED GLOBAL CONNECTIVITY OF WHALE SHARK POPULATIONS

Journal of Fish Biology 2013; 82:2, 595 – 617

SEQUEIRA, A.M.M. (Candidate)
Planned the article, performed the analysis, interpreted data, wrote the manuscript and acted as corresponding author.
I hereby certify that the statement of contribution is accurate.

Signed …………………………………………………………………………………Date………………

MELLIN, C.
Helped with data interpretation and provided critical evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed …………………………………………………………………………………Date………………

MEEKAN, M. G.
Supervised development of work and provided evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed …………… ................................Date………………
SIMS, D.W.
Provided evaluation of the manuscript.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed ……………………………………………………………………Date

BRADSHAW, C.J.A.
Supervised development of work, helped in data interpretation and with manuscript evaluation and writing.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed ……………………………………………………………………Date
Inferred global connectivity of whale shark populations

Abstract

Ten years have passed since the last synopsis of whale shark (Rhincodon typus, Smith 1828) biogeography. While a recent review of the species’ biology and ecology summarized the vast data collected since then, it is clear that information on population geographic connectivity, migration and demography of whale sharks is still limited and scattered. Understanding whale shark migratory behaviour is central to its conservation management considering the genetic evidence suggesting local aggregations are connected at the generational scale over entire ocean basins. By collating available data on sightings, tracked movements and distribution information, our review provides evidence for the hypothesis of broad-scale connectivity among populations, and generates a model describing how the world’s whale sharks might be part of a single, global meta-population. Whale shark occurrence timings and distribution patterns make possible a connection between several aggregation sites in the Indian Ocean. Our conceptual model and validating data lend support to the hypothesis that whale sharks are able to move among the three largest ocean basins with a minimum total travelling time of around two to four years. Our model provides a worldwide perspective of possible whale shark migration routes, and suggests a modified focus for additional research to test its predictions. Our framework can be used to trim the hypotheses for whale shark movements and aggregation timings, thereby isolating possible mating and breeding areas that are currently unknown. This will assist endeavours to predict the longer-term response of the species to ocean warming and changing patterns of human-induced mortality.
Introduction

The large and docile whale shark (*Rhincodon typus*, Smith 1828) is a filter-feeding chondrichthyan that can reach over 18 m in total length (Borrell *et al.*, 2011; Chen *et al.*, 1997; Compagno, 2001), making it the largest extant fish species. Its geographic range is more less known across all tropical and warm temperate waters (Compagno, 2001; Last & Stevens, 2009), especially around sites where individuals aggregate seasonally. Ironically, little is known about how these aggregations are connected via migration, apart from preliminary population genetic evidence suggesting connectivity at the generational scale (Castro *et al.*, 2007; Schmidt *et al.*, 2009). Further, their pelagic distribution is poorly described, as is the true extent of their range. Rising concerns about how warming seas will affect whale shark distributions are potentially being manifested already; for example, occasionally individuals are sighted at latitudes higher than their nominal 30 ° – in the Bay of Fundy, Canada 44 ° N (Turnbull & Randell, 2006), in the southern Azores islands, Portugal 41 ° N (Sa, 2008) or in the north-east of New Zealand > 35 ° S (Duffy, 2002). This lack of understanding of their distribution, connectivity and migration pathways therefore severely limits our capacity to broker international conservation plans for the species.

Whale shark sightings are most often reported near shore because they often aggregate in specific locations along the coast within warm/temperate waters at approximately the same time every year (*e.g.*, Rowat, 2007). The incentive for aggregating is not yet fully understood, but most studies suggest that such occurrences are related to
food blooms, currents or temperature variation (Cárdenas-Palomo et al., 2010; Colman, 1997; Kumari & Raman, 2010; Meekan et al., 2009; Sleeman et al., 2010b; Wilson, 2002).

In contrast, whale sharks are rarely recorded on the high seas, largely because of the challenges of sampling the ocean realm (Richardson & Poloczanska, 2008). Due to the drive to chase ever-more-elusive commercial fish species, the fishing industry seems to be the only sector capable of recording marine species regularly in the open ocean (Sequeira et al., 2012). Whale shark-targeted fisheries ran mainly in South-East Asia and India prior to the turn of the last century (Joung et al., 1996; Pravin, 2000; Stevens, 2007; White & Cavanagh, 2007). As a consequence of declining landings, the species was classified as Vulnerable in 2000 in the IUCN Red List (www.iucnredlist.org) and the fisheries were eventually banned – the last bans occurred in India in 2001 (as per the Indian Wildlife Protection Act, Schedule I, amended in 2001) and Taiwan in 2007 (Council of Agriculture, 2007). Even though most of the bans happened more than a decade ago, a recent review of this classification resulted in no change of status (IUCN, 2010). Together with the Taiwanese commercial catch lasting until 2007, the persistent high threat risk of the migratory whale shark (UNCLOS, 1982) might be due to a low scrutiny of the ban’s implementation (Stewart & Wilson, 2005), and continued illegal fishing in areas such as eastern Indonesia (White & Cavanagh, 2007) and the Maldives (Riley et al., 2009). The unintentional catch of whale sharks is also occurring in other fisheries of large ocean coverage (e.g., purse seiners; Romanov, 2002).

Whale shark-based tourism has proven the worth of live sharks (e.g., Anderson & Waheed, 2001), with related incomes replacing those once provided by targeted fisheries (Quiros, 2007). Established tourism operations continue to expand throughout the world at
known whale shark aggregation sites, such as in the Seychelles (Rowat & Engelhardt, 2007), Australia (Catlin & Jones, 2009) and Mexico (de la Parra Venegas et al., 2011), and they play an important conservation role because of the positive example of economic benefits from renewable tourism. Tourism also assists whale shark research because photographs taken during dive operations are used to identify returning individuals (Speed et al., 2007), thus providing data essential for estimating abundance, vital rates and persistence probability (Bradshaw, 2007; Meekan et al., 2006; Rowat et al., 2009a). By collating tourism-derived photographic data from different locations, research is currently pursuing evidence of inter-aggregation migration hypothesised to occur given the finding of ocean-scale genetic connectivity (Castro et al., 2007; Schmidt et al., 2009); however, none has yet been found (Brooks et al., 2010).

Examining whale sharks at the scale of single aggregations cannot adequately describe the species’ life history because it encapsulates only a small proportion of the life stages. Collecting data outside aggregation areas is therefore essential. Continual developments in tagging technology to improve estimates of home range size, movement patterns and habitat use (Hammerschlag et al., 2011) have been partially successful in this regard, although despite > 3000 whale shark tagged to date (e.g., www.whaleshark.org), trajectories have not revealed reliable evidence for inter-aggregation connection. This is not entirely surprising given the low probability of resighting migratory individuals in widely spaced aggregations. The lack of bio-logging efficiency is mostly due to premature detachment and limited spatial coverage of acquired data (e.g., Brunnschweiler et al., 2009), tag removal and/or damage (Fitzpatrick et al., 2006; Hays et al., 2007) which might result in part from attacks from other sharks or killer
whales (Fitzpatrick et al., 2006; Speed et al., 2008), and the accumulation of bio-fouling organisms causing tag malfunction (Hays et al., 2007). Moreover, even though tracking methods can provide some insight into the movement behaviour of whale sharks, they are still unlikely to encompass the full range of the population’s distribution.

To investigate putative links among aggregations, mitochondrial and microsatellite DNA studies have revealed that whale sharks from the three major ocean basins have low genetic differentiation (Schmidt et al., 2009), albeit sharks in the Atlantic Ocean had moderately different mitochondrial haplotype frequencies from those in the Indian and Pacific Oceans (Castro et al., 2007). These results partially validate the prediction that whale sharks have at least generational-scale migrations that connect populations among the world’s oceans. According to Hartl and Clark (1989), population differentiation can be prevented with only a few breeding migrants per generation, and panmixia can occur even with only about four breeding migrants per generation (Hartl & Clark, 1989; Morjan & Rieseberg, 2004). Given the species’ long generation time (ranging from 15 to 37 years; Bradshaw et al., 2007), only rare dispersal would be required to demonstrate equivalent gene flow. This suggests that current bottom-up approaches based on collecting difficult-to-obtain tagging data to estimate vital rates and life history traits have a low probability of characterizing broad-scale migratory patterns. Since Smith first described the species in 1828, we still know next to nothing about their physiology (e.g., growth rates) or reproduction (e.g., mating areas, variation in pup production, breeding frequency). The only clue that whale sharks are ovoviviparous was provided by a single pregnant female specimen carrying 300 embryos in different stages of development (Joung et al., 1996).
In this overview and analysis, we propose instead a top-down approach to infer whale shark occurrence probability and global patterns of movement. We suggest that the primary focus of research should be shifted to broader-scale approaches that generate hypotheses based on conceptual models of migration. We argue that testing these hypotheses indirectly through modelling approaches, or directly via improved technologies that provide higher-quality and longer-term migration data, will ultimately provide a more realistic picture of the species’ distribution throughout its life stages. This will assist in making predictions on population viability and redistribution resulting from warming oceans and changing patterns of human-induced mortality.

**Knowledge base**

Our understanding of whale shark ecology and biology has accelerated from an average of < 3 papers year\(^{-1}\) between 1992 and 2005 to a mean of 15.7 year\(^{-1}\) over the following 6 years (source: Thomson-Reuters ISI Web of Knowledge consulted 23 March 2011 using “whale shark” and “*Rhincodon typus*” as key words). A peak in the number of papers occurred in 2007 when a special issue on the subject was published (Fisheries Research, 2007) following the first International Whale Shark Conference in Perth, Western Australia (Irvine & Keesing, 2007).

Since whale sharks were first identified by Andrew Smith in Table Bay, South Africa (Smith, 1828), E. W. Gudger has still published the highest number of peer-reviewed papers on the species, mainly describing occurrences in Florida, the Gulf of California (Sea of Cortés), Seychelles, Galapagos, Hawaii, The Bahamas and the Caribbean Sea (as reviewed in Gudger, 1934), and was the first to attempt to portray
whale shark “migratory behaviour” partially as a function of ocean current patterns (e.g., Gudger, 1932, 1934). Iwasaki (1970) attempted to surmise the distribution of whale sharks based on data collected by Japanese tuna fishing vessels, from which he concluded that their occurrence around Japan was seasonal and depended mainly on the behaviour of the Kuroshio Current, temperature and prevailing winds. Uchida et al. (1984) suggested that whale sharks are annually regular in their appearance on the east coast of Taiwan, but that seasonality in presence on the west coast depended on temperature. Later, Taylor (1996) published the first paper describing the regular autumn aggregation of whale sharks at Ningaloo Reef, Australia, and associated it with possible zooplankton blooms following the coral spawn. Since then, Ningaloo Reef has been the site of the most targeted studies (Figure 4), even if much research has been done elsewhere relating mainly to occurrence, movement, tourism, mortality and feeding behaviour (Figure 4 and Table A1 in Appendix A). Few data are available on whale shark life-history traits such as reproduction (Chang et al., 1997; Joung et al., 1996; Schmidt et al., 2010) and growth (Pauly, 1997; Wintner, 2000), or internal biology (Dove et al., 2010; Wilson & Martin, 2001; Yopak & Frank, 2009), which reflects the difficulty of obtaining specimens for detailed examination and measurement. From the summary of recent studies per site (Figure 4; Table A1), the following trends emerge: (1) knowledge of whale shark ecology is most advanced in the Indian Ocean, and (2) similar work has been done in the different locations worldwide, but (3) mostly at the scale of individual aggregations.

Recent genetic studies (Castro et al., 2007; Schmidt et al., 2009) provide the only clue so far that the world’s whale shark populations are connected via dispersal occurring during some stage of their life. However, there has been no effort so far to synthesise all
the disparate information available into a conceptual model of dispersal and population connectivity, even after the conservation of whale sharks was highlighted as requiring both “regional cooperation and conservation initiatives” (Rowat, 2007).

Figure 4: Graphical illustration of the number of representative whale shark studies (since 1997). Grouped by (1) occurrence, population and correlates for occurrence, (2) movement, (3) tourism and conservation, (4) mortality and (5) foraging, in different locations worldwide. References provided in Table A1 in Appendix A.
Global occurrences

The first 76 recorded whale shark occurrences worldwide were compiled in the early 1930s by Gudger (e.g., Gudger, 1932). He suggested the Sulu Sea (between southern Philippines and northern Borneo) as the single location from which all whale sharks originate, dispersing from there depending on environmental conditions (Gudger, 1934). After Gudger, Wolfson and Notarbartolo-di-Sciara (1981) counted 345 records worldwide (including repeat sightings) until 1980. Currently, whale sharks have been recorded in the three major oceans spanning the Equator and have been seen, even if only occasionally, near the shore of over 100 countries in five continents (Figure 5) (Compagno, 2001; Martin, 2007; Rowat, 2007; Stacey et al., 2008).

Although existing photo-identification libraries have yet to report a match from sharks seen at distant locations, there is some evidence for short-distance movements. This is the case for resights of the same shark in Gladden Spit (Belize), Isla Contoy (Mexico) and Utila (Honduras) (Graham & Roberts, 2007), and a resighting in Mozambique of a shark tagged in the Seychelles 11 months earlier (Rowat & Gore, 2007).

Peaks in whale shark occurrence appear to happen synchronously in different locations around the world (Figure 6). For example, in January they occur in KwaZulu-Natal (South Africa), Djibouti and Christmas Island (Australia); in March to May, they occur at Gladden Spit (Belize), Gujarat (India), Ningaloo (Australia) and around the Philippines; in August to October, whale sharks have been reported in Portugal (around the Azores islands), Mozambique, Seychelles and Bahia de Los Angeles (Gulf of California, Mexico).
Figure 5: Compilation of worldwide whale shark occurrences. From Compagno (2001), Martin (2007), Stacey (2008) and Rowat (2007) overlaid with the IUCN (2001) distribution (hatched area). Square size represents size class, and colors represent gender (where available): blue = male, red = female and black = gender not specified. Square dots represent juvenile whale shark aggregations and the ★ represents the sighting of pregnant female with 300 embryos (Joung et al., 1996). Small black dots represent long-term (1980 to 2010) whale shark sightings from tuna fisheries.
However, when examining the timings over several years, a pattern of sequential monthly peaks at neighbouring locations emerges (Figure 6).

A highly skewed sex ratio is common in certain aggregations (e.g., Ningaloo, Seychelles, Mozambique, Belize) where mostly immature males are found (Graham & Roberts, 2007; Meekan et al., 2006; Rowat et al., 2011). Rowat et al. (2001) provided some evidence for size segregation, stating that individuals observed in the Djibouti aggregation (mainly immature males) are smaller than those from the Seychelles (and other locations in the Indian Ocean). They suggested that the Djibouti population might be a ‘staging’ group for other regional aggregations (i.e., segregation might therefore be another feature contributing to migration among sites). In the Gulf of California (Mexico), whale sharks segregate both by sex (Eckert & Stewart, 2001) and size, with larger adult females mainly seen in the southern part of the Sea of Cortés (Ramirez-Macias et al., 2007).

Smaller whale sharks (< 2.5 to 3 m total length) have been observed only rarely (Colman, 1997) and almost all neonates reported so far have been caught as by-catch in purse-seine fisheries (Wolfson, 1983). The smallest neonate ever recorded (46 cm total length) was recently found swimming freely in Sorsogon, Philippines (Aca & Schmidt, 2011), and the only fertility data available for the species is derived from a single, 10.6 m-long female caught off Taiwan containing 300 ‘ready-to-be-born’ embryos measuring 58 to 64 cm (Joung et al., 1996).
Tracking studies

Gunn et al. (1999) and Eckert and Stewart (2001) published the first whale shark tracking results derived from satellite tags deployed at Ningaloo Reef (Western Australia) and the Sea of Cortés (Gulf of California), respectively. Despite the > 3000 data-logging and satellite tags deployed to date (www.whaleshark.org, consulted 5 August 2011), we found only 109 whale shark tracks publicly available (Table 1, Table 2 and Table 3), with only 75 published in peer-reviewed papers.

Most of the tracking results available are from the Indian Ocean (47 tracks), with Ningaloo being the main tagging location. Next was the Pacific Ocean from where 30 tracks have been published (16 starting at the Sea of Cortés and 14 from the Galapagos Islands), followed by 22 in the Atlantic Ocean deployed mostly in Belize, and finally 10 in the Indo-Pacific area (from the Philippines, Malaysia and Taiwan). The sex ratio of these tracked individuals is 33 females:31 males:45 unknown (Tables 1-3). Average track duration is around 90 days, although they range from only a few hours (Gunn et al., 1999) to more than 3 years (Eckert & Stewart, 2001), with about 30 % of the tracks lasting < 1 month and only 2 lasting > 9 months. Distances travelled varied between 4 and 12,620 km and averaged about 1,545 km, with only 10 sharks travelling > 2,000 km.
Figure 6: Global sequence of whale shark monthly appearance timings (3 years presented) in different locations worldwide.

Table 1: List of published whale shark tracks from the Indian Ocean.
* indicates tracks recreated and displayed in Figure 7; numbers in superscript indicate the reference for each track where there are multiple references for the same tagging location.

<table>
<thead>
<tr>
<th>#</th>
<th>Length (m)</th>
<th>Sex</th>
<th>Tagging Date</th>
<th>Pop-up Date</th>
<th>Duration (days)</th>
<th>Distance (km)</th>
<th>Mean speed (km/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa - Cape Vidal (Gifford et al., 2007a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>7</td>
<td>F</td>
<td>08.03.1998</td>
<td>25.03.1998</td>
<td>17</td>
<td>199</td>
<td>-</td>
</tr>
<tr>
<td>2*</td>
<td>7</td>
<td>M</td>
<td>25.10.1998</td>
<td>01.11.1998</td>
<td>7</td>
<td>90</td>
<td>-</td>
</tr>
<tr>
<td>3*</td>
<td>8</td>
<td>M</td>
<td>23.01.1999</td>
<td>25.01.1999</td>
<td>&lt;3</td>
<td>65</td>
<td>-</td>
</tr>
<tr>
<td>Mozambique - Tofo (Brunnschweiller et al., 2009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4*</td>
<td>6-7</td>
<td>F</td>
<td>18.02.2006</td>
<td>16.5.2006</td>
<td>87</td>
<td>1200</td>
<td>31.2</td>
</tr>
<tr>
<td>Madagascar – Nosy Be (Graham et al., 2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>5-12.12.2005</td>
<td>15.04.2006</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>5-12.12.2005</td>
<td>07.02.2006</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>M</td>
<td>5-12.12.2005</td>
<td>24.11.2005</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>F</td>
<td>5-12.12.2005</td>
<td>19.12.2005</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>5-12.12.2005</td>
<td>15.03.2006</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Djibouti - Arta Beach (Rowat et al., 2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>M</td>
<td>22.01.2006</td>
<td>31.01.2006</td>
<td>9</td>
<td>-</td>
<td>10.1</td>
</tr>
<tr>
<td>Arabian Sea (MML, 2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11*</td>
<td>4.2</td>
<td>F</td>
<td>18.03.2010</td>
<td>20.04.2010</td>
<td>33</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Seychelles - Isle of Mahe (Rowat &amp; Gore, 2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12*</td>
<td>5</td>
<td></td>
<td>02.09.2001</td>
<td>17.10.2001</td>
<td>46</td>
<td>1422.74</td>
<td>33.1 (max)</td>
</tr>
<tr>
<td>13*</td>
<td>7</td>
<td></td>
<td>25.10.2001</td>
<td>12.11.2001</td>
<td>19</td>
<td>502.1</td>
<td>23.8 (max)</td>
</tr>
<tr>
<td>14*</td>
<td>6.5</td>
<td>M</td>
<td>25.10.2001</td>
<td>23.12.2001</td>
<td>60</td>
<td>3382.63</td>
<td>56.9 (max)</td>
</tr>
<tr>
<td>Western Australia - Ningaloo (CSIRO, 2005; Gunn et al., 1999; Wilson et al., 2006; Wilson et al., 2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>-</td>
<td></td>
<td>28.03.1994</td>
<td>-</td>
<td>&lt;1</td>
<td>-</td>
<td>60.5</td>
</tr>
<tr>
<td>161</td>
<td>-</td>
<td></td>
<td>29.03.1994</td>
<td>-</td>
<td>&lt;2</td>
<td>-</td>
<td>51.3</td>
</tr>
<tr>
<td>171</td>
<td>-</td>
<td></td>
<td>06.04.1997</td>
<td>-</td>
<td>&lt;2</td>
<td>-</td>
<td>64.8</td>
</tr>
<tr>
<td>181</td>
<td>-</td>
<td></td>
<td>10.04.1997</td>
<td>-</td>
<td>&lt;1</td>
<td>-</td>
<td>73.5</td>
</tr>
<tr>
<td>192</td>
<td>7</td>
<td>F</td>
<td>22.04.2002</td>
<td>17.07.2002</td>
<td>85</td>
<td>3130</td>
<td>36</td>
</tr>
<tr>
<td>202</td>
<td>4.2</td>
<td>M</td>
<td>05.05.2005</td>
<td>29.09.2005</td>
<td>147</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>212</td>
<td>4</td>
<td>F</td>
<td>06.05.2005</td>
<td>19.09.2005</td>
<td>136</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>222</td>
<td>7.5</td>
<td></td>
<td>05.05.2005</td>
<td>30.08.2005</td>
<td>118</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>232</td>
<td>4.2</td>
<td>M</td>
<td>07.05.2005</td>
<td>10.06.2005</td>
<td>34</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>242</td>
<td>-</td>
<td></td>
<td>02.05.2005</td>
<td>10.06.2005</td>
<td>39</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>252</td>
<td>4.2</td>
<td></td>
<td>06.05.2005</td>
<td>04.06.2005</td>
<td>29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>262</td>
<td>7.3</td>
<td>M</td>
<td>28.06.2005</td>
<td>31.07.2005</td>
<td>33</td>
<td>1740</td>
<td>52.8</td>
</tr>
<tr>
<td>273</td>
<td>8-10</td>
<td>F</td>
<td>02.05.2003</td>
<td>10.05.2003</td>
<td>8</td>
<td>85</td>
<td>-</td>
</tr>
<tr>
<td>ID</td>
<td>Age</td>
<td>Sex</td>
<td>Date First</td>
<td>Date Last</td>
<td>Tag</td>
<td>Length</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>------------</td>
<td>-----------</td>
<td>-----</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>28^3</td>
<td>6</td>
<td>F</td>
<td>03.05.2003</td>
<td>04.05.2003</td>
<td>1</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>29^3</td>
<td>4.5</td>
<td>F</td>
<td>05.05.2003</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>30^3</td>
<td>6-8</td>
<td>-</td>
<td>05.05.2003</td>
<td>11.11.2003</td>
<td>191</td>
<td>942</td>
<td></td>
</tr>
<tr>
<td>31^3</td>
<td>4.5</td>
<td>M</td>
<td>03.05.2004</td>
<td>03.05.2004</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>32^3</td>
<td>8.5</td>
<td>F</td>
<td>03.05.2004</td>
<td>03.05.2004</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>33^3</td>
<td>7.5</td>
<td>F</td>
<td>03.05.2004</td>
<td>07.05.2004</td>
<td>4</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>34^3</td>
<td>4.7</td>
<td>F</td>
<td>04.05.2004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>35^3</td>
<td>4.5</td>
<td>F</td>
<td>04.05.2004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>36^3</td>
<td>11</td>
<td>F</td>
<td>04.05.2004</td>
<td>31.08.2004</td>
<td>119</td>
<td>1302</td>
<td></td>
</tr>
<tr>
<td>37^3</td>
<td>7.3</td>
<td>F</td>
<td>04.05.2004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>38^3</td>
<td>4.7</td>
<td>-</td>
<td>05.05.2004</td>
<td>24.11.2004</td>
<td>203</td>
<td>673</td>
<td></td>
</tr>
<tr>
<td>39^3</td>
<td>7.6</td>
<td>F</td>
<td>05.05.2004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>40^3</td>
<td>5.6</td>
<td>F</td>
<td>07.05.2004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>41^3</td>
<td>7.6</td>
<td>F</td>
<td>07.05.2004</td>
<td>10.05.2004</td>
<td>3</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>42^3</td>
<td>6.2</td>
<td>F</td>
<td>08.05.2004</td>
<td>06.09.2004</td>
<td>119</td>
<td>738</td>
<td></td>
</tr>
<tr>
<td>43^3</td>
<td>7.2</td>
<td>M</td>
<td>08.05.2004</td>
<td>04.07.2004</td>
<td>57</td>
<td>1501</td>
<td></td>
</tr>
<tr>
<td>44^3</td>
<td>7</td>
<td>F</td>
<td>08.05.2004</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>45^3</td>
<td>5.3</td>
<td>F</td>
<td>09.05.2004</td>
<td>11.12.2004</td>
<td>216</td>
<td>443</td>
<td></td>
</tr>
<tr>
<td>46^4</td>
<td>4-5</td>
<td>M</td>
<td>05.05.2005</td>
<td>29.09.2005</td>
<td>147</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Christmas Island (Australian Institute of Marine Sciences ^)**

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Sex</th>
<th>Date First</th>
<th>Date Last</th>
<th>Tag</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>47^4</td>
<td>5</td>
<td>M</td>
<td>17.01.2008</td>
<td>14.04.2008</td>
<td>88</td>
<td>-</td>
</tr>
</tbody>
</table>

^ whale shark tagged by M.M., unpublished.
Table 2: List of published whale shark tracks from the Pacific Ocean and Indo-Pacific area.

* indicates tracks displayed in Figure 7.

<table>
<thead>
<tr>
<th>#</th>
<th>Length (m)</th>
<th>Sex</th>
<th>Tagging Date</th>
<th>Pop-up Date</th>
<th>Duration (days)</th>
<th>Distance (km)</th>
<th>Mean speed (km/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>-</td>
<td>M</td>
<td>28.06.2009</td>
<td>31.07.2009</td>
<td>33</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2*</td>
<td>4.5</td>
<td>M</td>
<td>08.04.2002</td>
<td>01.11.2002</td>
<td>208</td>
<td>5896</td>
<td>28.3</td>
</tr>
<tr>
<td>3</td>
<td>4.2</td>
<td>F</td>
<td>17.04.2002</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4*</td>
<td>4.2</td>
<td>M</td>
<td>13.11.2002</td>
<td>28.02.2003</td>
<td>108</td>
<td>3732</td>
<td>34.6</td>
</tr>
<tr>
<td>5*</td>
<td>4</td>
<td>M</td>
<td>09.04.2004</td>
<td>01.08.2004</td>
<td>114</td>
<td>3121</td>
<td>28.41</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>-</td>
<td>05.02.1997</td>
<td>-</td>
<td>10</td>
<td>900</td>
<td>18.9</td>
</tr>
<tr>
<td>7*</td>
<td>3</td>
<td>-</td>
<td>18.02.1997</td>
<td>02.05.1997</td>
<td>73</td>
<td>1567</td>
<td>31.9</td>
</tr>
<tr>
<td>8*</td>
<td>5</td>
<td>-</td>
<td>23.02.1997</td>
<td>03.03.1997</td>
<td>8</td>
<td>54</td>
<td>13.7</td>
</tr>
<tr>
<td>9*</td>
<td>7</td>
<td>-</td>
<td>03.02.1997</td>
<td>9.02.1997</td>
<td>6</td>
<td>220</td>
<td>6.9</td>
</tr>
<tr>
<td>10*</td>
<td>7</td>
<td>-</td>
<td>05.02.1997</td>
<td>12.06.1997</td>
<td>127</td>
<td>8025</td>
<td>23.3</td>
</tr>
<tr>
<td>11*</td>
<td>7</td>
<td>-</td>
<td>05.02.1997</td>
<td>15.02.1997</td>
<td>10</td>
<td>900</td>
<td>18.9</td>
</tr>
<tr>
<td>12*</td>
<td>3.7</td>
<td>F</td>
<td>06.10.1994</td>
<td>07.10.1994</td>
<td>1</td>
<td>57.5</td>
<td>-</td>
</tr>
<tr>
<td>13*</td>
<td>4.6</td>
<td>-</td>
<td>22.9.1994</td>
<td>14.10.1994</td>
<td>22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14*</td>
<td>3.0</td>
<td>F</td>
<td>19.9.1994</td>
<td>06.10.1994</td>
<td>17</td>
<td>320.3</td>
<td>21.7</td>
</tr>
<tr>
<td>15</td>
<td>4.6</td>
<td>-</td>
<td>20.9.1994</td>
<td>-</td>
<td>17</td>
<td>31.7</td>
<td>3.2</td>
</tr>
<tr>
<td>16*</td>
<td>4.3</td>
<td>-</td>
<td>20.9.1994</td>
<td>29.10.1994</td>
<td>39</td>
<td>818.3</td>
<td>23.5</td>
</tr>
<tr>
<td>17</td>
<td>4.0</td>
<td>-</td>
<td>23.9.1994</td>
<td>-</td>
<td>5-7</td>
<td>23.3</td>
<td>8.4</td>
</tr>
<tr>
<td>18</td>
<td>4.0</td>
<td>-</td>
<td>07.10.1994</td>
<td>-</td>
<td>1</td>
<td>11.2</td>
<td>-</td>
</tr>
<tr>
<td>19*</td>
<td>6.1</td>
<td>F</td>
<td>10.9.1995</td>
<td>22.09.1995</td>
<td>12</td>
<td>404.6</td>
<td>23.6</td>
</tr>
<tr>
<td>20*</td>
<td>3.7</td>
<td>-</td>
<td>10.9.1995</td>
<td>13.09.1995</td>
<td>3</td>
<td>8.5</td>
<td>11.1</td>
</tr>
<tr>
<td>22*</td>
<td>3.7</td>
<td>F</td>
<td>10.9.1995</td>
<td>28.09.1995</td>
<td>18</td>
<td>46.3</td>
<td>2</td>
</tr>
<tr>
<td>23*</td>
<td>-</td>
<td>-</td>
<td>09.9.1995</td>
<td>07.10.1995</td>
<td>28</td>
<td>199.6</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>15.0</td>
<td>F</td>
<td>20.6.1996</td>
<td>-</td>
<td>111</td>
<td>2863.6</td>
<td>28.8</td>
</tr>
</tbody>
</table>

Gulf of California – Sea of Cortez (Eckert & Stewart, 2001)

| 12*| 3.7        | F   | 06.10.1994     | 07.10.1994        | 1              | 57.5          | -                   |
| 13*| 4.6        | -   | 22.9.1994      | 14.10.1994        | 22             | -             | -                   |
| 14*| 3.0        | F   | 19.9.1994      | 06.10.1994        | 17             | 320.3         | 21.7                |
| 15 | 4.6        | -   | 20.9.1994      | -                 | 17             | 31.7          | 3.2                 |
| 16*| 4.3        | -   | 20.9.1994      | 29.10.1994        | 39             | 818.3         | 23.5                |
| 17 | 4.0        | -   | 23.9.1994      | -                 | 5-7            | 23.3          | 8.4                 |
| 18 | 4.0        | -   | 07.10.1994     | -                 | 1              | 11.2          | -                   |
| 19*| 6.1        | F   | 10.9.1995      | 22.09.1995        | 12             | 404.6         | 23.6                |
| 20*| 3.7        | -   | 10.9.1995      | 13.09.1995        | 3              | 8.5           | 11.1                |
| 22*| 3.7        | F   | 10.9.1995      | 28.09.1995        | 18             | 46.3          | 2                   |
| 23*| -          | -   | 09.9.1995      | 07.10.1995        | 28             | 199.6         | -                   |
| 24 | 15.0       | F   | 20.6.1996      | -                 | 111            | 2863.6        | 28.8                |

Galápagos – Total of 14 tracks (Green, 2011)

27* - - 07.2011 - - -
Table 3: List of published whale shark tracks from the Atlantic Ocean.

* indicates tracks displayed in Figure 7; numbers in superscript indicate the reference for each track where there are multiple references for the same tagging location.

<table>
<thead>
<tr>
<th>#</th>
<th>Length (m)</th>
<th>Sex</th>
<th>Tagging Date</th>
<th>Pop-up Date</th>
<th>Duration (days)</th>
<th>Distance (km)</th>
<th>Mean speed (km/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gulf of Mexico - Gladden Spit (*Graham et al., 2006); Utila, Honduras (*Gifford et al., 2007b); Florida (*MML, 2010); Quitana Roo (*Hueter et al., 2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1†</td>
<td>5.5</td>
<td>M</td>
<td>23.04.2000</td>
<td>07.05.2000</td>
<td>14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2†</td>
<td>3.6</td>
<td>M</td>
<td>23.04.2000</td>
<td>02.06.2000</td>
<td>40</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3†</td>
<td>9.7</td>
<td>M</td>
<td>11.04.2001</td>
<td>15.10.2001</td>
<td>188</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4†</td>
<td>6.7</td>
<td>M</td>
<td>11.04.2001</td>
<td>removed</td>
<td>248</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5†</td>
<td>6.7</td>
<td>M</td>
<td>21.04.2000</td>
<td>06.05.2000</td>
<td>14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6†</td>
<td>5.5</td>
<td>M</td>
<td>25.04.2000</td>
<td>08.08.2000</td>
<td>105</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7†</td>
<td>5.5</td>
<td>M</td>
<td>25.04.2000</td>
<td>15.06.2000</td>
<td>40</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8†</td>
<td>5.2</td>
<td>M</td>
<td>16.03.2001</td>
<td>31.07.2001</td>
<td>127</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9†</td>
<td>6.7</td>
<td>M</td>
<td>11.04.2001</td>
<td>30.06.2001</td>
<td>187</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10†</td>
<td>5.2</td>
<td>M</td>
<td>11.04.2001</td>
<td>-</td>
<td>163</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11†</td>
<td>5.5</td>
<td>M</td>
<td>10.05.2001</td>
<td>03.07.2001</td>
<td>249</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12†</td>
<td>8</td>
<td>M</td>
<td>18.02.1999</td>
<td>29.06.1999</td>
<td>132</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13†</td>
<td>8</td>
<td>M</td>
<td>31.12.1999</td>
<td>31.01.2000</td>
<td>31</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14*</td>
<td>7.5</td>
<td>F</td>
<td>28.05.2010</td>
<td>20.08.2010</td>
<td>84</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15*</td>
<td>7.5</td>
<td>M</td>
<td>18.06.2010</td>
<td>04.10.2010</td>
<td>108</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16*</td>
<td>6</td>
<td>M</td>
<td>31.08.2005</td>
<td>01.10.2005</td>
<td>31</td>
<td>889</td>
<td>28.7</td>
</tr>
<tr>
<td>17*</td>
<td>5.5</td>
<td>M</td>
<td>02.09.2005</td>
<td>01.10.2005</td>
<td>29</td>
<td>260</td>
<td>9.0</td>
</tr>
<tr>
<td>18*</td>
<td>7</td>
<td>M</td>
<td>23.07.2006</td>
<td>27.12.2006</td>
<td>157</td>
<td>970</td>
<td>6.2</td>
</tr>
<tr>
<td>20*</td>
<td>7</td>
<td>F</td>
<td>13.09.2006</td>
<td>25.11.2006</td>
<td>73</td>
<td>1027</td>
<td>14.1</td>
</tr>
<tr>
<td>21*</td>
<td>8</td>
<td>F</td>
<td>14.09.2006</td>
<td>10.01.2007</td>
<td>118</td>
<td>371</td>
<td>3.14</td>
</tr>
<tr>
<td>22*</td>
<td>7.5</td>
<td>F</td>
<td>31.08.2007</td>
<td>24.01.2008</td>
<td>150</td>
<td>7213</td>
<td>48.1</td>
</tr>
</tbody>
</table>

Compiling all published tracks at a global scale (Figure 7) shows that the current information on whale shark movement range and habitat use is still highly scattered. In the Atlantic ocean for example, there are tracks in the area of the Gulf of Mexico (Gifford et al., 2007b) that confirm the link between Honduras, Belize and Mexico suggested by Graham and Roberts (2007). There is also a > 7000 km track from a tag deployed at Quintana Roo (Mexico) finishing in the central Atlantic Ocean south of the Saint Peter and Saint Paul Archipelago (Brazil) about 5 months later (Hueter et al., 2008). However, there are no
tracks available from locations in the eastern Atlantic where whale sharks have been reported (e.g., Portugal or the west coast of Africa – Figure 5). In the Indian Ocean, existing tracks link different aggregations. This is the case for South Africa and Mozambique (Gifford et al., 2007b), Mozambique and Madagascar (Brunnschweiler et al., 2009), United Arab Emirates and Qatar (MML, 2010) and Ningaloo with Timor, Indonesia and Christmas Island (CSIRO, 2005; Wilson et al., 2006; Wilson et al., 2007). There is also a 60-day track suggesting a connection between Seychelles and Sri Lanka (Rowat & Gore, 2007), although it was inferred based on only one transmission obtained after an inactive period of 1.5 months (represented as a straight line in Figure 7). Likewise, the tags deployed in western Philippines (Eckert et al., 2002) show a possible connection with Malaysia and south Vietnam, while sharks tagged in Taiwan (Hsu et al., 2007) reached waters off east Philippines.

Several tags deployed in the Gulf of California (Eckert & Stewart, 2001) showed that most sharks remain in the Sea of Cortés; however, four left the Gulf and one was reported to have travelled almost 13,000 km from the east to the west Pacific Ocean in 1144 days (< 37 months). Eckert and Stewart (2001) asserted that due to data-storage limitations, sharks with longer tracks (and not surfacing regularly enough to transmit stored data), were located less frequently. However, whale sharks spend most of their time at the surface (Sleeman et al., 2010a; Sleeman et al., 2010b; Wilson et al., 2006), so such long periods without surfacing for satellite position acquisition are unlikely. For this reason, the 37-month track with an atypically straight trajectory, lack of surface intervals, and unprecedented rate and magnitude of travel, should be deemed biologically unrealistic.
Figure 7: Global overview of published whale shark tracks.

(Colour code indicates area of tag deployment. Tracks adapted from original sources as detailed in Tables 1-3.)
(e.g., it is more likely that the satellite tag had detached and was floating passively with the currents). Recently, some tags deployed near the Galapagos (Table 3 and Figure 7) have revealed that most sharks head west after congregating at this location, although the information available is still incomplete (www.galapagoswhaleshark.com).

**The big picture: formulating global hypotheses**

After collating all available whale shark occurrences and tracking data, the question of whether or not whale sharks are migrating across or within oceans remains unanswered. One must conclude therefore that such conventional methods to study whale shark movement are still of insufficient temporal and spatial coverage to provide useful conclusions at the ocean-basin scale. We thus propose a different approach by shifting research from a local to a global perspective.

The Indian Ocean is the region from which the most whale shark data have been collected; therefore, it is the best region in which to begin formulating hypotheses regarding whale shark movement patterns. We know that some locations are linked via direct observation such as Mozambique and Seychelles (Rowat & Gore, 2007) or Ningaloo and Christmas Island (Figure 7). It is clear that appearance timings occur sequentially (Figure 6). Therefore, we hypothesise that at least some sharks move in a general clockwise pattern from the southwest (KwaZulu-Natal) in January, to the southeast Indian Ocean (Ningaloo) in March over approximately two years. According to Eckert and Stewart (2001), the mean pace at which whale sharks travel varies from about 2 to 30 km day⁻¹, with more horizontal distance covered when closer to the surface. Average speeds of around 30 km day⁻¹ have also been reported by others (Brunnschweiler et al., 2009; Eckert
et al., 2002; Hsu et al., 2007; Hueter et al., 2008; Rowat & Gore, 2007), while faster speeds have been reported for short-term movements (Gunn et al., 1999). Assuming constant travel speeds, a dominance of surface swimming and otherwise ideal conditions, a shark moving at about 30 km day\(^{-1}\) would be able to travel > 10,000 km year\(^{-1}\) – approximately the distance between South Africa and southern Western Australia. Using this logic, two years to complete the longer clockwise migration from South Africa to Ningaloo is biologically plausible.

Once in Ningaloo, even though the strength of the southward-flowing Leeuwin Current along the western coast of Australia might influence the degree to which they penetrate southward, previous tagging results demonstrate that the predominant movements are northward towards Indonesia, and then either westward or eastward (CSIRO, 2005) (Figure 7). Tagging data show that whale sharks take about five months to return to the central Indian Ocean (passing near Christmas Island much earlier than the seasonal peak in abundance) and about four months to travel east as far as East Timor. Even though peaks of occurrence are not published for the southern areas of the Malay Archipelago/Timor, whale sharks are found occasionally in the Coral Sea, about four to five months after the Ningaloo peak (Figure 6). From the Coral Sea, environmental conditions are likely influential on the patterns of subsequent movement. Either they (1) move farther south towards the north coast of New Zealand (surpassing 35 ° S) where their occurrence peaks in February (Duffy, 2002), (2) return to the Indian Ocean or (3) migrate eastward (whale sharks are occasionally seen around the islands of the south west and central Pacific Ocean (Compagno, 2001; and data collected by purse seine fisheries) (Figure 5).
In the central Indo-Pacific, whale sharks tracked from Taiwan travelled close to the Philippines even outside the peak abundance season (March and June). Eckert et al. (2002) tracked whale sharks from the Philippines in February, from where they travelled west passing Malaysia and south Vietnam toward the Gulf of Thailand in about 4.5 months. There are many routes that potentially link the Pacific and Indian Oceans in the Malay Archipelago (e.g., Andaman and Java Seas), such that whale sharks can cross to the Indian Ocean towards Christmas Island or the north of Australia.

From a temporal perspective (Figure 6), there are no obvious paths from eastern locations in the Indian Ocean to the west through to Bangladesh. That lead us to the hypothesis that whale sharks could swim straight towards the Maldives (as was the case for a shark tagged off Ningaloo moving far into the central Indian Ocean). This would also be consistent with the lower numbers of whale sharks sighted on the eastern coast of India compared to the western coast (Pravin, 2000). Additionally, whale sharks could also move from the Maldives through to the Seychelles and then to South Africa based on observed temporal patterns of occurrence (Figure 6).

For there to be population connectivity between the Indian/Pacific and Atlantic Oceans, whale sharks would need to travel through the Cape of Good Hope in South Africa. Beckley et al. (1997) suggested that whale shark strandings on the east coast of South Africa might be associated with changes in temperature associated with Agulhas Current from the Indian Ocean mixing with the cold water upwelled by the Benguela Current in the southeast Atlantic. Beckley et al. (1997) also stated that stranded animals were generally small, which suggests that only larger animals would survive the crossing from the Indian to the Atlantic Ocean. The greater thermal inertia of larger sharks might
allow them to transit in the southern Atlantic Ocean until they reach the warmer waters of, for example, Gabon (peak of occurrences registered by purse seine fisheries in the Atlantic - Figure 5), or the region around the Brazilian Saint Peter and Saint Paul Archipelago where sharks are observed year-round, but mainly during the first semester with a peak in June (Hazin et al., 2008) – six months after the peak in South Africa.

Within the entire Atlantic Ocean, there are tracks available mostly for the area of the Gulf of Mexico/Caribbean Sea that show some linkages between the aggregations of Central America (Figure 7). This suggests that at least here there is a single meta-population. The odd long track starting from Mexico in August and finishing south of the Saint Peter and Saint Paul Archipelago in January suggests that whale sharks from this Central American population can travel great distances within the Atlantic Ocean. Year-round sightings of whale sharks in the Archipelago outside of peak months (Hazin et al., 2008) hint at the possibility of a trans-Atlantic thoroughfare.

