

**Reef manta rays, *Mobula afredi*, of the Chagos Archipelago: Habitat use and the effectiveness of the region's marine protected area.**

by

**Joanna Lee Harris**

Thesis submitted to the University of Plymouth  
in partial fulfilment of the requirements for the degree of

**MRes Applied Marine Science**

**University of Plymouth**  
**Faculty of Science & Engineering**

in collaboration with  
**Manta Trust, Dorset, UK.**

September 2019

Journal format: Marine Biodiversity Records

**MRes Dissertation licence**

This material has been deposited in the University of Plymouth Learning & Teaching repository under the terms of the student contract between the students and the Faculty of Science & Engineering. The material may be used for internal use only to support learning and teaching. Materials will not be published outside of the University and any breaches of this licence will be dealt with following the appropriate University policies.

## **Abstract**

*Background:* The subpopulation of reef manta ray (*Mobula alfredi*) of the Chagos Archipelago; British Indian Ocean Territory (BIOT) inhabit a marine environment which is almost completely devoid of anthropocentric stressors. As such, they offer the opportunity to study the natural processes which influence habitat use so to establish a natural baseline to inform conservation planning for the species. However, to date, they are one of the most understudied of the known subpopulations in the world. Here, photo-ID sighting records, satellite tag data, remote sensing and various modelling techniques are integrated to describe the current understanding of the subpopulation and to explore the influences of foraging and the utilisation of aggregation hotspots. Additionally, an assessment of the effectiveness of the BIOT marine protected area (MPA) is initiated.

*Results:* Photo-ID surveys (2006-2018) recorded 123 sightings of *M.alfredi* (female = 65, male=58) predominantly engaged in feeding at 17 locations across four atolls where some degree of demographic segregation was evident. In 2016, SPOT5 tags deployed to five *M.alfredi* tracked the individuals for 10-35 days (mean = 23 days). All tagged manta rays remained close to atolls aggregating predominantly at Egmont Atoll (hotspot). All remained within the boundary of the MPA where they travelled between 9.6-576 km (mean = 183.9). One manta ray travelled 548 km in 29 days. State-space modelling indicated the manta rays were predominantly engaged in foraging behaviour (ARS). Boosted regression trees identified a shallow mixed layer depth, depth and high sea surface temperature (SST) and low current speeds and high SST as important drivers of ARS and hotspot utilisation respectively. A specific chlorophyll-*a* niche (0.08-0.1 mg/m<sup>3</sup>) also appears to be important in both models. Diving behaviour ascertained from tag data showed a reverse diel vertical migration pattern, however, manta rays were also recorded at depths of up to 400m during the day.

*Conclusion:* The summation of the current understanding of *M.alfredi* of BIOT given here provides an essential baseline for future research into the subpopulation. The initial assessment of the effectiveness of the protection offered by the BIOT MPA can help to inform management planning of other MPAs in the Indian Ocean.

**Keywords:** Photo-identification, Satellite telemetry, Habitat use, Marine protected area, Site fidelity, State-space analysis, Boosted regression trees.

## 1. Introduction

The reef manta ray (*Mobula alfredi*) is one of the world's largest planktivorous elasmobranch of the family Mobulidae [1]. The species have conservative life-history strategies which are characterised by slow growth, late maturation [2] and low fecundity [3,4]. The population is widely distributed in highly fragmented subpopulations [5,6] which appear to rely on atoll ecosystems [7] throughout the tropical and sub-tropical waters of the Indo-West Pacific Oceans [5,6]. As a large-bodied planktivore, dense aggregations of prey that provide a high energy intake are essential to offset the energetic cost of feeding [8]. As the waters they inhabit are largely oligotrophic, the species display a distinct movement ecology which reflects resource and habitat requirements [9]. For example, satellite tagging, acoustic telemetry analysis and photographic identification studies have shown the species exhibit exceptional site fidelity [10–13]. Home range size varies extensively between subpopulations, but each appears to concentrate the majority of their activities at specific locations [14,15] with individuals observed making migrations of up to 500 km between sites [16,17]. As their aggregation behaviour is predictable, the species are particularly vulnerable to exploitation [18,19].

Fisheries are one of the greatest anthropogenic threats to mobulids [20], particularly in the Indian Ocean, where some bordering countries have large targeted fisheries [21]. These fisheries have little effective regulation or enforcement and have been described as the most exploited yet poorly managed in the world [22,23], an issue which is exacerbated by an increase in unregulated small-scale fisheries [24]. Commonly exploited for their gill rakers; pre-branchial appendages which are cartilaginous sieve-like structures used to filter zooplankton from the ocean [25,26], *M. alfredi* have been subject to catastrophic declines [18]. They are now listed as vulnerable to extinction on the IUCN Red List of Threatened Species [27], in Appendix II of the Convention of the International Trade in Endangered Species (CITES) [2] and on Appendix I and II of the Convention on the Conservation of Migratory Species of Wild Animals (CMS) [28].