Because the occurrence of whale sharks in the Azores islands in the eastern Atlantic is seasonal (mostly at the end of August and September) and does not occur every year (Machete & Afonso, pers. comm.), it is possible that whale sharks only travel there when environmental conditions (especially water temperature) are most suitable (Sa, 2008). The Azores current, flowing south of the Azores Archipelago (where most whale sharks are spotted by tuna fisheries), originates from the Gulf Stream (Klein & Siedler, 1989). When the Gulf Stream is strong, the warmer-than-usual waters near the Azores might encourage more northerly forays (Sa, 2008). We therefore hypothesise that individuals travelling to the Azores might originate from the Gulf of Mexico. Whale sharks are also occasionally seen in the Madeira Islands (Portugal) to the south-east of the
Azores (Wirtz et al., 2008). Because water temperatures in these parts of the Atlantic are generally cooler, movements toward the warmer southerly waters closer to the Equator near the Saint Peter and Saint Paul Archipelago and West Africa would be more typical – a prediction supported by observation from fishing fleets in those areas.

Breeding within the Atlantic is supported by the discovery of a single neonate (58 cm total length) specimen inside the stomach of a captured blue shark (Kukuyev, 1996; Martin, 2007). Other neonates have been caught near the Equator in this ocean (Wolfson, 1983) and an egg case was found in the Gulf of Mexico in the early 1950s (Baughman, 1955). Neonates have also been found in the eastern Pacific Ocean (Wolfson, 1983), where large adult females are often observed (Eckert & Stewart, 2001). Although more is known about whale sharks inhabiting the Indian Ocean, only 2 neonates (< 1 m) have ever been reported there (Rowat et al. 2008). Despite young juveniles (~ 1 m total length) occasionally observed swimming with larger individuals (e.g., Pillai, 1998), most of the Indian Ocean aggregations are comprised mainly of immature males, which makes it unlikely that the region includes a permanent breeding area. If there is a nursery in the Indo-Pacific area (e.g., around The Philippines), and if whale sharks are able to cross to the Indian Ocean, small juveniles could potentially occur in the latter even in the absence of a nursery.

Notwithstanding these hypothesised linkages, we know that whale sharks often return to the same aggregation at least semi-annually (e.g., Maldives, Ningaloo; Meekan et al., 2006) based on photographic matching (Speed et al., 2007). Such repeat sightings do not support the notion that all individuals regularly travel around ocean basins over two- to four-year cycles. We hypothesise instead that even though whale sharks can travel over
entire or even between ocean basins, many (perhaps most) remain near single aggregation sites for several months or years. For example in the Maldives, whale sharks might remain in the general area, travelling west (December to April) to east (May to November) of the islands over one or more years (Figure 6), or extend the travelling farther from Gujarat (northwest side of India) to west of the Maldives in the first half of the year, and then east of the Maldives and Tamil-Nadu (east side of India) in the second. From here they would have the option to move east (e.g., to Bangladesh in December) or back to the west of the Maldives, south India and Gujarat (Figure 6). This back-and-forth movement around India and the Maldives accords with the higher re-sighting rate in the Maldives relative to other nearby locations (Riley et al., 2010). Other examples of these hypothesised shorter migration routes and possible populations are depicted in Figure 8.
Figure 8: Whale shark migration patterns.

Depicting likely sub-populations (red circles). Blue arrows indicate possible links between sub-populations (aggregations). Black arrows denote movements revealed from tagging data, and dashed lines indicate no current evidence for a migratory pathway (except the questionable 13,000-km track across the Pacific from Eckert and Stewart (2001) shown as a back dotted line).
Foreseeing change

The global distribution of whale sharks lies predominantly within tropical and warm temperate waters between ~ 30 ° north and south (Compagno, 2001; Last & Stevens, 2009). Since the late 1980s, however, individual sharks have been observed well poleward of these latitudes. Indeed, Wolfson (1986) recorded whale shark at latitudes of 41 ° N and 36.5 ° S, and more recently, Duffy (2002) reported them in New Zealand south of 35 ° S. One whale shark has even been sighted as far north as 44 ° in the Bay of Fundy, Canada (Turnbull & Randell, 2006). Whale sharks are also occasionally sighted south of the Azores (~ 41 ° N), mainly by tuna fishermen (Machete & Afonso, pers. comm.) and recently, many sightings (> 400) have been made mainly around the island of Santa Maria from June to September (fisheries data), likely in response to an unusual pulse of warm water (Sa, 2008).

Whale sharks mainly stay within a narrow range of sea surface temperatures. In the Sea of Cortés, most tracked individuals were in waters between 28 and 32 ° C (Eckert & Stewart, 2001); in the Seychelles, most were between 25 and 35 °C (Rowat & Gore, 2007); in the north-western Pacific, most were between 23 and 32 ° C (Hsu et al., 2007); for pelagic sightings derived from fisheries records, > 90 % of 1185 records were in surface waters between 26.5 and 30 °C (Sequeira et al., 2012). If water temperature plays an important role in modifying whale shark distribution, the rapid warming of the world’s oceans arising from anthropogenic climate change (Hegerl & Bindoff, 2005) will likely affect their future distribution. Because whale sharks seem to avoid higher temperatures (e.g., around the Equator; Sequeira et al., 2012), poleward shifts are more probable than an overall expansion of their current distribution. In addition to these expected
distributional changes, shifts in both abundance and distribution are expected from previous (Bradshaw et al., 2007; Romanov, 2002) and ongoing (Riley et al., 2009; White & Cavanagh, 2007) commercial fishing, potentially from excessive disturbance arising from ecotourism at aggregation sites (Cárdenas-Torres et al., 2007; Pierce et al., 2010; Quiros, 2007; Rowat & Engelhardt, 2007) and other human-related disturbances (Haetrakul et al., 2007; World Wildlife Foundation, 2012).

Conclusions

Despite ample evidence that whale sharks are capable of long-distance travel, confirmed by genetic evidence that the world’s populations are connected at least at the generational scale, we know that many individuals spend long periods within the immediate surrounds of particular locations. We have thus provided a conceptual model based on known movement patterns, aggregation timings and oceanic structure demonstrating how within- and inter-ocean movements might occur, even if they do so rarely and without non-genetic evidence yet to test our hypotheses. This lack of evidence however does not counter our model. It mostly derives from the thousands (i.e., not millions or tens of millions) of photos used in photo identification-based studies. Photo-matching is only semi-automated (e.g., I3S; Van Tienhoven et al., 2007), thus limiting the probability of photographing ‘migrating’ individuals, and making the probability of resighting these individuals in other aggregations even smaller. Moreover, flank mismatch (right or left side only) can also limit the number of individuals that can be compared.

Viewing whale shark ecology from our global perspective therefore engenders many related hypotheses that can be tested as technology improves and genetic evidence
is refined, and it isolates the most important questions to determine the long-term conservation status of this species. A natural progression of our global movement model is the construction of a broad-scale population model that connects sub-populations based on inferred demographic and movement rate data. Application and continual refinement of biologically plausible rates will further hone the global model and its predictions, thereby unfolding key aspects of the species cryptic life history so that research can target the most relevant aspects. It is clear that more data are required from locations close to key aggregation sites where hypothesised connections can be validated. Tagging and photographic-identification data within these regions should be prioritized based on their putative connections predicted under the global model. For example, areas such as Madagascar, the Seychelles and Tanzania, or the Philippines, Malaysia and Gulf of Thailand require much broader research coverage to validate existing hypotheses of connectivity.

Tagging studies should also target many sharks within the same aggregation to chronicle the highest number of potential movement patterns (cf. Eckert & Stewart, 2001; Wilson et al., 2006), and tags should be deployed as close to the end of the peak aggregation season as possible to maximize the potential for measuring long-distance trajectories, especially given the short monitoring periods characteristic of such studies (Hammerschlag et al., 2011). Other improvements to maximize monitoring time include the application of anti-fouling paint to tracking equipment (Hammerschlag et al. 2011), using fast-loc GPS technology (Hays et al. 2007) and switching from steel to copper salt-water switches used to improve the efficiency of duty cycling (Hays et al. 2007). Improvements in
tag attachment methods are also needed to prevent premature detachment (e.g., Brunnschweiler et al., 2009).

Another important consideration when describing migratory patterns is that longer migrations might reflect sex- and age-specific behaviour. Indeed, 3 of the 4 whale shark tracks exceeding 3000 km were large females (~7 m) (Tables 1-3). As suggested previously, female natal philopatry might result in long migrations only for the purposes of occasional breeding (Ramírez-Macias et al., 2007). This sort of sex-specific behaviour would also explain the high resighting rates in some aggregations dominated by immature males. While a similar number of males and females tracks have been published to date, different size classes (≤ 6 m or > 6 m) have different average track lengths (200-300 and 1200 km, respectively). Of course, there are exceptions to this trend; some juveniles tagged in Taiwan had tracks > 3000 km even though they remained resident (Hsu et al., 2007). More tracking studies of longer duration will assist in describing these stage-specific movement trends and capacities.

Given the large distances between some of the known aggregations, another way to refine our knowledge of this species is to profit from the opportunities provided by certain pelagic fisheries. Expansive commercial fisheries (e.g., tuna purse seiners) do release whale sharks that are accidentally captured or encircled by their nets. Tagging endeavours associated with these releases could potentially provide the data necessary to estimate vital rates and a better description of pelagic movement patterns. Further, focussing tag deployments and tissue collections near potential thoroughfares, such as those hypothesised under our model like the Saint Peter and Saint Paul Archipelago and
Christmas Island, would likely increase the probability of capturing pan- or trans-oceanic movements.
CHAPTER III. Ocean-scale prediction of whale shark distribution

A. Sequeira\textsuperscript{a}, C. Mellin\textsuperscript{a,b}, D. Rowat\textsuperscript{c}, M. Meekan\textsuperscript{d}, C. Bradshaw\textsuperscript{a,e}

\textsuperscript{a}The Environment Institute and School of Earth and Environmental Sciences, The University of Adelaide, South Australia 5005, Australia

\textsuperscript{b}Australian Institute of Marine Science, PMB No. 3, Townsville MC, Townsville, Queensland 4810, Australia

\textsuperscript{c}Marine Conservation Society, Seychelles, P.O. Box 1299, Victoria, Mahe, Seychelles

\textsuperscript{d}Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling Hwy, Crawley, Western Australia 6009, Australia.

\textsuperscript{e}South Australian Research and Development Institute P.O. Box 120, Henley Beach, South Australia 5022, Australia

\textit{Diversity and Distributions 2012; 18 : 504 – 519}
Statement of authorship

OCEAN-SCALE PREDICTION OF WHALE SHARK DISTRIBUTION

Diversity and Distributions 2012; 18 : 504 – 519

SEQUEIRA, A.M.M. (Candidate)
Planned the article, performed the analysis, interpreted data, developed the models, wrote manuscript and acted as corresponding author.
I hereby certify that the statement of contribution is accurate.

Signed……………………………………………………………………Date………………

MELLIN, C.
Helped planning the article, helped with data interpretation, analysis and development of the models, and assisted the writing.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………Date………………

ROWAT, D.
Provided data and assisted interpretation.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………Date………………
MEEKAN, M. G.

Provided critical evaluation and interpretation.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………

BRADSHAW, C.J.A.

Supervised development of work, helped in data interpretation, models development and manuscript evaluation and writing.

I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………
Ocean-scale prediction of whale shark distribution

Abstract

Aim Predicting distribution patterns of whale sharks (*Rhincodon typus*, Smith 1828) in the open ocean remains elusive due 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 spatio-temporal variation in sampling effort and probability of detection. We applied generalised linear, spatial mixed-effect and Maximum Entropy models to predict seasonal variation in habitat suitability and produced maps of distribution.

Results The saturated generalised 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 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

**Introduction**

Species distribution modelling has been applied widely in conservation ecology (Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000; Hirzel et al., 2002; 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 (Beger & Possingham, 2008; Garza-Pérez et al., 2004; Mellin et al., 2010b; Tittensor et al., 2009), 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 & Solis, 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 programs can afford multiple ships operating simultaneously, and datasets 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), MaxEnt (Phillips & Dudík, 2008) — and produce acceptable results (Elith et al., 2006; Zaniewski et al., 2002), regression methods such as boosted regression trees (De’ath, 2007; Elith et al., 2008), multivariate adaptive regression splines (Leathwick et al., 2005), or generalised linear or additive models (Hastie & Tibshirani, 1986; Nelder & Wedderburn, 1972) tend to provide more realistic projections when using reliable and accurate absence data (Brotons et al., 2004; Zaniewski et al., 2002). 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 (Compagno, 2001; Last & Stevens, 1994) 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 (Brunnschweiler et al., 2009; Wilson et al., 2006). 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 (Riley et al., 2009; Stewart & Wilson, 2005; White & Cavanagh, 2007). 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.
(Meekan et al., 2009; Wilson, 2002). These fine-scale studies (100s m – 100s km) and other hypotheses have never been formally tested at the ocean-basin scale (1000s km) (Cárdenas-Torres et al., 2007; Rowat, 2007; Southall et al., 2006).

Access to a 17-year dataset 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: (i) determine the spatio-temporal patterns in habitat suitability for whale sharks over a broad (ocean basin) spatial scale and (ii) 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

Dataset 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 dataset consisted of a 17-year (1991-2007) time series of (i) whale shark sightings, date and location (longitude and latitude at 0.01° precision) for a total
of 1185 whale sharks; and (ii) tuna catch (tonnes) and fishing effort (days) pooled for each 5° grid cell, per month and per year (6618 records total).

The dataset covered 172,800 km² of the Indian Ocean between 30 °N to 30 °S and 35 to 100 °E (Figure 9a). When all years were combined, a total of 811 whale sharks were recorded in autumn (Apr- Jun; 68 % of records), 68 in winter (Jul- Sep; 6 %), 191 in spring (Oct-Dec; 16 %) and 115 in summer (Jan-Mar; 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 (Figure 9a). 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 dataset 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 dataset consisted of: total tuna catch (tonnes) and fishing effort (days) pooled for each 1° grid cell per month (Figure 9b). As for the 5° resolution dataset, 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 to March), autumn (April to June), winter (July to September) and spring (October to December).

**Pseudo-absence dataset generation**

Whale shark dataset were presence-only, as true absences were unknown. We therefore generated pseudo-absences for use in generalised linear models (GLM) and spatial generalised linear mixed-effect models (GLMM). The pseudo-absence locations were also included as background data in additional MaxEnt runs, to get 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: (i) 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), (ii) 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 (iii) selection weighted by the total tuna catch from the 1° dataset (tuna), assuming that whale sharks and tuna would have similar distribution patterns due to their association – it is 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 dataset at 9-km resolution over the study area that included bathymetry (including mean depth in m, and slope in °), 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 Figure B1 in Appendix B) across the study area at approximately 1.7-km resolution using the one-minute 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 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) and Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua), respectively (see Figure B2 and Figure B3). We averaged weekly daytime measures available since September 1997 for SeaWiFS and July 2002 for MODIS-Aqua for each month. For each season, we calculated the resulting Chl a and SST mean and standard deviation in R (R Development Core Team, 2010). Due to non-Gaussian data, we investigated monotonic relationships among environmental predictors using Spearman’s rank correlation coefficient (ρ).

Models

Generalised 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 ($\text{AIC}_c$) and the Bayesian information criterion (BIC). $\text{AIC}_c$ 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 ($w\text{AIC}_c$ and $w\text{BIC}$). We used the percentage of deviance explained ($De$) to quantify each model’s explanatory power.

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 dataset. We assessed the predictive power of the models according to Cohen’s Kappa statistics $\kappa$ that measures agreement/accuracy (Cohen, 1960) and varies from $\leq 0$ 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 $\kappa < 0.4$, good when $0.4 < \kappa < 0.75$, and excellent when $\kappa > 0.75$; these ranges were based on the $\kappa$ 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 dataset.

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., 2010b). We fit GLMM using the penalized quasi-likelihood (Venables & Ripley, 2002a) 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.

**MaxEnt (Maximum Entropy)**

We compared GLM results to predictions obtained using a presence-only modelling technique, Maximum Entropy(Phillips et al., 2006; Phillips & Dudík, 2008; Phillips et al., 2009; Phillips et al., 2004), using the software provided by the authors (MaxEnt version 3.3.3e November 2010, AT&T Labs Research). MaxEnt is a tool for generating species distribution models from presences-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 datasets 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). Due to the functionality of MaxEnt to use presences-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-absences datasets, we calculated the Kappa statistic to compare 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**

Due 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; Figure 9a), the aggregated normalized whale shark data was 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 to 160 × 10⁻³ in winter and 0.46 × 10⁻³ to 37.9 × 10⁻³ in spring.

Sighting (SPUE) patterns of whale sharks shifted between seasons (Figure 10); there was a non-random (Kruskal-Wallis H₃ = 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
dataset 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 10: 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 (Apr-Jun): 811 whale sharks recorded (68% of sightings); (b) winter (Jul-Sep): 68 whale sharks sighted (6% of sightings); (c) spring (Oct-Dec): 191 whale sharks sighted (16% of sightings); (d) summer (Jan-Mar): 115 whale sharks sighted (10% of total sightings).
**Generalised linear and mixed-effects models**

The percentage of deviance explained was highest for the generalised linear model including all environmental variables in all seasons, irrespective of the technique used for generating pseudo-absences (Table 4). 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* dataset (Table 4). The top-ranked model according to wBIC only matched the one that also maximized wAIC$_c$ in autumn for the three pseudo-absence datasets (Figure B4), and in winter and summer when using *random* pseudo-absences (Table 4). Both observations and GLM residuals were spatially autocorrelated ($P < 0.001$). 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 Figure B5).

The pseudo-absence dataset 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 5), 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).
Table 4: Summary of generalised 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 mean2), sea surface temperature standard deviation (SST SD) and chlorophyll a (Chl a). Shown for each model are the biased-corrected model evidence based on weights of Akaike’s information criterion corrected for small sample sizes (wAICc), 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 only for cases where wAICc > 0.001, and values ≤ 0.1 are shown in light grey. Note: slope, depth, distance to shore referred to together as physical variables (phys); mean sea surface temperature (SST mean), its quadratic term (SST mean2) and sea surface temperature standard deviation (SST SD) referred to together as SST variables (SST var).

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>autumn</th>
<th>winter</th>
<th>spring</th>
<th>summer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wAICc</td>
<td>wBIC</td>
<td>%DE</td>
<td>wAICc</td>
<td>wBIC</td>
</tr>
<tr>
<td>random</td>
<td>phys + SST var</td>
<td>0.25</td>
<td>0.07</td>
<td>0.15</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>saturated</td>
<td>1.00</td>
<td>1.00</td>
<td>57.05</td>
<td>0.99</td>
</tr>
<tr>
<td>IDW</td>
<td>slope</td>
<td>0.05</td>
<td>0.09</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>depth</td>
<td>0.02</td>
<td>0.08</td>
<td>0.55</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>shore</td>
<td>0.21</td>
<td>0.07</td>
<td>4.51</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>SST var</td>
<td>0.07</td>
<td>0.07</td>
<td>0.32</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Chl a</td>
<td>0.06</td>
<td>0.00</td>
<td>7.47</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>phys + SST variables</td>
<td>0.09</td>
<td>0.09</td>
<td>11.88</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>saturated</td>
<td>0.13</td>
<td>0.60</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>tuna</td>
<td>SST variables</td>
<td>0.06</td>
<td>0.90</td>
<td>17.92</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>phys + SST variables</td>
<td>0.26</td>
<td>0.00</td>
<td>20.60</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>saturated</td>
<td>1.00</td>
<td>1.00</td>
<td>42.61</td>
<td>0.86</td>
</tr>
</tbody>
</table>
For each season predictive maps derived from different pseudo-absence datasets resulted in similar patterns (Figure 11), while greater differences occurred among seasons (Figure 12). 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.

Figure 11: 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 generalised linear mixed-effects models, and those in the right panel are derived from MaxEnt 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 by iii) a probability directly weighted by total tuna catch (Tuna catch), respectively per row.
**Maximum Entropy**

The pseudo-absences datasets derived from the three different techniques resulted in similar maps in MaxEnt in each season (Figure 11). MaxEnt prediction maps were generally consistent with that predicted by the GLM and GLMM (Figure 11), except for spring - where MaxEnt presented the central equatorial area of the Indian Ocean as the most suitable habitat. Within 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-absences dataset used in the model, while the percent of variable contribution varied mainly within seasons (Table 6 and Supplementary information and Figure B6 in Appendix B). In autumn the more important variable was the quadratic term of sea surface temperature 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 6). The most important single variables were mainly the producing the poorest model results when excluded from the set of predictors.

AUC obtained with the MaxEnt models was generally low, varying from 0.574 to 0.721 (Table 6). In all seasons, the highest AUC values were obtained for models with the random, followed by tuna pseudo-absences datasets (around 0.7), while the
lower scores were obtained when the IDW dataset 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 5).

MaxEnt 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 \textit{random} and \textit{tuna} pseudo-absences in terms of variables percentage of contribution and permutation importance (Table 6 / Table B1). 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 6). Despite having the highest AUC (> 0.92), the predicted suitable area seemed to be more restricted with this model set.
Table 5: Comparison of results for the MaxEnt and generalised linear (GLM) and mixed effects (GLMM) models.

Where spatial autocorrelation was included for the three pseudo-absence techniques used (random, IDW and tuna). Shown for each model is 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 wAICc) is also shown for the GLM and GLMM.

<table>
<thead>
<tr>
<th>Season</th>
<th>autumn</th>
<th>winter</th>
<th>spring</th>
<th>summer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>κ</td>
<td>SD</td>
<td>%DE</td>
<td>κ</td>
</tr>
<tr>
<td><strong>random</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.54</td>
<td>0.02</td>
<td>-</td>
<td>0.57</td>
</tr>
<tr>
<td>GLM</td>
<td>0.79</td>
<td>0.09</td>
<td>58.0</td>
<td>0.51</td>
</tr>
<tr>
<td>GLMM</td>
<td>0.36</td>
<td>0.02</td>
<td>23.9</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>IDW</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.34</td>
<td>0.03</td>
<td>-</td>
<td>0.10</td>
</tr>
<tr>
<td>GLM</td>
<td>0.35</td>
<td>0.02</td>
<td>10.9</td>
<td>0.00</td>
</tr>
<tr>
<td>GLMM</td>
<td>0.21</td>
<td>0.02</td>
<td>4.9</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>tuna</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.64</td>
<td>0.02</td>
<td>-</td>
<td>0.37</td>
</tr>
<tr>
<td>GLM</td>
<td>0.68</td>
<td>0.02</td>
<td>42.3</td>
<td>0.47</td>
</tr>
<tr>
<td>GLMM</td>
<td>0.59</td>
<td>0.02</td>
<td>25.5</td>
<td>0.56</td>
</tr>
</tbody>
</table>

103
Table 6: 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 mean2), sea surface temperature standard deviation (SST SD) and chlorophyll a (Chl a). 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 to 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 alone) and which environmental variables decreased the gain the most when omitted (worst without) 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 B1).

<table>
<thead>
<tr>
<th>Season</th>
<th>autumn</th>
<th>winter</th>
<th>spring</th>
<th>summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>random – linear and quadratic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top ranked</td>
<td>SST mean²</td>
<td>Chl a</td>
<td>SST mean</td>
<td>depth</td>
</tr>
<tr>
<td>Best / Worst</td>
<td>shore</td>
<td>Chl a</td>
<td>Chl a / SST SD</td>
<td>depth</td>
</tr>
<tr>
<td>AUC</td>
<td>0.721</td>
<td>0.701</td>
<td>0.691</td>
<td>0.668</td>
</tr>
<tr>
<td>IDW – linear and quadratic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top ranked</td>
<td>SST mean²</td>
<td>depth</td>
<td>Chl a</td>
<td>depth</td>
</tr>
<tr>
<td>Best / Worst</td>
<td>depth</td>
<td>Chl a</td>
<td>Chl a</td>
<td>depth</td>
</tr>
<tr>
<td>AUC</td>
<td>0.625</td>
<td>0.574</td>
<td>0.627</td>
<td>0.607</td>
</tr>
<tr>
<td>tuna – linear and quadratic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top ranked</td>
<td>SST mean²</td>
<td>SST mean²</td>
<td>SST mean</td>
<td>depth</td>
</tr>
<tr>
<td>Best / Worst</td>
<td>shore</td>
<td>Chl a</td>
<td>Chl a / SST SD</td>
<td>depth</td>
</tr>
<tr>
<td>AUC</td>
<td>0.710</td>
<td>0.673</td>
<td>0.672</td>
<td>0.676</td>
</tr>
</tbody>
</table>

104
**background – linear and quadratic features**

<table>
<thead>
<tr>
<th>Top ranked</th>
<th>SST mean$^2$</th>
<th>SST mean$^2$</th>
<th>SST mean</th>
<th>SST mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best / Worst</strong></td>
<td>depth</td>
<td>Chl a</td>
<td>SST mean$^2$ / Shore</td>
<td>depth</td>
</tr>
<tr>
<td><strong>AUC</strong></td>
<td>0.928</td>
<td>0.837</td>
<td>0.858</td>
<td>0.843</td>
</tr>
</tbody>
</table>

**background – auto features**

<table>
<thead>
<tr>
<th>Top ranked</th>
<th>Chl a</th>
<th>Chl a</th>
<th>SST mean$^2$</th>
<th>Chl a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best / Worst</strong></td>
<td>Chl a</td>
<td>Chl a</td>
<td>SST mean / Chl a</td>
<td>Chl a</td>
</tr>
<tr>
<td><strong>AUC</strong></td>
<td>0.961</td>
<td>0.920</td>
<td>0.956</td>
<td>0.928</td>
</tr>
</tbody>
</table>
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 Jr. 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; Graham & Roberts, 2007; Jonahson & Harding, 2007; Meekan et al., 2006; Rowat et al., 2009a). 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 (Figure 12). 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 (Cárdenas-Palomo et al., 2010; Sleeman et al., 2007; Wilson, 2002; Wilson et al., 2001), 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 sea surface temperature patterns (Figure B3). Despite average temperatures ranging between 23 and 34 °C (Figure B3), 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 sea surface temperature 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 which 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), white sharks (10-14°C; Boustany et al., 2002). Our results are consistent with much of what is known about the occurrence of these sharks from finer-scale (km – 100s km) studies of nearshore aggregations in the Indian Ocean (Figure B7). 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 was 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 (Figure B7 - 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 sea surface temperature 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 (Figure B2).

Generally, we found that sea surface temperature was a better predictor of whale shark distributions than chlorophyll $a$. 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 to be time (and therefore, spatial) lags between peaks of chlorophyll $a$ 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 chlorophyll $a$ 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 (Sleeman et al., 2010b; Wilson et al., 2007) 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 suggest at least some migratory behaviour at the individual level, implying that broad-scale movements are possible such that individuals could visit several known aggregation sites as they follow suitable environmental conditions among seasons. We identified sea surface temperature 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 inter-annual variability in sightings at an ocean scale. Fluctuating conditions measured at inter-annual scales, using remote sensing can be used to infer inter-annual 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.

GLM 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 datasets. 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-absences 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 dataset 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 (Figure B7). This resulted in a reduced capacity to predict some known whale shark areas of occurrence relative to GLM and GLMM predictions (Figure B7). Additionally, MaxEnt identified depth, chlorophyll a and sea surface temperature 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 km²), 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 (Beger & Possingham, 2008; Lehmann et al., 2002; 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 eco-tourism on the probability of the long-term persistence of the species.
CHAPTER IV. Spatial and temporal predictions of inter-decadal trends in Indian Ocean whale sharks

A. Sequeira\textsuperscript{a}, C. Mellin\textsuperscript{a,b}, S. Delean\textsuperscript{a}, M. Meekan\textsuperscript{c}, C. Bradshaw\textsuperscript{a,d}

\textsuperscript{a}The Environment Institute and School of Earth and Environmental Sciences, The University of Adelaide, South Australia 5005, Australia

\textsuperscript{b}Australian Institute of Marine Science, PMB No. 3, Townsville MC, Townsville, Queensland 4810, Australia

\textsuperscript{c}Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling Hwy, Crawley, Western Australia 6009, Australia.

\textsuperscript{d}South Australian Research and Development Institute P.O. Box 120, Henley Beach, South Australia 5022, Australia

Marine Ecology Progress Series (\textit{in press})
Statement of authorship

SPATIAL AND TEMPORAL PREDICTIONS OF INTER-DECADAL TRENDS IN INDIAN OCEAN WHALE SHARKS

Text in manuscript

SEQUEIRA, A.M.M. (Candidate)
Planned the article, performed the analysis, interpreted data, developed the models, wrote manuscript and acted as corresponding author.
I hereby certify that the statement of contribution is accurate.

Signed……………………………………………………………………….Date……………

MELLIN, C.
Helped generating the hypothesis, developing the models and assisted writing.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………

DELEAN, S.
Assisted data interpretation and helped with the development of the models.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………

MEEKAN, M. G.
Supervised development of work and provided evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed:……………………………………………………………….. Date:……………….

BRADSHAW, C.J.A.
Supervised development of work, helped with data and results interpretation, and assisted writing.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed:……………………………………………………………….. Date:……………….
Spatial and temporal predictions of inter-decadal trends in Indian Ocean
whale sharks

Abstract

The processes driving temporal distribution and abundance patterns of whale sharks remain largely unexplained. We present a temporal trend analysis of whale shark occurrence at an ocean-basin scale, incorporating both spatial and temporal elements. We test the hypothesis that the average sighting probability of sharks in the western Indian Ocean has not changed over nearly two decades, and whether variance in this parameter can be partially explained by climate signals. We used a 17-year dataset (1991-2007, autumn only) of whale shark observations recorded in logbooks of tuna purse-seiners covering most of the western Indian Ocean. We randomly generated pseudo-absences both in space and time, and applied sequential generalised linear mixed-effects models within a multi-model information-theoretic framework, accounting for sampling effort and random annual variation, to evaluate relative importance of temporal and climatic predictors on sighting probability. After accounting for seasonal patterns in distribution, we found evidence that whale shark sighting probability varied with fishing effort, and increased slightly in the first half of the sampling interval (1991 – 2000) to decrease thereafter (2000 – 2007). The highest-ranked model (including a spatial predictor of occurrence, fishing effort, a quadratic term for time and a random spatial effect) explained ~ 60 % of the deviance in sighting probability. The inclusion of El Niño variation in the central Pacific resulted in the highest model rank, with a weak positive effect on the sighting probability in later years. Changes in seasonal distribution explained 27 % of the deviance in annual occurrence. We found that sighting probability increases slightly with rising sea surface temperature in the central Pacific Ocean and reduced temperatures in the Indian Ocean. The declining phase of
the peak, concurrent with recent accounts of declines in population size at near-shore aggregations and with the most pronounced global warming, deserves continued investigation. Teasing apart the legacy effects of past exploitation and those arising from on-going climate changes will be a major challenge for the successful long-term management of the species.

KEYWORDS temporal trends; *Rhincodon typus*; tuna purse-seine fisheries; generalised linear mixed-effects models; spatial distribution; satellite data

Introduction

Understanding the mechanisms driving observed patterns in species occurrence in space and time is a key and challenging objective in ecology (Gotelli *et al.*, 2010; Pimm *et al.*, 1995). While global declines in exploited marine fish species are well-documented (Jackson *et al.*, 2001; Pauly *et al.*, 2005; Roberts, 2002), the evidence for large shifts in distribution and abundance of much of the world’s biota arising from a warming climate is also mounting (e.g., Parmesan, 2006; Traill *et al.*, 2010; Walther *et al.*, 2002). There is, however, considerably less evidence for climate change-induced shifts in the marine environment (but see Hoegh-Guldberg & Bruno, 2010; Last *et al.*, 2011; Sumaila *et al.*, 2011; Wernberg *et al.*, 2011), principally due to the physical and economic constraints of collecting long-term datasets in the marine realm (Richardson & Poloczanska, 2008). These problems are exacerbated for elusive migratory marine species because the low probability of detection can become an important issue in quantifying trends (Gotelli *et al.*, 2010) and in disentangling spatial and temporal patterns.
The whale shark (Rhincodon typus Smith, 1828) is a highly migratory species (Sequeira et al., 2013) found in warm and temperate waters around the globe (Last & Stevens, 1994). Aggregating seasonally near the shore at specific coastal locations (e.g., Rowat, 2007), it became an important species both for fishing (Pravin, 2000) and ecotourism industries (Rowat & Engelhardt, 2007). Due to the species’ poorly quantified population size, demography and behaviour, as well as evidence for regional declines (Bradshaw et al., 2008; Bradshaw et al., 2007), the legal targeted commercial fisheries have been banned (Bradshaw et al., 2008; Theberge & Dearden, 2006), and whale sharks are now classified as Vulnerable in the IUCN Red List (www.iucnredlist.org). In contrast, whale shark ecotourism is growing worldwide (e.g. Hsu et al., 2012; Pierce et al., 2010; Rowat & Engelhardt, 2007), but is highly dependent on the expectation that animals will return to the same locations every year.

Reported whale shark sighting rates within these locations are highly variable, even within the same months (e.g., de la Parra Venegas et al., 2011). Such variation in local occurrence has been associated with fluctuation in climatic signals such as El Niño events and the Southern Oscillation index (Sleeman et al., 2010a; Wilson et al., 2001). There have also been several attempts to quantify trends in whale shark population size and abundance (Meekan et al., 2006; Rowat et al., 2009a), although these studies have been based mostly on data from near-shore aggregations composed largely of juvenile males (Meekan et al., 2006). Due to the transitory nature of whale shark occurrence, some have suggested that regional approaches should instead be used to quantify broader-scale patterns spatial patterns and temporal trends (Rowat et al., 2009b). However, expanding a study site from one aggregation to a region (which might include multiple aggregation sites) inevitably results in adding spatial complexity to the process.
Partitioning variance across spatial and temporal gradients to detect patterns in species occurrence is not usually straightforward, but can be addressed statistically through the use of random-effects models (see Ogle, 2009; Qian et al., 2010). These multilevel, mixed-effects or hierarchical models have been used extensively to understand temporal trends in species assemblages (Gotelli et al., 2010), community structure and patterns in the marine environment (e.g., MacNeil et al., 2009; Mellin et al., 2010a), abundance and biomass of species (Ruiz & Laplanche, 2010) and biological responses to different environmental conditions (Bedoya et al., 2011).

Using a long-term (1991 – 2007) and wide-extent dataset of whale shark sightings in the Indian Ocean collected by the tuna purse-seine industry, we present an analysis incorporating both spatial and temporal elements to examine temporal trends of this species at the ocean-basin scale. Following previous work quantifying the habitat suitability and seasonal variation in whale shark relative abundance in the Indian Ocean (Sequeira et al., 2012), our latest approach now partitions the complex spatial and temporal variation in whale shark occurrence patterns. Specifically, we test the hypothesis that sighting probability remains constant over time, and quantify the influence of global climatic signals on temporal patterns of occurrence at a broad spatial scale.

**Methods**

We developed generalised linear mixed-effects models (GLMM) sequentially, with the first step assessing the evidence for a temporal trend in whale shark occurrence, and a second testing the hypothesis that sighting probability is correlated with variation in
climatic indices. Below we detail the datasets used (presence/absence data and sampling effort) and the modelling steps (predictors and model development).

**Whale shark dataset**

We used data recorded in logbooks from purse-seine fishers registered under the Indian Ocean Tuna Commission (Pianet et al., 2009). These logbooks contain long-term (1991 - 2007) data on whale shark occurrences (hereafter referred to as ‘sightings’) derived from associated net-sets for tuna catch using whale sharks as fish aggregation devices. A total of 1185 sightings were recorded during the sampling period, including date and location at a 0.01° resolution (i.e., latitude and longitude data were collected using the GCS WGS84 system and made available in units of decimal degrees to a precision of 1/100th of a degree). The dataset provided no indication of gender for the sighted sharks. Due to substantial fluctuation over time and higher numbers of sightings occurring mostly in autumn (Figure 13 – top), we used data from this season to examine the inter-annual trends in autumn whale shark occurrence (Figure C1 in Appendix C shows spatial variation in occurrences). Seasonal patterns of whale shark occurrence in the same dataset were previously described by Sequeira et al. (2012).
Figure 13: Temporal variation in whale sharks sighted by the tuna purse-seine fisheries and comparison with climatic indices.

Top chart: Total number of whale sharks sighted by tuna purse-seine fisheries per month and per year. Grey bars represent autumn (Apr-Jun from 1991 to 2007). Bottom charts: Total number of whale sharks sighted in autumn per year, overlaid with the values for the Indian Ocean Dipole (IOD; left) and sea surface temperature variation in Region 4 of the central Pacific due to El Niño/La Niña events (NiNO4; right).

Pseudo-absence generation

Whale shark sightings were presence-only, so we generated pseudo-absences to produce the denominator of the logit function that allows for the binomial estimation in our GLMM detailed below. For each presence recorded, we randomly generated 100 pseudo-absences (1:100 ratio) both (i) in space by randomly choosing non-presence grid cells over the western Indian Ocean (function `srswor` – simple random sampling without replacement – from the `Sampling` package in R), and (ii) in time by randomly assigning to the selected point an autumn date within the 17-year interval (function
srswr – simple random sampling with replacement – from the \{Sampling\} package in R) (R Development Core Team, 2011). The high presence/pseudo-absence ratio (1:100), which inevitably results in low prevalence (0.01), allows a better representation of the ‘background’ available, which consisted of both spatial (each grid cell) and temporal (a specific date within the time period covered in the dataset) components.

**Fisheries effort data**

Effort data (number of fishing days per month) were recorded with a resolution of 5° within the area covered by the fisheries (30° N – 30° S and 35° – 100° E; grey area in Figure 14), giving a total of 1638 records (autumn only) and 13674 fishing days with associated net-sets. The variables *effort* and number of sightings per year are illustrated in Figure 14. Because the eastern part of the Indian Ocean (east of the Maldives) was sampled only in one year during autumn (1998 – Figure C1), we used only the western area of the Indian Ocean (west of the Maldives as depicted in Figure 15) in our temporal analysis.

![Figure 14: Example of variation in fishing effort (days – top row) and number of whale sharks sighted (bottom row).](image-url)
Only two years of autumn (April-June from 1991 to 2007) data when effort was highest (1997 - left) and lowest (2005 - right) shown. Grey area depicts the total area sampled by the tune purse-seiners.

Figure 15: Results of the fishing effort analysis across the western Indian Ocean where the purse-seine fisheries operated in more than one year.

Green line = limit of the western Indian Ocean area considered. Squares represent the 5-˚ area for which fishing effort data were available, and colours represent the effort analysis results based on the GLM evidence ratios: $\text{effort} \sim \text{Time}$ against the null model. Values inside each square are the estimates obtained for the coefficient of the $\text{Time}$ variable (when the null models were not ranked higher) after Holm correction.

Spatio-temporal variation in sampling effort can affect the ability to detect temporal trends in occurrence (Phillips et al., 2009), so we developed a series of generalised linear models (GLM) with a Poisson error distribution using time in years ($\text{Time}$) as a predictor for effort (in each of the 5-˚ grid squares sampled more than
once during autumn), to test if the spatial patterns in sampling effort were evenly distributed throughout the 17-year period. We compared the results of these models with a null (intercept-only) model by calculating the evidence ratio – a bias-corrected index of the likelihood of one model over another ($w_{AIC_c^{(GLM \text{ Time})}} / w_{AIC_c^{(GLM \text{ null})}}$) – for each model. To control for inflation of type I errors due to multiple testing across grid squares, we used the Holm correction through the *Bioconductor* (*multtest*) package (Pollard *et al.*, 2005) in R (R Development Core Team, 2011).

**Model predictors**

This section describes the predictors used in each model step, detailing how we first accounted for temporal variation both in effort and sightings (*Step 1*), and then tested for correlations between observed trends and changes in climatic predictors (*Step 2*).

**Step 1**

Because temporal and spatial variation are seldom dissociated, we needed to include a term covering the variation in spatial probability of whale shark occurrence to test the hypothesis that the average probability of sightings remained constant over time within the large area under study. For this we used the results derived from Sequeira *et al.* (2012), where we fitted models of seasonal spatial distribution of whale sharks in the Indian Ocean. Here we re-fitted the whale shark distribution model for autumn (Sequeira *et al.*, 2012), including some modifications to the likelihood estimation, pseudo-absence ratio, and covariate treatments (see Supplementary Information in Appendix C for a detailed description of these changes). We then used logit-scale predictions of the likelihood of whale shark occurrence in autumn (spatial probability -
SpatialP; Figure C2) within each 9-km grid cell (resolution used in Sequeira et al., 2012) as an explanatory variable of whale shark occurrence in the temporal models developed herein. The inclusion of SpatialP in the GLMM (see below) accounts for possible spatial autocorrelation.

Temporal changes in effort can explain some of the inter-annual variation detected in sightings; therefore, we also added effort as a predictor in our temporal models to account for its potential effect on the temporal patterns observed. Because the mean effort for autumn was already accounted for within the spatial predictor, we only included the temporal variation around the mean (i.e., the zero-centred inter-annual deviations around mean effort). We included both a fixed and a random effect for time (Time and year, respectively) to account for inter-annual variability in presences, allowing the random structure to contain only information that could not have been modelled with fixed effects (following Zuur et al., 2009).

Step 2

To test the hypothesis that inter-annual variation in whale shark sightings was correlated with variation in indices of sea surface temperature and air pressure as reported previously in a near-shore aggregation (Sleeman et al., 2010a; Wilson et al., 2001), we considered variation in climatic indices in both the Indian and the Pacific Oceans. We tested 4 indices: (i) the Indian Ocean Dipole (IOD – Saji et al., 1999), (ii) El Niño variation in the central Pacific Region 4, 160 °E – 150 °W / 5 °S – 5 °N (NINO4, Burgers & Stephenson, 1999), (iii) the Oceanic Niño Index, ONI – three-month moving averages of sea surface temperature in the Niño 3.4 Region -170 – 120 °W / 5 °S – 5 °N and the (iv) Southern Oscillation Index (SOI – Walker, 1925). We
collected the online climatic indices for the total period covered by the purse-seine fisheries from the Earth System Research Laboratory (U. S. Department of Commerce, National Oceanic and Atmospheric Administration, Boulder, Colorado, www.esrl.noaa.gov/index.html), the Japanese Agency for Marine-Earth Science and Technology (www.jamstec.go.jp/frsgc/research/d1/iod) and the Australian Government Bureau of Meteorology (www.bom.gov.au). To understand how the different indices are associated with each other and to assist interpretation of the model results, we investigated the correlation between the climatic predictors using the *pairs.panels* function in the package `{psych}` in R (R Development Core Team, 2011) and their monotonic relationships using Spearman’s rank correlation ($\rho$). IOD is weakly collinear with both NINO4 and ONI (Pearson coefficients = 0.06 and -0.06, respectively) while the two latter predictors are highly correlated (Pearson coefficient = 0.80) (Figure C3). Spearman coefficients also showed collinearity only between NINO4 and ONI ($\rho = 0.8$).

### Models

We applied generalised linear mixed-effects models (GLMM) with a binomial error distribution and a logit link function to compare the predictive ability of different combinations of the predictors. The mixed-effects models we developed in each step included all possible combinations of the fixed and random effects (Step 1), and each of the four individual climatic predictors (Step 2). By including climatic variables one at a time, we concurrently tested whether replacing *Time* by any of the climatic variables could explain away any trends observed in Step 1.