As well as target and bycatch fisheries, habitat degradation and extreme touristic pressure threaten the survivorship of *M. alfredi* [29–32]. The use of marine protected areas (MPA) as a conservation measure has proliferated in recent years [2,12,33]; however, the effectiveness of this conservation strategy is controversial [34–36]. Concern is expressed

over the limited size of many MPAs which can leave a substantial protection shortfall for highly mobile species [16,19,34–36]. Various management schemes that permit different activities of varying intensities within an MPA are also a concern [19,32]. The decision to prohibit certain activities is often to remediate visible adverse impacts on *M.alfredi* [37] for example, in the Maldives, increased disturbance by tourist led to the establishment of Hanifaru Bay MPA [32]. However, the apparent absence of causal evidence of anthropogenic impacts upon a species or an ecosystem is not reliable evidence of no effect [38]. Furthermore, many studies that examine the effect of anthropogenic impacts on *M.alfredi* focus on subpopulations that have been subjected to anthropogenic stressors for decades [11,32,36,39]. An essential missing component in such studies is a baseline that describes how the species functions in an undisturbed environment.

The Chagos Archipelago; British Indian Ocean Territory (BIOT), situated in the central Indian Ocean, has been virtually uninhabited for over 50 years [22]. Due to the lack of human influence, such as coastal development and anthropogenic pollution, the region is considered pristine [40]. It also has enhanced fish abundance, and there is evidence of more natural behaviours by some species [22]. For example, the moray eel (*Gymnothorax pictus*) feed during the day in BIOT and even venture onto dry land to capture their prey which is a natural dynamic which only occurs in the absence of human activities [41].

Although uninhabited, both in-shore and pelagic fisheries have impacted BIOT including a 90% decline in reef sharks between 1975 and 2006 [22], and approximately up to 300 tonnes of both tuna and reef fish were caught in the region annually until 2010 [22]. In 2010, owing to the regions unique and pristine marine environment a no-take MPA was established encompassing the regions entire exclusive economic zone (EEZ) (640,000km<sup>2</sup>) except for a 3nm exclusion around the boundary of Diego Garcia atoll [22] making it one of the largest MPAs in the world. The archipelago is host to a subpopulation of *M.alfredi* which offer the opportunity to study the species in a near-natural environment in the absence of anthropogenic influence where the association between environmental variables, distribution and behaviour can be developed to assist in establishing a near-natural baseline for the species. With a multitude of threats including widespread target and bycatch fisheries in the Indian Ocean [2,33], BIOT's vast no-take MPA may provide essential refuge for the regions *M.alfredi* subpopulation. Consequently, due to the remoteness of the location, its lack

of inhabitation and strict protection, the subpopulation is largely undocumented, as is the regions physical oceanographic environment [42].

This study aims to describe the current understanding of the *M.alfredi* population of BIOT and elucidate how they utilise their environment. Through the integration of satellite tagging data, remote sensing and various modelling techniques, the relationship between environmental influences and foraging and aggregation behaviour is explored. Furthermore, an assessment of the effectiveness of the region MPA is initiated so to begin to build a framework for future conservation management plans.

## **2. Method**

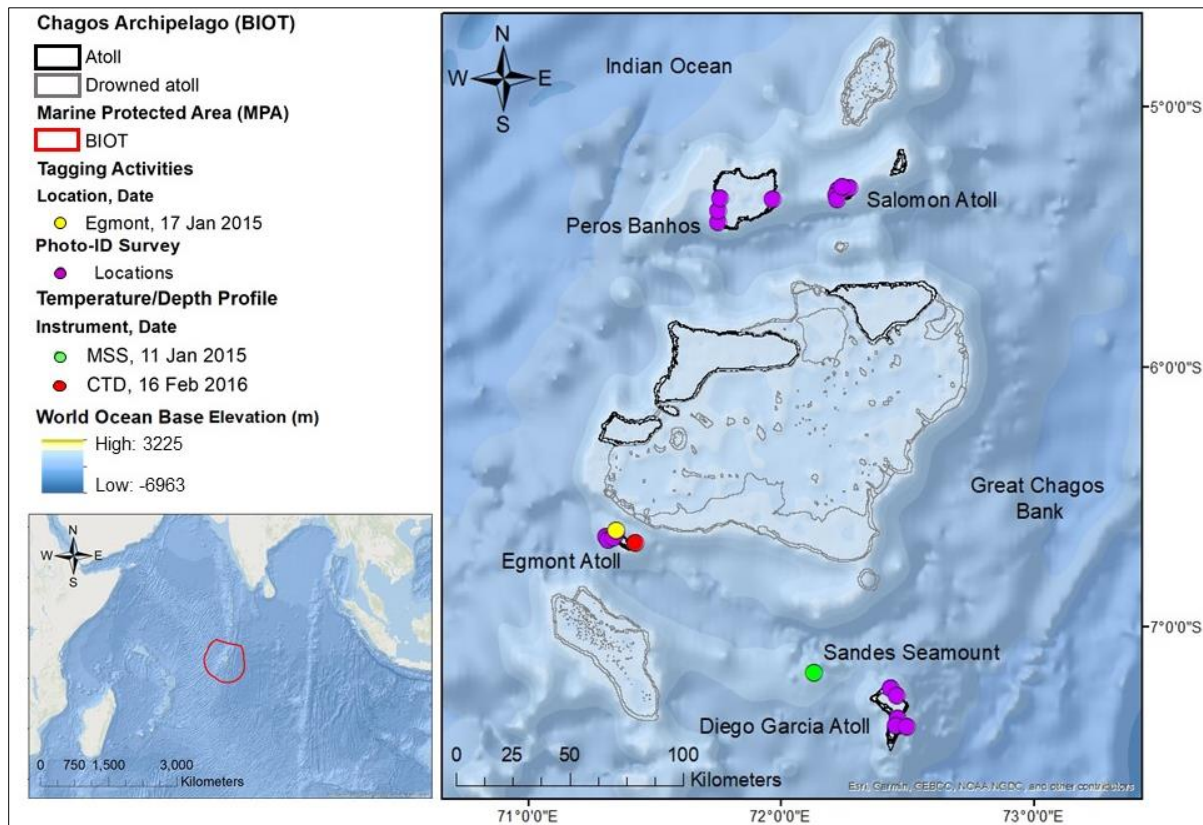
### *2.1. Location*

The Chagos Archipelago British Indian Ocean Territory (BIOT) is made up of a group of atolls and low lying islands located at the southernmost end of the Lakshadweep–Maldives–Chagos ridge, 500km south of the Maldives [22] (Fig. 1).

### *2.2. Photo Identification*

A database of sighting of individually identified *M.alfredi* in BIOT was created and has been maintained by the Manta Trust [44] since 2006. Contributions to the database include a total of 18 surveys conducted at 15 locations across four atolls between February 2006 and May 2018. Survey months vary each year, but all occur between January and September with the majority in February and May. All surveys were undertaken by trained staff members or volunteers from the Manta Trust while freediving from a dedicated research vessel and last on average 180 minutes. Other sightings that have been incorporated into the database include those from other research vessels operating in the region and a small number from private craft. These sightings occurred since 2011 and only during March and May. Sightings are only incorporated into the database if the data collected by the observer meets the Manta Trust protocol whereby a sighting is defined as a confirmed photo-ID of an individual *M.alfredi* on a given day at a defined location. Identification is confirmed via underwater photography of an individual's ventral side markings as well as sex and size; size is a proxy of maturity status [45] (age class). Where possible, the behavioural activity of each individual is also recorded. Behavioural activity is broken down into four major groups: (1) feeding, (2)

cleaning, (3) cruising, and (4) courtship [46]. During an encounter, an individual may undertake several different activities. In these situations, the activity which dominated the encounter is recorded as the primary behaviour.



**Figure 1** - Chagos archipelago British Indian Ocean Territory (BIOT) with tagging activity, photo identification (photo-ID) survey and temperature and depth profile locations. Inset: Indian Ocean with the BIOT marine protected area (MPA).

### 2.3. Satellite Tagging

Tagging activities were carried out at Egmont Atoll (Fig. 1) during the Chagos Pelagic Expedition on 17<sup>th</sup> January 2015 while freediving from a Fast Rescue Craft (FRC) deployed from the Pacific Marlin (BPV). Five smart position or temperature transmitting (SPOT5) tags, each tethered to a titanium anchor with a small diameter cable, were deployed on the right dorsal musculature using a modified Hawaiian hand sling while swimming behind the manta ray. Two of the manta rays were re-sighted within 15-30 minutes, having resumed their foraging activity and appearing unaffected by the tags.

The tags transmit near real-time position data via Argos satellite when the manta ray is at the surface. Tags also recorded the ambient temperature every ten seconds in 12 predefined

temperature bins (<5.9°C, 6-17.9°C in 3°C intervals, 18-29.9°C in 2°C intervals and >30°C). Time-at-temperature (TAT) histograms are then built over four time periods which represent dawn (03:00-09:00 h), day (09:00-15:00 h), dusk (15:00-21:00 h) and night (21:00-03:00 h) whereby the percentage of time spent within each temperature bin during each time period is recorded. The TAT data is limited to measurements recorded preceding a satellite transmission, thus large gaps in the data can occur [47].

The raw Argos data was filtered to remove transmissions received from tags which had detached from the manta ray. The time at which tags detached was determined via assessment of TAT data and the change in temporal consistency and quality of the transmission to Argos [48] (Appendix A). The remaining data were then filtered to include only the manta rays tag identification number, date, time, location and Argos location quality class (LC).

#### *2.4. Time-at-temperature (TAT) analysis*

To infer depth from the time-at-temperature (TAT) data recorded by the tags, vertical temperature and depth profiles were constructed from data obtained at two locations (Fig. 1). The first location (Egmont Atoll) was sampled in 2015 using an ISW Microstructure Sensing System (MSS) equipped with a temperature sensor with a cast depth of 300m. However, the temperature at the maximum depth did not match the minimum temperature observed in the TAT records. Therefore, the second location (Sandes), which was sampled using a compact multi-parameter conductivity, temperature, depth (CTD) system with a cast depth of 400m in 2016 was incorporated into the analysis. Both the 2015 and 2016 vertical temperature and depth profiles were first plotted in MATLAB (Appendix B(i), Fig. B1) to assess their similarity. Although temperatures were different in the surface layers, the profiles converged in deeper waters, and at 300 m the difference in the 2015 and 2016 profiles is <0.5°C (Fig. B1). The convergence occurs because deeper water is more stable, thus less susceptible to temperature fluctuations [49]. Therefore, to estimate the dive depth up to 300 m based on the temperature recorded by the tags in 2015, the Sandes (2015) profile is used as it is spatially and temporally consistent with the tagging data. For depths >300 m, depth from temperature is estimated from the Egmont Atoll (2016) profile which is spatially consistent with the tagging data and converges with the 2015 profile below this depth. All available TAT data was then incorporated into the MATLAB plots to ascertain the approximate depth



occupied by the manta (Fig. B1). After establishing the depth range of each temperature bin, the data was then plotted in R using the ggplot2 R package [50]. To elucidate the diel difference in vertical movement, only periods which were all daylight (09.00-15.00, day) or all darkness (21.00-15.00, night) are considered here. Dawn and dusk categories are not considered as the time range does not accurately represent these time periods; both include hours of daylight and night as well as sunrise and sunset.