In each step, we compared models based on the Akaike’s information criterion corrected for small sample sizes ($\text{AIC}_c$) (Burnham & Anderson, 2004), which favours
models with higher predictive capacity when sample sizes are large and tapering
effects exist – as expected to occur in our spatio-temporal models. We assessed each
model’s strength of evidence relative to the entire model set by calculating $AIC_c$ model
weights ($wAIC_c$) and used the percentage of deviance explained ($%De$) to quantify
each model’s goodness-of-fit. We retained the $AIC_c$ top-ranked model from the first
step and used it in the second step as a control model for the more complex
combinations with climatic predictors. We developed all models in R version 2.11.1 (R
Development Core Team, 2011).

Results

Fisheries effort analysis

The spatial analysis of effort (GLM models with Poisson distribution) per grid cell (5°
resolution) demonstrated no evidence of temporal trend in effort in approximately 60 %
of the grid cells (Figure 15). In the remaining cells, the evidence ratios of the models
including $Time$ were high ($> 150$; Figure 15), and effort generally increased over time
(estimated coefficients ranging from $0.03 \pm 0.007$ to $0.23 \pm 0.046$; Figure 15) with the
exception of only four cells where effort decreased with time (coefficients ranging from
$-0.04 \pm 0.005$ to $-0.12 \pm 0.02$; Figure 15). Two of these four declining-effort cells
corresponded to the area where most of the whale shark sightings occurred (compare
Figure 15 and Figure C1).
Models

Step 1

The GLMM with the highest information-theoretic support (waICc = 0.45; Step 1) included the spatial predictor (i.e., spatial predictions of the probability of whale shark occurrence in autumn on a logit scale – SpatialP; Figure C2), the zero-centred effort (inter-annual variation around the autumn mean, i.e., the variation not accounted for by the spatial predictor), and a quadratic temporal trend with a random intercept term accounting for the among-year variance (Table 7). This model explained almost 60 % of the deviance in whale shark sighting probability. We found evidence for a slight increase in the probability of whale shark sightings for the first half of the sampling period (1991-2000), and for a declining trend over the last half (2001-2007) (Figure 16). However, total waICc was shared approximately evenly among the two top-ranked models (Table 7; only one of them including the Time predictor), indicating that the temporal trend was only weak. Results for models not including the spatial term (SpatialP) generally performed poorly relative to the other models in the set, with this predictor
Table 7: Summary of the generalised linear mixed-effects models relating probability of whale shark occurrence to spatial and temporal predictors.

“SpatialP” – a spatial predictor derived from previous spatial distribution models (Sequeira et al., 2012), effort, Time fixed effect predictor for time (in years), and a random effect for year (Step 1) and global climatic predictors: Indian Ocean Dipole (IOD), Southern Oscillation Index (SOI), El Niño in the central Pacific - Region 4 (NINO4) and Oceanic Niño Index (ONI) (Step 2). Shown for each model are the number of parameters (k), log-likelihood (LL), biased-corrected model evidence based on weights of Akaike's information criterion corrected for small sample sizes (wAICc) and the percentage of deviance explained (%De). Top-ranking/best-performing models in each step are in bold and models are ordered by decreasing wAICc. Values < 0.001 not shown.

<table>
<thead>
<tr>
<th>Model</th>
<th>k</th>
<th>LL</th>
<th>wAICc</th>
<th>%De</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpatialP + effort + Time + Time² + (1</td>
<td>year</td>
<td>7</td>
<td>-1956.08</td>
<td>0.454</td>
</tr>
<tr>
<td>SpatialP + effort + (1</td>
<td>year</td>
<td>5</td>
<td>-1958.29</td>
<td>0.370</td>
</tr>
<tr>
<td>SpatialP + effort + Time + (1</td>
<td>year</td>
<td>6</td>
<td>-1958.03</td>
<td>0.176</td>
</tr>
<tr>
<td>SpatialP + effort</td>
<td>5</td>
<td>-2531.80</td>
<td>&lt;0.001</td>
<td>44.3</td>
</tr>
<tr>
<td>SpatialP + effort + Time</td>
<td>6</td>
<td>-2531.72</td>
<td>&lt;0.001</td>
<td>44.3</td>
</tr>
<tr>
<td>SpatialP + Time + Time² + (1</td>
<td>year</td>
<td>6</td>
<td>-2707.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SpatialP + (1</td>
<td>year</td>
<td>4</td>
<td>-2710.93</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SpatialP + Time + (1</td>
<td>year</td>
<td>5</td>
<td>-2710.87</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>effort + Time + Time² + (1</td>
<td>year</td>
<td>6</td>
<td>-3201.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>effort + (1</td>
<td>year</td>
<td>4</td>
<td>-3203.59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>effort + Time + (1</td>
<td>year</td>
<td>5</td>
<td>-3202.95</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SpatialP + Time</td>
<td>5</td>
<td>-3313.77</td>
<td>&lt;0.001</td>
<td>27.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------------</td>
<td>---</td>
<td>------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>SpatialP</td>
<td>4</td>
<td>-3317.50</td>
<td>&lt;0.001</td>
<td>27.0</td>
</tr>
<tr>
<td>Time + Time(^2) + (1</td>
<td>year)</td>
<td>5</td>
<td>-3995.69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1 + (1</td>
<td>year)</td>
<td>3</td>
<td>-3998.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Time + (1</td>
<td>year)</td>
<td>4</td>
<td>-3998.23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>effort</td>
<td>4</td>
<td>-4025.96</td>
<td>&lt;0.001</td>
<td>11.4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-4544.22</td>
<td>&lt;0.001</td>
<td>0</td>
</tr>
</tbody>
</table>

**Step 2**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialP + effort + Time + Time(^2) + NINO4 + (1</td>
<td>year)</td>
<td>8</td>
<td>-1861.75</td>
<td>0.81</td>
</tr>
<tr>
<td>SpatialP + effort + NINO4 + (1</td>
<td>year)</td>
<td>6</td>
<td>-1865.18</td>
<td>0.19</td>
</tr>
<tr>
<td>SpatialP + effort + Time + Time(^2) + ONI + (1</td>
<td>year)</td>
<td>8</td>
<td>-1926.15</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + ONI + (1</td>
<td>year)</td>
<td>6</td>
<td>-1929.97</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + Time + Time(^2) + IOD + (1</td>
<td>year)</td>
<td>8</td>
<td>-1954.44</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + IOD + (1</td>
<td>year)</td>
<td>6</td>
<td>-1956.63</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + Time + Time(^2) + (1</td>
<td>year)</td>
<td>7</td>
<td>-1956.08</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + (1</td>
<td>year)</td>
<td>5</td>
<td>-1958.29</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + Time + Time(^2) + SOI + (1</td>
<td>year)</td>
<td>8</td>
<td>-1956.00</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SpatialP + effort + SOI + (1</td>
<td>year)</td>
<td>6</td>
<td>-1958.16</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-4544.22</td>
<td>&lt;0.01</td>
<td>0</td>
</tr>
</tbody>
</table>
alone explaining 27% of the deviance (Table 7). A positive relationship with the zero-centred inter-annual deviations from mean fishing effort (coefficient estimate = 1.63 ± 0.06) indicates that the number of sightings is higher when effort is higher than average and vice versa.

Figure 16: Partial effect of time on whale shark presence.
(a) on the log-odds scale showing the rate of change and (b) on the probability scale, showing the effect of time on the probability of whale shark presence. Dotted lines represent 95% confidence interval.

Step 2

Including the index of sea surface temperature in the central Pacific (NINO4) in the highest-ranked model from Step 1 above resulted in the highest statistical support (wAICc = 0.81). However, the percentage of deviance explained (59%) was only slightly higher than the model excluding the climatic predictor (57%). All models that included a climatic predictor reflecting variation in sea surface temperature (NINO4, IOD, ONI) had higher support than those without climatic variables (from Step 1), while the model including SOI (relative to air pressure variation) performed poorly even when compared to models excluding climate signals (Table 7). The partial effects of
the climatic variables showed that an increase in NINO4 (reflecting higher sea surface temperature in the central Pacific Ocean) had a positive effect on whale shark probability of occurrence in the western Indian Ocean (Figure 17), and an increase in IOD (reflecting higher sea surface temperature in the western part of the Indian Ocean) had a negative effect (Figure C4).

Figure 17: Partial effects of NINO4 and Time (after accounting for climatic contributions) in the probability of whale shark presence.

Top row: Partial effect of NINO4 in the probability of whale shark presence. NINO4 is an index representing sea surface temperature variation in the central Pacific Region 4 (160 °E – 150 °W / 5 °S – 5 °N; Burgers and Stephenson, 1999). Bottom row: The partial effect of Time after accounting for climatic contributions to whale shark probability of presence is shown. Dotted lines represent 95 % confidence interval.
Discussion

Access to a long-term dataset of whale shark sightings covering the western sector of an entire ocean basin provided a unique opportunity to analyse temporal trends and variation in whale shark occurrence at a scale more likely to encompass the range of these highly migratory animals, and for a greater proportion of the population than in previous studies. Overall, our results highlighted a high inter-annual consistency in whale shark distribution patterns, with the spatial predictor (i.e., mean seasonal distribution) alone accounting for 27% of the deviance. We also found evidence for a modest peak in whale shark occurrences in the middle of the time series (~2000), which prevailed after accounting for changes in global climatic indices. To date, analogous temporal analysis of whale shark occurrence have only been done at the scale of single aggregations (~10s of km; Sleeman et al., 2010a) covering only a small part of this species’ range and only a small proportion of the population (mostly male juveniles, Meekan et al., 2006). Thus, our results are the first to estimate ocean-scale trends in occurrence that include several known aggregation sites for the species, and it is the first analysis to account simultaneously for spatial and temporal components, including effort, that can mask underlying trends in the probability of occurrence.

The high percentage of deviance explained by the spatial predictor (27%) implies that distribution patterns are annually consistent (autumn only), with a higher probability of occurrence around the Mozambique Channel. This consistency might be due to permanent characteristics of the area, such as physical features enhancing upwelling, local productivity and tolerable sea surface temperatures (Sequeira et al., 2012).
Modelling must ensure that variation in sampling effort does not confound results (Dennis et al., 1999). In our dataset, effort generally increased through time, although there was a decline in the Mozambique Channel (Figure 15) where most sightings occurred during the 17 years (compare with Figure C1). By including the zero-centred inter-annual deviations from mean fishing effort in our models (which represent temporal fluctuations from the mean within each grid cell) (Sequeira et al., 2012), we accounted for the contribution of effort in the temporal probability of occurrence. Combining the spatial predictor just with effort, our models explained ~ 44% of the deviance (Table 7), demonstrating the importance of incorporating both spatial and sampling effort components in temporal models.

We hypothesise that the non-linear temporal trend we observed (Figure 16), even though weak, reflects variation in whale shark abundance in the western area of the Indian Ocean. Such a trend, where the probability of occurrence first increases and then decreases, could represent an inter-decadal cycle (i.e., cycle over 15 years), similar to what has been observed for strandings of other large marine species in the Southern Ocean (Evans et al., 2005) and fish catches in the same region (Jury et al., 2010). In the latter study, such cycles were only just detected even with annual catch records spanning 40 years. Our dataset, covering < 20 years, revealed evidence (although weak) for a peak in whale shark occurrence that is in accordance with the longer cycles found for fish catches (Jury et al., 2010).

Focusing only on the declining segment of the identified peak, which might be a reflection of a long-term decline in relative abundance, we suggest three competing hypotheses to explain it: (i) a downward component of possible inter-decadal abundance (as presented above), (ii) a real decline in abundance, or (iii) a distributional shift due to changing habitat characteristics (e.g., via climate change).
Considering the decline observed as reflecting a real decline in abundance, our results support previous data (Bradshaw et al., 2008) and predictions (Bradshaw et al., 2007) from Ningaloo in Western Australia suggesting a decline in population size. In fact, the South-East Asian and Indian whale shark-targeted fisheries (Hsu et al., 2012; Pravin, 2000) were banned due to the low numbers captured even with increasing effort – a signal that can indicate population decline. Despite the bans, targeted fisheries still present a risk to whale sharks due to lack of enforcement (Stewart & Wilson, 2005) and there is evidence that illegal fishing is still occurring (Riley et al., 2009; White & Cavanagh, 2007). Such illegal fisheries could still be responsible for on-going declines, although it is currently impossible to determine to what extent this might be still occurring. The unintentional bycatch of whale sharks in other fisheries (e.g. purse seiners; Romanov, 2002) also remains unquantified.

The decline segment is also concurrent with the most pronounced warming observed over the last decade, such that a related hypothesis where latitudinal shifts in distribution (e.g., to areas not considered in this study) resulting from rising water temperatures (as suggested in Sequeira et al., 2012) driven by global climate change could be influential. Our models’ estimates suggest that sighting probability increases with rising sea surface temperature in the central Pacific Ocean (Figure 16) and reduced temperatures in the Indian Ocean (Figure C4). Despite IOD being a climatic index directly measured in the Indian Ocean, our models gave higher support for NINO4 as a predictor. However, there were some occasions where a few whale sharks were sighted despite high NINO4 values (Figure 13). This is particularly interesting for 1997 when not only the NINO4 index was high, effort was also maximal (Figure 14). Our data covered mainly the western Indian Ocean, but there was a peak in sightings at Ningaloo in 1997 (Wilson et al., 2001). The same occurred in 1992.
Wilson et al., 2001) and 2002 (see Sleeman et al., 2010a), when NINO4 was high and few sharks were observed in the western Indian Ocean. In both 1997 and 2002, IOD was positive (0.04 to 0.88) and coincided with a decrease in probability of occurrence in the western Indian Ocean (Figure C4). When IOD is positive, the thermocline deepens in the eastern Indian Ocean, resulting in more intense upwelling and greater surface productivity (Behera & Yamagata, 2003). Saji et al. (1999) first described IOD as independent of ENSO (El Nino / Southern Oscillation), although others suggest possible associations between the two (reviewed by Maity et al., 2007). Understanding how these coupled sea surface-atmosphere phenomena are related is not our aim, but the (weak) support for Pacific and Indian Ocean indices suggest some possible effects of broad-scale oceanic climate events influencing whale shark habitat use.

Despite cycles in fish abundance in the western Indian Ocean being directly related to environmental fluctuation, most of the variance in fish catch can be explained with estimates of local productivity (Jury et al., 2010). Changes in climate are already affecting the base of the principal food web – phytoplankton (e.g. Edwards & Richardson, 2004). This evidence, coupled with our demonstration of possible climate-influenced distribution of the world’s largest fish, demonstrates a need to predict how species will respond to future climate scenarios. As such, long-term management of whale sharks requires both a better understanding of basic ecology and demography, and how these will be altered as the oceans warm.
CHAPTER V. Inter-ocean asynchrony in whale shark occurrence

A. Sequeira\textsuperscript{a}, C. Mellin\textsuperscript{a,b}, M. Meekan\textsuperscript{c}, C. Bradshaw\textsuperscript{a,d}

\textsuperscript{a}The Environment Institute and School of Earth and Environmental Sciences, The University of Adelaide, South Australia 5005, Australia

\textsuperscript{b}Australian Institute of Marine Science, PMB No. 3, Townsville MC, Townsville, Queensland 4810, Australia

\textsuperscript{c}Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling Hwy, Crawley, Western Australia 6009, Australia.

\textsuperscript{d}South Australian Research and Development Institute P.O. Box 120, Henley Beach, South Australia 5022, Australia

\textit{(Manuscript in preparation)}
INTER-OCEAN ASYNCHRONY IN WHALE SHARK OCCURRENCES

Text in manuscript

SEQUEIRA, A.M.M. (Candidate)
Planned the article, performed the analysis, interpreted data, developed the models, wrote manuscript and acted as corresponding author.
I hereby certify that the statement of contribution is accurate.

Signed……………………………………………………………………….Date……………

MELLIN, C.
Helped planning the article, and assisted data interpretation and models development.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis

Signed……………………………………………………………………….Date……………

MEEKAN, M. G.
Provided evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis

Signed……………………………………………………………………….Date……………
BRADSHAW, C.J.A.
Supervised development of work, helped in model results interpretation and manuscript evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for the inclusion of the manuscript in the thesis.

Signed..........................................................Date..............
Inter-ocean asynchrony in whale shark occurrence

Abstract

The whale shark is a highly migratory species, with genetic evidence for inter-ocean connectivity. Given this migratory behaviour, population trends can only be understood by examining patterns of occurrence synchrony among aggregation locations. We present a two-step modelling approach of whale shark spatial and temporal occurrence in the Atlantic and Pacific Oceans by applying generalised linear mixed-effects models. To test the hypothesis that the probability of whale shark occurrence is asynchronous across oceans, as expected if inter-ocean migration is occurring, we used long-term datasets of whale shark sightings (April to June) derived from tuna purse-seine logbooks covering most of the central-east Atlantic (1980-2010) and western Pacific (2000-2010). We first predicted seasonal habitat suitability and produced maps of shark distribution in each area, and then evaluated the relative effect of time (year) on sighting probability. Additionally, we applied fast Fourier transforms to determine if any periodicity was apparent in whale shark occurrences in each ocean. After partialling out the effects of seasonal patterns in spatial distribution and sampling effort, we found no evidence for a temporal trend in the Atlantic. Conversely, we detected an increase in the whale shark probability of occurrence in the Pacific. The highest-ranked model for the latter included a spatial predictor of whale shark occurrence along with fishing effort, a linear term for time, and a random temporal effect (year) (15 % deviance explained). Fast Fourier transforms revealed a prominent 15-year cycle for shark occurrences in the Atlantic. We further reveal that the increase in probability of shark occurrence in the Pacific is concurrent with a decrease in the Indian Ocean. We conclude that cyclic patterns driven by migratory behaviour would better explain temporal trends in whale shark occurrence at the
oceanic scale. Despite cycles partially explaining observations of fewer sharks in some years, overall reported sighting rate has been decreasing. More data are needed to examine the flow of individuals across oceans to isolate mechanisms driving change.

KEYWORDS temporal trends; Rhincodon typus; tuna purse-seine fisheries; generalised linear mixed-effects models; spatial distribution; satellite data

Introduction

While species loss in terrestrial systems is strongly associated with habitat degradation (e.g., Pimm & Raven, 2000), most of humanity’s obvious negative impacts in marine ecosystems result from direct exploitation (Pauly et al., 1998; Worm et al., 2006) or related by-catch of marine life (Agardy, 2000; Hall et al., 2000), and now climate change is also beginning to impose its toll (e.g., Sumaila et al., 2011). Reported declines in marine species challenge the idea that extinctions in the oceans are unlikely (Hendriks et al., 2006), and raise awareness that understanding connectivity is paramount to assist sustainable management (Agardy, 2000).

Whale sharks (Rhincodon typus, Smith 1928) travel large distances (e.g., Rowat & Gore, 2007), and its sub-populations are expected to be connected across the world’s oceans (Castro et al., 2007; Sequeira et al., 2013). Due to the species’ highly migratory behaviour, concerns regarding the adequacy of current management measures have been raised (Rowat, 2007). These measures mostly concern confined areas where these sharks are economically important (e.g., for tourism) (Pierce et al., 2010; Quiros, 2007), and might neglect negative impacts occurring elsewhere (Bradshaw, 2007). Whale shark-based eco-tourism has been developed based on the anticipation that individuals return to the same location each year at approximately the
same time (Taylor, 1996); however, evidence for declining relative abundance has been reported at some of these locations (Bradshaw et al., 2008; Theberge & Dearden, 2006). There is also some quantitative support for a slight reduction in the probability of occurrence in the Indian Ocean during the last decade (Sequeira et al., in press).

Whale sharks are potentially affected by a range of human-driven alterations to the marine environment, including exploitation through direct commercial fisheries (last ban occurred in Taiwan only in 2008) (Council of Agriculture, 2007), poaching (Riley et al., 2009), by-catch (Romanov, 2002), and habitat disturbance (Heyman et al., 2010). Anthropogenic climate disruption is another potential stressor because temperature is an important predictor of whale shark distribution (Sequeira et al., 2012) and relative abundance (Sleeman et al., 2010a). Alternatively or synergistically, observed short-term declines could also be confounded by inter-decadal cycles in relative abundance (Sequeira et al., in press) associated with broad-scale migration patterns. Because this species is highly mobile and populations are connected across oceans at least at the generational scale (Castro et al., 2007; Schmidt et al., 2009), temporal trends can only be truly revealed by comparing the synchrony of occurrence patterns among locations across ocean basins.

Temporal trends in species occurrence are seldom dissociated from spatial processes. Even though statistical models have been mostly used to assess and predict the spatial distribution of species (Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000; Hirzel et al., 2002; Phillips et al., 2009) based on the ecological niche (Kearney & Porter, 2009), they can also be used to assess temporal trends (Gotelli et al., 2010). For example, species distribution models have indeed been used to estimate habitat suitability for highly migratory marine species (Elith & Leathwick,
Access to the logbooks compiled by tuna purse-seiners from the Atlantic and Pacific Oceans gave us the opportunity to use similar models to estimate broad-scale trends in whale shark occurrence to complement the assessment made previously for the Indian Ocean (Sequeira et al., in press). Here we: (1) predict whale shark habitat suitability within the areas covered by the tuna fisheries in the Atlantic and Pacific, (2) test the hypothesis of temporal asynchrony in probability of occurrence, and (3) assess possible cyclic patterns in occurrence. Our main objective is to provide a temporal assessment of whale shark occurrence probability encompassing most of the species’ known geographical range by comparing their probabilities of occurrence in different oceans.

**Methods**

With the main objective to assess temporal trends in whale shark occurrence in the Atlantic and Pacific oceans, and compare them with the results obtained previously for the Indian Ocean (Sequeira et al., in press; Sequeira et al., 2012), the models we developed herein follow a similar approach. First we developed habitat suitability models and used the resulting habitat suitability as part of the input data to model temporal trends in whale shark occurrence. Below we describe the data used (opportunistic fisheries-derived and environmental datasets), the modelling methods, including how we accounted for pitfalls in the opportunistic dataset used (e.g., presence only-data and sampling bias), and how we used the Fourier Transforms to look for possible cyclic patterns in the whale shark sightings dataset.
Data

We used whale shark occurrence data from the Atlantic and Pacific Oceans recorded in the logbooks of tuna purse-seiners. Because tuna aggregate underneath whale sharks (possibly foraging on similar preys), fishers deploy nets around the sharks to facilitate harvest of the target fish (Matsunaga et al., 2003). Hereafter, we use the term ‘sighting’ to describe logbook records of these net sets associated with whale sharks. The dataset made available by the Institut de Recherche pour le Développement (France) and the Secretariat of the Pacific Community comprises most of the central area of the Atlantic (21 °N to 15 °S and 34 °W to 14 °E) and central western Pacific (15 °N to 15 °S and 130 °E to 150 °W) (Figure 18). It includes date (month and year), longitude and latitude information for whale shark sightings (0.01° precision), and spatially aggregated information on sampling effort (number of days spent fishing) per month and per grid cell of 1° resolution in the Atlantic, and 5° resolution in the Pacific (Figure 18). No information on individual vessel or trip units was available. To keep the highest resolution possible in the input data for our models, effort was included as an offset - as detailed below. The data spanned 1980 to 2010 in the Atlantic (total of 18277 records), and 2000 to 2010 in the western Pacific (total of 2272 records provided by only part of the fleets registered with the Secretariat of the Pacific Community, but representative of the fisheries in the area – P. Williams pers. comm.). To compare our results for the Atlantic and Pacific with previous results for the Indian Ocean (Sequeira et al., in press), and due to the generally low number of sightings in other seasons for each ocean, we used data only for the months of April to June (austral autumn).
Figure 18: Whale shark (Rhincodon typus, Smith 1828) datasets recorded by tuna purse-seiners including location of net-sets in the (left) Atlantic (1980-2010, data provided by the Institut de Recherche pour le Développement - France) and (right) Pacific (2000-2010, data provided by the Secretariat of the Pacific Community). Top panel: area covered by the fisheries (grey) and location of whale shark sightings (blue dots). A total of 1030 sightings were recorded in the Atlantic, and 167 in the Pacific Ocean (between April and June). Bottom panel: number of fishing days (effort) with a resolution of 1° for the Atlantic and 5° for the Pacific Ocean.

We assembled environmental data on daytime sea surface temperature (SST in °C) and chlorophyll a (Chl a in mg m⁻³) at a 9-km resolution derived from the Advanced Very High Resolution Radiometer (AVHRR) PathFinder version 5.0 and Sea-viewing Wide Field-of-view Sensor (SeaWiFS) satellites, respectively. We used ArcToolBox functions (ArcGIS 9.3.1™ automated with Python scripts) to calculate mean and standard deviation of SST and Chl a per grid cell for all weekly composites between April and June for the time period of each ocean dataset. We also derived depth (m), slope (°) and distance to shore (km; using the Near tool in ArcGIS 9.3.1™ on an equidistant cylindrical coordinate system) from the General Bathymetry Chart of
the Oceans (GEBCO, 2003). We then collated the full dataset with a common resolution of 9-km including six predictors: mean depth, slope, distance to shore, mean SST, SST standard deviation and mean Chl a.

**Models**

We developed the modelling approach in two steps to (1) compare the spatial predictive ability of different combinations of the environmental predictors in modelling whale shark habitat suitability, and (2) assess evidence for a temporal trend in whale shark occurrence in each ocean. In both steps, we applied generalised linear mixed-effects models (GLMM) – with a binomial error distribution and a logit link function – to our presence-only data by generating pseudo-absences for binomial estimation.

The process of generating pseudo-absences differed in each modelling step. In the first step, we randomly generated 10 pseudo-absences per presence based on a spatially random distribution within the area covered by the fisheries (excluding all the presence cells). In the second step, we generated 100 pseudo-absences per presence based on both temporally and spatially random distributions, that is, randomly choosing a date within the temporal coverage of each dataset, and then randomly assigning to it a location within the area covered by the fisheries (for each ocean). In both steps, we generated the spatially random distributions with the `srswor` function (simple random sampling without replacement) from the `{sampling}` package in the R programming language (R Development Core Team, 2012). For the temporally random pseudo-absence distribution, we randomly selected a date within the temporal coverage of each dataset (April to June only) by using the `srswr` function (simple random sampling with replacement) from the same package in R. To assess the influence of the date and location associated with the selected pseudo-absences,
we generated each set of pseudo-absences 100 times prior to their use in the spatial models, and 10 times for the temporal models (due to computing time).

In the first modelling step (i.e., spatial models), we included in our list of predictors the quadratic terms (nonlinear response) for SST and depth to account for a possible higher suitability for whale shark occurrence within a range of these variables. We did this by using the `poly` function (with degree 2) from the `{stats}` package in R. We also included a code for each 1° grid cell as a spatial random effect to reduce spatial autocorrelation. To account for the sampling bias associated with effort (more whale shark sightings expected in areas where fishing effort was higher), and to keep the highest resolution possible in the input data, we have included effort as an offset in the models.

In the second modelling step (temporal models), we used as predictors the outputs from step 1 (i.e., the resulting spatial habitat suitability – `Hsuit`), a zero-centred `effort` term (i.e., inter-annual variation around the mean not accounted for within the spatial predictor) (Sequeira et al., in press) and time, both as a fixed (`time`) and random effect (`year`) to ensure that the random structure contains only information that could not have been modelled with fixed effects (following Zuur et al., 2009). To account for a parabolic-like dependence of occurrences with time, we also added the quadratic term for `time` using the `poly` function (with degree 2).

A generic way to write the GLMM (with a logit link function) is:

\[
\logit(Presence) = \alpha + \beta X_i + \gamma_i Z_i + \epsilon_i [ + \log(fishing\ effort)]
\]

where `Presence` is the expected mean probability of sighting occurrence. `X` and `Z` represent the fixed and random covariates used in the models: the environmental predictors and the spatial grid in the spatial models, and habitat suitability, zero-
centred effort and time, and year in the temporal models. The index $i$ corresponds to the number of observations among grid-cells in the spatial models, and years in the temporal models. $\beta$ and $\gamma$ represent the parameters of the fixed and random effects, respectively, and $\alpha$ is the intercept. The log of fishing effort was included as an offset only in the spatial models.

We compared the models relative strength of evidence by weighting each model’s Akaike’s information criterion corrected for small sample sizes ($wAIC_c$) (Burnham & Anderson, 2004), and assessed the goodness-of-fit by calculating the percentage of deviance explained ($\%De$) for each model. We calculated the 10-fold cross-validation error for the model with highest $wAIC_c$ support and assessed the model’s predictive power with the $\kappa$ statistic (Cohen, 1960). To build the habitat suitability maps (as result of the first spatial modelling step), we weighted each model’s predictions according to its weight of evidence ($wAIC_c$), and used the result of the weighted mean prediction for all models as input to the temporal models. We also calculated the weight of evidence for each predictor used in the temporal models by summing the $wAIC_c$ weights over all models in which each predictor appeared. This allows estimation of the predictor with the highest relative importance to explain the response variable.

**Fourier Transforms**

Fourier transforms allow for the decomposition of signals (i.e., time series) into the sum of sinusoidal curves with different frequencies – a highly useful approach for ecological time series (Platt & Denman, 1975). To analyse possible cyclic variation in whale shark occurrences, we applied the fast Fourier transform function (FFT;
following Moler, 2004) to the sightings time series after correcting for effort bias (i.e., standardizing sightings per unit effort [SPUE], with an effort unit = 1 fishing day). We used MatLab version 7.12.0.635 (R2011a) (The Mathworks Inc., Natick, MA), and interpreted the results of the FFT by plotting the periodogram for the function: power against the inverse of frequency (time). A cycle is then defined by identifying the strongest frequency.

To examine cyclic trends apart from the known annual cycles (i.e., number of sightings is usually higher in particular seasons each year), we started by only using data for the season in study (Apr-Jun), and defining frequency as year. As an alternative test to include all sightings data available (i.e., all months), and therefore increase the length of the vector used in the function, we also ran FFT for the sightings per monthly effort (i.e., defining frequency as month). In this test, we used a running average including the six months before and after each month to eliminate the known annual periodicity. The strength of each resulting cycle (periodicity) is given by the power of the corresponding frequency in the periodogram.

Results

Spatial patterns in whale shark occurrence

In the Atlantic Ocean, whale shark sightings occurred mostly off Gabon in equatorial eastern Africa, and at around 10 °N between Senegal and Sierra Leone – these areas comprise most of the fishing effort (Figure 18). Likewise in the western Pacific, more sightings were recorded around the area where more days were spent fishing, although sightings were relatively more spread out through the sampled area (Figure 18).
Our spatial models (step 1) revealed higher habitat suitability mostly corresponding to the areas where more whale sharks were spotted in the Atlantic, and to the northeast of Indonesia/ Papua New Guinea in the western Pacific also covering the area with more sightings (Figure 18 and Figure 19). The model with highest statistical support was the same for both oceans in step 1 ($wAIC_c = 0.83$ and 0.33 in the Atlantic and Pacific, respectively), and included all predictors except chlorophyll $a$. This model explained 50 % of deviance in the Atlantic and only ~ 6 % in the Pacific (Table 8). During the austral autumn, habitat was more suitable in the Pacific (habitat suitability > 0.7) than in the Atlantic (~ 0.5) (Figure 19), however, $\kappa$ was more than double in the Atlantic ($\kappa \sim 0.7$: ‘good’ performance compared to $\kappa \sim 0.3$ in the Pacific: ‘poor’ performance) (Table 8). The mean prediction error calculated through the 10-fold cross validation ($CV_e$) was relatively low in both oceans ($\leq 0.1$).

Figure 19: Predicted habitat suitability of whale sharks (Rhincodon typus, Smith 1828) in the Atlantic (left) and Pacific (right) Oceans during the months of April to June.

Values for habitat suitability vary from 0 (low) to 0.51 and 0.72 (high) in the Atlantic and Pacific, respectively.
Table 8: Summary of spatial generalised linear models (step 1) relating probability of whale shark (Rhincodon typus, Smith 1828) occurrence to ocean properties.

Ocean properties: slope, depth, depth2 and distance to shore (shore) - referred to together as physical variables (physvar), mean sea surface temperature (mSST) and its quadratic term (SST2), sea surface temperature standard deviation - referred to together as SST variables (SSTvar), and mean chlorophyll a (mChl a) in the Atlantic and west Pacific Oceans during April to June. We shown the biased-corrected model evidence based on weights of Akaike’s information criterion corrected for small sample sizes (wAICc; only values ≥ 0.001 shown), the percentage of deviance explained (%De), and the Cohen’s Kappa statistics (κ; only values ≥ 0.1 shown) for each model performing better than the null model (including only an offset for effort and the spatial random effect) in each ocean.

<table>
<thead>
<tr>
<th>Model</th>
<th>Atlantic Ocean</th>
<th>Pacific Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wAICc</td>
<td>%De</td>
</tr>
<tr>
<td>physvar + SSTvar</td>
<td>0.83</td>
<td>50.4</td>
</tr>
<tr>
<td>physvar + mSST + SST²</td>
<td>0.17</td>
<td>50.3</td>
</tr>
<tr>
<td>mSST + SST² + mChl a</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>physvar + mChl a</td>
<td>-</td>
<td>49.5</td>
</tr>
<tr>
<td>mChl a</td>
<td>-</td>
<td>47.0</td>
</tr>
<tr>
<td>mSST + SST²</td>
<td>-</td>
<td>47.4</td>
</tr>
<tr>
<td>physvar</td>
<td>-</td>
<td>48.3</td>
</tr>
<tr>
<td>SSTvar</td>
<td>-</td>
<td>47.1</td>
</tr>
<tr>
<td>shore</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>depth</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>depth²</td>
<td>-</td>
<td>49.5</td>
</tr>
<tr>
<td>slope</td>
<td>-</td>
<td>47.3</td>
</tr>
</tbody>
</table>
Temporal trends in whale shark occurrence

The total number of annual whale sharks sighted in the Atlantic varied from 0 to 137 (33 ± 39; Figure 20a) between April and June. Sightings peaked (~ 120 individuals) in 1995, even though total fishing effort (sum of fishing days spent fishing in all 1° grid cells) remained relatively constant between 1985 and 2000 at about 6000 fishing days (5.8 ± 2.6×10³; Figure 20c). During this peak, the number of occurrences was similar to the one recorded in the early 1980s (137 sightings) when effort was close to 10000 days (Figure 20). Between 2000 and 2010, effort dropped to ~ 2000 days in the Atlantic. In contrast, fishing effort increased steadily in the Pacific from year 2000 (Figure 20c). The number of sightings in the latter ocean also increased with time, from around 10 sightings per year in the early 2000s to a maximum of only 35 in 2010 (Figure 20a).

The temporal model (step 2) with highest statistical support differed between oceans (Table 9). In the Atlantic, this model (wAICc = 0.633) included as predictors the habitat suitability derived from step 1 (Hsuit), effort, and year (as a random intercept). The models including the linear and quadratic terms for time (time and time²) resulted in similar log-likelihood (Table 9), and explained similar percentage of deviance, but received lower support from wAICc. In the Pacific, the highest-ranked model (wAICc = 0.472) also included the time predictor, although it only explained < 15 % of the deviance. The partial effect of time on the probability of whale shark occurrence during the months of April to June in the western Pacific is depicted in Figure 21b, and shows an increase from 0.003 to 0.012 between 2000 and 2010 (Figure 21b).
Figure 20: Yearly variation in (a) number of whale sharks sighted, (b) relative whale shark occurrences, and (c) days spent fishing.

During the months of April to June in the Atlantic (1º resolution), Indian (5º resolution) (from Sequeira et al., in press; included here for comparison only), and western Pacific (5º resolution) Oceans.
Table 9: Summary of the temporal generalised linear mixed-effects models (step 2).

Relating probability of whale shark occurrence in the Atlantic and Pacific Oceans (from April to June) to: a spatial predictor derived from the spatial distribution models ($Hsuit$), effort (temporal variation in fishing effort), time (years), and a random effect for year. Shown for each model are the number of parameters ($k$), log-likelihood (LL), biased-corrected model evidence based on weights of Akaike’s information criterion corrected for small sample sizes ($wAIC_c$) and the percentage of deviance explained (%De). Top-ranked models in each step are in bold and ordered by %De. Models with $wAIC_c < 0.05$ not shown.

<table>
<thead>
<tr>
<th>Model</th>
<th>Atlantic Ocean</th>
<th>Pacific Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>LL</td>
</tr>
<tr>
<td>$Hsuit + effort + time + time^2 + (1</td>
<td>year)$</td>
<td>7</td>
</tr>
<tr>
<td>$Hsuit + effort + time + (1</td>
<td>year)$</td>
<td>6</td>
</tr>
<tr>
<td>effort + time + time^2 + (1</td>
<td>year)$</td>
<td>6</td>
</tr>
<tr>
<td>effort + time + (1</td>
<td>year)$</td>
<td>5</td>
</tr>
<tr>
<td>$Hsuit + effort + (1</td>
<td>year)$</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 10: Estimated weight of evidence for each temporal predictor used in the generalised linear mixed-effects models: $Hsuit$ - spatial predictor derived from the spatial distribution models, effort - temporal variation in fishing effort and a linear and quadratic term for time (years).

<table>
<thead>
<tr>
<th>Ocean</th>
<th>weight of evidence</th>
<th>$Hsuit$</th>
<th>effort</th>
<th>time</th>
<th>time^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic</td>
<td>1</td>
<td>0.999</td>
<td>0.367</td>
<td>0.133</td>
<td></td>
</tr>
<tr>
<td>Pacific</td>
<td>0.761</td>
<td>1</td>
<td>0.953</td>
<td>0.328</td>
<td></td>
</tr>
</tbody>
</table>
Figure 21: Partial effect of time on the probability of whale shark presence during the months of April to June.
(a) the western Indian Ocean (from Sequeira et al., in press; included here for comparison only); (b) the western Pacific Ocean according to the model with highest information-theoretic support: Presence ~ Hsuit + effort + time + (1 | year) (Table 2). Dashed lines indicate 95% confidence intervals. Note: the increasing trend in the Pacific Ocean starts from 2000, being concurrent with the declining segment previously observed in the Indian Ocean.

The weight of evidence we estimated for each temporal predictor is shown in Table 10. Habitat suitability (Hsuit) and effort received the highest weight (1 and 0.999, respectively) in the Atlantic Oceans, while in the Pacific we found the highest weight for effort (weight of evidence = 1) followed by the linear term for time (time: 0.953).

Cyclic patterns in whale shark occurrence
The fast Fourier transform applied to the time series of sightings per unit effort in the Atlantic (Figure 22a, left) revealed a prominent peak at 15 years (Figure 22a, centre), both with yearly and monthly frequencies. This periodicity also received the highest power when plotting the possible cycles present in the dataset (Figure 22a, right); however, its relative magnitude was similar to other possible cycles present in the time series. In the Pacific, even though some cycles were identified (Fig 5b), their power was consistently low (< 1×10⁻³).
Figure 22: Whale shark sightings per unit effort during the months of April to June in the Atlantic (top row; 1981 - 2010) and Pacific (bottom row; 2000 - 2010).

Left - variation in sightings per unit effort with time; centre – strongest frequency observed; and right – cyclic description in whale shark sightings per unit effort as result of the Fourier transforms.
Discussion

The detrimental effects of a changing climate expected for some species (Walther et al., 2002) has stimulated a recent boom in efforts to examine temporal patterns in species occurrence (e.g., Gotelli et al., 2010). We have mapped habitat suitability for whale sharks and quantified temporal variation in their occurrence both in the Atlantic and Pacific Oceans for the first time. Our modelling results showed no linear (or quadratic) temporal trends in the Atlantic, although there was some evidence for a cyclic pattern, and a slight increase (linear trend) in the Pacific since 2000. The latter result contrasts with the decreasing trend detected in the Indian Ocean during the last decade (Sequeira et al., in press). Given these opposing results, the evidence for an inter-decadal occurrence pattern in the Indian Ocean (Sequeira et al., in press), the low genetic differentiation of whale sharks sampled in different oceans (Castro et al., 2007; Schmidt et al., 2009), and the implicit notion that whale sharks travel across oceans (Hueter et al., 2008; Rowat & Gore, 2007), we hypothesise that whale shark occurrence might be asynchronously cyclical across oceans. Such a pattern would be consistent with the inter-ocean migratory behaviour we have previously hypothesised (Sequeira et al., 2013).

Despite genetic clues (Castro et al., 2007; Schmidt et al., 2009) being the only evidence to date corroborating inter-ocean migration, these results could also have arisen even if only a small number of individuals were moving between ocean basins (Hartl & Clark, 1989). However, the number of whale sharks occurring in aggregation locations within the predictable season is highly variable among years (Rowat et al., 2009a; Wilson et al., 2001), and this variability holds for the number of whale sharks sighted at the ocean-basin scale (Figure 20a and b). The opposing trends we detected between the Indian and the western Pacific Oceans during the last decade might
therefore reflect a migration of sharks between oceans. Similar distributional shifts, promoting asynchrony in occurrences, have also been suggested for other filter-feeding sharks (basking sharks *Cetorhinus maximus*) between the west coast of Republic of Ireland and the Norwegian Sea (Sims & Reid, 2002), and were potentially associated with spatial changes in foraging conditions. If whale sharks are migrating between oceans (Sequeira *et al.*, 2013), a cyclic pattern such as the one observed in the Atlantic Ocean would explain both the interannual variation in whale sharks numbers at the ocean-basin scale (Figure 20a and b) and the asynchrony in their occurrences in the Indian and Pacific Oceans. Moreover, such large-group and multi-year migrations would still be consistent with the low differentiation found in genetic studies (Castro *et al.*, 2007; Schmidt *et al.*, 2009). These whale shark migrations could be associated with changes in environmental conditions (Sleeman *et al.*, 2010b; Wilson *et al.*, 2001), or with sex/age or reproduction-related behaviour (Ramírez-Macías *et al.*, 2007; Sequeira *et al.*, 2013). Reproduction-associated multi-year cycles have been observed in anadromous fish (e.g., sockeye salmon *Oncorhynchus nerka*) (Dingle, 1996). Multi-year cycles have also been reported for fish catches in the western Indian Ocean (Jury *et al.*, 2010), and for other marine megafauna, such as the decadal cycle in cetacean strandings in southeast Australia (Evans *et al.*, 2005).

We found a stronger cycle with a rhythmicity of 15 years in the Atlantic Ocean using the fast Fourier transform, which corresponds to half of the full dataset’s temporal span, and is therefore close to the highest frequency that can be detected by this method (Nyquist limit) (Moler, 2004). Analogous studies reporting decadal cycles in marine species have used much longer datasets (~ 40 years; Evans *et al.*, 2005), and so obtaining longer time series for whale sharks in the Atlantic will be required to reveal whether this cycle persists. The western Pacific dataset was only a third the
span of the Atlantic dataset, and so we did not expect to reveal meaningful cycles. Nonetheless, we detected possible cycles for the Pacific at 5.5 and 2.2 years (Figure 22), but with only weak differences in power among them (< 10^{-3}). Our hypothesis of cyclic occurrence among ocean basins could mean that the light increase in relative abundance observed in the Pacific Ocean over the last 10 years might correspond to only a segment of an inter-decadal cycle (> 15 years), but this would be undetectable with such short dataset.

The results from the spatial models (step 1) agreed well with the data in the Atlantic (κ ~ 0.7) where higher-resolution data (1° grid cell) were available. In this ocean, we predicted higher suitable habitat mostly close to shore around Gabon, Congo and Equatorial Guinea, and between Coté d’Ivoire and Mauritania. Due to the paucity of studies on whale sharks in the Atlantic (excluding the Gulf of Mexico and Caribbean Sea which were not covered in our dataset), we found no independent data to validate these results. Higher suitability was also predicted further from shore at around the Equator and between 15–20 °W (Figure 19, left panel). Whale sharks occur close to this area in the Saint Peter Saint Paul archipelago, peaking in occurrence at the end of June (Hazin et al., 2008).

In the western Pacific, the resulting habitat maps show mostly suitable habitat within the area covered by the fisheries, but despite the higher suitability (up to 0.72), model accuracy was poor (κ ~ 0.3). Possible reasons for the poor performance include poorer resolution of the data or unmeasured environmental covariates that could be better predictors, such as current or wind conditions (Wilson et al., 2001). During the season we considered (April to June), the number of sightings in the western Pacific (167) was about 6 times lower than in the Atlantic (1030; Figure 18), which affects the performance of the western Pacific model. Because the data available for this ocean
represent only partial coverage, access to the full dataset (giving more sightings in this season), would likely improve model performance by refining the distinction between presence and absence locations. Nevertheless, the available data are representative of the total fisheries in the area (P. Williams pers. comm.), and so application of the same models to other seasons when more sightings are recorded might improve results. However to compare synchronous seasonal occurrences in the three major oceans, data covering the months of April to June were required. Due to the lack of independent studies reporting whale shark occurrence off Indonesia / Papua New Guinea, we could not externally validate this model's predictions.

We found a lack of support for the linear or quadratic time predictor in the Atlantic Ocean from April to June (Table 9) where habitat suitability (Hsuit) and effort received the highest weight of evidence (Table 10). At least for the last decade in the western Pacific, there was evidence for a linear, positive trend in the probability of whale shark occurrence (Table 9; Figure 21). In the latter ocean, even though habitat suitability was also ranked as an important variable (weight of evidence = 0.761), we found the highest weight of evidence for the predictors effort and time (Table 10). Despite the modest trend we found (from 0.003 to 0.012), the difference between maximum and minimum values (~0.009) is similar to the prevalence assumed in our temporal models (0.01, corresponding to 100 absences to each presence). Although the Pacific positive trend is concurrent with the increase observed in fishing days, we accounted for both spatial and temporal effort bias. Because the habitat suitability outputs from the spatial model were inputs for the temporal model, the increase in the whale shark probability of occurrence in the Pacific should be interpreted with caution. However, if this increase is real, the trend is opposite to that identified in the Indian Ocean over the same interval (Sequeira et al., in press) (Figure 21), and could indicate
a quasi-decadal shift in sharks between oceans as referred above. A slight lag in occurrences between the Indian and Atlantic oceans is also evident in some years.

Even if whale shark occurrence is cyclic, which could explain some of the declines observed in some years / locations, the total number of sightings reported in the fisheries datasets (ocean-scale) is also lower in the last decade (< 50 per year in all oceans; Figure 20a). In the 1990s, there were a total of about 500 sightings in the Atlantic Ocean, and in the early 2000s, only around 150 (both in the Atlantic and west Pacific) during the study season. We observed a similar pattern in the Indian Ocean (Sequeira et al., in press), with 600 sightings reported in the 1990s and ~ 200 in following years (Figure 20a). The reduced sightings in the last decade agree with the recent accounts of declines from near-shore aggregations (Bradshaw et al., 2008).

The datasets we used were opportunistically collected and therefore lead to inherent complications. Differential sampling effort is a major issue for which we accounted at least partially in both modelling steps to reduce potential bias in spatial and temporal effort. However, an underlying assumption of our models is that failure to report a whale shark presence (whether due to a failure in detecting a shark or in reporting a detection) is evenly distributed across the sampling area and study period. Access to higher-resolution data would likely improve the results (exemplified by the Atlantic Ocean models), and can be achieved through establishing collaboration between researchers and fisheries-management organisations. Commercial-in-confidence restrictions on access to fisheries data (especially sensitive data such as fish catch) delay scientific research from providing more accurate models on fisheries impacts. Such collaborations would also allow researchers to give greater insight on how better datasets could be collated for specific objectives using the resources currently in place. However, while data collected by fisheries might not be as precise
as those derived from scientifically designed surveys, they are still essential sources of
information given the logistic challenges of surveying widely distributed species over
their entire range. Fisheries data therefore present unique opportunities to advance
our understanding of such species.

Because the datasets we used contained only recorded presences, the
generation of pseudo-absences was necessary to allow for the binomial estimation in
the models. Differences associated with the method to generate pseudo-absences
was assessed in a previous study using similar datasets (Sequeira et al., 2012) where
we found that model performance was not reduced by using random pseudo-absences
selection, and was therefore the method used in this analysis. This accords with the
results from another recent publication specifically addressing pseudo-absence
selection issues (Barbet-Massin et al., 2012). The number of pseudo-absences
selected is another issue to consider when generated. Barbet-Massin et al. (2012)
reported that model accuracy increases until the presence to pseudo-absence ratio
(i.e., prevalence) equals 0.1, and remains constant for lower ratios. In the models we
developed, prevalence was 0.1 in the spatial assessment, and 0.01 in the temporal
assessment (to allow enough points for the spatially explicit temporal analysis);
therefore in both cases, we do not expect the ratio presences:pseudo-absences to
have affected model accuracy.

Another important aspect of the input data we used is that both sightings and
environmental variables correspond mostly to the ocean surface layer. Three-
dimensional environmental data at an adequate resolution for regional analyses such
as that presented here are currently not available. Even though whale sharks spend
most time at the surface (e.g., Rowat et al., 2007) they also dive frequently, and
therefore, assessment of how they explore the vertical habitat and estimation of its
suitability at a broad scale is also of importance. However, this will only be possible when new technologies are developed to allow vertical data collection at broad spatial scales.

Despite the limitations associated with both the opportunistic dataset and species distribution models, we have demonstrated a powerful way to use such datasets to identify both spatial and temporal trends of highly migratory species. We also revealed that longer-term and higher-resolution sightings datasets are still required to tease apart potentially confounding aspects of inter-ocean migration in whale sharks. Continued investigation of these connections is necessary to assess temporal trends in occurrence, and to reveal the suspected impacts of human modifications to the wider marine realm.
CHAPTER VI. Predicting current and future global distributions of
whale sharks

A. Sequeira, C. Mellin, D. Fordham, M. Meekan, C. Bradshaw

\(^a\)The Environment Institute and School of Earth and Environmental Sciences, The
University of Adelaide, South Australia 5005, Australia

\(^b\)Australian Institute of Marine Science, PMB No. 3, Townsville MC, Townsville,
Queensland 4810, Australia

\(^c\)Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling
Hwy, Crawley, Western Australia 6009, Australia.

\(^d\)South Australian Research and Development Institute P.O. Box 120, Henley Beach,
South Australia 5022, Australia

(Manuscript in preparation)
PREDICTING CURRENT AND FUTURE GLOBAL DISTRIBUTION OF WHALE
SHARKS

Text in manuscript

SEQUEIRA, A.M.M. (Candidate)
Planned the article, performed the analysis, interpreted data, developed the models,
wrote manuscript and acted as corresponding author.
I hereby certify that the statement of contribution is accurate.

Signed……………………………………………………………………….Date……………

MELLIN, C.
Helped with data interpretation, models development and assisted writing.
I hereby certify that the statement of contribution is accurate and I give permission for
the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………

FORDHAM, D.
Assisted with climate change models development and incorporation of results.
I hereby certify that the statement of contribution is accurate and I give permission for
the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………
MEEKAN, M. G.

Supervised development of work and provided evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for
the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………

BRADSHAW, C.J.A.

Supervised development of work, helped in model results interpretation and
manuscript evaluation.
I hereby certify that the statement of contribution is accurate and I give permission for
the inclusion of the manuscript in the thesis.

Signed……………………………………………………………………….Date……………
Predicting current and future global distributions of whale sharks

Abstract

The highly migratory whale shark, spanning the warm and temperate waters around the globe, is classed as Vulnerable by the IUCN. However, no one has yet predicted their present-day or possible future global distribution. We used 30 years (1980-2010) of whale shark observations recorded by tuna purse-seiners fishing in the Atlantic, Indian and Pacific Oceans, and applied generalised linear mixed-effects models to test the hypothesis that similar environmental covariates predict whale shark occurrence in all major ocean basins. We derived global data for potential predictors from satellite images for chlorophyll a (an index of primary productivity) and sea surface temperature, and bathymetric charts for depth, bottom slope and distance to shore. We randomly generated a set of pseudo-absences within the total area covered by the fisheries, and included fishing effort as an offset to account for potential sampling bias. We also predicted sea surface temperatures for 2070 using an ensemble of 5 global climate models under a no-climate-policy reference scenario, and used these temperatures to predict changes in whale shark distribution. The full model for all oceans (excluding standard deviation of sea surface temperature) had the highest relative statistical support (\(w_{AIC_c} = 0.99\)) and explained ~ 60 % of the deviance in the presence-absence data. Whale shark habitat suitability was mainly driven by spatial variation in bathymetry and sea surface temperature among oceans, although we found evidence for the effects of these variables to differ slightly among oceans. Changes in temperature to 2070 resulted in a slight shift of suitable habitat towards the poles (~ 5 °N in the Atlantic Ocean and 3-8 °S in the Indian Ocean) accompanied by an overall range contraction (~ 6%). Assuming that whale shark environmental requirements, and human disturbances (i.e., in a scenario of no-stabilization of
greenhouse gas emissions) remain similar in the near future, we show that changes derived from warming oceans might promote a net retreat from current aggregation areas and an overall redistribution of the species.

KEYWORDS *Rhincodon typus*; Indian Ocean; Atlantic Ocean; Pacific Ocean; remote sensing; sea surface temperature; tuna purse-seine fisheries; species distribution models; global warming; climate change; conservation

**Introduction**

Changes in climate are expected to alter the current distribution of species (Parmesan, 2006; Thomas *et al.*, 2004; Wernberg *et al.*, 2011). Despite a bias of studies towards terrestrial habitats (Richardson & Poloczanska, 2008), there is mounting evidence for distributional shifts in the marine environment (*e.g.*, Edwards & Richardson, 2004; Hiddink & ter Hofstede, 2008), which generally result in pole-ward latitudinal (*e.g.*, Perry *et al.*, 2005) or depth (*e.g.*, Dulvy *et al.*, 2008) shifts. These shifts have been reported in species as divergent as plankton (Beaugrand *et al.*, 2009; Southward *et al.*, 1995), mussels (Berge *et al.*, 2005), capelin (Rose, 2005), demersal fish (Beaugrand & Kirby, 2010); (Dulvy *et al.*, 2008; Perry *et al.*, 2005) and macroalgae (Wernberg *et al.*, 2011). However, assessing how much the distribution of species will be affected by climate change implies that the current distribution of the species is adequately identified.

A species’ distribution and its habitat requirements can be estimated by modelling information on occurrence together with environmental correlates (Kearney & Porter, 2009). Despite the diverse use of species distribution models in terrestrial
environments, until recently only a handful of these models had been developed for the marine environment (Elith & Leathwick, 2009; Robinson et al., 2011), but now they have been applied to many marine species such as corals (e.g., Beger & Possingham, 2008; Garza-Pérez et al., 2004), molluscs (Mellin et al., 2011), reef fish (Mellin et al., 2010), algae (e.g., Kelly et al., 2001; Sérgio et al., 2007), mammals (e.g., Oviedo & Solís, 2008; Praca & Gannier, 2007) and sharks (McKinney et al., 2012; Sequeira et al., in press; Sequeira et al., 2012). However, applying such models to highly migratory species is particularly challenging (e.g., Kreakie et al., 2012), and the problem is exacerbated in the marine environment mostly due to poor detection (Elith & Leathwick, 2009), lack of recorded occurrences in the open ocean (McKinney et al., 2012), and the large geographical range of the species under consideration (Kaschner et al., 2006; Sequeira et al., in press; Sequeira et al., 2012).

The whale shark (Rhincodon typus Smith 1828), the largest of all fish (> 18 m; Borrell et al., 2011; Chen et al., 1997; Compagno, 2001), has a circumglobal geographical range between 30 °N and 35 °S (Compagno, 2001; Last & Stevens, 2009). This range has been defined based on occasional occurrences (Compagno, 2001), but most importantly, the range of temperatures where the species is expected to occur: tropical/ warm temperate (Compagno, 2001; Last & Stevens, 2009). Being ectotherms, ambient temperatures directly influence their metabolic processes (Sims, 2003). Whale sharks occur regularly at the ocean surface (Gunn et al., 1999; Rowat et al., 2007; Thums et al., 2012) in association with sea surface temperatures (Sequeira et al., 2012) possibly to assist thermoregulation (Thums et al., 2012). They can also dive deeply, opening speculation on this species’ potential to modify its diving behaviour with increasing surface water temperatures associated with climate change (Hoegh-Guldberg & Bruno, 2010), and use deeper, cooler waters more frequently.
However, oxygen limitation might pose a problem for extended time spent in deeper waters (Graham et al., 2006; Pörtner, 2001; Prince & Goodyear, 2006) so for this reason, expected warming will likely affect the spatial distribution of whale sharks possibly by promoting shifts in their present geographical range. Moreover, because whale sharks are filter-feeders, these shifts might follow the simultaneous redistribution of plankton (Beaugrand et al., 2009).

Whale sharks have been regularly recorded in locations near shore (Sequeira et al., 2013) to where they attract substantial tourism interest due to their innocuous behaviour and large size (e.g., Cárdenas-Torres et al., 2007; Rowat & Engelhardt, 2007). Concerns about population declines, possibly driven by illegal direct catches elsewhere (Riley et al., 2009) or habitat disturbance, such as derived from tourism activities (Heyman et al., 2010), have prompted a research focus on documenting migratory pathways to determine whether partial-range protection is sufficient to ensure persistence locations (Bradshaw et al., 2008; Rowat, 2007; Sequeira et al., 2013). In addition to these direct threats, climate change might also already be affecting whale sharks. However, the species’ high mobility and the lack of extensive occurrence data mean that teasing apart spatial and temporal patterns is problematic (Sequeira et al., in press). Thus, determining whether departures from current distributional limits (Duffy, 2002; Rodrigues et al., 2012; Sa, 2008; Turnbull & Randell, 2006) or temporal patterns in relative abundance (e.g., Bradshaw et al., 2008; Theberge & Dearden, 2006) are linked to changing temperatures (e.g., Sleeman et al., 2007) is equally difficult.

Fortunately, statistical tools exist to facilitate inference in this regard. For example, known occurrences can be used to define the current distribution of a species, and future distributions can be projected by coupling climate forecasts
derived from global circulation models (e.g., Araújo et al., 2005; Cheung et al., 2010). Despite the known limitations of these models, such as biases derived from the input data or from the geographical extent available (Barbet-Massin et al., 2012; Barbet-Massin et al., 2010; Phillips et al., 2009), and uncertainties derived from the models used (both species distribution models and global circulation models) (Buisson et al., 2010), confidence in their application can be improved by using an ensemble result (Araújo & New, 2006). This ensemble result takes into account variability among models used (Hijmans & Graham, 2006; Kearney et al., 2010; Knutti et al., 2010; Lawler et al., 2006). Another important limitation, especially from a conservation perspective, is the coarse resolution of predictions derived from global circulation models, and their consequently limited biological relevance (Hannah et al., 2002; Seo et al., 2009). However, this can be improved by down-scaling the forecasts, coupling them to ecological processes (Araújo et al., 2005; Fordham et al., 2011; Fordham et al., 2012; Hannah et al., 2002).

For the first time, we used a global dataset of pelagic whale shark sightings derived from the tuna purse-seine fishery operating in the Atlantic, Indian and Pacific Oceans to: (1) predict the spatial distribution of whale shark habitat suitability across three ocean basins, and (2) extend our projections to a future scenario of ocean warming based on a no-climate-policy reference (no stabilisation of greenhouse gas emissions). Following our previous work where we derived ocean-basin scale predictions for whale shark distribution and temporal trends (Sequeira et al., 2012; Sequeira et al., 2013), herein our overarching aim was to derive realistic predictions of the species current geographical range (circumglobal) and provide a modelling tool for exploring the species’ future distribution under anticipated climate change.
Methods

Data

We obtained whale shark presence data (henceforth, ‘sightings’) from the logbooks of tuna purse-seine fisheries, where whale sharks associated with net sets (Matsunaga et al., 2003) were recorded. Data were provided by the Institut de Recherche pour le Développement (France) for the Atlantic and Indian Oceans, and by the Secretariat of the Pacific Community for the western Pacific Ocean (Table 11 and Figure 23).

The full dataset spans the three major oceans with a maximum temporal coverage of 30 years for the Atlantic Ocean (1980 – 2010) from 21 °N to 15 °S and 34 °W to 14 °E, 17 years (1991 – 2007) in the Indian Ocean from 30 °N to 30 °S and 35 to 100 °E and 11 years (2000 – 2010) in the western Pacific Ocean from 15 °N to 15 °S and 130 °E to 150 °W (Table 11 and Figure 23). Data included (i) the number of whale shark sightings (Table 11 and Figure 23a) and location (longitude and latitude at 0.01° precision) and (ii) fishing effort (days; Figure 23b) per month and per year pooled for each 5-° grid cell both in the Indian (6618 records) and western Pacific (2272 records) Oceans, and for both 5-° and 1-° grid cells in the Atlantic Ocean (maximum of 18277 records). No information was made available on vessel or trip units for any ocean. The data from the western Pacific Ocean, even though representative of total tuna purse-seine fisheries in the area (P. Williams pers. comm.), were incomplete as data access was not granted from all fleets registered with the Secretariat of the Pacific Community.

Because whale shark numbers fluctuate seasonally (Table 11), but not necessarily in a spatially consistent manner (Sequeira et al., in press), we used the season in which the highest number of whale sharks was recorded for each ocean. A seasonal maximum of 314 whale sharks was recorded in the western Pacific between
January and March, 811 in the Indian Ocean between April and June, and 1153 between July and September in the Atlantic Ocean.

Binary data (presence and absence) are required for binomial estimation in spatial generalised linear mixed-effect models (GLMM); therefore, we generated pseudo-absences (10:1 presence ratio, following Barbet-Massin et al., 2012) based on a random spatial distribution within the area covered by the fisheries (excluding presence locations). We used the `srswor` function (simple random sampling without replacement) from the `{Sampling}` package in the `R` programming language (R Development Core Team, 2012), and generated random pseudo-absences among non-presence cells in each ocean within the same season for which the presences were being considered. To assess the influence of the chosen pseudo-absences, we generated each set 10 times prior to their inclusion in the spatial GLMM (described below).

We collated an environmental dataset at a 9-km resolution composed of six physical variables (Phys\text{var}): (1) distance to shore (\text{shore}; \text{km}) calculated with the `Near` tool in ArcGIS 9.3.1\textsuperscript{TM} using a world equidistant cylindrical coordinate system, (2) mean depth (\text{depth}; \text{m}) and (3) slope (\text{slope}; \text{o}) derived from the one-minute grid of the General Bathymetry Chart of the Oceans (GEBCO, 2003), (4) mean and (5) standard deviation of sea surface temperature (in °C) (\text{SST}, \text{SSTsd} or together \text{SSTvar}) derived from daytime measures from the Advanced Very High Resolution Radiometer (AVHRR) PathFinder version 5.0 and (6) mean concentration of chlorophyll a (Chl a in mg m\textsuperscript{-3}) derived from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS). We calculated averages and standard deviations of chlorophyll a and sea surface temperature based on weekly satellite measures available within the total period covered by the fisheries in each ocean using `ArcToolBox` functions from ArcGIS.
9.3.1™ automated with Python scripts. The environmental dataset for each ocean matched the ocean-specific season of maximum whale shark abundance.
Table 11: Number of whale shark sightings in each ocean per decade as recorded by tuna purse-seine fisheries.

Source: Indian Ocean Tuna Commission and the Institut de Recherche pour le Développement, France for the Atlantic [Atl] and Indian [Ind], and Secretariat of the Pacific Community for the western Pacific [Pac]. See also Figure 23. Numbers of sightings are split by trimester: Jan – Mar, Apr – Jun, Jul – Sep and Oct – Dec. Boldface indicates data used to fit models. Numbers for the western Pacific are derived from observer data only.

<table>
<thead>
<tr>
<th>Time period</th>
<th>Jan - Mar</th>
<th>Apr – Jun</th>
<th>Jul - Sep</th>
<th>Oct - Dec</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atl</td>
<td>Ind</td>
<td>Pac</td>
<td>Atl</td>
<td>Ind</td>
</tr>
<tr>
<td>1980-1990</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>405</td>
<td>-</td>
</tr>
<tr>
<td>1991-2000</td>
<td>2</td>
<td>92</td>
<td>-</td>
<td>518</td>
<td>599</td>
</tr>
<tr>
<td>2000-2010</td>
<td>3</td>
<td>23*</td>
<td>314</td>
<td>147</td>
<td>212*</td>
</tr>
<tr>
<td>Totals</td>
<td>7</td>
<td>115</td>
<td>314</td>
<td>1070</td>
<td>811</td>
</tr>
</tbody>
</table>

* Indian Ocean data for the last decade only for the period 2000 – 2007.
Figure 23: Fisheries data derived from tuna purse-seine fishery logbooks in the Atlantic, Indian and western Pacific Oceans.

(a) Sightings from 1980 to 2010: 4336 whale sharks records, with 53% of sightings in the Atlantic, 27% in the Indian from 1991 to 2007, and 20% in the western Pacific from 2000 to 2010. Shaded area shows whale shark geographical range (WS range) defined by the International Union for Conservation of Nature, and thick, grey lines mark current latitudinal range (Compagno, 2001).

(b) Tuna purse-seine fishing effort for the Atlantic, Indian and western Pacific Oceans (the common resolution of 5° was used for easier comparison). Data from the eastern Pacific (within inset) consisted of locations without information on year or season of sighting, but with no associated temporal information (NA in Map b) and were thus not included in the analysis.
**Generalised linear mixed-effects models**

We applied generalised linear mixed-effects models with a binomial error distribution and a logit link function to the full, three-ocean dataset. To account for spatial bias in sampling effort without reducing the spatial resolution of the dataset, we included this variable as an offset, and used the highest resolution available in each ocean. The mixed-effects models included several possible combinations of the fixed (environmental predictors – see Table 12) and a spatial random effect (1° grid cells) to remove spatial autocorrelation (Sequeira et al., 2012). We also included the interaction terms ocean x depth, ocean x slope and ocean x temperature, to test the hypothesis that these predictors had similar predictive power in every ocean. To account for non-linear dependencies of occurrence with both bathymetric features and sea surface temperature, we included depth and mean sea surface temperature as a second-order polynomial function using `poly` in the `{stats}` package in R.

We used Akaike’s information criterion corrected for small sample sizes (AICc) (Burnham & Anderson, 2004) and their weights (wAICc) to compare model probabilities. We used percentage of deviance explained (\%De) to quantify goodness-of-fit, and Cohen’s Kappa statistics (κ) to assess the model’s predictive power (Cohen, 1960). We calculated mean prediction error for the model with highest support using a 10-fold cross-validation (Davison & Hinkley, 1997).
Table 12: Summary of generalised linear mixed-effects models relating probability of whale shark occurrence to ocean properties in the Atlantic, Indian and western Pacific during the season of peak occurrences for each ocean.

Slope, distance to shore (shore), depth and its quadratic term referred to together as physical variables (Physvar); mean sea surface temperature (SST), its quadratic term (SST2) and standard deviation referred to together as SST variables (SSTvar). Average chlorophyll a is represented as Chl a. Shown for each model are biased-corrected model probabilities based on weights of Akaike's information criterion corrected for small sample sizes (wAICc, only > 0.0001 shown), percentage of deviance explained (%De), 10-fold cross validation error (CVerror) and kappa statistics. Note: All models included an offset term for effort and a spatial random effect (1-º grid cell). An interaction term between ocean and depth, slope or SST was also included whenever these predictors were present.

The GLMM (with a logit link function) can be expressed as: \( \text{logit}(\text{Presence}) = \alpha + \beta X_i + \gamma_i Z_i + \epsilon_i + \log(\text{fishing effort}) \), where Presence is the expected mean probability of sightings occurrence. \( X \) and \( Z \) represent the fixed and random covariates used in the models: the environmental predictors and the spatial grid, respectively. \( \beta \) and \( \gamma \) represent the coefficients associated with the fixed and random effects, respectively, and \( \alpha \) is the intercept. The log of fishing effort was included as an offset. The index \( i \) corresponds to the number of observations among grid-cells.

<table>
<thead>
<tr>
<th>Model</th>
<th>wAICc</th>
<th>%De</th>
<th>CVerror</th>
<th>kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physvar + SST + SST²</td>
<td>0.99</td>
<td>57.9</td>
<td>0.09</td>
<td>0.64 ± 0.03</td>
</tr>
<tr>
<td>Physvar + SSTvar</td>
<td>-</td>
<td>57.8</td>
<td>0.09</td>
<td>0.64 ± 0.03</td>
</tr>
<tr>
<td>Physvar + Chl a</td>
<td>-</td>
<td>57.3</td>
<td>0.09</td>
<td>0.64 ± 0.03</td>
</tr>
<tr>
<td>Physvar</td>
<td>-</td>
<td>56.9</td>
<td>0.09</td>
<td>0.59 ± 0.03</td>
</tr>
<tr>
<td>depth + depth²</td>
<td>-</td>
<td>56.8</td>
<td>0.12</td>
<td>0.58 ± 0.03</td>
</tr>
<tr>
<td>SST + SST²</td>
<td>-</td>
<td>56.5</td>
<td>0.09</td>
<td>0.57 ± 0.03</td>
</tr>
<tr>
<td>SST + SST² + Chl a</td>
<td>-</td>
<td>56.5</td>
<td>0.09</td>
<td>0.55 ± 0.04</td>
</tr>
<tr>
<td>SSTvar</td>
<td>-</td>
<td>56.5</td>
<td>0.09</td>
<td>0.54 ± 0.04</td>
</tr>
<tr>
<td>depth</td>
<td>-</td>
<td>55.9</td>
<td>0.09</td>
<td>0.22 ± 0.04</td>
</tr>
<tr>
<td>slope</td>
<td>-</td>
<td>55.9</td>
<td>0.09</td>
<td>0.30 ± 0.04</td>
</tr>
<tr>
<td>Chl a</td>
<td>-</td>
<td>50.8</td>
<td>0.11</td>
<td>0.33 ± 0.03</td>
</tr>
<tr>
<td>Shore</td>
<td>-</td>
<td>48.6</td>
<td>0.09</td>
<td>0.07 ± 0.03</td>
</tr>
</tbody>
</table>

To account for variability in habitat suitability resulting from different modelling approaches, we have also run some additional analysis using the BIOMOD modelling platform (Thuiller, 2003). However, being a model-comparison framework (rather than hypothesis-based) BIOMOD does not allow for direct input on the contribution of each variable. Details are presented in supplementary information (Appendix D).

Global projection
We compiled a worldwide environmental dataset as described above and including the same predictors. Here, we used seasonal averages for sea surface temperature and chlorophyll a for the last decade (2000 – 2012) by using data derived from the Moderate Imaging Spectroradiometer (MODIS) Aqua available online (seasonal climatology maps; http://oceancolor.gsfc.nasa.gov/cgi/l3). After fitting the GLMM with the tuna purse-seine data and running predictions for the entire area covered by the fisheries, we used a model-averaging approach (based on wAICc) to project habitat suitability by including the worldwide dataset under current (2000 – 2012) conditions. As a qualitative validation, we overlaid the currently known locations for whale shark seasonal occurrence (Sequeira et al., 2013) onto these global habitat suitability maps.

Climate change projections
Using the free software package MAGICC/SCENGEN5.3 (a model for the assessment of greenhouse-gas induced climate change and a regional climate scenario generator; for details refer to www.cgd.ucar.edu/cas/wigley/magicc/), we generated monthly forecasts of expected change in sea surface temperature in 2070 (relative to 1995) under a no-climate-policy reference scenario (no stabilization of greenhouse gas
emissions) (cf. MiniCAM Ref. in Clarke et al., 2007). We made forecasts using five atmospheric-ocean general circulation models chosen based on their skill (Fordham et al., 2011) in predicting global sea surface temperatures for the baseline period 1981 – 2000. The chosen models were CCSM-3, MRI-CGCM2.3.2, ECHAM5/MPI-OM, MIROC3.2 (hi-res) and UKMO-HadCM3 (terminology follows CMIP3 Multi-Model Dataset Archive naming convention; www-pcmdi.llnl.gov/ipcc/about_ipcc.php). We used bilinear interpolation of the GCM data (2.5° grid cells) to a finer resolution of 0.5° longitude/latitude to reduce discontinuities in the perturbed climate at the GCM grid-box boundaries (Fordham et al., 2011). We calculated an average change expected per grid cell across months for each quarter of the year (Jan – Mar, Apr – Jun, Jul – Sep and Oct – Dec). We then downscaled the predicted climate anomalies (multi-model average) using the ‘change factor’ empirical method, where the low-resolution climate anomaly is added directly to a high resolution (9-km grid cells) observed sea surface temperature baseline. In our case, this high resolution baseline derived from the MODIS Aqua satellite composites for 1995 (the mid-year in our time series from 1980 – 2010).

We used these forecasts of sea surface temperature to generate predictions of worldwide whale shark habitat suitability using our species distribution models. We considered the range of predictor values used to fit the model, and masked areas where model predictions occurred outside the environmental space used for the original model fit.
Results

Spatial patterns of occurrence and sampling effort

Whale shark sightings in the Atlantic and Pacific Oceans occurred mostly in the areas where sampling effort was highest (Figure 23), whereas most records in the Indian Ocean fell within the Mozambique Channel, which had lower sampling effort than in the central-west area (Figure 23b). There was generally higher effort in the Pacific Ocean (Figure 23b; > 40 000 fishing days in some 5-º cells), even though the Pacific time series covered only the last decade (i.e., only one third of the duration of the Atlantic Ocean dataset; Table 11). Within the area covered by the fisheries, mean depth in whale shark sightings locations was similar in all three oceans (~ 3800 m; Figure 24), but the Pacific bathymetric range was higher and average distances to shore were slightly shorter (~ 265 km), and in the Indian Ocean average distance to shore was greater (~ 435 km). Sea surface temperature varied among the three oceans (Figure 24) with cooler temperatures in the Atlantic ranging from 19.4 to 28.3 ºC (± 0.31 – 7.2; Jul – Sep), between 25.2 and 31.2 ºC (± 0.39 – 2.57, Apr - Jun) in the Indian, and between 26.7 and 30.1 ºC (± 0.19 - 1.36; Jan - Mar) in the western Pacific, with the latter having the lowest standard deviation. Chlorophyll a values averaged 0.75, 0.18 and 0.11 mg m⁻³ in the Atlantic, Indian and Pacific, respectively.
Habitat suitability prediction

The percentage of deviance explained was highest (58 %) for the model including all physical and temperature variables (except standard deviation for sea surface temperature), a spatial random effect (1-o grid cell) and an offset for effort (Table 12). Statistical support (wAICc) was highest for the same model (0.99; Table 12). As a result of the interaction terms (sea surface temperature, slope and depth interactions with ocean) within the model with the highest statistical support, we found evidence for a different effect of temperature and depth in each ocean, with both predictors...
affecting whale shark occurrence in the Atlantic, only temperature in the Indian, and only depth in the Pacific Ocean.

The resulting prediction maps (Figure 25) show suitable habitat in different areas in the Atlantic from July to September, with higher suitability (> 0.8) around Gabon, Congo and Equatorial Guinea. The model also predicted high suitability (> 0.6) around Côte d’Ivoire, Ghana and Bénin, Cape Verde and Mauritania. In the Indian Ocean, the area around the Mozambique Channel and close to shore in the south-eastern side of the African continent had the highest suitability (~ 0.3) during the months of April to June. In the western Pacific, the central portion of the total area covered by the fisheries had the highest suitability for the first trimester of the year, but predicted suitability there was the lowest of all three oceans examined (~ 0.1). Similar areas with higher habitat suitability were obtained with BIOMOD (Figure D2, Spearman $\rho \sim 0.6$; logit transformed predictions).

Figure 25: Predicted whale shark habitat suitability within the area covered by tuna purse-seine fisheries.
Representation only for the peak whale shark occurrence seasons: Jul – Sep in the Atlantic, Apr – Jun in the Indian, and Jan – Mar in the western Pacific.
**World projection of whale shark habitat suitability**

Only the area between ~ 40 °N and ~ 35 °S (Figure 26) had environmental variables within the range used to fit the GLMM. Exceptions occur in some seasons for areas around the west coast of South America and southwest Africa (Figure 26). In the Atlantic Ocean, highest habitat suitability was patchily distributed in the north (e.g., Gulf of Mexico, Mediterranean Sea, North West coast of Africa) and south (between South America and Africa, but not farther south than 35 °S). In the Indian Ocean, areas with highest habitat suitability were mostly near shore; for example: in the Bay of Bengal, Arabian Sea, Mozambique Channel and off Ningaloo Reef (Western Australia). In the western Pacific, highest habitat suitability (albeit lower than in the other oceans) occurred throughout the year (mostly in the central area covered by the tuna purse-seine fisheries).

**Projections of future whale shark distribution**

The average anomaly forecast for each season in 2070 resulted in an increase of around 0.8 to 2 °C (Figure D1 in Appendix D), mostly within the region between 30 °N and 35 °S (commonly considered the whale shark’s current latitudinal range). Forecasting whale shark suitability based on the 2070 sea surface temperature scenario resulted in a weak but evident pole-ward shift of habitat (Figure 27 and compare with Figure 26), mostly in the Atlantic and Indian Oceans. We also observed a contraction in habitat suitability in 2070 resulting in a loss of 5.0 % of suitable area (in 9-km cells, as per model resolution) in Jan-Mar, 3.3 % in Apr-Jun, 4.7 % in Jul-Sep and 6.1 % in Oct-Dec (Figure 27). In the Pacific, the projected increase in temperature resulted in similar habitat suitability (still < 0.1) within roughly the same region, although a ‘corridor’ of suitable habitat linking the western to the eastern Pacific Ocean was strengthened for all seasons (Figure 27).
Figure 26: Global predictions of current seasonal habitat suitability for whale sharks.

Prediction maps generated from generalised linear mixed-effects models fit with the sightings and effort data collected by tuna purse-seine fisheries. Where environmental inputs fell outside the environmental space used for the original statistical fit results are shown as “out of range” in the map. ★ indicates known aggregation locations within the seasons represented in each map (symbol size proportional to relative size of aggregation). Areas where some environmental predictors were not available (e.g., due to cloud cover) are shown in white (no result). To aid visualization, black line delineates areas where habitat suitability > 0.1 was predicted.
Figure 27: Predicted shift in global of whale sharks habitat suitability for 2070 under a no-climate-policy reference (no greenhouse gas emissions) scenario.

Black line delineates areas where higher habitat suitability (> 0.1) had been predicted under current environmental conditions (Figure 26), for visual assessment of the habitat contraction (reduction in the number of suitable 9-km cells of 5.0 % in Jan-Mar, 3.3 % in Apr-Jun, 4.7 % in Jul-Sep and 6.1 % in Oct-Dec) and pole-ward shift (~ 5 °N in Jan-Sep in the Atlantic Ocean, and 3-8 °S in the Indian Ocean). ★ indicates known aggregation locations within the seasons represented in each map, with symbol size proportional to relative size of aggregation (Sequeira et al., 2013)
Discussion

Species distribution models can provide information on suitable habitat for species occurrence (e.g., Guisan & Thuiller, 2005), and coupling them to climate forecasts provides a strong approximation of potential future distributional shifts in response to warming (Araújo et al., 2005; Hannah et al., 2002). Access to a unique global dataset of at-sea whale shark sightings provided an opportunity to model both current and future habitat suitability over the entire range of this pan-oceanic species. With sea surface temperatures predicted to increase by at least 2 °C on average by 2070 (under a no-climate-policy reference scenario), our models forecast both a (weak) pole-ward shift and contraction of suitable area, with loss of habitat (~ 6%) occurring mainly in the warmer equatorial region of the Atlantic and Indian Oceans (Figure 27). Such a contraction is congruent with predictions for other warm-water plankton-feeding species (e.g., Southward et al., 1995).

The current range of whale sharks is thought to be predominantly between 30 °N and 35 °S, which generally accords with our current worldwide habitat suitability predictions for the austral spring and summer (Figure 26). However, the environmental envelope predicted during the boreal spring and summer shifts to latitudes as high as 40 °N – latitude where whale sharks have been reported more recently (e.g., Portugal) (Rodrigues et al., 2012).

Within the Indian Ocean, our current predictions agree with the distribution of observations in that basin (Figure 26) (Rowat, 2007; Sequeira et al., 2013). For example, we predicted high suitability during the first three months of the year around India, Maldives and Bangladesh, Djibouti, Kenya and Mozambique, which concurs with seasonal observations (Rowat, 2007). In the second quarter of the year, we also predicted high suitability at Ningaloo Reef (Australia), the Mozambique Channel and
Gujarat (India), again agreeing with seasonal observations (Pravin, 2000; Rowat, 2007). From July to September, suitability was highest around the Mozambique Channel (Pierce et al., 2010) and also Seychelles (Rowat, 2007), although we lacked enough data to predict suitability in the northern part of the Indian Ocean. Finally, we predicted high suitability in the last three months of the year along the east coast of Africa, Bangladesh and Thailand, which also reflects observations (Rowat, 2007).

In the western Pacific, the relatively low habitat suitability is partially a function of the higher sighting (fishing) effort there, which we offset in our models. Regardless, predicted suitability was slightly higher in the Philippines from April to June (Quiros, 2007), and around Taiwan from July to September (Chang et al., 1997) (Figure 26). Additionally, our models suggest the existence of a ‘corridor’ of suitable habitat, especially between July and September, linking the eastern and western Pacific. This longitudinal pattern is possibly associated with warmer temperatures, because the predicted habitat suitability within this ‘corridor’ was slightly stronger within our scenario for 2070. Although whale sharks occur in the Gulf of California (Cárdenas-Torres et al., 2007) and the Galapagos during most of the year, our model failed to predict higher suitability there – additional data from the eastern Pacific would probably help improve model performance.

The highest suitability worldwide was in the Atlantic Ocean, which is also the region for which the highest-resolution and longest-running (since 1980) data were available. Apart from the Gulf of Mexico and the Azores, there is little other information available for whale sharks in this region. In the Azores, occurrences peak at the end of August and September (Sa, 2008), and in the Gulf of Mexico, they peak in the second quarter of the year (Heyman et al., 2001). These are well predicted in our global model. However, our models did not predict high suitability during the following quarter
in the Gulf of Mexico where some sightings are known to occur (Motta et al., 2010). Other suitability congruent with fisheries observations includes the western coast of Africa mostly between April and September. However, the high suitability predicted in the southern Atlantic cannot yet be validated due to a lack of observations. The high suitability predicted off South Africa in the first quarter is congruent with the hypothesis that whale sharks can cross from the Indian to the Atlantic Ocean through the Cape of Good Hope (Sequeira et al., 2013).

The top-ranked model for all three oceans included all physical variables (excluding standard deviation for sea surface temperature) previously identified (Sequeira et al., in press; Sequeira et al., 2012), but the interaction terms with ocean demonstrated some differences among oceans. We found evidence for sea surface temperature to affect occurrence of whale sharks differently in each ocean, being an important predictor both in the Atlantic and Indian, but with no discernable effect in the Pacific Ocean. The range of temperatures for January to March in the Pacific (26.7 – 30 °C) was narrowest compared to other oceans (25.2 – 31.2 °C in the Indian, and 19.4 – 28.3 °C in the Atlantic; Figure 24), and falls entirely within the range reported for whale sharks surface sightings (26.5 – 30 °C) (Sequeira et al., 2012). This might explain why our models did not show evidence that sea surface temperature affects whale shark occurrence in the Pacific Ocean, and might also be part of the reason why the habitat suitability in the entire Pacific Ocean was generally low.

Whale shark occurrence is often associated with productive areas (Hsu et al., 2007; Taylor & Pearce, 1999), and chlorophyll a, a proxy for phytoplankton concentration, has been used successfully to predict occurrence (Kumari & Raman, 2010; McKinney et al., 2012). However, we found no evidence that chlorophyll a contributed to the suitability predictions as we determined previously for the Indian
Ocean (Sequeira et al., 2012), possibly due to the temporal/spatial lags associated with primary production and secondary consumers (dilution and downstream effects) (Bradshaw et al., 2004; Wafar et al., 1984). The broad spatial scale of investigation might also have contributed to reducing the importance of primary production (cf. Bradshaw et al., 2004), especially in the open ocean where variation in chlorophyll a concentration is low.

According to Thuiller (2004), projections on climate change impacts in species occurrence should be developed considering the species' entire range. In our case, this range is circumglobal, which is the reason behind our extrapolation to the total area where predictor values are within the range used to calibrate the model. The climate-induced impact on species occurrence is usually investigated by the coupling climate forecasts with species distribution models (Araújo et al., 2005; Hannah et al., 2002). However, this coupling adds extra limitations (as specified in the Introduction), and results from such modelling frameworks are now considered to provide only a first approximation of possible changes. In the same way, our predictions of future whale shark habitat changes are intended only to be informative, providing only a baseline for temperature-dependent predictions. For a more comprehensive study on the influence of temperature in the occurrence of this species, not only should more climate-change scenarios be tested (including multiple global circulation model outputs), other species distribution models could be generated to assess variability in habitat suitability results, examine error and determine confidence intervals (Araújo & New, 2006; Thuiller, 2003). We extended our main results using the BIOMOD modelling platform (see Appendix D – BIOMOD run). The ensemble habitat suitability results for the area covered by the fisheries we obtained were similar to those derived from GLMM (Figure D2 in Appendix D).
Moreover, to improve our model predictions, more sightings data (ideally with circumglobal coverage) should be used. We have used the largest, most extensive dataset derived from tuna fisheries logbooks; however, it only covered a portion of areas managed by Regional Fisheries Management Organisations (New Zealand Ministry for Primary Industries website at fs.fish.govt.nz/Page.aspx?pk=103&tk=319.). Therefore, there is scope to develop our models based on data from other areas, such as the eastern Pacific Ocean (managed by the Inter-American Tropical Tuna Commission), which could assist improving the overall model results. As suggested in the previous chapter, a more collaborative effort between researchers and fisheries management organisations would assist improving scientific research and, in this specific case, potential improving our model fitting.

The opportunistic dataset we used only contained whale shark presences relative to the sea surface only. The latter also applies to the set of environmental predictors we used to build the models. While the first drawback could be addressed by randomly generating pseudo-absences (Barbet-Massin et al., 2012; Phillips et al., 2009) and repeating this procedure multiple times to account for any biases derived from pseudo-absence selection, the second is not simple to address. Environmental (and sightings) data spanning subsurface dynamics are not currently available for practical model development, especially where the study area covers most of the global ocean extent. Further, the whale shark sightings dataset I used included recent sightings (up to 2010) that might herald future patterns of distribution change. Recently, whale sharks have been seen in locations where they were not expected to occur (Duffy, 2002; Rodrigues et al., 2012; Turnbull & Randell, 2006), and so the data I used could already epitomise shifts in this species’ occurrence. It is therefore possible that my results show an already biased range of suitable habitat for this
species. However, due to the large area under study and the limited data on presences, excluding part of the dataset available was not an option.