### 2.5. Bayesian joint estimation state-space model (hSSM)

The hierarchical version of the Bayesian state-space model of the first-difference correlated random walk with switching, also known as a joint estimation state-space model (hSSM) [51] was fitted to the SPOT5 tag data. The model accounts for location errors in Argos derived position estimates [52], regularises position estimates and estimates the manta ray's behavioural states [51–53]. The model jointly estimates behavioural state across all animals which compensates for the limited number of Argos position estimates for several of the manta rays [54]. The hSSM model was fitted in R 3.5.2 using code adapted from Jonsen *et al.* [55], which is accomplished using the bsam R package [56]. The model is further described with the accompanying R script in Appendix C and Appendix E(iii), respectively. Briefly, the code implements the model using Markov Chains Monte Carlo (MCMC) via JAGS software [57,58]. To best match the mean temporal resolution of the Argos position estimates, the model was fitted with a 12-hour time step to ensure there was at least one observation per timestep. The model assumes that the manta rays travel in a straight line between locations, and the MCMC describe the evolution of the animal's behaviour which is estimated from changes in speed and direction [53]. Two different behavioural states: transiting and area-restricted search (ARS) are defined [53]. Faster more direct movements characterise a transiting state while during ARS, movements are slower with frequent changes in direction [53]. Behavioural state is reported as a discrete parameter; however, the MCMC samples provide means of these variables, which are continuous from 1 to 2 [59]. These means assist in identifying behaviour switches whereby those closer to 1 represent transiting and closer to 2 indicate ARS [53]. Rather than applying a single cut off point of 1.5, the current study applies a more conservative approach of <1.25 and >1.75 indicating transiting and ARS respectively. The approach has been adopted in various studies including of leatherback turtles (*Dermochelys coriacea*) [53], sei whales (*Balaenoptera borealis*) [60] and killer whales (*Orcinus*

*orca*) [59]. Values between these thresholds are regarded as undetermined due to insufficient evidence to distinguish between behaviours [53].

Two MCMC chains of 40,000 samples were run with the first 10,000 from each chain discarded as burn-in then retaining every 30<sup>th</sup> sample of the remaining 30,000 resulting in a total of 1000 iterations. Model convergence was assessed by visual assessment of diagnostic plots to ensure posterior samples were stationary, MCMC chains were well mixed [55], in-chain autocorrelation was satisfactory and the Gelman-Rubin-Brooks potential scale reduction factor ( $\hat{r}$ ) for all parameters were  $<1.1$  [54,55,61] (Appendix C(ii), Fig. C2).

## 2.6. Track analysis

The great circle distance was calculated between hSSM position estimates using the *geosphere* R package [62] to obtain daily and total distance of horizontal swim tracks. The method shares the assumption of the hSSM that the manta rays travel in a straight line (Euclidean distance), thus resulting in the minimum possible distance travelled between points [62]. Swim speed between these points was then calculated using the *dplyr* R package [63]. Track analysis R scripts can be found in Appendix E(iv). The hSSM position estimates and tracks of the individual manta rays were then projected in ArcGIS 10.6.

## 2.7. Hotspot analysis

To identify areas which are predominantly utilised (hotspots), two methods were applied. The first was Kernel density estimation (KDE) which identifies key regions of occupancy using the Kernel Density function of spatial analysis tools in ArcGIS. Secondly, the variation in the intensity at which space is utilised was assessed via Kernel utilization distribution (KUD) modelling [64] using the *kernelUD()* function of the *adehabitatHR* R package [65] (Appendix E(ii)). The model describes the animal's use of space using a bivariate probability density function (the UD) which correspond to the probability density for recapture according to coordinates [66]. A suitable bandwidth (smoothing parameter) was determined manually [67], and 0.2 was selected as the most appropriate for identifying areas of intensive use. The resulting model is then passed to the *getverticeshr()* function of the *adehabitatHR* R package [65]. The function estimates home range size within a specified threshold (e.g. 95%) by dividing the region into a grid then calculating the probability by volume associated with each grid cell [68] to produce probability contours. As the technique cumulates KDE from highest

to lowest density areas, lower probability contours (e.g. 10%) represent areas of the highest observed densities [69]. Therefore, the 10% KUD was used to define the hotspot.

### *2.8. Environmental influences: Boosted regression trees*

Boosted regression trees (BRT) were used to investigate the relationship between (1) hotspot utilisation, (2) behaviour states (transiting and ARS) and environmental variables. As different environmental variables have been shown to influence ARS and transiting behavioural states, each were modelled separately [70]. In the current study, there was an insufficient number of transiting states for analysis; therefore, only ARS was considered.

For the hotspot utilisation model (hereafter the hotspot model), the response variable is all hMMS position estimates observed inside (present) or outside (absent) the hotspot area identified by 10% KUD (Fig. 4).

To elucidate the relationship between environmental variables and ARS behaviour, a binomial response variable of ARS (1) and non-ARS (0) was established. For ARS, all the hSSM ARS position estimates which occurred throughout BIOT were used (Fig. 4). For non-ARS, 171 (1:1) pseudo-absences [71] were generated in ArcGIS. Selecting pseudo-absences from areas of unsuitable habitat may overinflate model performance results [72] therefore, a buffer zone of the maximum daily distance travelled by a manta ray during ARS (23km) was applied to each hSSM ARS position estimate in ArcGIS. The buffer zone represents areas the manta rays could have occupied but did not at the time of observation. Pseudo-absences that were temporally paired with the presence points were then randomly generated [71] within these buffer zones.

The chosen explanatory variables have previously been suggested to influence the foraging and movement ecology of planktivorous marine megafauna as they affect the spatial and temporal distribution of their prey. For both models, a total of nine variables were initially included: chlorophyll-*a* concentration (CHL-A) [73], sea surface temperature (SST) [13,36], sea level anomalies (SLA) [74], distance from Cayula-Cornillon fronts (FRONT) [75], mixed layer depth (MLD) [76], ocean current speed (C-SPD) and direction (C-DIR) [77] and bathymetric variables – depth (DEPTH) and seabed slope (SLOPE) [78]. All data were extracted in ArcGIS using data sources and tools depicted in Table 1. Both CHL-A and SST had a substantial number of pixels with missing data due to cloud cover. Missing data were filled using Marine

Geospatial Ecology Tools (MGET) package [79] to a maximum fill region of 30 pixels using the Del2a method [80,81]. The tool calls MATLAB 2007b via Python from within ArcGIS, which interpolates missing data using Laplacian interpolation [82], which reduced the amount of missing data by 10%.

The MLD was identified using the Hybrid Coordinate Ocean Model (HYCOM). The model incorporates three mixing models as described by Chassignet *et al.* [83] to diagnose mixed layer thickness using the method described by Kara *et al.* [84]. Briefly, using SST and sea surface salinity, a defined temperature jump is converted to an equivalent density jump using the equation of state [83,84]. During seasons when surface heating is strong and in equatorial regions, a criterion of 0.2°C is applied [84,85]. The base of the MLD is identified at the depth where density differs from layer one density by 0.2°C [83,84,86]. Moving down from layer one, the first model layer where the density exceeds layer one density by more than 0.2°C is identified [83,84]. Given the central depth and density of this layer along with the central depth and density of the layer above, linear interpolation is then used to estimate the thickness of the mixed layer [83,84].

For FRONT, thermal fronts were detected in Group for High-Resolution Sea Surface Temperature (GHR SST) Level 4, 8-day SST images hosted by NASA's Physical Oceanography Distributed Active Archive Center [87] using Cayula and Cornillon (1992) single image edge detection (SIED) algorithm with a minimum frontal edge detection threshold of 0.45°C (SST) via MGET [79]. The resulting raster images were incorporated into ArcGIS where they were converted to polyline and distance from the nearest fronts was calculated using the Near Analysis Tool.

**Table 1** - Environmental data used for BRT models with a description of the variables, their source and processing technique.

Variable abbreviation and name		Unit	Spatial/ Temporal Resolution	Description	Source
DEPTH	Depth (Bathymetry)	m	30 arc-second	GEBCO_2014 Grid bathymetry data	GEBCO via BODC [89]
SLOPE	Seabed slope	°	30 arc-second	Derived from GEBCO_2014 Grid using Surface Tools in ArcGIS 10.6	N/A
CHL-A	Chlorophyll- <i>a</i> concentration	mg/m <sup>3</sup>	4 x 4 km / 8-day	Level 3 composites	NASA MODIS Aqua [90]
SST	Sea surface temperature	°C	4 x 4 km / 8-day	Level 3 composites	NASA MODIS Aqua [90]
SLA	SLA	m	0.25° x 0.25° / Daily	Sea surface height anomaly relative to the multi-year mean sea surface height.	DUACS 2014 delayed time, AVISO hosted by CMEMS [91] - Accessed using MGET [92]
FRONT	Distance from Cayula-Cornillon fronts	m	4 x 4 km / 8-day	Binary response rasters layers showing GHRSSST L4 product as time series of SST images with the Cayula and Cornillon SIED algorithm applied to identify fronts (0.45°C threshold). Distance from calculated using Near Analysis Tools in ArcGIS 10.6.	GHRSSST hosted by PO.DAAC [93] - Accessed using MGET [92]
C-DIR	Ocean surface current direction	Cardinal directions	0.33° x 0.33° / 5-day	Surface current data averaged over the top 30m of the ocean calculated via diagnostic model using satellite-derived sea surface height, wind, and temperature and quasi-linear and steady flow momentum equations [94].	OSCAR hosted by PO.DAAC [95] - Accessed using MGET [92]
C-SPD	Ocean surface current speed	m/s	0.33° x 0.33° / 5-day	See above (C-DIR)	OSCAR [95] - Accessed using MGET [92]
MLD	Mixed layer depth	m	8.9 x 8.9 km / Daily	Mixed layer thickness defined as the depth at which the temperature change from the surface temperature is 0.2°C [84,85]. Derived from a 3D ocean environment model established using Navy Coupled Ocean Data Assimilation (NCODA) which assimilates vast amounts of in situ and <i>ex-situ</i> data to accurately nowcast and forecast oceanographic conditions [96].	HYCOM [97] GLBa0.08 - Accessed using MGET [92]