Despite the current lack of alternatives to fisheries-collected sightings in the open ocean (Jessup, 2003), combining remotely sensed data with opportunistically recorded occurrences has, however, proved its worth for estimating whale shark distributions (Kumari & Raman, 2010; McKinney et al., 2012; Sequeira et al., 2012). Indeed, most of our understanding of whale shark ecology and biology hails from opportunistically collected datasets by the eco-tourism (e.g., Wilson et al., 2001) or fishing industries near shore at aggregation sites (e.g., McKinney et al., 2012). As such, we should endeavour to analyse all available datasets to conserve this pan-oceanic species and largest remaining fish – refusal to examine multiple lines of evidence risks imperilling the species further through lack of ecological understanding.
CHAPTER VII. GENERAL DISCUSSION

Overview of my research

My research provided the first ocean-scale and global distribution models of the world’s largest fish, the whale shark (*Rhincodon typus*). Here, I applied a novel and comprehensive modelling framework that took spatial variation and sub-optimal sampling effort data into account while teasing out temporal patterns. I showed how appropriate statistical approaches can be applied to oceanic species for which data are incomplete, fragmented and uneven, such as for the whale shark, highlighting the value of unique datasets collected opportunistically. My approach also provides a mathematical platform for marine climate change predictions that have typically lagged behind terrestrial systems, and shows the possible impacts of a warming climate on a pan-oceanic shark species.

In terms of the modelled organism on which I focussed by analytical framework, my results highlight the need to account for global migratory behaviour both in whale shark research programs, and conservation management decisions. As a starting point, I suggest that the whale shark’s current IUCN Red List status of *Vulnerable* (Norman, 2005) be revised. According to my spatially explicit trend results, the number of whale sharks observed by tuna fisheries, which cover a large extent of the three major oceans, has been reduced to about 50% in the last decade, both in the Atlantic and Indian Oceans (there are no data available prior to 2000 for the Pacific Ocean). The causes for this reduction at such a broad spatial scale (possibly reflecting a global population trend) are not yet understood and are potentially still occurring and possibly irreversible. These characteristics are captured within the IUCN Red List
Criterion A for *Endangered* species. Moreover, considering the life span of whale sharks, 10 years might correspond to only 10% of a whale sharks lifetime (or even up to 25% of a single generation). As I suggested in Chapter V, despite the possible cyclical occurrence of whale sharks in each ocean basin, the observed reduction in sightings is occurring simultaneously in a large part of the species known geographical range. Due to their long life span, late age at maturity, and the relatively little information available on their reproduction, this species might actually already be facing a high risk of extinction in the wild from past and ongoing exploitation and climate warming. Therefore, I recommend that it should be considered at least for the status of *Endangered*.

Other criteria from the IUCN Red List class of *Endangered* might also apply. For example, Criterion C relates to the total number of mature individuals. Most whale sharks spotted in specific aggregations are immature (Graham & Roberts, 2007; Heyman *et al.*, 2001; Rowat *et al.*, 2011; Wilson *et al.*, 2001), and available population estimates are around 300-500 individuals for a single site at Ningaloo, Australia (Meekan *et al.*, 2006). Moreover, whale sharks are thought to attain maturity at ~ 9 metres in total length (Colman, 1997), and there is evidence that the average size of sighted sharks is declining (Bradshaw *et al.*, 2008). Considering that whale sharks are only rarely sighted, that most of the sightings are immature animals, and viewing aggregations as possible sub-populations, Criterion C2ai could also apply: “population size estimated at < 2500 mature individuals, in continuing decline (observed, projected or inferred) and with < 250 mature individuals per subpopulation.”

In addition to the suggestion of status change from Vulnerable (currently based on the IUCN Red List Criterion A2bd+3d) to Endangered (based on Criterion A2ad + 4d or C2a1), my research also provides an updated global map reflecting
seasonality in whale shark habitat suitability, which can now be added to the whale shark section of the IUCN Red List. My thesis answers some of the most important questions about the ecology of whale sharks to assist with their long-term conservation. In summary, I proposed how whale sharks are connected at a global scale, how their temporal patterns of occurrence probably reflect their migratory behaviour, and how their distribution depends on environmental conditions that shift seasonally. The latter can, to some extent, explain the known patterns of the species’ seasonal occurrence, and also how a principally warm-water organism can travel beyond its ideal thermal nice (i.e., to latitudes > 30 °N and 30 °S) (e.g., Rodrigues et al., 2012). The underlying hypothesis that climate change is already affecting this species’ distribution underscores an urgent need to measure their global occurrence and inter-ocean connectivity.

While focusing on the global patterns of whale shark ecology and connectivity, my research highlighted the following specific findings:

1. Timing and distribution patterns of whale shark occurrence suggest a connection between several aggregation sites among the three largest ocean basins (Chapter II);

2. Whale shark habitat suitability in the Indian Ocean is correlated mainly with spatial variation in sea surface temperature, and shifts seasonally (Chapter III);

3. The temporal variation in whale shark probability of occurrence could be representative of inter-decadal cycles (i.e., cycles ≥ 15 years) (Chapter IV);
4. The probability of whale shark occurrence is asynchronous across oceans, but even being cyclical, the decreasing number of sightings reported in the fisheries datasets (at an ocean-basin scale) agrees with the reported declines from near-shore aggregations (Chapter V).

5. With increasing sea surface temperature from anthropogenic climate disruption, whale shark habitat suitability is expected to shift polewards, and currently suitable habitat within the warmer equatorial region of the Atlantic and Indian Oceans will potentially be lost (Chapter VI).

A worldwide perspective

My thesis’ global perspective indicates that a more comprehensive focus can, and should be adopted in future research endeavours. This shift of focus from local aggregations to wider areas (e.g., regional or global) engenders new hypotheses about the ecology of this species that can be tested as technology improves, more tagging results are obtained and genetic evidence is refined.

One of the first questions that follows naturally from my global connectivity approach (Chapter II) is: If individual whale shark movements do span the world’s oceans, how are their global movement patterns defined? My results show that environmental conditions, such as temperature and depth, are important determinants of whale shark occurrence, and could therefore be related to their migratory behaviour. For example, the seasonal clockwise shift in habitat suitability I found in the Indian Ocean (Chapter III) could partially explain why re-sightings occur in the same near-shore locations in following years. My temporal analysis describing possible cyclic
patterns (Chapter IV and V) could help explain the variation recorded in inter-annual occurrences in several locations (Graham & Roberts, 2007; Rowat et al., 2009a; Wilson et al., 2001). When expanded from the regional (ocean basins) to a global scale (Chapter VI), my analysis revealed seasonal suitable habitat in some critical areas, such as around South Africa. This is the first insight into how migration between the Indian and Atlantic is possible.

Despite the role environmental conditions play in determining whale shark occurrence, to clarify movement patterns requires an understanding of population structure. It has been suggested that whale shark migratory patterns could reflect sex- and age-specific behaviour (Borrell et al., 2011; Eckert & Stewart, 2001; Ramírez-Macías et al., 2007; Rowat et al., 2011). In Chapter II where I collated and reviewed all available whale shark movement data, including published tracks, I highlighted two points. First, the longest whale shark tracks were mostly reported for females, and second, most of the larger sharks (> 6 metres in total length) travelled on average greater distances than smaller sharks (Chapter II). These findings add further support to the hypothesis that whale shark movement patterns might be associated with behavioural segregations.

Consequently, we might ask: If movement patterns are associated with population structure and environmental changes, will different whale shark classes (length/age or sex) be affected differently by climate change? Although my research did not focus on defining patterns for different classes (the focus was on occurrences, with no data available on sex or length/age), my global model suggests that whale shark habitat suitability will be reduced in locations of the central Indian and Atlantic Oceans as the climate warms. In the Indian Ocean aggregation sites in particular, mostly immature males are observed in the area which I projected as having the
highest probability of warming – such as the Seychelles (Rowat & Gore, 2007) and Maldives (Riley et al., 2010). If this immature male class is constrained to using these areas, then climate change will possibly have a stronger effect on them than other components of the population. In the Atlantic Ocean, the predicted reduction in habitat suitability in the central area might impose a separation (north and south) of the population in the future. If the longer migrations, such as the one between the Caribbean Sea to near the Saint Peter Saint Paul archipelago reported by Hueter (2008), are limited to females, and are associated with specific behaviours such as reproduction, then climate change might affect this behaviour (e.g., by imposing a physical barrier to these migrations), and therefore lead to severe impacts on the global whale shark population.

A related question is whether fishing and tourism are also impacting only specific classes. If sex and age segregation is indeed happening, with juvenile males occurring near shore and females being more oceanic, then any habitat disturbance caused by coastal development, shipping and/or tourism could have a greater impact on juvenile males, while broad-scale tuna fisheries might affect mostly females. If true, this notion highlights the need to measure whale shark population structure, and to define the extent of disturbance, damage and survival rates derived from human-related activities (Bradshaw et al., 2007).

**Better management**

My results provide a strong incentive for current whale shark management policies to be revised. There are currently no global measures in place to limit the exploitation of live whale sharks (Rowat & Brooks, 2012). Whale sharks were recently included in
Appendix II of the Convention on Migratory Species (CMS, 2010), which could lead to the development of international conservation measures (Rowat & Brooks, 2012).

However, current management of whale sharks occurs mostly within confined tourist locations through the imposition of a maximum number of tourism operator licenses, and by implementing codes of conduct to minimise disturbance in some countries (Cárdenas-Torres et al., 2007; Pierce et al., 2010; Quiros, 2007). In the 30 years after the first initiative for international management of marine resources (UNCLOS, 1982), and despite whale sharks being classed as Vulnerable in the IUCN Red List (www.iucn.org), a global effort to protect whale sharks is only indirectly supported through control of international trade of specimens (Appendix II of the Convention of Trade in Endangered Species of Wild Fauna and Flora; www.cites.org/eng/app/appendices.php). Also, national bans have been imposed on targeted commercial fisheries of whale sharks, but they lack enforcement (Riley et al., 2009), and the global-scale fisheries interaction with whale sharks is currently not regulated in any way.

My global connectivity model implies that whale sharks cross multiple national and high-seas fisheries jurisdictions (whales sharks occur near the shore of > 100 countries; Figure 5 in Chapter II), so if long-term conservation of this species is to be achieved, active global conservation measures need to be implemented and enforced. The large extent at which interactions between tuna fisheries and whale sharks occurrences manifest (see Figure 23 in Chapter VI), and the evidence that whale shark numbers are declining in both touristic locations (Bradshaw et al., 2008; Bradshaw et al., 2007) and in fisheries interactions (Chapter V) further highlight the need for global conservation measures. For example, in the Western and Central Pacific there is on-going discussion on the implementation of a ban on using the whale
shark as tuna aggregation device. However, the implementation of this ban has been hindered mostly by a lack of support from Japan (see reports of regular sessions of the Western and Central Pacific Fisheries Commission; www.wcpfc.int/meetings/1). Given the relatively lower sighting probability in the three oceans in the last decade, this measure should be not only enforced in the Western and Central Pacific, but also embraced by all the Regional Fisheries Management Organizations.

**Future improvements**

**Input data**

As I mention throughout the thesis, the datasets I used were collected opportunistically, and therefore they bring with them a series of inherent complications. I addressed and accounted for some of these imperfections during model development. For example, I addressed the temporal and spatial bias of sampling effort within each grid-cell by including fishing effort in the models as a 'weight' (Chapter III) or as an 'offset' (Chapters IV, V and VI). The latter accounts for the fact that more whale shark sightings are expected in areas where fishing effort was higher, and seems more appropriate to reflect the different exposure of each area to the event of interest (*i.e.*, a sighting), while the former weights the model's response according to the amount of effort observed in each grid cell. However, other problems such as the coarse resolution of the effort data are also problematic, and can only be addressed by gaining access to higher-resolution data. Access to better datasets can be achieved if collaborations are established between researchers and fisheries management.

---

1 Following examiners’ comments, parts of this section were included in the chapters and therefore the reader will find some repetition.
organisations. Such collaborations would allow researchers to give greater insight on how a better dataset could be collated for specific objectives using the resources currently in place. Commercial-in-confidence restrictions on access to fisheries data (especially sensitive data such as fish catch) delay scientific research from providing more accurate models on fisheries impacts.

Another important aspect of the data I used, including both whale shark sightings and environmental variables, is that they pertain mostly to the sea surface. Environmental data collection at depth, and at an adequate resolution for a regional or global analysis such as presented here, are currently not available. Therefore, even though whale sharks dive frequently, assessment of how they explore the vertical habitat and estimation of its suitability at a broad scale will only be possible when new technologies are developed to allow vertical data collection at a broad scale.

The datasets I used contained only recorded presences, and therefore two disadvantages are associated with such data. The first is related to the probability of detecting a whale shark. There is a high possibility that in any given area, whale sharks are present but not observed, and so an occurrence is not recorded (e.g., if shark was below the surface). The second disadvantage relates to the lack of information on absences. To account for this drawback, I used different methods to select absence points within the area covered by the fisheries and include them in the models as pseudo-absences; this is known to affect model results, which I discuss below. However, even datasets containing absence locations can be problematic when using them as inputs for species distribution models (Elith et al., 2006; Zaniewski et al., 2002). This is the case when the reason for recorded absences is not correctly identified (i.e., they might be associated with poor detection, habitat disturbance, competition or inappropriate timing rather than with an unsuitable habitat). Therefore,
both the inclusion of ‘real’ absences or pseudo-absences can lead to biased results. With no current alternatives to fisheries-collected sightings in the open ocean (Jessup, 2003), the regional and global analysis presented here could not have been done without using these uniquely available datasets.

In the context of climate change, there is yet another point to consider. The whale shark sightings dataset I used included recent sightings (up to 2010). Recently, whale sharks have been seen in locations where they were not expected to occur (Duffy, 2002; Rodrigues et al., 2012; Turnbull & Randell, 2006), and so the data I used in my models could already epitomise shifts in this species occurrence. It is therefore possible that my results show an already biased range of suitable habitat for this species. However, due to the large area under study and the limited data on presences, excluding part of the dataset available was not an option.

Models

Despite being partially limited by data collected opportunistically, the species distribution models themselves have several limitations. Although I used presence and pseudo-absence data in my linear models, I also developed distribution models in MaxEnt2 (Elith et al., 2011; Phillips & Dudík, 2008) using presence-only data. Even though the resulting maps of whale shark probability of occurrence were relatively similar using both methods (Chapter III), the linear models generally provide more realistic projections (Brotons et al., 2004; Elith et al., 2006; Zaniewski et al., 2002).

---

2 MaxEnt also requires background data to estimate the species distribution; therefore, it cannot be considered a ‘true’ presence-only model such as rectilinear envelopes.
Also, Tsoar et al. (2007) concluded that presence-only models are more accurate for species with narrow niches (not expected to be the case for whale sharks).

To allow binomial estimation in the generalised linear models, I generated pseudo-absences, but these can influence results and predictions (Barbet-Massin et al., 2012; VanDerWal et al., 2009; Wisz & Guisan, 2009). It is therefore important to acknowledge the assumptions made during the generation of pseudo-absences. Following published suggestions on how to select pseudo-absences accounting for the same sampling bias in presences (Phillips et al., 2009), or by using pre-defined areas within some distance from the recorded presences (Engler et al., 2004; Lobo et al., 2010; Zaniewski et al., 2002), I tried different methods for pseudo-absence selection. These included a selection of grid cells with a probability weighted by the inverse distance to the presences, weighted by the total tuna catch, and based on a random selection of non-presence grid cells (Chapter III). In subsequent analysis, I used only random pseudo-absence generation, which, being simple to apply, did not produce prediction maps dissimilar to the ones obtained by the other methods, and it did not decrease model performance (Chapter III). The last point accords with the results from a recent publication specifically addressing pseudo-absence selection issues (Barbet-Massin et al., 2012). Also, the only method that showed better performance (in some cases only) was dependent on data from the tuna catch, which was only available at a much coarser resolution.

The number of pseudo-absences chosen can also affect models results. I made this choice based on the number of presences available for each model. I used presence:pseudo-absence ratios (i.e., prevalence) of 1:1 and 1:10 in the spatial analysis, and 1:100 in the temporal models. I also averaged results of multiple model runs with different sets of pseudo-absences in each model. Barbet-Massin et al.
(2012) tested the influence of prevalence in models accuracy, and concluded that for
generalised linear models, accuracy increased until prevalence was 0.1 \((i.e., 1\text{ presence : 10 pseudo-absences})\) as used in my spatial analysis; Chapter IV and V and VI), and is constant when a higher number of pseudo-absences are used. Therefore, using a large number of pseudo-absences is not expected to decrease models accuracy, but allows enough points for the spatially explicit temporal analysis.

Another related issue is the decision regarding which probability value should be used to convert the model results into presences or absences (Jiménez-Valverde & Lobo, 2007; Liu et al., 2005). This value is commonly assumed to be 0.5, but for presence: absence ratios different to 1:1, this assumption results in incorrect classification of absences (if prevalence is low, i.e., more absences than presences) or presences (if prevalence is high) (Jiménez-Valverde & Lobo, 2007). Several methods can be used to specify this threshold (see Liu et al., 2005), such as averaging the predicted probabilities, using sensitivity and specificity, the prevalence value, or the value that maximises \(\kappa\). I considered this threshold equal to the prevalence used, which is considered a good approach (Liu et al., 2005).

In the final chapter (Chapter VI), I coupled my distribution models with the results of models projecting future climate (derived from global circulation models) (e.g., Araújo et al., 2005; Cheung et al., 2010) to produce worldwide predictions of future whale shark occurrence. According to Thuiller et al. (2004b), projection of climate change impacts should be developed when considering the entire range for the species. In the case of whale sharks, I have shown that this can only be achieved when considering the worldwide distribution. Although the coupling of climate forecast and species distribution models is considered to be a powerful way to investigate possible climate-induced distributional shifts in species occurrence (Araújo et al.,
it incurs additional limitations, and is now considered to provide only a first approximation of possible changes. For example, there is considerable variation in the results of different climate-forecasting models (e.g., Araújo et al., 2005; Thuiller, 2004), and the spatial scale of the resulting predictions is usually too coarse to have any biological relevance (Hannah et al., 2002). Following recent published suggestions (Fordham et al., 2011; Fordham et al., 2012), I used averaged results of five climate-forecasting models to minimise uncertainty related to the underlying generation of the climate prediction. I also down-scaled the climate predictions before coupling them to my distribution models. However, my predictions of future habitat changes are intended only to be informative, showing a possible way by which whale shark distribution might be affected by changes in temperature. For a more coherent study on the influence of temperature in the occurrence of this species, more climate change scenarios should be tested along with the inclusion of other global circulation model results in the ensemble approach.

**Results**

Validation of model performance is an important component of any modelling procedure (Franklin, 2010). I validated my models through the use of the 10-fold cross validation technique (Davison & Hinkley, 1997), leaving a tenth of the dataset out while running the models, and then using it to validate the results. As a qualitative validation, I also compared the predicted probabilities of occurrence both to the original data (to match higher probability areas with number of sightings), and with current knowledge on whale shark seasonal occurrences in near-shore locations. Independent validation by using independently collected data could be useful. This will be possible as more data are collected, and then used to validate my prediction maps. However, this sort of
validation can only be used for predictions of current habitat suitability. Validation of future projections is, of course, not possible; however, validation of model performance can be achieved by hind-casting and validating the results with already collected data (Araújo et al., 2005). As pointed out above, independent model validation in my whale shark case study was not possible due to data limitations.

**Future directions**

My research has identified some key aspects of whale shark ecology that should be analysed further to advance our knowledge of whale sharks. Below, I describe these aspects presenting some of the gaps that still remain, and I give some suggestions on where I think further research should be taken to allow timely conservation of this impressive species.

In Chapter II, where I formulated my hypothesis about global connectivity of the whale shark meta-population, I suggested that some locations could be whale shark ‘thoroughfares’. Relatively little research has been done in these locations (e.g., Christmas Island and Saint Peter Saint Paul archipelago), and deployment of tags and tissue collections near potential thoroughfares would likely increase the probability of capturing pan- or trans-oceanic movements. Throughout my thesis I also highlighted the value of using opportunistic data, considering the logistic challenges of surveying widely distributed species over their entire range. Opportunistically collected data, especially wide-ranging and long-term data such as the fisheries datasets I used, are essential sources of information to assist with the understanding of the ecology of such a highly migratory species. The data I used covered only sections of the area managed by Regional Fisheries Management Organisations (Figure 28a), and so
there is still scope to develop similar models based on data from other areas, such as the eastern Pacific Ocean (managed by the Inter-American Tropical Tuna Commission). Also, some of the other Regional Fisheries bodies (see Figure 28b) might possess data of importance for whale shark studies. A way to collect additional data that can be useful for similar studies is simply by adding records of whale shark sightings within catch logbooks also recording the search effort. A more elaborate approach is also to include sightings of other species. This could be used as a proxy for absences when a shark is not recorded.

Figure 28: Map of world organisations managing fisheries.

a) Regional Fisheries Management Organisations for highly migratory fish stocks (Tuna and Tuna-like). Map obtained from the New Zealand Ministry for Primary Industries website at

NOTE:
This figure/table/image has been removed to comply with copyright regulations. It is included in the print copy of the thesis held by the University of Adelaide Library.
Despite finer-scale evidence that chlorophyll a (used as a proxy for food availability) is an important predictor of whale shark occurrence (Kumari & Raman, 2010; McKinney et al., 2012), my results revealed that at least at broad spatial scales, its importance does not persist. Availability of data describing the distribution of whale shark prey, such as zooplankton abundance (Clark & Nelson, 1997; Jarman & Wilson, 2004), might improve the results of the distribution models. The lack of these data at spatial scales relevant to oceanic patterns is an important gap that hinders further insight on how whale shark movement is related to food abundance. Although availability of such data depends on technological advances, data collection on zooplankton concentration in limited areas (such as aggregation locations) could be included in future research programs.

Another gap in our knowledge of whale sharks relates to population structure. The construction of broad-scale (meta-) population models to connect whale shark sub-populations should follow the global conceptual movement model presented in Chapter II. Such population models could be based on inferred demographic
(Bradshaw et al., 2007) and movement rate data, but key aspects of whale shark life history still need to be measured. More tagging and genetic studies will assist in describing those aspects, and in quantifying if there are any stage-specific movement trends. Studies using ‘proxy’ organisms, such as region-specific parasites (Braicovich & Timi, 2008; Lester, 1990), could also assist in describing movement for this species. One way to test the validity of my connectivity model is by focusing tagging and photographic-identification studies on ‘connectivity regions’ - regions where evidence for connection already exists (Chapter II). Examples of these regions include Madagascar, the Seychelles and Tanzania, or the region including the Philippines, Malaysia and Gulf of Thailand.

Tagging approaches can provide insight into the movement behaviour of migratory species, however they bring their own problems (e.g., Brunnschweiler et al., 2009). For example, despite the long tracks obtained for some of the species in the successful TOPP program (tagging of Pacific predators; Block et al., 2011), most of the temporal coverage of the individual tracks was less than one year (maximum duration for different species ranged from 19 to 358 days) (Block et al., 2011 supplementary material). This multinational, collaborative program (part of the Census of Marine Life) received funding from several organisations to allow the deployment of several thousands of tags, of which only less than half resulted in individual tracks (Block et al., 2011). However, some of the problems associated with tagging procedures can be improved as technology progresses. Some suggestions to maximise the duration of tracking have already been made, and should be adopted when planning new tagging studies. These suggestions include using fast-loc GPS technology (Hays et al., 2007), applying anti-fouling paint (Hammerschlag et al., 2011), and use copper salt-water switches (Hays et al., 2007). But improvements to prevent
premature detachment (Brunnschweiler et al., 2009) are also needed, and new attachment options should be considered. To maximise the potential for measuring long-distance migrations, I suggest that tagging studies target multiple sharks simultaneously within the same aggregation location (cf. Eckert & Stewart, 2001; Wilson et al., 2006), and that deployments are made close to the end of the peak aggregation season. Additional tagging studies on whale sharks released from pelagic fisheries nets would also provide information on oceanic movements to complement results obtained from known aggregations.

In the context of future distribution shifts, my results provide only a baseline for temperature-dependent predictions, and to provide an initial picture of the most plausible and/or extreme predictions under warmer conditions, multiple climate change scenarios should be tested along with multiple global circulation models outputs. However, other factors such as changes in zooplankton abundance might also influence the future distribution of whale sharks. Climate change is currently affecting plankton distributions (Edwards & Richardson, 2004; Hinder et al., 2012), and so my prediction of future suitable habitat might be further restricted by food limitations. The long-term management of whale sharks will therefore require a better understanding of their basic ecology and demography, but also of how these will be altered under a warming climate.

So, for timely conservation of this iconic species, I suggest that the following measures should be taken:

- Develop tagging and photo-identification studies covering ‘thoroughfares’ and ‘connectivity regions’. This can be done by establishing international collaborations. To avoid divergence of methods used, guidelines for analysis should be shared or similar data should be analysed by the same group.
• Establish links with fisheries management organisations to:
  
  o give insight into data collection and gain access to better data
  
  o participate in observer and/or research programmes including tagging and identification of whale sharks used in fisheries interactions

• Include sampling of zooplankton abundance in forthcoming research programs

• Consider the development of models including data on whale shark prey and population structure

• Develop active global management strategies on the use of whale sharks by commercial fisheries.

The last measure would have the most immediate positive impact on whale shark populations and would, in turn, allow more time to collect new data to assist with the conservation of this species. Safe Release Guidelines have been suggested at the regular sessions of the Western and Central Pacific Fisheries Commission, as well as the inclusion of 100 % observer programmes and tagging studies. Such data are fundamental to generate appropriate management strategies. Also, the IUCN status for the species should be revised and changed from Vulnerable to Endangered.

In conclusion, I have demonstrated how opportunistic data can be used in species distribution modelling by addressing specific problems such as effort bias and spatial autocorrelation. By using similar modelling tools with new input data, including food availability and information on population structure, the timing of whale shark appearances at specific sites and their movements could be predicted, and subsequently used to examine the drivers of population trends. As a conservation
ecologist, I believe we now have a last (and time-constrained) opportunity to preserve this unique species, and action is needed immediately.
AFTERWORD

On a logistic note, it is worth referring that while doing my Ph.D., I have participated in four whale shark field trips where I collected data to identify individuals (photo-identification, stereo-measurement and gender), collected biopsy samples (for genetic analyses) and deployed tags (to track movement). Due to delays in recovering tags, movement data could not be included in this thesis. Other field work components, as derived from the individuals’ data collection (identification, measurements and genetics) are still being analysed (and compared with data collected in other study sites), and results will soon be evaluated in the context of my global approach.
### Table A1: Summary of representative whale shark studies (since 1996) included in Figure 4.

Grouped by (1) occurrence, population and correlates for occurrence, (2) movement, (3) tourism and conservation, (4) mortality and (5) foraging, in different locations worldwide.

<table>
<thead>
<tr>
<th>Local</th>
<th>Occurrence Population</th>
<th>Movements</th>
<th>Eco-tourism / conservation</th>
<th>Mortality</th>
<th>Foraging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic Ocean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>(Turnbull &amp; Randell, 2006)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td>(Hoffmayer et al., 2006)</td>
<td>-</td>
<td>(Suárez et al., 2007)</td>
<td>-</td>
<td>(Hoffmayer et al., 2007)</td>
</tr>
<tr>
<td></td>
<td>(Burks et al., 2006)</td>
<td></td>
<td></td>
<td></td>
<td>(Hueter, 2007)</td>
</tr>
<tr>
<td></td>
<td>(Cárdenas-Palomo et al., 2010)</td>
<td>-</td>
<td></td>
<td></td>
<td>(Motta et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>(McKinney et al., 2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caribbean Sea</td>
<td>(Graham &amp; Roberts, 2007)</td>
<td>(Graham et al., 2006)</td>
<td>(Quiros, 2005)</td>
<td>(Diaz, 2007)</td>
<td>(Heyman et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>(de la Parra Venegas et al., 2011)</td>
<td>(Gifford et al., 2007b)</td>
<td>(Cárdenas-Torres et al., 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Graham &amp; Roberts, 2007)</td>
<td>(Heyman et al., 2010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Ziegler et al., 2011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>(Gadig, 2007)</td>
<td>-</td>
<td>(Gadig, 2007)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(Hazin et al., 2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angola/Nigeria</td>
<td>(Weir, 2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KwaZulu-Natal</td>
<td>(Cliff et al., 2007)</td>
<td>(Gifford et al., 2007a)</td>
<td>-</td>
<td></td>
<td>(Beckley et al., 1997)</td>
</tr>
<tr>
<td>(South Africa)</td>
<td></td>
<td>(Gifford et al., 2007b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozambique</td>
<td>-</td>
<td>(Brunnschweiler et al., 2009)</td>
<td>(Pierce et al., 2010)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Brunnschweiler &amp; Sims, 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>-</td>
<td>-</td>
<td>(Bassen, 2007)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Njonjo, 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madagascar</td>
<td>(Jonahson &amp; Harding, 2007)</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Seychelles</td>
<td>(Rowat et al., 2009a)</td>
<td>(Rowat &amp; Gore, 2007)</td>
<td>(Rowat, 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Rowat et al., 2009b)</td>
<td>(Rowat et al., 2011)</td>
<td>(Rowat &amp; Engelhardt, 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>(Kumari &amp; Raman, 2007)</td>
<td>-</td>
<td>(Pravin, 2000)</td>
<td>(Hanfee, 1997)</td>
<td>(Borrell et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>(Kumari &amp; Raman, 2010)</td>
<td></td>
<td>(Joshi et al., 2007)</td>
<td>(Hanfee, 2001)</td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>Reference 1</td>
<td>Reference 2</td>
<td>Reference 3</td>
<td>Reference 4</td>
<td>Reference 5</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td><strong>Indian Ocean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KwaZulu-Natal</td>
<td>(Riley et al., 2010)</td>
<td>-</td>
<td>-</td>
<td>(Riley et al., 2009)</td>
<td>-</td>
</tr>
<tr>
<td>Djibouti</td>
<td>(Rowat et al., 2007)</td>
<td>(Rowat et al., 2011)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pakistan and Bangladesh</td>
<td>(Rowat et al., 2008)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thailand</td>
<td>(Theberge &amp; Dearden, 2006)</td>
<td>-</td>
<td>-</td>
<td>(Haetrakul et al., 2007)</td>
<td>-</td>
</tr>
<tr>
<td>Christmas Island (Australia)</td>
<td>(Hobbs et al., 2009)</td>
<td>-</td>
<td>(Davies, 2007)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ningaloo (Western Australia)</td>
<td>(Wilson et al., 2001)</td>
<td>(Meekan et al., 2006)</td>
<td>(Bradshaw, 2007)</td>
<td>(Gunn et al., 1999)</td>
<td>(Davis et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>(Bradshaw et al., 2007)</td>
<td></td>
<td>(Bradshaw et al., 2008)</td>
<td>(Wilson et al., 2006)</td>
<td>(Bradshaw, 2007)</td>
</tr>
<tr>
<td></td>
<td>(Norman &amp; Stevens, 2007)</td>
<td></td>
<td>(Hollmberg et al., 2008)</td>
<td>(Wilson et al., 2007)</td>
<td>(Mau &amp; Wilson, 2007)</td>
</tr>
<tr>
<td></td>
<td>(Bradshaw et al., 2008)</td>
<td></td>
<td>(Hollmberg et al., 2009)</td>
<td></td>
<td>(Catlin &amp; Jones, 2009)</td>
</tr>
<tr>
<td></td>
<td>(Holmberg et al., 2008)</td>
<td></td>
<td>(Sieeman et al., 2010a)</td>
<td></td>
<td>(Catlin et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>(Sieeman et al., 2010b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>(Stacey et al., 2008)</td>
<td>(Eckert et al., 2002)</td>
<td></td>
<td>(Quitos, 2005)</td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>(Aca &amp; Schmidt, 2011)</td>
<td>(Eckert et al., 2002)</td>
<td></td>
<td>(Quitos, 2007)</td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
<td>(Eckert et al., 2002)</td>
<td></td>
<td>(Pine et al., 2007)</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td></td>
<td></td>
<td>(Hsu et al., 2007)</td>
<td>(Hsu et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
<td></td>
<td>(Chen et al., 1997)</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td>(Li et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>(Duffy, 2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Regional / Global</strong></td>
<td>(Graham, 2007)</td>
<td>(Brooks et al., 2010)</td>
<td>(Sequeira et al., 2012)</td>
<td>(Norman, 2007)</td>
<td>(Fordham, 2007)</td>
</tr>
</tbody>
</table>
APPENDIX B – Ocean-scale prediction of whale shark distribution

Table B1: 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 mean2), sea surface temperature standard deviation (SST SD) and chlorophyll a (Chl a). 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 to 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 alone) and which environmental variables decreased the gain the most when omitted (worst without) are also shown together with the value obtained for the area under the curve (AUC) test. Shown for each model are the percent of contribution (cont(%) and the permutation importance (Imp) - results = 0 are shown as '-', values ≤ 5 in light grey, and highest values in bold.

<table>
<thead>
<tr>
<th>Season</th>
<th>autumn</th>
<th>winter</th>
<th>spring</th>
<th>summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Cont(%)</td>
<td>Imp</td>
<td>Cont(%)</td>
<td>Imp</td>
</tr>
<tr>
<td>random</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SST mean</td>
<td>4.9</td>
<td>53.4</td>
<td>48.9</td>
<td>40.4</td>
</tr>
<tr>
<td>SST mean</td>
<td>3.2</td>
<td>22.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SST SD</td>
<td>10.2</td>
<td>3.6</td>
<td>9.4</td>
<td>3.2</td>
</tr>
<tr>
<td>Chl a</td>
<td>5.6</td>
<td>1.7</td>
<td>37.5</td>
<td>42</td>
</tr>
<tr>
<td>depth</td>
<td>12.9</td>
<td>3.8</td>
<td>0.5</td>
<td>2.6</td>
</tr>
<tr>
<td>shore</td>
<td>62.2</td>
<td>14.7</td>
<td>2.4</td>
<td>9.9</td>
</tr>
<tr>
<td>slope</td>
<td>1.2</td>
<td>-</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td>Best alone</td>
<td>shore</td>
<td>Chl a</td>
<td>Chl a</td>
<td>Depth</td>
</tr>
<tr>
<td>Worst without</td>
<td>shore</td>
<td>Chl a</td>
<td>SST SD</td>
<td>Depth</td>
</tr>
<tr>
<td>AUC</td>
<td>0.721</td>
<td>0.701</td>
<td>0.691</td>
<td>0.668</td>
</tr>
</tbody>
</table>
### IDW – linear and quadratic features

<table>
<thead>
<tr>
<th></th>
<th>Depth</th>
<th>Chl a</th>
<th>Chl a</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SST mean$^2$</strong></td>
<td>1.1</td>
<td>42.6</td>
<td>8.3</td>
<td>2.1</td>
</tr>
<tr>
<td><strong>SST mean</strong></td>
<td>1.9</td>
<td>29.8</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td><strong>SST SD</strong></td>
<td>3.0</td>
<td>0.4</td>
<td>-</td>
<td>30</td>
</tr>
<tr>
<td><strong>Chl a depth</strong></td>
<td>12.6</td>
<td>3.5</td>
<td>52.3</td>
<td>27.2</td>
</tr>
<tr>
<td><strong>shore slope</strong></td>
<td>45.6</td>
<td>2.8</td>
<td>1</td>
<td>9.1</td>
</tr>
<tr>
<td><strong>Best alone</strong></td>
<td>Depth</td>
<td>Chl a</td>
<td>Chl a</td>
<td>Depth</td>
</tr>
<tr>
<td><strong>Worst without</strong></td>
<td>Depth</td>
<td>Chl a</td>
<td>Chl a</td>
<td>Depth</td>
</tr>
</tbody>
</table>

| AUC | 0.625 | 0.574 | 0.627 | 0.607 |

### tuna – linear and quadratic features

<table>
<thead>
<tr>
<th></th>
<th>Depth</th>
<th>Chl a</th>
<th>Chl a</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SST mean$^2$</strong></td>
<td>7.6</td>
<td>59.2</td>
<td>55.8</td>
<td>61.6</td>
</tr>
<tr>
<td><strong>SST mean</strong></td>
<td>1.7</td>
<td>18.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>SST SD</strong></td>
<td>1.1</td>
<td>-</td>
<td>8.9</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Chl a depth</strong></td>
<td>9.3</td>
<td>3.6</td>
<td>29.8</td>
<td>34.2</td>
</tr>
<tr>
<td><strong>shore slope</strong></td>
<td>65.4</td>
<td>12.7</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td><strong>Best alone</strong></td>
<td>shore</td>
<td>Chl a</td>
<td>Chl a</td>
<td>Depth</td>
</tr>
<tr>
<td><strong>Worst without</strong></td>
<td>shore</td>
<td>Chl a</td>
<td>SST SD</td>
<td>Depth</td>
</tr>
</tbody>
</table>

| AUC | 0.710 | 0.673 | 0.672 | 0.676 |

### background – linear and quadratic features

<table>
<thead>
<tr>
<th></th>
<th>Depth</th>
<th>Chl a</th>
<th>SST SD</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SST mean$^2$</strong></td>
<td>8.9</td>
<td>51.8</td>
<td>48.3</td>
<td>47.6</td>
</tr>
<tr>
<td><strong>SST mean</strong></td>
<td>3.6</td>
<td>17.9</td>
<td>-</td>
<td>16.3</td>
</tr>
<tr>
<td><strong>SST SD</strong></td>
<td>4.8</td>
<td>5.2</td>
<td>3.8</td>
<td>8.4</td>
</tr>
<tr>
<td><strong>Best alone</strong></td>
<td>shore</td>
<td>Chl a</td>
<td>SST SD</td>
<td>Depth</td>
</tr>
<tr>
<td><strong>Worst without</strong></td>
<td>shore</td>
<td>Chl a</td>
<td>SST SD</td>
<td>Depth</td>
</tr>
</tbody>
</table>

<p>| AUC | 0.710 | 0.673 | 0.672 | 0.676 |</p>
<table>
<thead>
<tr>
<th></th>
<th>Depth</th>
<th>Chl a</th>
<th>SST mean</th>
<th>Depth</th>
<th>Chl a</th>
<th>SST mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best alone</td>
<td>28.3</td>
<td>33.2</td>
<td>2.3</td>
<td>54.7</td>
<td>47.4</td>
<td>41.3</td>
</tr>
<tr>
<td>Worst without</td>
<td>39.8</td>
<td>41.6</td>
<td>9.6</td>
<td>20.4</td>
<td>23.7</td>
<td>17.8</td>
</tr>
<tr>
<td>AUC</td>
<td>0.928</td>
<td>0.837</td>
<td>0.858</td>
<td>0.843</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*background – auto features*

<table>
<thead>
<tr>
<th></th>
<th>Depth</th>
<th>Chl a</th>
<th>Chl a</th>
<th>SST mean</th>
<th>Chl a</th>
<th>Chl a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best alone</td>
<td>21.2</td>
<td>2.4</td>
<td>0.6</td>
<td>0.9</td>
<td>2.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Worst without</td>
<td>35.8</td>
<td>35.9</td>
<td>8.8</td>
<td>8</td>
<td>2.6</td>
<td>0.5</td>
</tr>
<tr>
<td>AUC</td>
<td>0.961</td>
<td>0.928</td>
<td>0.956</td>
<td>0.928</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure B1: Bathymetry of the Indian Ocean as per the General Bathymetric Chart of the Oceans (1 minute grid ~ 1.8 km).

Figure B2: Seasonal chlorophyll a concentration (Chl a) obtained from averaged of weekly SeaWiFS satellite composites at a 9 km spatial resolution from 1997 to the end of 2007.
Figure B3: 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 B4: 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 B5: Moran’s I plots showing the reduction in spatial autocorrelation in the GLM residuals. When a random effect was included to the models, Moran’s I is shown in the y axis and the lags in the x axis. Results for only one technique (tuna) in autumn and spring (when more presence data were available) are shown due to space constraints. Model 1: Presence ~ Slope; Model 2: Presence ~ Depth, Model 3: Presence ~ Shore, Model 4: Presence ~ SST variables*, Model 5: Presence ~ Chl a, Model 6: Presence ~ Physical variables + SST variables*, Model 7: Presence ~ Physical variables + SST variables + Chl a.
Figure B6: 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 B7: Seasonal habitat suitability of whale sharks (Rhincodon typus, Smith 1828) in the Indian Ocean.

Maps show the GLM/GLMM ensemble result of three pseudo-absences generation techniques used per season in the top row and the best-performing (according to AUC) MaxEnt model in the bottom row. The 0.4 estimated probability of whale shark occurrence contour is shown for the ensemble maps for clarity. Black dots represent real whale shark locations as summarised by Rowat (2007); dot size is proportional to frequency of sightings.
Receiving operator characteristic curve / area under the curve (ROC/AUC)

To measure models performance we also calculated the receiving operator characteristic curve / area under the curve (ROC/AUC) and results are shown in Table B2. The trend arising when comparing the AUC results between each MaxEnt and GLM model is similar to the one I obtained previously when comparing \( \kappa \) – i.e., when using the random technique to generate pseudo-absences, GLM was scored higher by both measures (AUC and \( \kappa \)).

Table B2: Comparison of the Area Under the Curve (AUC) and Kappa results for the two modelling approaches used in chapter III: MAxEnt and GLM/GLMM for each of the pseudo-absence dataset generated method (P/A) and for each season: Autumn (Aut), Winter (Win), Spring (Spr) and Summer (Sum).

<table>
<thead>
<tr>
<th>P/A</th>
<th>Approach</th>
<th>Aut</th>
<th>Win</th>
<th>Spr</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random MaxEnt</td>
<td>Kappa</td>
<td>0.54</td>
<td>0.57</td>
<td>0.50</td>
<td>0.49</td>
</tr>
<tr>
<td>GLM</td>
<td>Kappa</td>
<td>0.79</td>
<td>0.51</td>
<td>0.59</td>
<td>0.51</td>
</tr>
<tr>
<td>AUC</td>
<td>0.72</td>
<td>0.70</td>
<td>0.69</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.87</td>
<td>0.61</td>
<td>0.79</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>IDW MaxEnt</td>
<td>Kappa</td>
<td>0.34</td>
<td>0.10</td>
<td>0.25</td>
<td>0.31</td>
</tr>
<tr>
<td>GLM</td>
<td>Kappa</td>
<td>0.35</td>
<td>0.0</td>
<td>0.33</td>
<td>0</td>
</tr>
<tr>
<td>AUC</td>
<td>0.63</td>
<td>0.57</td>
<td>0.63</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.82</td>
<td>0.72</td>
<td>0.80</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Tuna MaxEnt</td>
<td>Kappa</td>
<td>0.64</td>
<td>0.37</td>
<td>0.48</td>
<td>0.43</td>
</tr>
<tr>
<td>GLM</td>
<td>Kappa</td>
<td>0.68</td>
<td>0.47</td>
<td>0.51</td>
<td>0.56</td>
</tr>
<tr>
<td>AUC</td>
<td>0.90</td>
<td>0.75</td>
<td>0.79</td>
<td>0.77</td>
<td></td>
</tr>
</tbody>
</table>
Boyce index comparison between MaxEnt and GLM/GLMM models

We assessed the monotonic relationship (Spearman’s rank correlation) between habitat suitability classes estimated by the models and the proportion of evaluation points adjusted by the area covered (Boyce et al., 2002) by calculating the Boyce index (autumn only). According to this index (Table B3) the results among pseudo-absence selection techniques and among modelling tools are similar, showing that the predictions from all models considered are consistent with the presences distribution in the sampled dataset. Results for the GLM predictions were generally slightly higher.

Table B3: Boyce’s index for the three pseudo-absence generation techniques (P/A) for each modelling technique (MaxEnt and GLM) for autumn. Standard deviation was always < 0.001.

<table>
<thead>
<tr>
<th>P/A</th>
<th>Tool</th>
<th>Boyce index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>MaxEnt</td>
<td>0.827</td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>0.845</td>
</tr>
<tr>
<td>IDW</td>
<td>MaxEnt</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>0.911</td>
</tr>
<tr>
<td>Tuna</td>
<td>MaxEnt</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>0.952</td>
</tr>
</tbody>
</table>
BIOMOD

We set the models according to the modelling summary shown in Table B4, and details on the best model selected according to True Skill Statistics and Receiver Operator Curve are shown in Table B5. The variable importance ranking for each model considered in BIOMOD is detailed in Table B6.

Table B4: BIOMOD modeling summary.

Modeling whale shark distribution in the Indian Ocean, using the same dataset used in Chapter III. ANN – Artificial Neural Networks; CTA – Classification Tree Analysis; GAM – Generalised Additive Models; GBM – Generalised Boosted Models (or BRT – Boosted Regression Trees); GLM – Generalised Linear Models; MARS – Multivariate Adaptive Regression Splines; FDA – Flexible Discriminate Analysis; RF – Random Forest.

<table>
<thead>
<tr>
<th>Number of species modelled</th>
<th>1 (whale shark)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical variables</td>
<td>Chla_av.In</td>
</tr>
<tr>
<td></td>
<td>Dist2Shore.In</td>
</tr>
<tr>
<td></td>
<td>Depth.c</td>
</tr>
<tr>
<td></td>
<td>Slope.In</td>
</tr>
<tr>
<td></td>
<td>SST_av</td>
</tr>
<tr>
<td></td>
<td>SST_sd</td>
</tr>
<tr>
<td></td>
<td>logEff (note: logEff could not be included as an offset within BIOMOD)</td>
</tr>
<tr>
<td>Number of evaluation repetitions</td>
<td>3</td>
</tr>
<tr>
<td>Number of pseudo-absences runs</td>
<td>2</td>
</tr>
<tr>
<td>Models selected</td>
<td>ANN, CTA, GAM, GBM, GLM, MARS, FDA, RF, SRE</td>
</tr>
<tr>
<td>Total number of model runs</td>
<td>72</td>
</tr>
</tbody>
</table>
Table B5: Best model selection in BIOMOD

According to True Skill Statistic (TSS) and Receiver Operator Curve (ROC) for each pseudo-absence dataset (PA), and each repetition considered (Rep). Random Forests were not considered for this classification. Bold italic letters indicate when GLM (Generalised Linear Models) was ranked as “best model” within the model runs considered (i.e., only 2 PA datasets). Generally GBM (Generalised Boosted Models) were ranked higher. CTA (Classification Tree Analysis) was also ranked higher in the first repetition of the run with the second set of pseudo-absences.

<table>
<thead>
<tr>
<th></th>
<th>Best Model</th>
<th>Cross-validation</th>
<th>Independent data</th>
<th>Total score</th>
<th>Cut off</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA1</td>
<td>TSS</td>
<td>0.859</td>
<td>0.864</td>
<td>0.888</td>
<td>436.5</td>
<td>96.1</td>
<td>92.7</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.977</td>
<td>0.973</td>
<td>0.982</td>
<td>532.9</td>
<td>94.3</td>
<td>94.3</td>
</tr>
<tr>
<td>PA1</td>
<td>TSS</td>
<td>0.863</td>
<td>0.860</td>
<td>0.887</td>
<td>494.2</td>
<td>95.2</td>
<td>93.5</td>
</tr>
<tr>
<td>Rep 1</td>
<td>ROC</td>
<td>0.975</td>
<td>0.973</td>
<td>0.981</td>
<td>539.3</td>
<td>94.0</td>
<td>94.0</td>
</tr>
<tr>
<td>PA1</td>
<td>TSS</td>
<td>GBM</td>
<td>0.841</td>
<td>0.840</td>
<td>0.853</td>
<td>658.7</td>
<td>91.5</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.974</td>
<td>0.972</td>
<td>0.981</td>
<td>496.4</td>
<td>93.3</td>
<td>93.4</td>
</tr>
<tr>
<td>PA1</td>
<td>TSS</td>
<td>GBM</td>
<td>0.876</td>
<td>0.857</td>
<td>0.886</td>
<td>502.9</td>
<td>94.6</td>
</tr>
<tr>
<td>Rep 3</td>
<td>ROC</td>
<td>0.981</td>
<td>0.973</td>
<td>0.982</td>
<td>510.3</td>
<td>94.1</td>
<td>94.2</td>
</tr>
<tr>
<td>PA2</td>
<td>TSS</td>
<td>GBM</td>
<td>0.832</td>
<td>0.865</td>
<td>0.883</td>
<td>482.0</td>
<td>96.3</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.961</td>
<td>0.975</td>
<td>0.980</td>
<td>547.8</td>
<td>93.8</td>
<td>93.9</td>
</tr>
<tr>
<td>PA2</td>
<td>TSS</td>
<td>CTA</td>
<td>0.830</td>
<td>0.865</td>
<td>0.887</td>
<td>320.0</td>
<td>94.7</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.951</td>
<td>0.973</td>
<td>0.977</td>
<td>570.5</td>
<td>92.7</td>
<td>92.8</td>
</tr>
<tr>
<td>PA2</td>
<td>TSS</td>
<td>GBM</td>
<td>0.848</td>
<td>0.863</td>
<td>0.881</td>
<td>531.6</td>
<td>95.0</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.964</td>
<td>0.974</td>
<td>0.979</td>
<td>560.3</td>
<td>93.3</td>
<td>93.3</td>
</tr>
<tr>
<td>PA3</td>
<td>TSS</td>
<td>GBM</td>
<td>0.851</td>
<td>0.867</td>
<td>0.879</td>
<td>546.4</td>
<td>95.0</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.968</td>
<td>0.975</td>
<td>0.979</td>
<td>580.8</td>
<td>93.5</td>
<td>93.6</td>
</tr>
</tbody>
</table>

The GLM model formula considered within the BIOMOD platform for each of the pseudo-absence dataset were as follow:

1) Presence ~ logEff + SST_av^3 + Depth.c + Depth.c^3 + Chla_av.ln^3 + Chla_av.ln + SST_sd^3 + Slope.ln + SST_av + SST_av^2

2) Presence ~ logEff + SST_av^3 + SST_av + Depth.c^3 + Depth.c + Chla_av.ln^3 + Chla_av.ln + Slope.ln + SST_av^2 + Dist2Shore.ln^3 + Dist2Shore.ln^2 + logEff^2 + logEff^3
Note that the model formulas differ in each run with different pseudo-absences considered.

The BIOMOD ensemble result for probability of whale shark distribution during autumn in the Indian Ocean according to the models considered is depicted in Figure B8, and it shows good agreement with the maps obtained both in Chapter III and Chapter IV (after re-fitting the model for autumn).

Table B6: Ranking of variable importance for each model considered within the BIOMOD platform.

Effort resulted in the variables with highest importance in all models, as expected. Note that this variable could not have been included as an offset in BIOMOD. The models developed throughout the thesis aim at understanding which variables (other than fishing effort) can be used to explain the occurrence of whale sharks. Following effort, depth, SST_{av} and Chl \textit{a}, were the variables with the highest importance.

ANN – Artificial Neural Networks; CTA – Classification Tree Analysis; GAM – Generalised Additive Models; GBM – Generalised Boosted Models (or BRT – Boosted Regression Trees); GLM – Generalised Linear Models; MARS – Multivariate Adaptive Regression Splines; FDA – Flexible Discriminate Analysis; RF – Random Forest. Chl \textit{a} – mean concentration of chlorophyll \textit{a}; Shore – distance to shore (logarithm); Depth – centred depth; Slope – logarithm of slope; SST_{av} – mean sea surface temperature; SST_{sd} – sea surface temperature standard deviation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Chl \textit{a}</th>
<th>Shore</th>
<th>Depth</th>
<th>Slope</th>
<th>SST_{av}</th>
<th>SST_{sd}</th>
<th>Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.010</td>
<td>0.011</td>
<td>0.376</td>
<td>0.016</td>
<td>0.003</td>
<td>0.008</td>
<td>0.594</td>
</tr>
<tr>
<td>CTA</td>
<td>0.000</td>
<td>0.295</td>
<td>0.275</td>
<td>0.020</td>
<td>0.024</td>
<td>0.014</td>
<td>0.573</td>
</tr>
<tr>
<td>GAM</td>
<td>0.079</td>
<td>0.018</td>
<td>0.043</td>
<td>0.013</td>
<td>0.024</td>
<td>0.010</td>
<td>0.509</td>
</tr>
<tr>
<td>GBM</td>
<td>0.016</td>
<td>0.010</td>
<td>0.024</td>
<td>0.003</td>
<td>0.057</td>
<td>0.002</td>
<td>0.614</td>
</tr>
<tr>
<td>GLM</td>
<td>0.090</td>
<td>0.019</td>
<td>0.062</td>
<td>0.012</td>
<td>0.015</td>
<td>0.000</td>
<td>0.531</td>
</tr>
<tr>
<td>MARS</td>
<td>0.189</td>
<td>0.000</td>
<td>0.111</td>
<td>0.001</td>
<td>0.133</td>
<td>0.007</td>
<td>0.589</td>
</tr>
<tr>
<td>FDA</td>
<td>0.021</td>
<td>0.010</td>
<td>0.067</td>
<td>0.013</td>
<td>0.066</td>
<td>0.000</td>
<td>0.660</td>
</tr>
<tr>
<td>RF</td>
<td>0.106</td>
<td>0.035</td>
<td>0.114</td>
<td>0.013</td>
<td>0.128</td>
<td>0.044</td>
<td>0.485</td>
</tr>
<tr>
<td>SRE</td>
<td>0.048</td>
<td>0.023</td>
<td>0.024</td>
<td>0.015</td>
<td>0.021</td>
<td>0.023</td>
<td>0.031</td>
</tr>
</tbody>
</table>
Figure B8: Whale shark autumn distribution in the Indian Ocean as a result of the ensemble of models in BIOMOD.

Colour gradient reflects probability of whale shark occurrence with red indicating higher probability values. Only the area covered by the fisheries is shown in the maps. The ensemble result is calculated by averaging across all the models for each run. Shown are: (top left to right order) the presence locations used, and prob.mean and prob.mean.weighted – which use the mean and weighted mean of thresholds (respectively) to convert probabilities into presence/absence data; (bottom left to right order) median – used the median of the thresholds, and Kappa.mean and TSS.mean – use 0.5 as the threshold to distinguish between presence and absence.
BRT / GBM

We adapted the code from Elith et al. (2008) to run the BRT (Boosted Regression Trees) analysis with our data from the Indian Ocean (autumn season only). The tree complexity was set to 5, the learning rate was 0.005, and the bag fraction was considered at 0.5. Effort was included as an offset (logEff).

The maximum number of trees fitted was 1800, with the best number of trees defined at 1350. The evaluation procedure resulted in a ROC (Remote Operated Curve) of 0.956. The function "gbm.perspec" was used to check for pairwise interactions, and the results are shown in Table B7. 3D figures that assist interpreting the interactions found are depicted in Figure B9. The interaction values represent the contribution of the pairwise interaction to the predictive performance of the model (Williams et al., 2010).

The map for distribution of whale sharks obtained by BRT is similar to the maps obtained by the other modelling techniques used (Figure B10).

Table B7: Analysis of pairwise interactions between variables using BRT

<table>
<thead>
<tr>
<th></th>
<th>Chl a</th>
<th>Shore</th>
<th>Depth</th>
<th>Slope</th>
<th>SST_av</th>
<th>SST_sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl a</td>
<td>0</td>
<td>5.07</td>
<td>13.35</td>
<td>2.85</td>
<td>12.12</td>
<td>40.44</td>
</tr>
<tr>
<td>Shore</td>
<td>0</td>
<td>0</td>
<td><strong>124.73</strong></td>
<td>1.04</td>
<td>14.57</td>
<td>44.15</td>
</tr>
<tr>
<td>Depth</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.04</td>
<td><strong>71.99</strong></td>
<td>43.49</td>
</tr>
<tr>
<td>Slope</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13.24</td>
<td>10.48</td>
</tr>
<tr>
<td>SST_av</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td><strong>39.45</strong></td>
</tr>
<tr>
<td>SST_sd</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
When running the model using the GBM package, we used the model formula including all the predictors as linear terms, quadratic terms for both temperature and depth, and effort as an offset. This package allows for computation of the Friedman’s H-statistic through the function interact.gbm. The Friedman’s H statistic, which is returned by that function, provides an assessment of the relative strength of interaction effects in the models. The value returned by interact.gbm was $2.9 \times 10^{-14}$, confirming the low strength of interaction effects in the model.

Figure B9: 3D interpretation of the two top ranked pairwise interactions found with BRT.
Left: Interaction between distance to shore (logarithm; Dist2Shore.Ln) and centred depth (Depth.c). Right: Interaction between sea surface temperature (SST_av) and centred depth (Depth.c). The maximum values for the interactions are: 0.02 (left) and 0.06 (right).
Figure B10: Map with probability of whale shark occurrence in the Indian Ocean obtained by BRT.

Colour gradient reflects probability of whale shark occurrence, with red indicating higher probabilities. Only the area covered by the fisheries is shown in the map. Higher probabilities of occurrence are similar to the ones described previously in chapter 3.
APPENDIX C – Spatial and temporal predictions of decadal trends in Indian Ocean whale sharks

Model predictors (Step 1)

Here we re-fitted the whale shark distribution model for autumn (Sequeira et al., 2012), using generalised linear mixed-effects models estimated with a Laplace approximation to the likelihood (lmer function; Bates et al., 2011) rather than using a penalized quasi-likelihood approach (glmmPQL function in R; Venables & Ripley, 2002b) as we used in Sequeira et al. (2012). Using the lmer approach we can now obtain information criteria output to rank models. When re-running the autumn spatial model, we also used a different response variable for these models by increasing the pseudo-absence:presence ratio from 1:1 (Sequeira et al., 2012) to 1:10 to ensure that the variable coded as a random effect (spatially aggregated cells) contained at least one point for the analysis. We included an additional quadratic term for depth (bathymetry) because although whale sharks spend most of their time at the surface (Sleeman et al., 2010b) they frequently make sub-surface ‘dives’ (Wilson et al., 2006). Including this quadratic term tests the hypothesis that sightings occur mostly in locations where depth is within a specific range. Moreover, depth can also be viewed as a proxy for a set of environmental conditions differentiating the conditions encountered in shallow versus deep habitats (e.g., upwelling and primary productivity, hydrodynamics), which could play a role in influencing whale shark occurrence at the surface. We also standardized some explanatory variables used in Sequeira et al. (2012) to stabilize parameter estimation within the lmer function: we centred depth and mean sea surface temperature, and log-transformed mean chlorophyll a, distance to shore, slope
and effort (i.e., autumn mean value) (Table C1). In this model update, we have included effort (autumn mean value) as an offset term to account directly for the proportional increase in sightings with proportional increases in spatial fishing effort.

Table C1: Treatment given to each explanatory variable (standardization) used to obtained the spatial predictor (SpatialP) (Sequeira et al., 2012) for the first step of our temporal models; SST – sea surface temperature; Chl a – chlorophyll a.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment</th>
<th>Range</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>depth</td>
<td>centred</td>
<td>-3745.4 – 2444.4</td>
<td>0.00</td>
</tr>
<tr>
<td>slope</td>
<td>log</td>
<td>-5.21 – 3.43</td>
<td>-1.25</td>
</tr>
<tr>
<td>distance to shore</td>
<td>log</td>
<td>0.10 – 14.17</td>
<td>12.61</td>
</tr>
<tr>
<td>mean SST*</td>
<td>centred</td>
<td>-3.97 – 2.40</td>
<td>0.00</td>
</tr>
<tr>
<td>mean Chl a</td>
<td>log</td>
<td>-2.63 – 2.91</td>
<td>-1.93</td>
</tr>
<tr>
<td>effort</td>
<td>log (used as an offset)</td>
<td>-0.18 – 7.56</td>
<td>3.30</td>
</tr>
</tbody>
</table>
Figure C1: Purse-seine fishing effort (in days) during autumn for each year from 1991 to 2007.
Figure C2: Habitat suitability of whale sharks (*Rhincodon typus* Smith 1828) in the Indian Ocean during autumn.

Updated model from Sequeira et al. 2012 (see supplementary information in Appendix C for details). Areas with higher probability of whale shark sightings are similar to the ones previously described and show that mostly the western Indian Ocean – specifically the Mozambique Channel – is more suitable for whale sharks in autumn, although lower probabilities (around 0.3) were obtained when compared to Sequeira et al. (2012). Black dots represent known occurrences (Rowat, 2007), with size representing strength of occurrence (i.e., larger dots correspond to a higher number of expected sightings).
Figure C3: Correlation between climatic predictor variables.

Top panel: Scatter plot of matrices, showing: below the diagonal – the bivariate scatter plots with linear fits and correlation ellipses, and above the diagonal – the Pearson’s correlation results (top) and Spearman’s $\rho$ (bottom).
Figure C4: Partial effect of the Indian Ocean Dipole (IOD) on the probability of whale shark sightings.


Bottom row: Partial effect of Time after accounting for climatic contributions to whale shark sighting probability. Dotted lines represent 95 % confidence interval.
Prevalence test (down-weighting pseudo-absences)

We have re-run the temporal models after down-weighting the 100 absences (from 1 to 0.01) to get a 1:1 ratio of presences to absences (i.e., a presence:pseudo-absence ratio of 0.5), as suggested. Because the partial effect of time (Figure 16, in Chapter IV) is highly dependent on prevalence, the change in the presence to absence ratio resulted in a much more pronounced curve reflecting the dependency of occurrence with time. This is depicted in Figure C5.

![Figure C5: Partial effect of time on whale shark presence, using a presence:pseudo-absence ratio of 0.5 (prevalence).](image)

(a) on the log-odds scale showing the rate of change and (b) on the probability scale, showing the effect of time on the probability of whale shark presence. Dotted lines represent 95% confidence interval.

Although this change in prevalence increased the estimated percentage of deviance explained (71.1% for the best ranked model), it did not affect the model ranking based on w$\text{AIC}_c$ (Table 7, in Chapter IV).
APPENDIX D – Predicting current and future global distributions of whale sharks

Figure D1: Global sea surface temperature averages from April to June. Top panel: current temperatures as seasonal averages for 2002 – 2012 derived from MODIS-Aqua satellite. Bottom panel: increase in sea surface temperatures predicted by 2070. The climate change scenario for 2070 was obtained by adding the predicted increase calculated by an ensemble forecast of climate change derived from five global circulation models: CCSM-3, MRI-CGCM2.3.2, ECHAM5/MPI-OM, MIROC3.2 ( hires) and UKMO-HadCM3. These GCM were chosen based on their global skill in predicting recent SST temperatures (1981 – 2000). Model terminology follows CMIP3 Multi-Model Dataset Archive naming convention (www-pcmdi.llnl.gov/ipcc/about_ipcc.php), and baseline observed temperatures used were from 1995 (AVHRR Pathfinder).
BIOMOD – Global model

In the preliminary run in the BIOMOD platform, the models were set according to the summary shown in Table D1, and details on the top-ranked model according to true skill statistics (TSS) and receiver operator curve (ROC) are shown in Table D2. The variable importance ranking for each model considered in BIOMOD is detailed in Table D3.

Table D1: BIOMOD modelling summary for the runs with the global dataset.

Modeling whale shark distribution in the three major oceans using the dataset described in Chapter VI. Due to the complexity of the dataset and the computing time associated when running multiple models with large datasets, the preliminary run was made with only one evaluation run and one repetition of pseudo-absences selection. Fishing effort (logEff) was included in this run as a ‘weight’. ANN – artificial neural networks; CTA – classification tree analysis; GAM – generalised additive models; GBM – generalised boosted models (or BRT – boosted regression trees); GLM – generalised linear models; MARS – multivariate adaptive regression splines; FDA – flexible discriminate analysis; RF – random forest.

<table>
<thead>
<tr>
<th>Number of species modelled</th>
<th>1 (whale shark)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical variables</td>
<td>Chla_av.In</td>
</tr>
<tr>
<td></td>
<td>Dist2Shore.In</td>
</tr>
<tr>
<td></td>
<td>Depth.c</td>
</tr>
<tr>
<td></td>
<td>Slope.In</td>
</tr>
<tr>
<td></td>
<td>SST_av</td>
</tr>
<tr>
<td></td>
<td>SST_sd</td>
</tr>
<tr>
<td>Factorial variables</td>
<td>Ocean</td>
</tr>
<tr>
<td>Number of evaluation repetitions</td>
<td>1</td>
</tr>
<tr>
<td>Number of pseudo-absences runs</td>
<td>1</td>
</tr>
<tr>
<td>Models selected</td>
<td>ANN, CTA, GAM, GBM, GLM, FDA, RF (MARS and SRE were automatically excluded due to the inclusion of a factorial variable)</td>
</tr>
<tr>
<td>Total number of model runs</td>
<td>14</td>
</tr>
</tbody>
</table>
Table D2: Model selection in BIOMOD

According to true skill statistic (TSS) and receiver operator curve (ROC) for the pseudo-absence dataset generated and the one replicate considered (Rep). Random forests were not considered for this classification, and GBM (generalised boosted models) were ranked higher.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cross-validation</th>
<th>Independent data</th>
<th>Total score</th>
<th>Cut off</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA1 TSS</td>
<td>GBM</td>
<td>0.732</td>
<td>-</td>
<td>0.742</td>
<td>559.3</td>
<td>85.2</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.925</td>
<td>-</td>
<td>0.939</td>
<td>519.9</td>
<td>86.7</td>
</tr>
<tr>
<td>PA1 TSS</td>
<td>GBM</td>
<td>0.732</td>
<td>-</td>
<td>0.739</td>
<td>535.1</td>
<td>85.7</td>
</tr>
<tr>
<td>ROC</td>
<td>GBM</td>
<td>0.925</td>
<td>-</td>
<td>0.938</td>
<td>506.5</td>
<td>86.7</td>
</tr>
</tbody>
</table>

Table D3: Ranking of variable importance for the global model considered within the BIOMOD platform.

Effort resulted in the variables with highest importance in all models, as expected. Note that this variable could not have been included as an offset in BIOMOD. The models developed throughout the thesis aim at understanding which variables (other than fishing effort) can be used to explain the occurrence of whale sharks. Following effort, depth, SST_av and Chl a, were the variables with the highest importance.

ANN – artificial neural networks; CTA – classification tree analysis; GAM – generalised additive models; GBM – generalised boosted models (or BRT – boosted regression trees); GLM – generalised linear models; MARS – multivariate adaptive regression splines; FDA – flexible discriminate analysis; RF – random forest.

Chl a – mean concentration of chlorophyll a; Shore – distance to shore (logarithm); Depth – centred depth; Slope – logarithm of slope; SST_av – mean sea surface temperature; SST_sd – sea surface temperature standard deviation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Ocean</th>
<th>Chl a</th>
<th>Shore</th>
<th>Depth</th>
<th>Slope</th>
<th>SST_av</th>
<th>SST_sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.364</td>
<td>0.704</td>
<td>0.363</td>
<td>0.187</td>
<td>0.045</td>
<td>0.682</td>
<td>0.284</td>
</tr>
<tr>
<td>CTA</td>
<td>0.000</td>
<td>0.472</td>
<td>0.425</td>
<td>0.193</td>
<td>0.005</td>
<td>0.491</td>
<td>0.000</td>
</tr>
<tr>
<td>GAM</td>
<td>0.101</td>
<td>0.053</td>
<td>0.424</td>
<td>0.093</td>
<td>0.047</td>
<td>0.455</td>
<td>0.002</td>
</tr>
<tr>
<td>GBM</td>
<td>0.001</td>
<td>0.251</td>
<td>0.376</td>
<td>0.076</td>
<td>0.001</td>
<td>0.244</td>
<td>0.000</td>
</tr>
<tr>
<td>GLM</td>
<td>0.161</td>
<td>0.121</td>
<td>0.377</td>
<td>0.141</td>
<td>0.059</td>
<td>0.620</td>
<td>1.313</td>
</tr>
<tr>
<td>FDA</td>
<td>0.000</td>
<td>0.178</td>
<td>0.394</td>
<td>0.141</td>
<td>0.059</td>
<td>0.620</td>
<td>1.313</td>
</tr>
<tr>
<td>RF</td>
<td>0.012</td>
<td>0.427</td>
<td>0.331</td>
<td>0.191</td>
<td>0.046</td>
<td>0.131</td>
<td>0.135</td>
</tr>
</tbody>
</table>

The GLM formula considered within the BIOMOD platform was:

\[
\text{Presence} \sim \text{Dist2Shore,ln}^3 + \text{Dist2Shore,ln} + \text{Chla_av,ln}^3 + \text{Depth,ln} +  \\
\text{Depth,ln}^3 + \text{SST_sd}^3 + \text{Chl a} + \text{Ocean} + \text{Slope,ln} + \text{Slope,ln}^3 + \text{Slope,ln}^2 + 
\]
Chla$_\text{av}.\ln^2 + \text{Depth}.c^2 + \text{Chla}_\text{av}.\ln + \text{SST}_\text{sd}^2 + \text{SST}_\text{av}.c + \\
\text{Dist2Shore}.\ln^2$

where: Chla$_\text{av}.\ln$ – logarith of mean concentration of chlorophyll $a$; Dist2Shore – distance to shore (logarithm); Depth.c – centred depth; Slope.$\ln$ – logarithm of slope; SST$_\text{av}.c$ – centred mean sea surface temperature; SST$_\text{sd}$ – sea surface temperature standard deviation. Effort (logEff) was included as a weight in this preliminary run.

The BIOMOD ensemble result for the preliminary model run is depicted in Figure D2, showing good agreement with the map obtained in Chapter VI (for the area covered by the tuna-purse seine fisheries in the three oceans), with similar areas with habitat suitability.

![BIOMOD ensemble prediction](image1)

![GLMM prediction](image2)

Figure D2: Comparison of GLMM and BIOMOD ensemble prediction for the area covered by the tuna fisheries (mean probability).

The ensemble result is calculated by averaging across all the models used in the BIO)MOD run.
ADDENDUM – Thesis amendments

Outcome of thesis examination

Letter from Examiner 1

December 6th 2012
Examiner’s Report for PhD Candidate Ana Martin Sequeira – J.K. Baum, U. Victoria

Summary
This is an interesting and timely PhD thesis, which makes an important contribution to marine ecology and conservation. I enjoyed reading it. The research is of a high standard, and three of the five main research chapters are already published in the peer-reviewed literature.

Overall, I find Chapter 3, the investigation of ‘Ocean-scale prediction of whale shark distribution’, which has already been published (in Diversity and Distributions), to be the strongest piece of research. Understanding the spatial and temporal drivers of whale shark abundance is an enormous challenge given the vast scales involved and paucity of data covering these required scales. This thesis focuses on tackling various aspects of this challenge; here the candidate does so masterfully, employing multiple modeling approaches and carefully comparing outcomes. I strongly recommend the candidate adopt this approach for Chapters 4 to 6, and provide more detailed comments to this effect below.

Technically, the thesis appears to be fairly sound. Given that the thesis is based entirely on synthesizing and/or modeling already-collected data, this is of utmost importance. Comments on some technical weaknesses are below.

Conceptually, however, the thesis is rather weak. I would have liked to see the candidate link her research to the relevant ecological theory and clearly explicate her hypotheses relating to each model covariate. Much of the thesis is predicated on: 1) the idea that the whale shark may be comprised of a single global meta-population, with a sub-population or multiple sub-populations in each ocean basin, yet the candidate does not draw upon (or demonstrate knowledge) of the vast and relevant ecological theory pertaining to metapopulations; 2) the idea that climate change will significantly impact whale sharks by altering their distributions. Here, I would have liked to see a deeper discussion of the potential implications of climate change, including the potential for this species to utilize different depths, the potential for altered predator-prey dynamics, and consideration of the possibility of beneficial effects of climate change (not all effects will be negative!). The thesis chapters which are not yet published would benefit from these additions.

I suggest changing the overall thesis title to:
‘Regional and Global Distribution Models for Whale Sharks (Assessing Occurrence Trends of a Highly Migratory Marine Species)’
or
‘Ocean-basin Scale and Global Distribution Models for Whale Sharks (Assessing Occurrence Trends of a Highly Migratory Marine Species)’
to more accurately reflect the thesis contents; only Chapters 2 and 6 are global in scale.
The thesis is reasonably well-written, but would benefit from a careful final proofreading and grammatical check throughout by a native English speaker. Mistakes are minor but persistent. I have noted a few of these below in my detailed comments.

Below are my detailed comments about each chapter. I sincerely hope that these are helpful for strengthening your thesis (and related publications) through the suggested minor revisions. Great work!

With best wishes,
Julia K. Baum

Detailed Comments

Summary
- remove 'temporal' from first line, (or add 'spatial'),
- suggest to reword 'life cycle' here and throughout; sounds awkward. E.g. 'life stages'

Chapter 1 –
- The first paragraph is unnecessarily broad. Instead, you might lead with some of the relevant ecological theory. I would definitely like to see some discussion of the ecological theory, questions and hypotheses before diving into the methods (even in a methods driven thesis);
- p. 28, 2nd line – 'and might change' should be 'how it might change'
- p. 31 – tourism industry – If you mean whale sharks, then be specific. Otherwise this is incorrect i.e. do not generalize about 400+ shark species, when only a handful are relevant to any tourism industry;
- p. 31 – References here seem a bit dated (e.g. see recent report by Jessica Meeuwis's group on the value of sharks to tourism (from Palau?), as well as a key one on shark life histories by Luis Lucifora, Veronica Garcia and Jeffrey Hutchings from 2012 Ecology;
- p. 31/32 – Does Essington et al. 2002 really demonstrate that sharks play a keystone role (this is not what I recall from the pair of Essington / Schindler CNP Ecopath models). Other more recent refs you could cite here?
- p. 32 'study case' should be 'case study' throughout thesis (e.g also on p. 38 etc)
- p. 32 – The statement that polynuclitherms are a challenging case study is weakly argued. Many such species are the subject of climate modeling studies. Could you not draw upon some knowledge from the terrestrial literature to help inform this statement?
- p. 35 top paragraph: Please clarify here why it is that tuna aggregate near whale sharks - because they share prey or because the tuna prey upon species with which the whale sharks share prey?
- p. 35 'economic importance worldwide' This is an overstatement. I think you mean in 'select coastal locations around the world'
- p. 36 'in the near future' Is this specifically stated in the IUCN Red List Guidelines? If so, cite them. If not, and you infer it, then please remove.
- p. 38 Strikes me as odd to refer to the world's largest fish as 'elusive'
Chapter 2
-Abstract, p. 45 – half way down: "generates a model describing... are part of a single, global meta-population", change to "might be part of ...."
-Introduction p. 46 - 'Being large and docile...' Sentence does not follow logically from this beginning.

-This chapter provides an interesting overview of whale shark research. I don't, however, feel that here (e.g. pages 71-72) or elsewhere in the thesis (Chapter 6) you have made a convincing argument as to why we should be particularly concerned about the potential impacts of climate change on this wide-ranging, highly mobile marine species.

-p. 48 – life cycle (Do you mean e.g. juveniles vs. adults) – think you mean 'life stages'
-p. 48/49 – I'm also surprised by the argument that tagging data couldn't be used to estimate 'broad-scale migratory patterns' in this species. This is exactly what the TOPP (www.topp.org) program achieved over the past decade + for many large, highly mobile marine species.

Chapter 3
Excellent work! Well-written and beautifully executed.

Chapter 4
Already published in MEPS.

Chapters 5 and 6
These two chapters would benefit from implementation of an additional (and different) modeling approach (e.g. models with a truncated distribution or boosted regression trees) and/or some sensitivity analyses (e.g. re: generation of the pseudo-absences in for the Atlantic and Pacific Ocean data sets as in Chapter 3).

At minimum, I would like the candidate to include some discussion of the potential biases and limitations arising from the data used and/or the modeling approaches undertaken in these two chapters. Such admissions and discussions would strengthen (not weaken) these chapters and their related manuscripts intended for publication in the peer-reviewed literature. The candidate does an excellent job of reviewing these in Chapter 7, her General Discussion, but this chapter will not be published in the primary literature, and this material is critical to each of Chapters 5 and 6 in order for readers to be able to evaluate the research.

In each of these chapters, please also:
- show the equations for the models being used, so that the reader can more clearly see what is being done;
- provide some more detail on the exact nature of the logbook records of whale shark occurrences, including quantitative information or comment on the per cent of logbooks in which whale shark occurrences were consistently, or ever, recorded (i.e. does the entire fleet report whale sharks?). How might this alter your interpretation of the zeroes (i.e. do zeroes only reflect true absences and failure to detect whale sharks, or do zeroes reflect these two cases as well as 'failure to report' whale sharks) How might this alter how you treat 'effort' in your models? OR Are these data not provided on
a trip by trip, or vessel by vessel basis, but merely aggregated as per the spatial precision described (e.g. p. 139). These types of detail need to be made clear for readers.

Clearly explicate the relationship between the sightings and effort data (i.e. the fact that the two are not paired, and hence candidate cannot implement traditional 'CPUE' or 'SPUE' models, in which the two are available at the same scale);

Rather than modeling only certain components of the data (i.e. April to June), I would prefer the candidate to model all of the data, including seasonal covariate(s) in the model (e.g. modeling season as continuous variable e.g. as in Baum & Blanchard 2010 Fisheries Research);

Explain more clearly the justification for modeling time/year as both a fixed and a random effect;

Discussion p. 150 'we hypothesize that whale shark occurrence might be asynchronously cyclic across oceans.' – This statement is speculative and unconvincing. There is a huge difference between a small number of individuals moving between ocean basins to ensure 'low genetic differentiation of whale sharks sampled in different oceans' versus entire (sub)-populations moving across ocean basins to produce asynchronous cycles across oceans.


Chapter 7 – General Discussion

-p. 183 'partialled' –Correct.

-p. 183-184: I urge you to carefully reconsider this recommendation. Small fluctuations in relative abundance, of less than the span of 1 generation, do not fit the IUCN Red List Criterion (A1 or A2). Such changes could be spurious or cyclical, but clearly should not be extrapolated to the required period of three generations. We, in the conservation community, need to be vigilant about not 'crying wolf' about every observed or inferred population fluctuation.

-p. 184 'strongly suggest a connection' This overstates what your evidence shows. Please reword accordingly.

-p. 188 'preserve' Please distinguish between 'preserve' and 'conserve'. Do you really mean preserve?

-p. 188. What about the Convention on Migratory Species?

-p. 189 – 195 This is an excellent discussion of potential biases and limitations, and should be worked into each of the relevant chapters (or at minimum the relevant manuscripts for submission to peer-reviewed journals), as previously discussed.

Chapter 7 – Nicely wrapped up.
Letter from Examiner 2

Report on Ana Martin Sequeira’s PDH thesis

The PHD is generally well written, clear and well-illustrated. The figures are of very good quality and the legends informative. I really appreciated reading the overall thesis, especially given my strong ignorance but also strong interest on whale sharks. The Introduction and second chapter were thus very interesting for me and get the overall idea about what we know and do not know on this species.

Three papers are already in press and two in preparation. I have made more comments on the two papers in preparation to help improving the text and methods. Having three papers published demonstrates the quality of the research undertaken by Ms Sequeira. I am quite satisfied by her work and suggest acceptance of the thesis with minor corrections notably on Chap V. To me, the PhD will not have to be re-evaluated. I trust Ms Sequeira will be able to address most of my comments.

Find below few comments on the specific chapters:

Chap 3 (MS published with DDI)

I do not have much to say given the paper is already published in a respectful journal. The literature cover is relatively good although some key papers are ignored.

The methods are not very well-explained to me. There are even some misconceptions. Maxent is NOT a presence-only algorithm, not more than any GLM for instance. I am not entirely convinced by the classification of models as well. Maxent and BRT are not a machine-learning method. BRT are only classification trees in a boosting framework.

Why using Kappa for GLMs? The use of Kappa should be stopped now because it is strongly biased with prevalence, which is definitely the case here given the use of pseudo-absence. Tru Skill Statistics or Boyce Index should be preferred.

There is no justification on the use of Maxent or GLM? Why these two approaches? Only because the authors like GLM and because Maxent is over-used?

Why not including an autoregressive component to Maxent?

I do not understand why the ROC curve is not estimated for GLM/GLMM instead of keeping only with the deviance. It will have been easier to compare with Maxent. Explained deviance measure the explanatory power of a model, not is predictive accuracy. This is a quite different.

There is a problem of rows (the third row) in Table 5

Figure 12 is very nice and quite informative.

Discussion: end of Page 104: The Boyce Index allows to discount for pseudo-absence and to compare different models or repetitions with the same set of presence.

Line 105: Something is missing in the discussion about a missing factor. I do not buy the fact that
spatial autocorrelation bias the models. Given most environmental variables are spatially auto-
correlated, the fact that the residuals are also spatially auto-correlated strongly suggest a missing
element or a missing variable. What about dispersal, what about another variable? It could also be
because of interactions between factors that are not taking into account here.

Line 106: Maxent can use interactions between variables, this is why the set of important
variables are not the same. It comes back to my previous comment that GLM should also include
some interactions between crucial variables.

CHAP 4 (MS in press with Marine Ecology Progress Series)

Page 116-117: To artificially increase the prevalence of the species and thus the generated
probability of presence, this is generally recommended to down-weight the absence to have the
same weight than the presence.

Page 120: there is a mistake. Year is the random factor, and not time. (modify the order in the
parenthesis). Else this is contradictory with Table 1.

CHAP 5 CHAPTER V. Inter-ocean asynchrony in whale shark occurrence (MS in prep)

This is an interesting complementary paper to the previous one. It was somehow difficult to
pinpoint the differences between the two (excepted the two oceans).
My main concern relates to the Method part, which is not always understandable for non-
aficionados. For instance, the Fourier transformation is absolutely not explained. You cannot
guess that all readers will know what is it about and how it works. More explanations are needed
here. Explain also why you chose this framework.

I am a bit puzzled by the use of AICw without further attention (the same applied to the previous
chapter as well). What is nice with this framework is also to extract a weight of evidence of each
variable from the list of models.
Indeed, although the weight of evidence in favour of one given model is of interest (it allows to
highlight the potential problem of stepwise if several models have similar $\Delta I$ (close to 0), the
general interest for ecologists is a step further. By analogy to weight of evidence in favour of
model, the weight of evidence of each predictor ($wpi$) can be simply estimated as the sum of the
model AICs weights ($wpi$) over all models in which the selected predictor appeared. The
predictor with the highest $wpi$ (the closest to 1) gets the highest weight of evidence to explain the
response variable (the highest relative importance).
This approach based on set of multiple models is far more robust than inferring variable
importance based on a single selected model.

Similarly, instead of keeping one model at the end, I would strongly recommend to use a
weighted-average. These predictions are simply the weighted-averaged predictions of all sub-
models. Sub-models with low weight of evidence will have basically not power, where sub-
models with similar AIC will gave even weight.
\[ P = \sum_{i} w_i P_i \]

Where \( w_i P_i \) is the prediction from sub-model \( i \), weighted by the weight of evidence in favour of this sub-model.

Associated with this weighted mean, a weighted variance or standard deviation can be easily estimated which allows a useful estimation of the uncertainty associated with the modelling approaches. Unconditional interval of confidence can then be drawn.


CHAPTER VI. Predicting current and future global distributions of whale sharks

The beginning of the Introduction is more or less a copy-paste from the other articles submitted or published. I understand this not always easy to find inspiration but please try to change the sentence and the associated references (This whole sentence "Species distribution models have been used widely to estimate habitat requirements and range sizes (Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000; Hirzel et al., 2002; Phillips et al., 2009). Despite their diverse use, until recently only a handful of these models had been developed for the marine environment (Elith & ...") appears several times in the PhD thesis.

This whole part is not clear: "Despite the known limitations of these models, such as broad spatial scales of prediction (Hannah et al., 2002; Seo et al., 2009) and considerable variation among climate forecasts (Arévalo et al., 2005; Hannah et al., 2002; Knutti et al., 2010), their biological relevance can be improved by down-scaling and coupling to ecological processes (Arévalo et al., 2005; Fordham et al., 2011; Fordham et al., 2012; Hannah et al., 2002)."

There are plenty of more relevant references concerning uncertainty, lack of mechanisms etc. Here is the a non-exhaustive list:


This also not very clear what the authors want to explain. I do not understand "such as broad spatial scales of prediction (Hannah et al., 2002; Seo et al., 2009)". I do not see what is the limitation here. Try to be clearer and more precise.

I have a general problem with the overall methodology here. The authors state that there are known limitations and especially variations when projecting into the future, but they use only one approach without controlling whether the results are strongly biased toward the selected approach. Although I like GLMM, I do not think they are the most suitable tools for such a risk assessment and that I suggest a careful examination of the uncertainty if such results have to be used.

I would recommend the authors to have a look at packages such as dismo or biomod2 in R to allow running a set of distribution models and then having a careful examination of the error, confidence intervals. The biomod2 package, which I know most, allows such an approach. In a climate change context, this is something we should always to given the potential use of outputs. Ecologists find logical to use a set of climate change models, but still doubt they have to do the same with ecological models. This is out of sense.

I am also concerned about the use of spatial model in a climate change context. The spatial autocorrelation is certainly due to either dispersal or a missing variable. Both of them are not explicitly modelled here and there is no way to know how they will change into the future. This is impossible to "project" a spatial component and can only lead to spurious projection.

I would strongly reconsider the spatial component here, and rather try to build a simplistic diffusion model to account for dispersal.

Fig 26: This is surprising to see some red crosses (around Japan and Americas) that are supposedly in habitats suitability close to 0. Does it mean that the models fail to detect those zones? How would you explain the high suitability in Atlantic whereas very few known aggregations have been spotted?

The discussion misses a critical assessment on the quality and reliability of the models in respect to uncertainty (what about plotting the confidence interval or coefficient of variation?), robustness and potential use of such projections.

CHAPTER VII. GENERAL DISCUSSION

Page 183: Although I agree that the PhD is a very interesting and well-conducted one, I found the first paragraph quite over-stated. Cheung and colleagues have developed far more advanced species distribution models based on Metabolic Theory, physiology and trophic interactions.