All BRT models were fitted using the `gbm.fit()` function of the `dismo` R package [98]. During preliminary analysis, suitable tree complexity (TC) and learning rates (LR) were determined by fitting models with progressively increasing TC (2, 3, 4 and 5) and decreasing LR (0.01, 0.005 and 0.001) [99,100] (Table D1 and D3). Ten-fold cross-validation (CV) was applied to identify the optimal number of trees [101]. These processes are further described in Appendix D.

Two measures of BRT model performance are common: 1) area under the receiver operating characteristic curve (AUC) test statistic [100,102] (hereafter the AUC method), and 2) proportion (%) of total deviance explained by the fitted model ( $D^2$ ) [99,103] (hereafter the  $D^2$  method). The AUC method provides a threshold-independent metric which reflects the model's ability to accurately classify observations [104]. Two mean AUC values are given for each model, one for its performance classifying the primary sample (training AUC, TAUC) and another for the hold-out sample (cross-validation AUC, CVAUC) [102,105]. The AUC classification ranges from 0-1 whereby:  $\leq 0.5$  (fail), 0.6–0.7 (poor), 0.7–0.8 (acceptable), 0.8–0.9 (good),  $> 0.9$  (excellent) [106]. The difference between the TAUC and the CVAUC ( $\Delta$ AUC) indicates the level of overfitting of the primary sample [102]. Therefore, better model performance is categorised by higher AUC values for both TAUC and CVAUC but a lower  $\Delta$ AUC [102]. The  $D^2$  corresponds to the percentage of deviance for the null model explained by the fitted model [103], which can be calculated using the following form [99]:

$$D^2 = \frac{(\text{Null deviance} - \text{CV residual deviance})}{\text{Null deviance}}$$

For the current study, the AUC method is the primary measure of performance, and the  $D^2$  method is used as a secondary measure to distinguish between models with the same AUC values.

Following preliminary analysis, the hotspot model was fitted with a TC of 2, and an LR of 0.005 and the ARS model was fitted with a TC of 5 and an LR of 0.001 (Table D1 and D3). Both models had a bag fraction of 0.5. To ensure model performance was not being hindered by non-informative predictor variables, the `gbm.simplify()` function of the `dismo` R package [98] was applied to these models [101]. The function tests the effect of the elimination of predictor variables, excluding up to a specified maximum [101]. Here, the maximum was set to seven, which is the most allowed when the model contains nine predictor variables.

The results of the models reflect the relative influence of predictor variables, which is measured by averaging the number of times a variable is chosen for splitting and the squared improvement resulting from these splits [101]. The result is then scaled to 100 across all the variables [101]. The predictor variables with the highest numbers indicate a stronger influence on the response variable [101].

Partial dependency plots were generated using the `gbm.plot()` function. The plot represents the effect of the explanatory variable after accounting for the mean effects of all other explanatory variables [107]. The source code for the function was adapted to display rugs as percentiles for presence and absence separately. The BRT also models interactions during the course of building the model [101]. Important interactions were ascertained via the `gbm.interactions()` function which performs predictions on a linear scale for a grid of paired variable with all other variables held to their respective mean [101]. The nature and magnitude of these interactions [101] were visualised using `gbm.perspec()` which requires the `dismo` [98] and `gbm` [108] R packages. A marginal effect of zero indicates that no interactions were fitted [101]. Appendix D provides a detailed description of all methods used to fit and assess the BRT models and accompanying R scripts are in Appendix E(vi).