I found the conclusion very well written and summarizing well all advances and results of the PhD thesis.
GENERAL COMMENTS:

Examiner 1

... This thesis focuses on tackling various aspects of this challenge; here the candidate does so masterfully, employing multiple modelling approaches and carefully comparing outcomes. I strongly recommend the candidate adopt this approach for Chapters 4 to 6, and provide more detailed comments to this effect below.

Conceptually, however, the thesis is rather weak. I would have liked to see the candidate link her research to the relevant ecological theory and clearly explicate her hypotheses relating to each model covariate. Much of the thesis is predicated on: 1) the idea that the whale shark may be comprised of a single global meta-population, with a sub-population or multiple subpopulations in each ocean basin, yet the candidate does not draw upon (or demonstrate knowledge) of the vast and relevant ecological theory pertaining to metapopulations; 2) the idea that climate change will significantly impact whale sharks by altering their distributions. Here, I would have liked to see a deeper discussion of the potential implications of climate change, including the potential for this species to utilize different depths, the potential for altered predator-prey dynamics, and consideration of the possibility of beneficial effects of climate change (not all effects will be negative!). The thesis chapters which are not yet published would benefit from these additions.
I suggest changing the overall thesis title to:

'Regional and Global Distribution Models for Whale Sharks

(Assessing Occurrence Trends of a Highly Migratory Marine Species')

or

'Ocean-basin Scale and Global Distribution Models for Whale Sharks

(Assessing Occurrence Trends of a Highly Migratory Marine Species)'

to more accurately reflect the thesis contents; only Chapters 2 and 6 are global in scale.”

The thesis is reasonably well-written, but would benefit from a careful final proofreading and grammatical check throughout by a native English speaker. Mistakes are minor but persistent. I have noted a few of these below in my detailed comments.

The title reflects the main aim of this PhD, which is conceptualised in Chapter II, tested throughout Chapters III-V in different oceans, and achieved in Chapter VI. For this reason I think the current title (below) is more adequate.

'Global distribution models for whale sharks: assessing occurrence trends of highly migratory marine species'.

I address the other suggested amendments within each chapter as specified below.
SUMMARY

Examiner 1

remove 'temporal' from first line, (or add 'spatial),

Changed.

suggest to reword 'life cycle' here and throughout; sounds awkward. E.g. 'life stages'

Changed.
CHAPTER I

Examiner 1

The first paragraph is unnecessarily broad. Instead, you might lead with some of the relevant ecological theory. I would definitely like to see some discussion of the ecological theory, questions and hypotheses before diving into the methods (even in a methods driven thesis);

I have now included some of the relevant ecological theories – namely spatial ecology and metapopulation theories in the General Introduction section. However, I kept the first paragraph because it relates to the general background (i.e., general importance of the oceans) putting the relevance of my thesis into perspective. This paragraph is important to set the stage for my thesis within the major role of the oceans and ocean life, and for bringing to evidence the lack of knowledge about most marine species driven mainly from sampling limitations. This assists the reader in understanding the difficulties of studying a pan-oceanic shark species, and the limitations associated with the variables that can be used while designing distribution models for such species.

I highlight the changes to this section in the new text below:

“...However, this technology applies only to the sea surface, and so most of the ocean is still largely unsampled. The latter also applies to sampling marine species from higher trophic levels, which still relies on expedition surveys. In spatial ecology (Tilman & Kareiva, 1997), the distribution of species can be interpreted as a match between their ecological requirements and the environmental conditions available. Therefore, the difficulties of sampling the marine environment (through spatial surveys) hamper detection of the relevant spatial coverage to interpret these associations correctly. Modelling approaches
are used to assist the understanding of the relationships between fundamental biological conditions (or processes) and species occurrence - as I discuss below.

The distribution of marine populations is also strongly associated with movement, whether through migration (e.g., herring; Dickey-Collas et al., 2009) or through larval dispersal (e.g., reef fishes; Jones et al., 2009). This leads many marine ecology studies to embrace and test metapopulation concepts and hypotheses (Hanski, 1998; Roughgarden et al., 1985), even if not specifically acknowledged. Metapopulation dynamics are the processes associated with population turnover, i.e., extinction versus colonisation (based on Levin’s model from 1969), between patchily distributed subpopulations with some degree of connectivity between them (Hanski & Gilpin, 1991). In terrestrial systems, this theory has been widely applied to a range of species including butterflies (Hanski, 1994), rabbits (Nielsen et al., 2008), rats (O’Brien et al., 2008) and bears (Hellgren et al., 2005). Indeed, fragmentation of terrestrial habitats makes the application of applied metapopulation theory highly relevant. In the marine environment, metapopulation concepts have been applied mostly to invertebrates and reef fishes (Sale & Kritzer, 2003); however, testing hypotheses at broader spatial scales might not be suitable or even practical for many marine species that are either highly migratory or widely dispersing (Kritzer & Sale, 2006)."

and:

“...Despite the importance of this taxon, the basic biology and ecology of most sharks are still poorly quantified. For example, the economically important (tourism, fishing) whale shark – is still largely unknown. There is some evidence
that whale sharks can travel across oceans, which raises the possibility of applying concepts and testing hypotheses associated with metapopulation theory. Depending on true connectivity (currently unquantified) of whale shark sub-populations worldwide, they might actually be part of a unique global metapopulation (Sequeira et al., 2013), such that localised perturbations might have detrimental flow-on effects to other sub-populations in distant waters."

p.28, 2nd line - 'and might change' should be 'how it might change'

Changed.

p. 31 - tourism industry - If you mean whale sharks, then be specific. Otherwise this is incorrect i.e. do not generalize about 400+ shark species, when only a handful are relevant to any tourist industry;

Agreed and changed to:

“Appropriate management of shark populations is also a commercially lucrative ambition because of the high revenue some shark species bring to the tourism industry.”

p. 31 - References here seem a bit dated (e.g. see recent report by Jessica Meeuwig's group on the value of sharks to tourism (from Palau?), as well as a key one on shark life histories by Luis Lucifora, Veronica Garcia and Jeffrey Hutchings from 2012 Ecology;

The suggested references were now added.
“Due to their slow life histories (delayed maturity, long life spans, few offspring),
sharks are particularly vulnerable to exploitation (Baum et al., 2003; Field et al.,
2009; Garcia et al., 2007; Hutchings et al., 2012; Musick et al., 2000; Schindler
et al., 2002). Consequently, many recent management strategies, increasingly
reliant on good biological information and model projections, have been devised
to maximise the long-term persistence of shark populations in the face of these
pressures (Barker & Schluessel, 2005).

Appropriate management of shark populations is also a commercially lucrative
ambition because of the high revenue some shark species bring to the tourism
industry (Vianna et al., 2012). Sharks are often the subject of major tourism
ventures, such as shark watching and diving (Anderson & Waheed, 2001;
Brunnschweiler, 2010; Catlin & Jones, 2009), and some have argued that
certain species are worth more alive than dead – a recurrent source of profit
instead of a one-off income (Topelko & Dearden, 2009; Vianna et al., 2012).”

p.31/32 - Does Essington et al 2002 really demonstrates that sharks play a
keystone role (this is not what I recall from the pair of Essington / Schindler CNP
Ecopath models). Other more recent refs you could cite here?

Agreed. Essington et al. (2002) report that sharks play a key role, but do not describe
anything about this process. References from studies on cascading fishing effects
directly describe how the food web structure and ecosystem dynamics are affected by
removal of predatory fish. I have now included the references Scheffer et al. (2005) on
the consequences of the elimination of large predatory fish, and Myers et al. (2007) on
the consequences of apex predatory sharks loss. I have also included another
reference about modelling possible indirect effects of shark removal from ecosystems (Frid et al., 2008). I changed the references used in the sentence the examiner highlighted to:

“Moreover, sharks are also thought to play a key role in controlling the food web structure and the dynamics of marine ecosystems (Frid et al., 2008; Myers et al., 2007; Scheffer et al., 2005).”

p.32 ‘study case’ should be ‘case study’ throughout thesis (e.g also on p. 38 etc)

Changed.

p.32 - The statement that poikilotherms are a challenging case study is weakly argued. Many such species are the subject of climate modeling studies. Could you not draw upon some knowledge from the terrestrial literature to help inform this statement?

This text now reads:

“For example, the economically important (tourism, fishing) whale shark – the largest fish species – is still largely unstudied. There is some evidence that whale sharks can travel across oceans, which raises the possibility of applying concepts and testing hypotheses associated with metapopulation theory. Depending on true connectivity (currently unquantified) of whale shark sub-populations worldwide, they might actually be part of a unique global metapopulation, such that localised perturbations might have detrimental flow-on effects to other sub-populations in distant waters. Whale shark distribution is also strongly associated with temperature; they are generally found in warm and
temperate seas (Last & Stevens, 2009) and spend most of their time at the surface (Gunn et al., 1999; Sleeman et al., 2010a; Sleeman et al., 2010b; Wilson et al., 2006) within a narrow range of temperatures (Sequeira et al., 2012). This limited temperature range suggests that their time in deeper waters is limited by thermoregulatory constraints (Thums et al., 2012) or minimises the time spent in deep, oxygen-poor water (Graham et al., 2006). Changes in water temperature associated with climate change (Hoegh-Guldberg & Bruno, 2010), are thus likely to change whale shark distribution or abundance as they re-equilibrate to new temperature regimes and adapt to changing prey distribution (Gutierrez et al., 2008).

Being filter feeders, whale sharks are also likely to be affected by the expected climate-driven changes in ocean productivity and food-web dynamics (Hoegh-Guldberg & Bruno, 2010). Plankton distribution is expected to shift and its abundance will likely change in response to the thermal properties of changing oceans (Reygondeau & Beaugrand, 2011). This logically suggest that such changes will propagate through to higher trophic levels (Kirby & Beaugrand, 2009). This led Hayes et al. (2005) to refer to plankton monitoring programmes as “sentinels” to identify changes in marine ecosystems. Whale sharks have the capacity to travel large distances, so altered plankton distribution will repeatedly affect whale shark migration patterns (Chin et al., 2010). Moreover, whale sharks aggregations are usually associated with specific productivity events, such as coral spawn in Ningaloo, Western Australia (Taylor, 1996) that promotes an increase in zooplankton production. Changes in the timing of these events, as well as other impacts of climate change on coral reefs (e.g., habitat
destruction within whale shark feeding areas), might therefore also impact whale shark occurrence (Chin et al., 2010).

Despite whale sharks spending the majority of their time at the surface (Gunn et al., 1999; Rowat et al., 2007; Thums et al., 2012), they can also dive deeply. This opens speculation to the potential for this species to use different depths (where water is cooler) as surface water temperatures increase. However, oxygen limitation might pose a problem for extended time spent in deeper waters (Pörtner, 2001; Prince & Goodyear, 2006). For most of the reasons given above, whale sharks have been referred to “as potentially the most vulnerable species [to climate change] in the pelagic ecological group” (Chin et al., 2010; in their assessment of the vulnerability of sharks and rays in the Great Barrier Reef), and present a challenging case study in the context of climate change and expected shifts in marine species distribution.”

p. 35 top paragraph: Please clarify here why it is that tuna aggregate near whale sharks - because they share prey or because the tuna prey upon species with which the whale sharks share prey?

I have changed this section to:

“Aggregations of whale sharks usually coincide with higher concentrations of plankton in surrounding waters (Heyman et al., 2001; Sleeman et al., 2007). Being filter feeders, they generally only consume small secondary grazers (e.g., krill or crab larvae; small fish) (Meekan et al., 2009; Wilson & Newbound, 2001), and therefore other species are commonly associated with them, including commercially valuable anchovy (Duffy, 2002) and herring (Wilson, 2002). This
characteristic has been particularly helpful in oceanic tuna fisheries where whale sharks are used to spot tuna (Matsunaga et al., 2003) commonly aggregating under the shark (possibly foraging on similar prey) – in other words, whale sharks are used as fish aggregating devices.”

p. 35 'economic importance worldwide' This is an overstatement. I think you mean in 'select coastal locations around the world'

It now reads:

“Due to their predictable aggregation behaviour, together with their harmless characteristics, whale sharks have become economically important in several coastal locations around the world.”

p. 36 'in the near future' is this specifically stated in the IUCN Red List Guidelines? If so, cite them. If not, and you infer it, then please remove.

It now reads:

“Since 1990, whale sharks have been listed on the IUCN Red List of Threatened Species (IUCN, 2010), and since 2000 are classified as Vulnerable (i.e., at risk of extinction in the wild).”

p. 38 Strikes me as odd to refer to the world's largest fish as 'elusive'

Despite being the largest fish, whale shark patterns of occurrence near shore are not defined, and are still largely unknown in the open ocean, i.e., they are elusive. This term has been used in reference to whale sharks in several publications: Taylor, 2007;
Riley *et al.*, 2010; Rowat & Brooks, 2012; Sequeira *et al.*, 2012; Eckert and Stewart, 2001. Therefore, I believe the reference to whale shark as ‘elusive’ is appropriate in the following paragraph:

“Besides presenting a challenging case study in the context of climate change and likely shifts in marine species distribution, the elusive whale shark also represents a good model to understand biotic responses to oceanographic patterns; being a filter feeder, they should track primary production more closely than higher-order predators (Grémillet *et al.*, 2008).”
CHAPTER II

Examiner 1

Note: This chapter is now published in the Journal of Fish Biology, 2013 (82:2, 595-617), and therefore any changes made in the thesis based on examiner’s reports cannot be included in the published manuscript. Also, due to the journal requirements, the format of the text presented in the article (e.g., use of passive voice) differs from the contents in the thesis chapter.

Abstract, p. 45 -half way down: "generates a model describing... are part of a single, global meta-population', change to "....might be part of ...."

Changed.

Introduction p. 46 -'Being large and docile....' Sentence does not follow logically from this beginning.

Should now read:

“The large and docile whale shark (Rhincodon typus, Smith 1828) is a filter-feeding chondrichthyan that can reach over 18 m in total length (Borrell et al., 2011; Chen et al., 1997; Compagno, 2001), making it the largest extant fish species. Its geographic range is more less known across all tropical and warm temperate waters...”

This chapter provides an interesting overview of whale shark research. I don’t, however, feel that here (e.g. pages 71-72) or elsewhere in the thesis (Chapter VI)
you have made a convincing argument as to why we should be particularly concerned about the potential impacts of climate change on this wide-ranging, highly mobile marine species.

I have made changes both in the General Introduction (see answer above to the examiner’s similar comment re. p.32) and in the (not-yet-published) Chapter VI sections of the thesis to include a more convincing argument about how whale sharks will be impacted by climate change, and why this is of major concern if we are to conserve this species in the long term. Chapter I is currently published (Journal of Fish Biology, 82:2, 597-617), and because the emphasis of this chapter is on the potential of whale sharks to form a unique global population, and not on the importance of whale sharks as a case study for impacts of climate change, these changes were not included in this chapter.

p. 48 - life cycle (Do you mean e.g. juveniles vs. adults) - think you mean 'life stages'

Changed.

p.48/49 - I’m also surprised by the argument that tagging data couldn't be used to estimate 'broad-scale migratory patterns' in this species. This is exactly what the TOPP (www.topp.org) program achieved over the past decade + for many large, highly mobile marine species.

I state, in the pages referred by the examiner, that tracking methods can provide insight into the movement behaviour of whale sharks; therefore, they can be used to estimate ‘broad-scale migratory patterns’. However, in the context of the work being
presented in this chapter (entitled “Inferred global connectivity of whale shark populations”), I detail that due to the difficulties associated with collecting long-term and wide-coverage tagging data for whale sharks, these methods are unlikely to gather data encompassing the full (global) range of the species’ distribution and greatest individual movements:

“Examining whale sharks at the scale of single aggregations cannot adequately describe the species’ life history because it encapsulates only a small proportion of the life stages. Collecting data outside aggregation areas is therefore essential. Continual developments in tagging technology to improve estimates of home range size, movement patterns and habitat use (Hammerschlag et al., 2011) have been partially successful in this regard, although despite > 3000 whale shark tagged to date (e.g., www.whaleshark.org), trajectories have not revealed reliable evidence for inter-aggregation connection. This is not entirely surprising given the low probability of resighting migratory individuals in widely spaced aggregations. The lack of bio-logging efficiency is mostly due to premature detachment and limited spatial coverage of acquired data (e.g., Brunnschweiler et al., 2009), tag removal and/or damage (Fitzpatrick et al., 2006; Hays et al., 2007) which might result in part from attacks from other sharks or killer whales (Fitzpatrick et al., 2006; Speed et al., 2008), and the accumulation of bio-fouling organisms causing tag malfunction (Hays et al., 2007). Moreover, even though tracking methods can provide some insight into the movement behaviour of whale sharks, they are still unlikely to encompass the full range of the population’s distribution.”
The TOPP program was a multinational collaborative effort aiming at explaining the dynamics of marine predators occurring in the North West coast of America, as part of the Census of Marine Life. It received funding from several organisations what allowed the deployment of several thousands of tags, of which less than half resulted in individual tracks (Block et al., 2011). Also, despite the long tracks obtained for some of the species, most of the temporal coverage of the individual tracks was less than one year (maximum duration for different species ranged from 19 to 358 days) (Block et al., 2011 supplementary material). So, the same difficulties I highlighted in this chapter regarding collection/recovery of tag data were also highlighted in the study by Block et al. (2011). It was the amount of both collaborative effort and funding (together with the duration of the project itself – 10 years), that allowed TOPP to become a successful program. I have now included a reference to this program in the thesis General Discussion:

“Tagging approaches can provide insight into the movement behaviour of migratory species, however they bring their own problems (e.g., Brunnschweiler et al., 2009). Even in the successful TOPP program (tagging of Pacific predators; Block et al., 2011), despite the long tracks obtained for some of the species, most of the temporal coverage of the individual tracks was less than one year (maximum duration for different species ranged from 19 to 358 days) (Block et al., 2011 supplementary material). This multinational, collaborative program (part of the Census of Marine Life) received funding from several organisations to allow the deployment of several thousands of tags, of which only less than half resulted in individual tracks (Block et al., 2011). However, some of the problems associated with tagging procedures can be improved as technology progresses. Some suggestions to maximise the duration of tracking
have already been made, and should be adopted when planning new tagging
studies. These suggestions include using fast-loc GPS technology (Hays *et al.*, 2007),
applying anti-fouling paint (Hammerschlag *et al.*, 2011), and use copper
salt-water switches (Hays *et al.*, 2007). But improvements to prevent premature
detachment (Brunnschweiler *et al.*, 2009) are also needed, and new attachment
options should be considered. To maximise the potential for measuring long-
distance migrations, I suggest that tagging studies target multiple sharks
simultaneously within the same aggregation location (cf. Eckert & Stewart,
2001; Wilson *et al.*, 2006), and that deployments are made close to the end of
the peak aggregation season. Additional tagging studies on whale sharks
released from pelagic fisheries nets would also provide information on oceanic
movements to complement results obtained from known aggregations.

Despite whale sharks occurring mostly near the shore of developing nations (*i.e.*,
funding mainly provided via investment from other countries), > 3000 tags have been
deployed on whale sharks, but only a few resulted in usable track data (Sequeira *et
al.*, 2013; *i.e.*, this chapter). This confirms that using the same approach to study
whale sharks (and the same success rate in recovering data) would make it unlikely
that tracking data will provide insight into the movement behaviour of whale sharks
encompassing the full range of their global distribution.
CHAPTER III

Examiner2

Note: This chapter is published in *Diversity and Distributions*, 2012 (18, 504-519), and therefore any changes made in the thesis based on examiner’s reports were not included in the published manuscript.

I do not have much to say given the paper is already published in a respectful journal. The literature cover is relatively good although some key papers are ignored.

The methods are not very well-explained to me. There are even some misconceptions. MaxEnt is NOT a presence-only algorithm, not more than any GLM for instance. I am not entirely convinced by the classification of models as well. MaxEnt and BRT are not a machine-learning method. BRT are only classification trees in a boosting framework.

Even though I referred to MaxEnt as a presence-only model in the Introduction of this chapter, in the Methods I specified the inclusion of background data, and have discussed different ways to include background data in the discussion section. In Chapter IV, I have also corrected the MaxEnt description by including the need for background data in its definition:

“MaxEnt is a tool for generating species distribution models from presences-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).”
As for MaxEnt being a machine-learning method, my reference to this was in Chapter I (General Introduction), and was followed with a reference to Phillips et al. (2004) - where they present MaxEnt’s value for species distribution modelling contextualising it within the machine-learning community. I could not find my reference to BRT as a machine-learning method or as classification trees.

Why using Kappa for GLMs? The use of Kappa should be stopped now because it is strongly biased with prevalence, which is definitely the case here given the use of pseudo-absence. True Skill Statistics or Boyce Index should be preferred.

To account for the fact that Kappa is strongly biased with prevalence, I used the observed prevalence (presence:pseudo-absence ratio) as the value to optimise the threshold used in the cmx function (from the package PresenceAbsenceBo) to estimate Kappa. This I discussed in Chapter VII (General Discussion) as shown below:

“Another related issue is the decision regarding which probability value should be used to convert the model results into presences or absences (Jiménez-Valverde & Lobo, 2007; Liu et al., 2005). This value is commonly assumed to be 0.5, but for presence: absence ratios different to 1:1, this assumption results in incorrect classification of absences (if prevalence is low, i.e., more absences than presences) or presences (if prevalence is high) (Jiménez-Valverde & Lobo, 2007). Several methods can be used to specify this threshold (see Liu et al., 2005), such as averaging the predicted probabilities, using sensitivity and specificity, the prevalence value, or the value that maximises $\kappa$, I considered this threshold equal to the prevalence used, which is considered a good approach (Liu et al., 2005).”
Also, within the function `optimal.thresholds()` from the R package `{PresenceAbsence}`, the threshold that maximises the sum of sensitivity and specificity (equivalent to True Skill Statistics; Freeman, 2012) could have been chosen (instead of observed prevalence). However, while being independent from prevalence, this index can over-predict the species’ distribution (Freeman & Moisen, 2008). I have now calculated the Boyce index (Boyce et al., 2002) – this is shown below (answer to comment on page 104). Results are similar for all pseudo-absences techniques (especially for MaxEnt), and are generally higher (0.83 - 0.95) than the ones obtained with kappa (0.34 - 0.79).

**There is no justification on the use of MaxEnt or GLM? Why these two approaches? Only because the authors like GLM and because MaxEnt is over-used? Why not including an autoregressive component to MaxEnt?**

The choice of GLM derived from their ability to analyse non-Gaussian data, their flexibility and the relative ease with which relationships can be identified and hypotheses tested using the linear imposition. GLM can also be used in a hypothesis-testing framework using AIC and BIC as indices for model support (Burnham & Anderson, 2004). Moreover, this type of regression model generally results in more realistic projections (Brotons et al., 2004; Zaniewski et al., 2002). But no modelling approach is infallible, so we compared GLM results to MaxEnt, which deals with presence-only data (and background information) (Elith et al., 2011), and has been increasingly applied to model species distribution. Following the examiner’s comments, I have re-run the models within a BIOMOD framework (Thuiller, 2003) to check which modelling approach performed better – I give some of these details below, answering the examiner’s comment on pages 105-106.
Inclusion of an autoregressive component which describes time-varying processes could possibly assist to describe lags between whale shark occurrence and the predictors used. In relation to my study, this could be particularly useful in relation to chlorophyll $a$. However, due to the little variation in the concentration of the latter predictor (in open ocean) I believe it would be difficult to obtain a result. In the sense that autoregressive components represent random processes, some of these were included in the MaxEnt models in two ways. First, by using different pseudo-absences sets generated from the background data using each of the three methods tested ($\textit{random, IDW}$ and $\textit{tuna}$), and by including a random 10000 background cells (maximum allowed in MaxEnt) from the entire area considered in the models. Second, the type of variables considered for modelling was also varied by specifying the “features” allowed in MaxEnt runs – \textit{i.e.,} through the selection of “linear and quadratic features” or the use of “auto-features”. Moreover, the work presented in this chapter was further developed in the Chapter IV to include an analysis of temporal trends in whale shark occurrence.

I do not understand why the ROC curve is not estimated for GLM/GLMM instead of keeping only with the deviance. It would have been easier to compare with MaxEnt.

Following the examiner’s suggestion, I have now calculated the ROC/AUC value for the GLM/GLMM models used in Chapter III, and I have included a table with results in Appendix B (Table B2, copied below) together with the following text:

“Receiving operator characteristic curve / area under the cure (ROC/AUC)
To measure models performance we also calculated the receiving operator characteristic curve / area under the curve (ROC/AUC) and results are shown in Table B2. The trend arising when comparing the AUC results between each MaxEnt and GLM model is similar to the one I obtained previously when comparing \( \kappa \) – i.e., when using the random technique to generate pseudo-absences, GLM was scored higher by both measures (AUC and \( \kappa \)).

Using ROC/AUC to measure performance of species distribution models has been criticised for being influenced by the total extent to which the models are applied (Lobo et al., 2007), and especially when the objective of the models is to determine the potential distribution of species (Jiménez-Valverde, 2011). Therefore, to compare MaxEnt, GLM and GLMM results, I have used the \( \kappa \) statistic (accounting for the fact that this measure depends on prevalence), which some authors suggest (Manel et al., 2001). The percentage of deviance explained was only used for GLM and GLMM.

### Comparison of the Area Under the Curve (AUC) and Kappa results for the two modelling approaches used in chapter III: MaxEnt and GLM/GLMM for each of the pseudo-absence dataset generated method (P/A) and for each season: Autumn (Aut), Winter (Win), Spring (Spr) and Summer (Sum).

<table>
<thead>
<tr>
<th>P/A</th>
<th>Approach</th>
<th>Aut</th>
<th>Win</th>
<th>Spr</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random MaxEnt</td>
<td>Kappa</td>
<td>0.54</td>
<td>0.57</td>
<td>0.50</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>0.72</td>
<td>0.70</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>GLM</td>
<td>Kappa</td>
<td>0.79</td>
<td>0.51</td>
<td>0.59</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>0.87</td>
<td>0.61</td>
<td>0.79</td>
<td>0.63</td>
</tr>
<tr>
<td>IDW MaxEnt</td>
<td>Kappa</td>
<td>0.34</td>
<td>0.10</td>
<td>0.25</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>0.63</td>
<td>0.57</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td>GLM</td>
<td>Kappa</td>
<td>0.35</td>
<td>0</td>
<td>0.33</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td>0.82</td>
<td>0.72</td>
<td>0.80</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Explained deviance measure the explanatory power of a model, not is predictive accuracy. This is a quite different.

This was now changed to:

“We used the percentage of deviance explained (De) to quantify each model's explanatory power.”

There is a problem of rows (the third row) in Table 5

I have now corrected the problem with the rows in Table 5.

Discussion: end of Page 104: The Boyce Index allows to discount for pseudo-absence and to compare different models or repetitions with the same set of presence.

I have now calculated the Boyce index for the three pseudo-absence generation methods for each modelling tool – MaxEnt and GLM (Autumn runs only; Table B3). The Boyce index assesses the monotonic relationship (Spearman’s rank correlation) between habitat suitability classes and the proportion of evaluation points adjusted by the area covered (Boyce et al., 2002). At the end of page 104, I discuss that “because pseudo-absences locations vary within each model, the resulting deviance explained
and AUC are not directly comparable across models”, and that the three methods used to choose pseudo-absences are most useful to determine how absence locations influence model results. According to the Boyce index calculated for autumn, the results among pseudo-absence selection techniques and among modelling tools are similar, showing that the predictions from all models considered are consistent with the presences distribution in the sampled dataset. Results for the GLM predictions were generally slightly higher. I have included a table with results in Appendix B (Table B2, copied below) together with the following text:

"Boyce index comparison between MaxEnt and GLM/GLMM models

We assessed the monotonic relationship (Spearman’s rank correlation) between habitat suitability classes estimated by the models and the proportion of evaluation points adjusted by the area covered (Boyce et al., 2002) by calculating the Boyce index (autumn only). According to this index (Table B3) the results among pseudo-absence selection techniques and among modelling tools are similar, showing that the predictions from all models considered are consistent with the presences distribution in the sampled dataset. Results for the GLM predictions were generally slightly higher."

<table>
<thead>
<tr>
<th>P/A</th>
<th>Tool</th>
<th>Boyce index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>MaxEnt</td>
<td>0.827</td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>0.845</td>
</tr>
<tr>
<td>IDW</td>
<td>MaxEnt</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>0.911</td>
</tr>
<tr>
<td>Tuna</td>
<td>MaxEnt</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>0.952</td>
</tr>
</tbody>
</table>
Page 105: Something is missing in the discussion about a missing factor. I do not buy the fact that spatial autocorrelation bias the models. Given most environmental variables are spatially autocorrelated, the fact that the residuals are also spatially auto-correlated strongly suggest a missing element or a missing variable. What about dispersal, what about another variable? It could also be because of interactions between factors that are not taking into account here.

The models I developed in Chapter III (and throughout the thesis) could possibly be improved by the addition of other predictor variables. This was discussed in the General Discussion, where I suggest that a predictor directly describing whale shark prey (zooplankton abundance) could improve the results of the distribution models. There are also other important caveats; applied only to the sea surface, my models do not consider any variables describing the vertical distribution of this species, nor do they account for whale shark population dynamics and/or structure. However, data on the key aspects of whale shark life history (including dispersal) are still to be measured, and therefore no further data are currently available to add to the predictor set.

I have, however, included an additional term (depth\(^2\)) when re-fitting the models developed in Chapter III – detailed in the supplementary information for Chapter IV (autumn model only). This term, accounting for whale shark occurrence mostly in locations where depth is within a specific range, did improve the model fit; however, the residuals' spatial autocorrelation was reduced only after the inclusion of the spatial random effect, as in Chapter III.

Moreover, I found no support for any interactions between the predictor variables used based on model runs made:
i. in BIOMOD (Thuiller, 2003), none of the GLM model equations (one for each pseudo-absence repetition) included interactions between predictors

ii. using generalised boosting models (GBM), the interact.gbm object returned a Friedman’s statistic $= 2.9 \times 10^{-14}$, which indicates weak interaction effects (Ridgeway, 2013; GBM package)

iii. using boosted regression trees (BRT; Elith et al., 2008), where even though I found some support for an interaction between depth (centred) and distance to shore, and between depth (centred) and mean sea surface temperature, the coefficients for these were low (0.02 and 0.06, respectively) and did not contribute much to the models’ explanatory power.

I have now included a synthesis of the results obtained for these modelling tests (with BIOMOD and GBM/BRT) in Appendix B (“BIOMOD, GBM and BRT models for the Indian Ocean”).

Page 106: MaxEnt can use interactions between variables, this is why the set of important variables are not the same. It comes back to my previous comment that GLM should also include some interactions between crucial variables.

MaxEnt can use interactions between variables only when “auto-features” is selected. In Chapter III, I ran MaxEnt models mostly using the “linear and quadratic” features to make their results more comparable with those derived using GLM. I only selected “auto-features” in an extra test run. I discussed the results from each MaxEnt output within that chapter. Moreover, as discussed in response to the previous examiner’s comment, I found no support for interactions between predictor variables.
CHAPTER IV

Examiner2

Note: This chapter is now published in *Marine Ecology Progress Series* (doi: 10.3354/meps10166), and therefore, any changes made in the thesis based on examiner’s reports were not included in the published manuscript.

Page 116-117: To artificially increase the prevalence of the species and thus the generated probability of presence, it is generally recommended to down-weight the absence to have the same weight than the presence.

Following the examiner’s comment, I have now re-run the temporal models with down-weighted pseudo-absences and I have included the results in Appendix C (“Prevalence test – down-weighting pseudo-absences”) as follows:

“We have now re-run the models after down-weighting the 100 absences (from 1 to 0.01) to get a 1:1 ratio of presences to absences (*i.e.*, a presence:pseudo-absence ratio of 0.5), as suggested. Because the partial effect of time (Figure 16, in Chapter IV) is highly dependent on prevalence, the change in the presence to absence ratio resulted in a much more pronounced curve reflecting the dependency of occurrence with time. This is depicted in Figure C5. Although this change in prevalence increased the estimated percentage of deviance explained (92.3 % for the best ranked model), it did not affect the model ranking based on $wAIC_c$ (Table 7, in Chapter IV).”

Page 120: there is a mistake. Year is the random factor, and not time. (modify the order in the parenthesis). Else this is contradictory with Table 1.
Changed. It now reads:

“We included both a fixed and a random effect for time (*Time and year*, respectively) to ...”
CHAPTER V

Examiner 1

Note: Here, I address Examiner 1’s comments made on Chapter V only. I will address comments to Chapter VI in the following section. Several changes were made to Chapter V – as specified below, and this chapter is now being prepared for submission to the *Journal of Experimental Marine Biology and Ecology*.

**These two chapters would benefit from implementation of an additional (and different) modeling approach (e.g. models with a truncated distribution or boosted regression trees) and/or some sensitivity analyses (e.g. re: generation of the pseudo-absences in for the Atlantic and Pacific Ocean data sets as in Chapter III).**

Chapter V consolidates the methods used both in Chapters III and IV, but applies them to the Atlantic and Pacific Oceans instead. Due to the complexity of the analyses, and because the main objective was to obtain results comparable to those described in the previous chapter (mostly in terms of the partial effect of time), I used the same modelling approaches.

Similar to what I did in Chapter III, I included a ‘sensitivity’ analysis in each chapter, as multiple pseudo-absence datasets were always generated prior to the modelling to assess the influence of pseudo-absences location choice in the models results. I have now added the following text to make this clear:

“To assess the influence of the date and location associated with the selected pseudo-absences, we generated each set of pseudo-absences 100 times prior...
to their use in the spatial models, and 10 times for the temporal models (due to computing time).

In Chapter III, I have made an assessment of the influence of different pseudo-absences generation methods (such as using random selection, distance to whale sharks or based on tuna catch). Because the results from each pseudo-absence generation method used in Chapter III were similar, I decided to keep using the simpler method (random selection; as discussed in Chapter VII – see next answer) in the following chapters. This is in accordance with the results from a recent study on selection of pseudo-absences (Barbet-Massin et al., 2012), where the authors recommend a random selection of pseudo-absences.

At minimum, I would like the candidate to include some discussion of the potential biases and limitations arising from the data used and/or the modelling approaches undertaken in these two chapters. Such admissions and discussions would strengthen (not weaken) these chapters and their related manuscripts intended for publication in the peer-reviewed literature. The candidate does an excellent job of reviewing these in Chapter VII, her General Discussion, but this chapter will not be published in the primary literature, and this material is critical to each of Chapters 5 and 6 in order for readers to be able to evaluate the research.

Following the examiner’s advice, I have now included in Chapter V further discussion of the potential biases and limitations arising from the data and modelling approaches I used, as previously discussed in Chapter VII:
“The datasets we used were opportunistically collected and therefore lead to inherent complications. Differential sampling effort is a major issue for which we accounted at least partially in both modelling steps to reduce potential bias in spatial and temporal effort. However, an underlying assumption of our models is that failure to report a whale shark presence (whether due to a failure in detecting a shark or in reporting a detection) is evenly distributed across the sampling area and study period. Access to higher-resolution data would likely improve the results (exemplified by the Atlantic Ocean models), and can be achieved through establishing collaboration between researchers and fisheries-management organisations. Commercial-in-confidence restrictions on access to fisheries data (especially sensitive data such as fish catch) delay scientific research from providing more accurate models on fisheries impacts. Such collaborations would also allow researchers to give greater insight on how better datasets could be collated for specific objectives using the resources currently in place. However, while data collected by fisheries might not be as precise as those derived from scientifically designed surveys, they are still essential sources of information given the logistic challenges of surveying widely distributed species over their entire range. Fisheries data therefore present unique opportunities to advance our understanding of such species.

Because the datasets we used contained only recorded presences, the generation of pseudo-absences was necessary to allow for the binomial estimation in the models. Differences associated with the method to generate pseudo-absences was assessed in a previous study using similar datasets (Sequeira et al., 2012) where we found that model performance was not reduced by using random pseudo-absences selection, and was therefore the
method used in this analysis. This accords with the results from another recent publication specifically addressing pseudo-absence selection issues (Barbet-Massin et al., 2012). The number of pseudo-absences selected is another issue to consider when generated. Barbet-Massin et al. (2012) reported that model accuracy increases until the presence to pseudo-absence ratio (i.e., prevalence) equals 0.1, and remains constant for lower ratios. In the models we developed, prevalence was 0.1 in the spatial assessment, and 0.01 in the temporal assessment (to allow enough points for the spatially explicit temporal analysis); therefore in both cases, we do not expect the ratio presences:pseudo-absences to have affected model accuracy.

Another important aspect of the input data we used is that both sightings and environmental variables correspond mostly to the ocean surface layer. Three-dimensional environmental data at an adequate resolution for regional analyses such as that presented here are currently not available. Even though whale sharks spend most time at the surface (e.g., Rowat et al., 2007) they also dive frequently, and therefore, assessment of how they explore the vertical habitat and estimation of its suitability at a broad scale is also of importance. However, this will only be possible when new technologies are developed to allow vertical data collection at broad spatial scales.

Despite the limitations associated with both the opportunistic dataset and species distribution models, we have demonstrated a powerful way to use such datasets to identify both spatial and temporal trends of highly migratory species. We also revealed that longer-term and higher-resolution sightings datasets are still required to tease apart potentially confounding aspects of inter-ocean migration in whale sharks. Continued investigation of these connections is
necessary to assess temporal trends in occurrence, and to reveal the suspected impacts of human modifications to the wider marine realm.”

In each of these chapters, please also:

show the equations for the models being used, so that the reader can more clearly see what is being done;

The equations for Chapter V, in terms of variables used, are detailed in Table 8 and Table 9. I have now also included a generic description of the equations in the Methods:

“A generic way to write the GLMM (with a logit link function) is:

\[
\text{logit}(\text{Presence}) = \alpha + \beta X_i + \gamma_i Z_i + \epsilon_i \times [+ \log(\text{fishing effort})]
\]

where Presence is the expected mean probability of sighting occurrence. \(X\) and \(Z\) represent the fixed and random covariates used in the models: the environmental predictors and the spatial grid in the spatial models, and habitat suitability, zero-centred effort and time, and year in the temporal models. The index \(i\) corresponds to the number of observations among grid-cells in the spatial models, and years in the temporal models. \(\beta\) and \(\gamma\) represent the parameters of the fixed and random effects, respectively, and \(\alpha\) is the intercept. The log of fishing effort was included as an offset only in the spatial models.”

provide some more detail on the exact nature of the logbook records of whale shark occurrences, including quantitative information or comment on the percent of logbooks in which whale shark occurrences were consistently, or ever,
recorded (i.e. does the entire fleet report whale sharks?). How might this alter your interpretation of the zeroes (i.e. do zeros only reflect true absences and failure to detect whale sharks, or do zeros reflect these two cases as well as 'failure to report whale sharks)? How might this alter how you treat effort' in your models? OR Are these data not provided on a trip by trip, or vessel by vessel basis, but merely aggregated as per the spatial precision described (e.g. p. 139). These types of detail need to be made clear for readers.

The logbook records made available were merely aggregated as per the spatial precision described. No information was available on vessel or trip units, and therefore, we could not estimate quantitative information on the percent of logbooks in which whale shark occurrences were consistent. To clarify this point, I have changed the text to:

“The dataset made available by the Institut de Recherche pour le Développement (France) and the Secretariat of the Pacific Community comprises most of the central area of the Atlantic (21°N to 15°S and 34°W to 14°E) and central western Pacific (15°N to 15°S and 130°E to 150°W) (Figure 18). It includes date (month and year), longitude and latitude information for whale shark sightings (0.01° precision), and spatially aggregated information on sampling effort (number of days spent fishing) per month and per grid cell of 1° resolution in the Atlantic, and 5° resolution in the Pacific (Figure 18). No information on individual vessel or trip units was available.”

Regarding the interpretation of zeros in the data, they can reflect both situations highlighted by the examiner, i.e., they can be either a real absence or simply a failure to report a real presence (whether due to a failure in detecting a whale shark or not reporting detection). An underlying assumption of my models is that this potential bias
(failure to report presence) is evenly distributed across the sampling area and study period. Because there is no way to distinguish these types of information for zeros, I generated pseudo-absences, and accounted for biases arising from “this choice of points as absences” by generating several presence / pseudo-absence datasets (see previous responses). I could have only treated effort differently if detailed information had been available on the number of fleets, and trips recording whale shark sightings.

Due to commercial-in-confidence agreements, the datasets I used in this chapter cannot be displayed as supplementary material for the benefit of readers. However, public-domain data (similarly aggregated, but at coarser spatial grids) for both the Western Pacific and Atlantic Oceans can be obtained online: www.wcpfc.int/science-and-scientific-data-functions/public-domain-data and www.iccat.int/en/accesingdb.htm, respectively.

Following the examiner’s comment, I have now added the following sentence to the Discussion:

“Differential sampling effort is a major issue for which we accounted at least partially in both modelling steps to reduce potential bias in spatial and temporal effort. However, an underlying assumption of our models is that failure to report a whale shark presence (where due to a failure in detecting a shark or in reporting detection) is evenly distributed across the sampling area and study period.”

clearly explicate the relationship between the sightings and effort data (i.e. the fact that the two are not paired, and hence candidate cannot implement
traditional 'CPUE' or 'SPUE' models, in which the two are available at the same scale);

The reason why I did not implement traditional 'SPUE' models was because the spatial resolution at which data were made available differed for sightings and effort, with spatial information for sightings available at 0.01 ° precision, and effort data aggregated at 1 ° or 5 °. Normalising the sightings-per-unit-effort (SPUE) would therefore decrease the resolution of my input (sightings) data to 100 or 500 times lower (1 ° or 5 ° instead of 0.01°). Results from models applied to such coarse dataset, where environmental data would also need to be aggregated to those lower resolutions, would not result in meaningful information. For example, the relationship between whale shark occurrences and the (narrow) range of sea surface temperatures would not have been detected if the temperature values had been averaged within each 5-° grid cells (i.e., ~ 555 × 555 km grid cells at the Equator).

To keep the highest resolution possible in my dataset, I used the resolution of the sightings data (0.01-° resolution), and included effort as an offset. This allows accounting for higher sighting probabilities to be expected when more days were spent fishing, thereby reflecting how sightings and fishing effort are paired. To clarify this point, I have now added the following sentence to the Methods:

“To account for the sampling bias associated with effort (more whale shark sightings expected in areas where fishing effort was higher), and to keep the highest resolution possible in the input data, we have included effort as an offset in the models.”
Rather than modeling only certain components of the data (i.e. April to June), I would prefer the candidate to model all of the data, including seasonal covariate(s) in the model (e.g. modeling season as continuous variable e.g. as in Baum & Blanchard 2010 Fisheries Research);

The main objective of Chapter V is to obtain results comparable to those from the Indian Ocean described in the two previous chapters – i.e, spatial distribution of whale sharks at the ocean-basin scale and estimation of temporal trends. In the conclusions of Chapter III (published in Diversity and Distributions), I discussed that because 70% of the sightings occurred during the months of April to June, our spatial predictions are more robust for that season. For this reason, I did the temporal analysis only for these months in Chapter IV.

As I state in the abstract: “[whale shark] population trends can only be understood by examining patterns of occurrence synchrony among locations”. Therefore, to obtain results comparable to the trend observed in the Indian Ocean, I used data for that same season (i.e., April to June) to calculate trends in the Atlantic and Pacific Oceans.