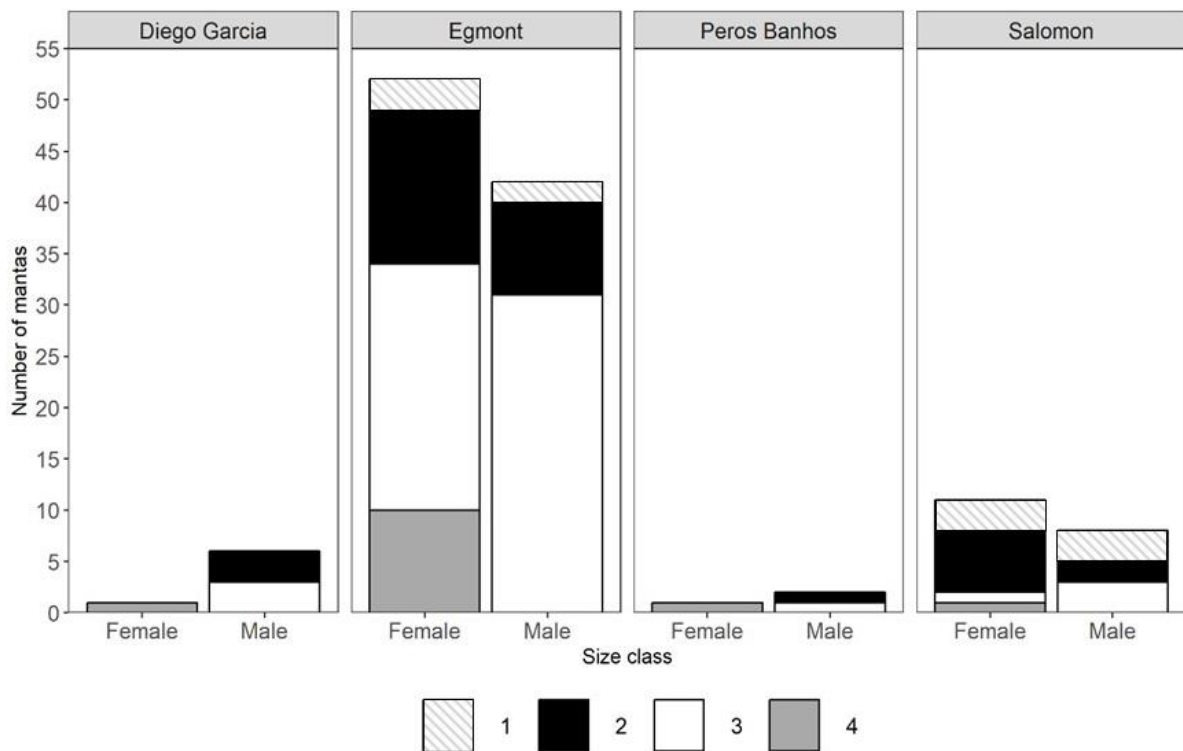
### **3. Results**

#### *3.1. Photo Identification*

Overall, there were 129 sightings of 108 (female = 55, male = 47, unknown = 6) individually identified *M. alfredi*. The six sightings where sex was not recorded have been excluded from the results. As size class and behaviour can change between sightings, the remaining 123 observations are described. A total of 65 female and 58 male *M. alfredi* were recorded, of which there was a higher proportion of adult males (42, 72%), than adult females (12, 18%) (Fig.2 and Table 2).

The highest numbers of *M. alfredi* were sighted at Egmont Atoll where there were more females (52) which were predominantly juvenile (42, 81%) than males (42) of which the majority were adult (33, 79%) (Fig. 2 and Table 2). At both Diego Garcia and Peros Banhos Atoll, more males have been sighted than females, and all females were adults while there was an even number of subadult and adult males. At Salomon Atoll, more females (11) than

males (8) have been recorded, and all but one female were juvenile while the males were almost evenly split among size classes.



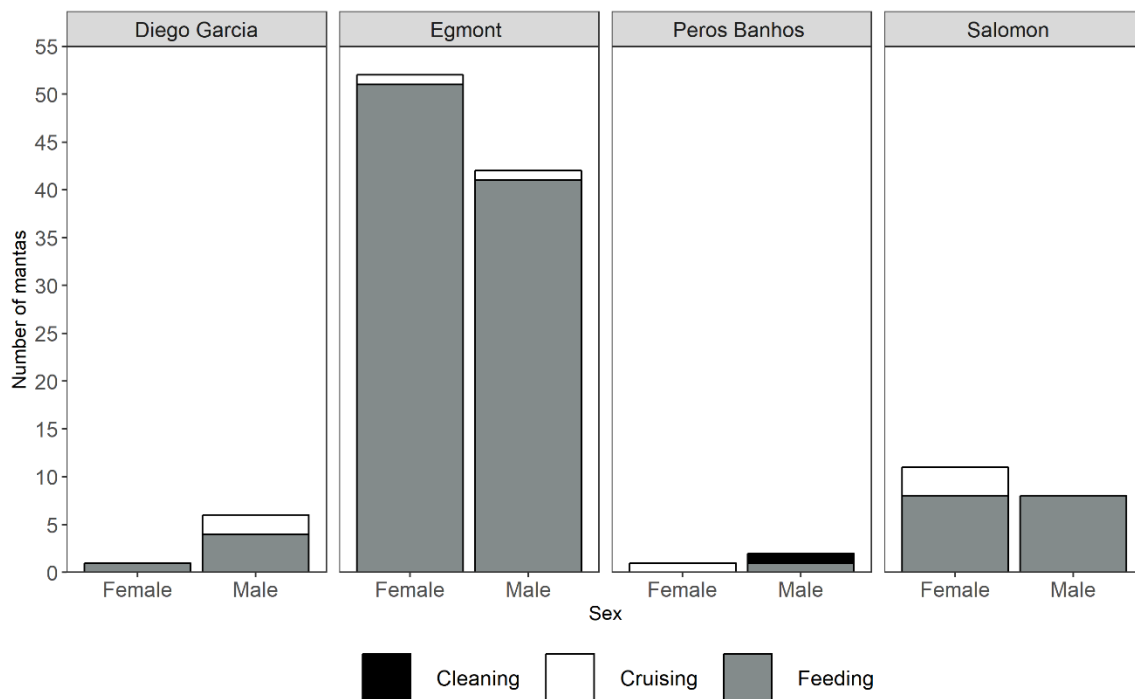
**Figure 2** - Total number of male and female *M. alfredi* recorded at each of the four atolls surveyed between 2006-2018 divided into size categories from which the manta rays life stage can be inferred. Size categories range from 1-4 for females and 1-3 for males (Table 2).