I did, however, include models for some other seasons in Chapter VI, where I modelled global whale shark habitat suitability using data from the season where more sightings occurred within each ocean. These were austral winter in the Atlantic (with 1153 total sightings), austral autumn in the Indian (with 811 total sightings), and austral summer in the Pacific (with 314 total sightings). Due to the generally low number of sightings reported in seasons not considered in my thesis (e.g., only 7 sightings were reported in the Atlantic during the months of January to March), including all data in the models would not add important information to improve their
predictive performance. To make this point clearer in the manuscript, I have now added the following:

“To compare our results for the Atlantic and Pacific with previous results for the Indian Ocean (Sequeira et al., in press), and due to the generally low number of sightings in other seasons for each ocean, we used data only for the months of April to June (austral autumn).”

Furthermore, all likelihood-based binomial models are based on the concept that grid cells must have unique information regarding the presence or absence of the species to enable discrimination of a covariate’s effects on presence. Including data for more than one season would mean repeating grid cell information, because the same grid cell that has been assigned a presence in one season could be selected as either a presence or absence in another.

Using season as a continuous variable applying sines and cosines with periodicity of 0.5 and 1 year (Baum & Blanchard, 2010) seems like an interesting approach for studies considering year-long variation. However, the method used by Baum and Blanchard (2010) did not include a likelihood-based method, and therefore did not allow for estimation of the AIC and BIC ranks to compare models: My study is entirely designed around the observation that whale shark occurrence is seasonal (Heyman et al., 2001; Taylor, 1996).

**explain more clearly the justification for modeling time/year as both a fixed and a random effect;**

The random effect for time (year) accounts for the fact that same-year observations are not independent (i.e., year effect). Such analysis of the conditional modes (i.e.,
random effects) can be useful when generating hypotheses for future studies or for qualitative assessment of any pattern (Gelman & Hill, 2007; MacNeil et al., 2009). However, for making statistical inferences, any trend observed in these conditional modes should be included in the fixed part of the models (following Zuur et al., 2009). This ensures that the random structure contains only information that could not have been modelled with fixed effects. Therefore, to describe correctly a possible linear or parabolic-like trend in the whale shark probability of occurrence, the variable Time was included as a fixed effect through the polynomial function poly() assuming a polynomial degree of 2 (to allow the detection of non-linear temporal trends). By accounting for this fixed effect, the resulting plot of the conditional modes of the year random effect (i.e., conditional on the observations and predictor values) only show random variation remaining over years (e.g., Figure 29).

Figure 29: Conditional modes of the random year intercept term for models including time as fixed effect (quadratic term; left), and without the fixed temporal trend (right).

To clarify the justification for inclusion of time, both as fixed (Time) and random (year) effects, I added the following sentence:

“...and time, both as a fixed (time) and random effect (year) to ensure that the random structure contains only information that could not have been modelled with fixed effects (following Zuur et al., 2009). To account for a parabolic-like
dependence of occurrences with time, we added the quadratic term for time using the poly function (with degree 2)."

Discussion p. 150 'we hypothesi[z]e that whale shark occurrence might be asynchronously cyclical across oceans.'

Typo hypothesi[z]e is now corrected.

This statement is speculative and unconvincing. There is a huge difference between a small number of individuals moving between ocean basins to ensure 'low genetic differentiation of whale sharks sampled in different oceans' versus entire (sub)-populations moving across ocean basins to produce asynchronous cycles across oceans.

The statement is now discussed as:

"Given these opposing results, the evidence for an inter-decadal occurrence pattern in the Indian Ocean (Sequeira et al., in press), the low genetic differentiation of whale sharks sampled in different oceans (Castro et al., 2007; Schmidt et al., 2009), and the implicit notion that whale sharks travel across oceans (Hueter et al., 2008; Rowat & Gore, 2007), we hypothesise that whale shark occurrence might be asynchronously cyclical across oceans. Such a pattern would be consistent with the inter-ocean migratory behaviour we have previously hypothesised (Sequeira et al., 2013).

Despite genetic clues (Castro et al., 2007; Schmidt et al., 2009) being the only evidence to date corroborating inter-ocean migration, these results could also have arisen even if only a small number of individuals were moving between
ocean basins (Hartl & Clark, 1989). However, the number of whale sharks occurring in aggregation locations within the predictable season is highly variable among years (Rowat et al., 2009a; Wilson et al., 2001), and this variability holds for the number of whale sharks sighted at the ocean-basin scale (Figure 20a and b). The opposing trends we detected between the Indian and the western Pacific Oceans during the last decade might therefore reflect a migration of sharks between oceans. Similar distributional shifts, promoting asynchrony in occurrences, have also been suggested for other filter-feeding sharks (basking sharks *Cetorhinus maximus*) between the west coast of Republic of Ireland and the Norwegian Sea (Sims & Reid, 2002), and were potentially associated with spatial changes in foraging conditions. If whale sharks are migrating between oceans (Sequeira et al., 2013), a cyclic pattern such as the one observed in the Atlantic Ocean would explain both the interannual variation in whale sharks numbers at the ocean-basin scale (Figure 20a and b) and the asynchrony in their occurrences in the Indian and Pacific Oceans. Moreover, such large-group and multi-year migrations would still be consistent with the low differentiation found in genetic studies (Castro et al., 2007; Schmidt et al., 2009). These whale shark migrations could be associated with changes in environmental conditions (Sleeman et al., 2010b; Wilson et al., 2001), or with sex/age or reproduction-related behaviour (Ramírez-Macías et al., 2007; Sequeira et al., 2013). Reproduction-associated multi-year cycles have been observed in anadromous fish (*e.g.*, sockeye salmon *Oncorhynchus nerka*) (Dingle, 1996). Multi-year cycles have also been reported for fish catches in the western Indian Ocean (Jury et al., 2010), and for other marine
megafauna, such as the decadal cycle in cetacean strandings in southeast Australia (Evans et al., 2005)."
Examiner 2

This is an interesting complementary paper to the previous one. It was somehow difficult to pinpoint the differences between the two (except the two oceans).

This chapter consolidates the spatial and temporal analysis presented in the two previous chapters (Chapters III and IV; Indian Ocean) applying the same methods to the Atlantic and Pacific Ocean. The main objective was to get results comparable to the ones obtained in the previous chapters (mostly in terms of time partial effect), with the general aim of getting an overview of the trend in whale sharks occurrence in the three major ocean-basins (as opposed to the ‘traditional’ analyses limited to aggregation areas).

My main concern relates to the Method part, which is not always understandable for non-afficionados. For instance, the Fourier transformation is absolutely not explained. You cannot guess that all readers will know what it is about and how it works. More explanations are needed here. Explain also why you chose this framework.

I have now updated the Methods detailing the Fourier Transforms (including the reasons to choose this approach):

“Fourier transforms allow for the decomposition of signals (i.e., time series) into the sum of sinusoidal curves with different frequencies – a highly useful approach for ecological time series (Platt & Denman, 1975). To analyse possible cyclic variation in whale shark occurrences, we applied the fast Fourier transform function (FFT; following Moler, 2004) to the sightings time series after
correcting for effort bias (i.e., standardizing sightings per unit effort [SPUE], with an effort unit = 1 fishing day). We used MatLab version 7.12.0.635 (R2011a) (The Mathworks Inc., Natick, MA), and interpreted the results of the FFT by plotting the periodogram for the function: power against the inverse of frequency (time). A cycle is then defined by identifying the strongest frequency.”

I am a bit puzzled by the use of AICw without further attention (the same applied to the previous chapter as well). What is nice with this framework is also to extract a weight of evidence of each variable from the list of models. Indeed, although the weight of evidence in favour of one given model is of interest (it allows to highlight the potential problem of stepwise if several models have similar Ai (close to 0), the general interest for ecologists is a step further. By analogy to weight of evidence in favour of model, the weight of evidence of each predictor (wpi) can be simply estimated as the sum of the model AICs weights (wpi) over all models in which the selected predictor appeared. The predictor with the highest wpi (the closest to 1) gets the highest weight of evidence to explain the response variable (the highest relative importance). This approach based on set of multiple models is far more robust than inferring variable importance based on a single selected model.

I agree with the examiner, and the weight of evidence for each temporal predictor was now calculated, and I have included them in the Results as a new table (copied below). I have also included the following text in the:

Methods
“We also calculated the weight of evidence for each predictor used in the temporal models by summing the wAIC weights over all models in which each predictor appeared. This allows estimation of the predictor with the highest relative importance to explain the response variable.”

Results:

“The weight of evidence we estimated for each temporal predictor is shown in Table 10. Habitat suitability (Hsuit) and effort received the highest weight (1 and 0.999, respectively) in the Atlantic Oceans, while in the Pacific we found the highest weight for effort (weight of evidence = 1) followed by the linear term for time (time: 0.953).”

Discussion:

“We found a lack of support for the linear or quadratic time predictor in the Atlantic Ocean from April to June (Table 9) where habitat suitability (Hsuit) and effort received the highest weight of evidence (Table 10). At least for the last decade in the western Pacific, there was evidence for a linear, positive trend in the probability of whale shark occurrence (Table 9; Figure 21). In the latter ocean, even though habitat suitability was also ranked as an important variable (weight of evidence = 0.761), we found the highest weight of evidence for the predictors effort and time (Table 10).”

Estimated weight of evidence for each temporal predictor used in the generalised linear mixed-effects models: Hsuit - spatial predictor derived from the spatial distribution models, effort - temporal variation in fishing effort and a linear and quadratic term for time (years).

<table>
<thead>
<tr>
<th>Ocean</th>
<th>weight of evidence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hsuit</td>
<td>effort</td>
<td>time</td>
</tr>
<tr>
<td>Atlantic</td>
<td>1</td>
<td>0.999</td>
<td>0.367</td>
</tr>
<tr>
<td>Pacific</td>
<td>0.761</td>
<td>1</td>
<td>0.953</td>
</tr>
</tbody>
</table>
Similarly, instead of keeping one model at the end, I would strongly recommend to use a weighted-average. These predictions are simply the weighted-averaged predictions of all submodels. Sub-models with low weight of evidence will have basically not power, where submodels with similar AIC will gave even weight.

\[ \overline{P} = \sum_{i=1}^{R} w_i p_i \]

Where \( w_i p_i \) is the prediction from sub-model \( i \), weighted by the weight of evidence in favour of this sub-model. Associated with this weighted mean, a weighted variance or standard deviation can be easily estimated which allows a useful estimation of the uncertainty associated with the modelling approaches.

Unconditional interval of confidence can then be drawn.

The predictions resulting from the spatial models were always calculated using model-averaging – in all chapters. To clarify this point, I have added the following sentence to the Methods:

“We compared the models relative strength of evidence by weighting each model’s Akaike’s information criterion corrected for small sample sizes (\( wAIC_c \)) (Burnham & Anderson, 2004), and assessed the goodness-of-fit by calculating the percentage of deviance explained (\( %De \)) for each model. We calculated the 10-fold cross-validation error for the model with highest \( wAIC_c \) support and assessed the model’s predictive power with the \( \kappa \)-statistic (Cohen, 1960). To build the habitat suitability maps (as result of the first spatial modelling step), we weighted each model’s predictions according to its weight of evidence (\( wAIC_c \)), and used the result of the weighted mean prediction for all models as input to the temporal models.”

Note: The references indicated by the examiner were already included in the thesis.
CHAPTER VI

Examiner 1

Note: Here I address the examiner 1’s comments made in the previous section, which also applied to this chapter. Those comments are repeated here under “[Repetition of comments]”, and identified with the preceding word: “[Repetition]”.

p.162 'penury' Reword.

Now reads:

“However, the species’ high mobility and the lack of expansive occurrence data mean that teasing apart spatial and temporal patterns is problematic (Sequeira et al., in press).”

Introduction pgs. 162-163 This chapter is tightly linked to the previous modelling chapters in the thesis, and the candidate should be clear about this in the Introduction. i.e. When building upon previous research, it is important to denote what was done (and published) previously and what is new and distinct in the current piece of work.

I have now changed the end of the Introduction to include a reference to my previous work:

“For the first time, we used a global dataset of pelagic whale shark sightings derived from the tuna purse-seine fishery operating in the Atlantic, Indian and Pacific Oceans to: (1) predict the spatial distribution of whale shark habitat suitability across three ocean basins, and (2) extend our projections to a future
scenario of ocean warming based on a no-climate-policy reference (no stabilisation of greenhouse gas emissions). Following our previous work where we derived ocean-basin scale predictions for whale shark distribution and temporal trends (Sequeira et al., 2012; Sequeira et al., 2013), herein our overarching aim was to derive realistic predictions of the species current geographical range (circumglobal), and provide a modelling tool for exploring the species’ future distribution under anticipated climate change.”

[Repetition of comments]

[Repetition] This chapter [II] provides an interesting overview of whale shark research. I don’t, however, feel that here (e.g. pages 71-72) or elsewhere in the thesis (Chapter VI) you have made a convincing argument as to why we should be particularly concerned about the potential impacts of climate change on this wide-ranging, highly mobile marine species.

The Introduction of Chapter VI where the examiner’s comment applies now reads:

“... between 30 ºN and 35 ºS (Compagno, 2001; Last & Stevens, 2009). This range has been defined based on occasional occurrences (Compagno, 2001), but most importantly, the range of temperatures where the species is expected to occur: tropical/ warm temperate (Compagno, 2001; Last & Stevens, 2009). Being ectotherms, ambient temperatures directly influence their metabolic processes (Sims, 2003). Whale sharks occur regularly at the ocean surface (Gunn et al., 1999; Rowat et al., 2007; Thums et al., 2012) in association with sea surface temperatures (Sequeira et al., 2012) possibly to assist thermoregulation (Thums et al., 2012). They can also dive deeply, opening
speculation on this species’ potential to modify its diving behaviour with increasing surface water temperatures associated with climate change (Hoegh-Guldberg & Bruno, 2010), and use deeper, cooler waters more frequently. However, oxygen limitation might pose a problem for extended time spent in deeper waters (Graham et al., 2006; Pörtner, 2001; Prince & Goodyear, 2006) so for this reason, expected warming will likely affect the spatial distribution of whale sharks possibly by promoting shifts in their present geographical range. Moreover, because whale sharks are filter-feeders, these shifts might follow the simultaneous redistribution of plankton (Beaugrand et al., 2009)."

[Repetition] These two chapters would benefit from implementation of an additional (and different) modeling approach (e.g. models with a truncated distribution or boosted regression trees) and/or some sensitivity analyses (e.g. re: generation of the pseudo-absences in for the Atlantic and Pacific Ocean data sets as in Chapter III).

Following the comments of both examiners, I will include an additional modelling approach in this chapter by using the BIOMOD modelling framework. This framework includes multiple habitat suitability modelling techniques, such as generalised linear, additive and boosting models (Thuiller et al., 2009 for details), allowing the explicit treatment of model uncertainties. Results obtained so far with this modelling framework are similar to the predictions obtained with my previous models (see details in Appendix D). However, my models were developed in a hypothesis-based framework, while BIOMOD is instead a model-comparison framework not allowing for direct input on the contribution of each variable (i.e., variables can be included as 2nd- or 3rd-order polynomials). Therefore, I chose to keep GLMMs as the main modelling
technique. I have now added the following sentence to the Methods (and in the
Results I point the reader to the results presented in supplementary material):

“To account for variability in habitat suitability resulting from different modelling
approaches, we have also run some additional analysis using the BIOMOD
modelling platform (Thuiller, 2003). However, being a model-comparison
framework (rather than hypothesis-based) BIOMOD does not allow for direct
input on the contribution of each variable. Details are presented in
supplementary information (Appendix D).”

Regarding the sensitivity analysis on pseudo-absence datasets, as answered
previously (comments to Chapter V), to assess the influence of pseudo-absences
location choice in the models results, I always generated multiple pseudo-absence
datasets prior to modelling. I have now clarified this in the Methods:

“To assess the influence of the chosen pseudo-absences, we generated each
set 10 times prior to their inclusion in the spatial GLMM.”

[Repetition] At minimum, I would like the candidate to include some discussion
of the potential biases and limitations arising from the data used and/or the
modelling approaches undertaken in these two chapters. Such admissions and
discussions would strengthen (not weaken) these chapters and their related
manuscripts intended for publication in the peer-reviewed literature. The
candidate does an excellent job of reviewing these in Chapter VII, her General
Discussion, but this chapter will not be published in the primary literature, and
this material is critical to each of Chapters 5 and 6 in order for readers to be able
to evaluate the research.
I have now included a discussion of the potential biases and limitations arising from the data and approach used. Text added to the discussion section is copied below:

“According to Thuiller (2004), projections on climate change impacts in species occurrence should be developed considering the species’ entire range. In our case, this range is circumglobal, which is the reason behind our extrapolation to the total area where predictor values are within the range used to calibrate the model. The climate-induced impact on species occurrence is usually investigated by the coupling climate forecasts with species distribution models (Araújo et al., 2005; Hannah et al., 2002). However, this coupling adds extra limitations (as specified in the Introduction), and results from such modelling frameworks are now considered to provide only a first approximation of possible changes. In the same way, our predictions of future whale shark habitat changes are intended only to be informative, providing only a baseline for temperature-dependent predictions. For a more comprehensive study on the influence of temperature in the occurrence of this species, not only should more climate-change scenarios be tested (including multiple global circulation model outputs), other species distribution models could be generated to assess variability in habitat suitability results, examine error and determine confidence intervals (Araújo & New, 2006; Thuiller, 2003). We extended our main results using the BIOMOD modelling platform (see Appendix D – BIOMOD run). The ensemble habitat suitability results we obtained were similar to those derived from GLMM (Figure D2 in Appendix D).

Moreover, to improve our model predictions, more sightings data (ideally with circumglobal coverage) should be used. We have used the largest, most extensive dataset derived from tuna fisheries logbooks; however, it only covered
a portion of areas managed by Regional Fisheries Management Organisations (New Zealand Ministry for Primary Industries website at fs.fish.govt.nz/Page.aspx?pk=103&tk=319). Therefore, there is scope to develop our models based on data from other areas, such as the eastern Pacific Ocean (managed by the Inter-American Tropical Tuna Commission), which could assist improving the overall model results. As suggested in the previous chapter, a more collaborative effort between researchers and fisheries management organisations would assist improving scientific research and, in this specific case, potential improving our model fitting.

The opportunistic dataset we used only contained whale shark presences relative to the sea surface only. The latter also applies to the set of environmental predictors we used to build the models. While the first drawback could be addressed by randomly generating pseudo-absences (Barbet-Massin et al., 2012; Phillips et al., 2009) and repeating this procedure multiple times to account for any biases derived from pseudo-absence selection, the second is not simple to address. Environmental (and sightings) data spanning subsurface dynamics are not currently available for practical model development, especially where the study area covers most of the global ocean extent. Further, the whale shark sightings dataset I used included recent sightings (up to 2010) that might herald future patterns of distribution change. Recently, whale sharks have been seen in locations where they were not expected to occur (Duffy, 2002; Rodrigues et al., 2012; Turnbull & Randell, 2006), and so the data I used could already epitomise shifts in this species’ occurrence. It is therefore possible that my results show an already biased range of suitable habitat for this species.
However, due to the large area under study and the limited data on presences, excluding part of the dataset available was not an option.

Despite the current lack of alternatives to fisheries-collected sightings in the open ocean (Jessup, 2003), combining remotely sensed data with opportunistically recorded occurrences has, however, proved its worth for estimating whale shark distributions (Kumari & Raman, 2010; McKinney et al., 2012; Sequeira et al., 2012). Indeed, most of our understanding of whale shark ecology and biology hails from opportunistically collected datasets by the eco-tourism (e.g., Wilson et al., 2001) or fishing industries near shore at aggregation sites (e.g., McKinney et al., 2012). As such, we should endeavour to analyse all available datasets to conserve this pan-oceanic species and largest remaining fish – refusal to examine multiple lines of evidence risks imperilling the species further through lack of ecological understanding.”

[Repetition] -show the equations for the models being used, so that the reader can more clearly see what is being done;

I have now added the following text to the caption for Table 11:

“The GLMM (with a logit link function) can be expressed as:

\[ \text{logit}(\text{Presence}) = \alpha + \beta X_i + \gamma Z_i + \epsilon_i + \log (\text{fishing effort}) \]

where Presence is the expected mean probability of sightings occurrence. \(X\) and \(Z\) represent the fixed and random covariates used in the models: the environmental predictors and the spatial grid, respectively. \(\beta\) and \(\gamma\) represent the coefficients associated with the fixed and random effects, respectively, and
$\alpha$ is the intercept. The log of fishing effort was included as an offset. The index $i$ corresponds to the number of observations among grid-cells."

[Repetition] -provide some more detail on the exact nature of the logbook records of whale shark occurrences, including quantitative information or comment on the per cent of logbooks in which whale shark occurrences were consistently, or ever, recorded (i.e. does the entire fleet report whale sharks?). How might this alter your interpretation of the zeroes (i.e. do zeros only reflect true absences and failure to detect whale sharks, or do zeros reflect these two cases as well as 'failure to report whale sharks)? How might this alter how you treat effort' in your models? OR Are these data not provided on a trip by trip, or vessel by vessel basis, but merely aggregated as per the spatial precision described (e.g. p. 139). These types of detail need to be made clear for readers.

The previous answer to the same comment in Chapter V applies here, and because the dataset was already described in more detail in that chapter, I have only included the following sentence in the data description section:

“No information was made available on vessel or trip units for any ocean.”

[Repetition] -clearly explicate the relationship between the sightings and effort data (i.e. the fact that the two are not paired, and hence candidate cannot implement traditional 'CPUE'or'SPUE' models, in which the two are available at the same scale); "

The previous answer to the same comment in Chapter V also applies here. The following change was made in the text:
“To account for spatial bias in sampling effort without reducing the spatial resolution of the dataset, we included this variable as an offset, and used the highest resolution available in each ocean.”
The beginning of the Introduction is more or less a copy-paste from the other articles submitted or published. I understand it is not always easy to find inspiration but please try to change the sentence and the associated references (This whole sentence "Species distribution models have been used widely to estimate habitat requirements and range sizes (Guisan & Thuiller, 2005; Guisan &, Zimmermann, 2000; Hirzel et al., 2002; Phillips et al., 2009).

This sentence was now changed to:

“...However, assessing how much the distribution of species will be affected by climate change implies that the current distribution of the species is adequately identified.

A species' distribution and its habitat requirements can be estimated by modelling information on occurrence together with environmental correlates (Kearney & Porter, 2009). Despite the diverse use of species distribution models in terrestrial environments, until recently only a handful of these models had been developed for the marine environment...”

Despite their diverse use, until recently only a handful of these models had been developed for the marine environment (Elith &...") appears several times in the PhD thesis.

Changed as outlined above. However, each chapter needs to stand alone as an independent publication, and therefore, general content reflecting the same information (e.g., the current and previous use of species distribution models) are
necessarily repeated within the thesis. This was highlighted in the General Introduction:

“All my data chapters were written as ‘stand-alone’ manuscripts, and therefore some of the material presented in the Introduction and Methods sections are repeated in each chapter.”

This whole part is not clear: "Despite the known limitations of these models, such as broad spatial scales of prediction (Hannah et al., 2002; Seo et al., 2009) and considerable variation among climate forecasts (Araújo et al., 2005; Hannah et al., 2002; l{n}uTt{t}i et al., 2010), their biological relevance can be improved by down-scaling and coupling to ecological processes (Araújo et al., 2005; Fordham et al., 2011; Fordham et al., 2012; Hannah et al., 2002)."

I have now changed the text to:

“Fortunately, statistical tools exist to facilitate inference in this regard. For example, known occurrences can be used to define the current distribution of a species, and future distributions can be projected by coupling climate forecasts derived from global circulation models (e.g., Araújo et al., 2005; Cheung et al., 2010). Despite the known limitations of these models, such as biases derived from the input data or from the geographical extent available (Barbet-Massin et al., 2012; Barbet-Massin et al., 2010; Phillips et al., 2009), and uncertainties derived from the models used (both species distribution models and global circulation models) (Buisson et al., 2010), confidence in their application can be improved by using an ensemble result (Araújo & New, 2006). This ensemble result takes into account variability among models used (Hijmans & Graham,
Another important limitation, especially from a conservation perspective, is the coarse resolution of predictions derived from global circulation models, and their consequently limited biological relevance (Hannah et al., 2002; Seo et al., 2009). However, this can be improved by down-scaling the forecasts, coupling them to ecological processes (Araújo et al., 2005; Fordham et al., 2011; Fordham et al., 2012; Hannah et al., 2002).

There are plenty of more relevant references concerning uncertainty, lack of mechanisms etc. Here is a non-exhaustive list:

I have now included most of these references in the Introduction of this chapter (as detailed above).

It is also not very clear what the authors want to explain. I do not understand "such as broad spatial scales of prediction (Hannah et al., 2002; Seo et al., 2009)". I do not see what is the limitation here. Try to be clearer and more precise.

I have now changed this sentence to:

"Another important limitation, especially from a conservation perspective, is the coarse resolution of predictions derived from global circulation models, and their consequently limited biological relevance (Hannah et al., 2002; Seo et al., 2009)."
I have a general problem with the overall methodology here. The authors state that there are known limitations and especially variations when projecting into the future, but they use only one approach without controlling whether the results are strongly biased toward the selected approach. Although I like GLMM, I do not think they are the most suitable tools for such a risk assessment and that I suggest a careful examination of the uncertainty if such results have to be used. I would recommend the authors to have a look at packages such as dismo or biomod2 in R to allow running a set of distribution models and then having a careful examination of the error, confidence intervals. The biomod2 package, which I know most, allows such an approach. In a climate change context, this is something we should always do given the potential use of outputs. Ecologists find logical to use a set of climate change models, but still doubt they have to do the same with ecological models. This is out of sense.

I agree that using a set of ecological models allows for examination of error between different approaches, and I am now including results obtained from the BIOMOD package (see Appendix D). However my study was driven by a hypothesis-testing idea, where the aim was to assess the extent to which the predictor variables were useful to define global suitable habitat for whale sharks by understanding which contribution derives from each predictor and how they might relate to the biology of the species in study. The modelling framework used in BIOMOD is not tailored for testing hypotheses on the predictors used. It instead uses the best fit to the data in each model, which might include multi-order polynomial functions for some of the predictors. Such functions are not easy to interpret in a biological context.
I am also concerned about the use of spatial model in a climate change context. The spatial autocorrelation is certainly due to either dispersal or a missing variable. Both of them are not explicitly modelled here and there is no way to know how they will change into the future. It is impossible to "project" a spatial component and can only lead to spurious projection. I would strongly reconsider the spatial component here, and rather try to build a simplistic diffusion model to account for dispersal.

As stated in my previous answer (Examiners' comment on pp.105, Chapter III), because data on the key aspects of whale shark life history (including dispersal) are still to be measured, no more data are currently available to use as predictor, nor to develop even a simplistic diffusion model (which would at least require a diffusion coefficient). While this has been done at a fine scale (Sleeman et al., 2010b), it is beyond reasonable biological inference to do this on a global scale. As the examiner states, even if these data were available, they could not be explicitly modelled in a climate change context. The random effects added to my models represent unobserved random variation.

However, I disagree with the impossibility to use the spatial component in the models within a forecast framework. The spatial random effect used in my models is simply a code for each 1-° grid cell in the area covered by the fisheries. By including this random effect, I am accounting for the correlation between the environmental predictors within each 1-° grid cell, and excluding its contribution from the models. When using the forecast results from the global circulation models – where resolution was increased from 2.5-° to 0.5-° grid cells using bilinear interpolation – the correlation between points within each 1-° grid cell will still be present. Including this random
The effect in the model set in a forecast framework continues reducing possible bias derived from within 1° grid cells correlation.

**Fig 26:** It is surprising to see some red crosses (around Japan and Americas) that are supposedly in habitats suitability close to 0. Does it mean that the models fail to detect those zones? How would you explain the high suitability in Atlantic whereas very few known aggregations have been spotted?

The models simply did not predict high habitat suitability for those areas, as I state in the Discussion:

“Although whale sharks occur in the Gulf of California (Cárdenas-Torres et al., 2007) and the Galapagos during most of the year, our model did not predict higher suitability there – additional data from the eastern Pacific would probably help improve model performance.”

As I further discussed in that chapter, the relatively low habitat suitability found in the Pacific reflects the much higher effort in the area, and relatively few whale sharks sighted. Because sea surface temperature variation was small within the calibrated area for that ocean (mostly within the range of sea surface temperatures where whale sharks occur) (Sequeira et al., 2012), this predictor had a lower effect in estimating habitat suitability in the Pacific Ocean, which might explain the lower habitat suitability revealed by the models. This was now added to the Discussion as shown below:

“This might explain why our models did not show evidence that sea surface temperature affects whale shark occurrence in the Pacific Ocean, and might
also be part of the reason why the habitat suitability in the entire Pacific Ocean was generally low.”

Moreover, because areas off the coast of America (and Japan) are outside the area for which our model was fitted, and even though we restricted the prediction area to the same environmental envelope used in the model, predictions outside the study area were geographically extrapolated and should be considered carefully. With this chapter, I aimed to provide a modelling tool for exploring the future distribution of whale sharks under anticipated climate change. The final maps of habitat suitability show that I have developed that tool; however, better input data will be needed to improve the realism of the predictions. I have now made a reference to this point in the Discussion (please refer to my answer to the Examiner 1’s comment regarding inclusion of a discussion on the potential biases and limitations of the approach used).

The discussion misses a critical assessment on the quality and reliability of the models in respect to uncertainty (what about plotting the confidence interval or coefficient of variation?), robustness and potential use of such projections.

I have now included an assessment of the potential uses and abuses of the model projections in the Discussion as mentioned previously. However, because I am currently revising the manuscript based on this chapter, this work is still in progress and additional results (e.g., from BIOMOD) will be added to the published version - including an assessment of uncertainty and robustness of the models used in comparison to other modelling tools, and plots of confidence intervals in the outputs generated.
CHAPTER VII

Examiner 1

p. 183 'partialled' -Correct.

Now reads:

“According to my spatially explicit trend results, the number of whale sharks observed by tuna fisheries,...”

p. 183-184: I urge you to carefully reconsider this recommendation. Small fluctuations in relative abundance, of less than the span of 1 generation, do not fit the IUCN Red List Criterion (41 or A2). Such changes could be spurious or cyclical, but clearly should not be extrapolated to the required period of three generations. We, in the conservation community, need to be vigilant about not ‘crying wolf about every observed or inferred population fluctuation.

Following the examiners’ comment, I have now developed on my recommendation (which I keep, and detail below). For the first time and based on direct observations, I have shown evidence that whale shark sightings are declining, at a scale that largely encompasses the species’ known geographical range. Total sightings in both the Atlantic and Indian Oceans have declined by about 50% in the last decade, and in some seasons the reduction reached ~ 70% (e.g., sightings during the second quarter of the year decreased to about 28% in the Atlantic and 35% in the Indian Oceans in the last few decades).
The current whale shark IUCN status (VU A2bd + 3d) reflects that this taxon is vulnerable (i.e., faces a high risk of extinction in the wild) based on the (best) evidence available showing a:

- (A) reduction in population size derived from
- (2) observed, estimated, inferred or suspected population size reduction of ≥ 30% over the last 10 years or three generations (whichever is the longer), where the reduction or its causes may not have ceased OR may not be understood OR may not be reversible, based both on
  - (b) an index of abundance appropriate to the taxon, and on the
  - (d) actual or potential levels of exploitation

I have developed on my recommendation that whale sharks should be considered for the status of Endangered, specifying that the criteria used to recommend this status are:

- (A) reduction in population size derived from:
  - (2) observed, estimated, inferred or suspected population size reduction of ≥ 50% over the last 10 years or three generations (whichever is the longer), where the reduction or its causes may not have ceased OR may not be understood OR may not be reversible, based on
- (a) direct observation and
- (d) actual or potential levels of exploitation

and from:

- (4) observed, estimated, inferred, projected or suspected population size reduction of ≥ 50% over any 10 year or three generation period, whichever is longer (up to a maximum of 100 years in the future), where the time period must include both the past and the future, and where the reduction or its causes may not have ceased OR may not be understood OR may not be reversible, based on
  - (d) actual or potential levels of exploitation

Criterion C might also apply:

- (C) population size estimated to number fewer than 2500 mature individuals and
  - (2) continuing decline, observed, projected, or inferred, in numbers of mature individuals and
  - (a) Population structure in the form of:
    - (i) no subpopulation estimated to contain more than 250 mature individuals, or

I suggest therefore a change of status from **VU A2bd + 3d** to **EN A2ad + 4d** or possibly, **EN C2ai**. Below I detail the changes made to the General Discussion:

“According to my spatially explicit trend results, the number of whale sharks observed by tuna fisheries, which cover a large extent of the three major oceans, has been reduced to about 50% in the last decade, both in the Atlantic
and Indian Oceans (there are no data available prior to 2000 for the Pacific Ocean). The causes for this reduction at such a broad spatial scale (possibly reflecting a global population trend) are not yet understood and are potentially still occurring and possibly irreversible. These characteristics are captured within the IUCN Red List Criterion A for *Endangered* species. Moreover, considering the life span of whale sharks, 10 years might correspond to only 10% of a whale sharks lifetime (or even up to 25% of a single generation). As I suggested in Chapter V, despite the possible cyclical occurrence of whale sharks in each ocean basin, the observed reduction in sightings is occurring simultaneously in a large part of the species known geographical range. Due to their long life span, late age at maturity, and the relatively little information available on their reproduction, this species might actually already be facing a high risk of extinction in the wild from past and ongoing exploitation and climate warming. Therefore, I recommend that it should be considered at least for the status of *Endangered*.

Other criteria from the IUCN Red List class of *Endangered* might also apply. For example, Criterion C relates to the total number of mature individuals. Most whale sharks spotted in specific aggregations are immature (Graham & Roberts, 2007; Heyman *et al.*, 2001; Rowat *et al.*, 2011; Wilson *et al.*, 2001), and available population estimates are around 300-500 individuals for a single site at Ningaloo, Australia (Meekan *et al.*, 2006). Moreover, whale sharks are thought to attain maturity at ~ 9 metres in total length (Colman, 1997), and there is evidence that the average size of sighted sharks is declining (Bradshaw *et al.*, 2008). Considering that whale sharks are only rarely sighted, that most of the sightings are immature animals, and viewing aggregations as possible sub-
populations, Criterion C2ai could also apply: “population size estimated at < 2500 mature individuals, in continuing decline (observed, projected or inferred) and with < 250 mature individuals per subpopulation.

In addition to the suggestion of status change from Vulnerable (currently based on the IUCN Red List Criterion A2bd+3d) to Endangered (based on Criterion A2ad + 4d or C2a1), my research also provides an updated global map reflecting seasonality in whale shark habitat suitability, which can now be added to the whale shark section of the IUCN Red List.”

p. 184 ‘strongly suggest a connection’ This overstates what your evidence shows. Please reword accordingly.

Deleted “strongly”. Now reads:

“While focusing on the global patterns of whale shark ecology and connectivity, my research highlighted the following specific findings:

1. Timing and distribution patterns of whale shark occurrence suggest a connection between several aggregation sites among the three largest ocean basins (Chapter II)”

p. 188 ‘preserve’ Please distinguish between ‘preserve’ and ‘conserve’. Do you really mean preserve?

In the text below, I used the word ‘preserve’ to mean the intention of keeping whale sharks safe (i.e., protected from danger). Although the word ‘conserve’ would also
apply (meaning that the whale shark stock should be managed wisely, i.e., preventing depletion), I have replaced ‘preserve’ with ‘protect’, as shown below:

“...and despite whale sharks being classed as Vulnerable in the IUCN Red List (www.iucn.org), a global effort to protect whale sharks is only indirectly supported through control of international trade of specimens (Appendix II of the Convention of Trade in Endangered Species of Wild Fauna and Flora; www.cites.org/eng/app/appendices.php).”

p. 188. What about the Convention on Migratory Species?

I have now included a reference to the Convention on Migratory Species. Text now reads:

“My results provide a strong incentive for current whale shark management policies to be revised. There are currently no global measures in place to limit the exploitation of live whale sharks (Rowat & Brooks, 2012). Whale sharks were recently included in Appendix II of the Convention on Migratory Species (CMS, 2010), which could lead to the development of international conservation measures (Rowat & Brooks, 2012). However, current management of whale sharks occurs mostly within confined tourist locations through the imposition of a maximum number of tourism operator licenses, and by implementing codes of conduct to minimise disturbance in some countries (Cárdenas-Torres et al., 2007; Pierce et al., 2010; Quiros, 2007). In the 30 years after the first initiative for international management of marine resources (UNCLOS, 1982), and despite whale sharks being classed as Vulnerable in the IUCN Red List (www.iucn.org), a global effort to protect whale sharks is only indirectly
supported through control of international trade of specimens (Appendix II of the Convention of Trade in Endangered Species of Wild Fauna and Flora; www.cites.org/eng/app/appendices.php). Also, national bans have been imposed on targeted commercial fisheries of whale sharks, but they lack enforcement (Riley et al., 2009), and the global-scale fisheries interaction with whale sharks is currently not regulated in any way.”

p. 189 - 195 This is an excellent discussion of potential biases and limitations, and should be worked into each of the relevant chapters (or at minimum the relevant manuscripts for submission to peer-reviewed journals), as previously discussed.

I have now included this in Chapters V and VI, as indicated in the answers to examiner comments made on each section of the thesis.
Page 183: Although I agree that the PhD is a very interesting and well-conducted one, I found the first paragraph quite over-stated. Cheung and colleagues have developed far more advanced species distribution models based on Metabolic Theory, physiology and trophic interactions.

The work by Cheung et al. (2011; 2009; 2010) has provided great advances in projecting possible impacts of climate change on exploited marine fish and invertebrates by using a model integrating population dynamics with a bioclimatic envelope. Because the models I developed here were species-specific, and due to the lack of knowledge on the biology of whale sharks, coupling the habitat suitability and temporal models to population dynamics was not possible. I have now changed the first paragraph to:

“My research provided the first ocean-scale and global distribution models of the world’s largest fish, the whale shark (*Rhincodon typus*). Here, I applied a novel and comprehensive modelling framework that took spatial variation and sub-optimal sampling effort data into account while teasing out temporal patterns. I showed how appropriate statistical approaches can be applied to oceanic species for which data are incomplete, fragmented and uneven, such as for the whale shark, highlighting the value of unique datasets collected opportunistically. My approach also provides a mathematical platform for marine climate change predictions that have typically lagged behind terrestrial systems, and shows the possible impacts of a warming climate on a pan-oceanic shark species.”
REFERENCES


Baum JK, Blanchard W (2010) Inferring shark population trends from generalized linear mixed models of pelagic longline catch and effort data. Fisheries Research, 102, 229-239.


Bradshaw CJA, Fitzpatrick BM, Steinberg CC, Brook BW, Meekan MG (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.


Field IC, Meekan MG, Buckworth RC, Bradshaw CJA (2009) Susceptibility of Sharks, Rays and Chimaeras to Global Extinction (Chapter 4). In: *Advances in Marine Biology*.


Occurrences of whale shark (Rhincodon typus Smith, 1828) in the Saint Peter and Saint Paul archipelago, Brazil. Brazilian Journal of Biology, 68, 385-389.


and Management. Conference Overview, Abstracts and Supplementary

Proceedings - CSIRO Marine and Atmospheric Research, Australia, 14-19.

(Rhincodon typus) is a livebearer: 300 embryos found in one ‘megamamma’


fish catch. Marine Environmental Research, 70, 162-170.


marine mammal species using a relative environmental suitability (RES)


Kearney MR, Wintle BA, Porter WP (2010) Correlative and mechanistic models of

species distribution provide congruent forecasts under climate change.


Kelly NM, Fonseca M, Whitfield P (2001) Predictive mapping for management and

conservation of seagrass beds in North Carolina. Aquatic conservation:


of the Royal Society B, 276, 4095-4103.


Research, 94, 6159-6168.


Machete M, Afonso P (pers. comm.) Azores Fisheries Observer Programme (*Programa de Observação para as Pescas dos Açores - POPA*). Dept. of Oceanography and Fisheries; IMAR - University of the Azores.


Mooloolaba, Australia. BBRG-12; National Research Institute of Far Seas Fisheries, Fisheries Research Agency, Japan.


Paternity analysis in a litter of whale shark embryos. *Endangered Species 

Schmidt JV, Schmidt CL, Ozer F, Ernst RE, Feldheim KA, Ashley MV, Levine M (2009) 
Low genetic differentiation across three major ocean populations of the whale 
doi:4910.1371/journal.pone.0004988.


models: implications for conservation planning under climate change. *Biology 
Letters, 5*, 39-43.

Sequeira A, Ferreira JG, Hawkins AJS et al. (2008) Trade-offs between shellfish 
aquaculture and benthic biodiversity: A modelling approach for sustainable 

Sequeira A, Mellin C, Delean S, Meekan MG, Bradshaw CJA (in press) Spatial and 
temporal predictions of decadal trends in Indian Ocean whale sharks. *Marine 
Ecology Progress Series*.

Sequeira A, Mellin C, Rowat D, Meekan MG, Bradshaw CJA (2012) Ocean-scale 
prediction of whale shark distribution. *Diversity and Distributions, 18*, 504-518.

Sequeira AMM, Mellin C, Meekan MG, Sims DW, Bradshaw CJA (2013) Inferred 
global connectivity of whale shark *Rhincodon typus* populations. *Journal of 
Fish Biology, 82*, 367-389.
distribution based on ecological information for extent of occurrence
assessment. Biological Conservation, 135, 341-351.

Sims DW (2003) Tractable models for testing theories about natural strategies:
foraging behaviour and habitat selection of free-ranging sharks. Journal of
Fish Biology, 63, (Supplement A) 53-73.

Sims DW, Reid PC (2002) Congruent trends in long-term zooplankton decline in the
north-east Atlantic and basking shark (Cetorhinus maximus) fishery catches

Sims DW, Southall EJ, Richardson AJ, Reid PC, Metcalfe JD (2003) Seasonal
movements and behaviour of basking sharks from archival tagging: no
evidence of winter hibernation. Marine Ecology Progress Series, 248, 187-
196.

Sims DW, Southall EJ, Tarling GA, Metcalfe JD (2005) Habitat-specific normal and
reverse diel vertical migration in the plankton-feeding basking shark. Journal

Sleeman JC, Meekan MG, Fitzpatrick BJ, Steinberg CR, Ancel R, Bradshaw CJA
(2010a) Oceanographic and atmospheric phenomena influence the
abundance of whale sharks at Ningaloo Reef, Western Australia. Journal of

Sleeman JC, Meekan MG, Wilson SG et al. (2007) Biophysical correlates of relative
abundances of marine megafauna at Ningaloo Reef, Western Australia.
Marine and Freshwater Research, 58, 608-623.

Sleeman JC, Meekan MG, Wilson SG, Polovina JJ, Stevens JD, Boggs GS, Bradshaw
CJA (2010b) To go or not to go with the flow: environmental influences on


Weng KC, Castilho PC, Morrissette JM et al. (2005) Satellite tagging and cardiac physiology reveal niche expansion in Salmon sharks. Science, 310, 104.


conducted on a whale shark (*Rhincodon typus*) in the Indian Ocean. *Fisheries Oceanography, 16*, 547-554.


World Widlife Foundation (2012) Dead whale shark found off Manila Bay. In: *WWF - Philippines, Newsroom*. Available at: 