**Table 2** - Size class of manta rays with the corresponding disc width (cm) and life stage.

Size class	Disc width (cm)	Female life stage	Male life stage
1	< 210	Juvenile	Juvenile
2	210 - 270	Juvenile	Juvenile / Subadult
3	271- 320	Juvenile	Adult
4	> 320	Adult	N/A

The behaviour of *M. alfredi* of both sexes was most frequently recorded as feeding (females = 60, males = 52) (Fig. 3). Cleaning behaviour was observed for a single juvenile *M. alfredi* which was sighted at Peros Banhos Atoll. The remainder of the individuals were cruising when sighted (female = 5, male = 5). None were observed while engaged in courtship behaviour.





**Figure 3** – Total number of male and female *M. alfredi* recorded at each of the four atolls surveyed between 2006-2018 divided into the three behavioural categories observed. The fourth category of ‘courtship’ has not been observed.

Eighteen of the 108 *M. alfredi* (female = 8, male = 10) were sighted more than once. All but one of the females were juvenile while six of the males were adults, and four were subadults. All re-sightings occurred at the same atoll as the initial sighting. Almost all re-sightings occurred in the same year on consecutive days at the same location or locations <2 km apart. Three *M. alfredi* were re-sighted in different years; one juvenile male, first documented in 2013 was observed at the same location five years later and one male and one female, both adults, were sighted in 2015 and 2018 at Egmont Atoll at locations approximately 5 km apart.

### 3.2. Satellite Tags

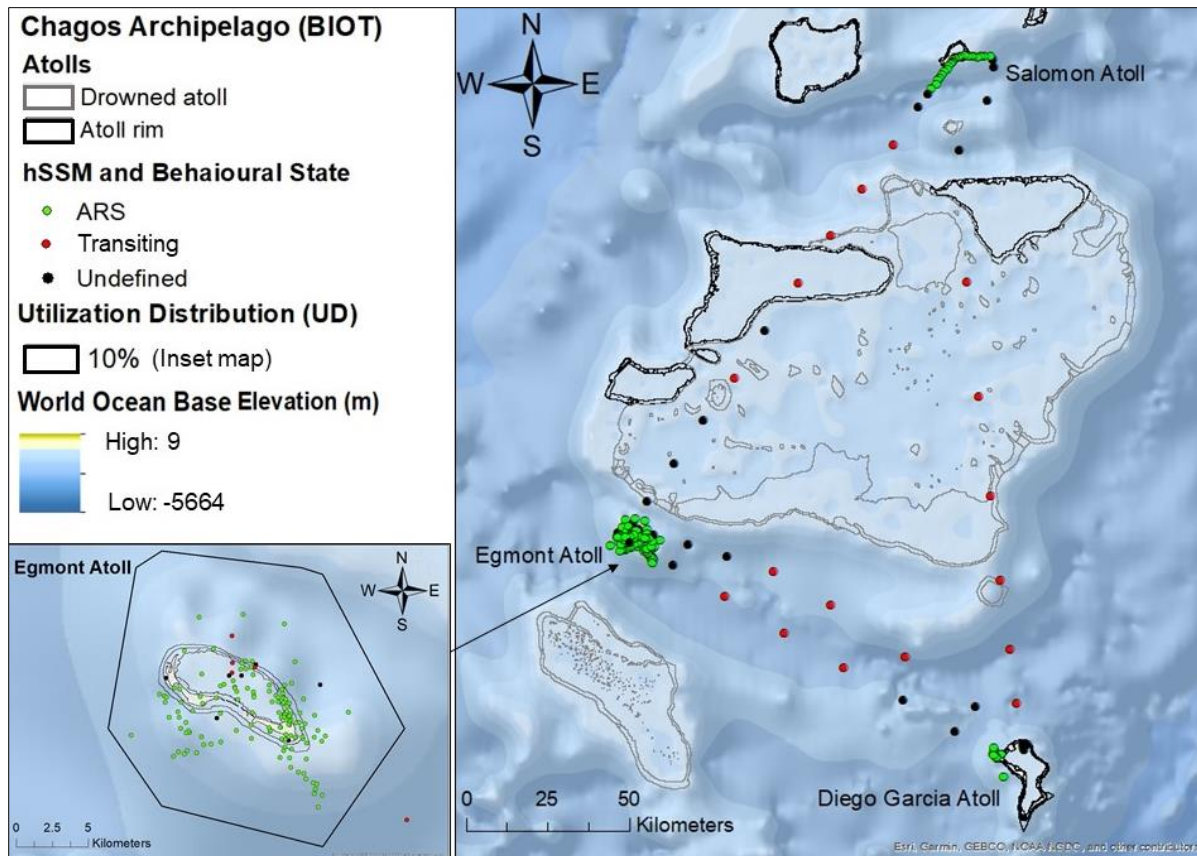
Satellite tags were deployed on five *M. alfredi* in January 2015 (Table 3). Tracking duration ranged from 10-35 days (mean  $23.2 \pm 10.28$  days). A total of 199 Argos position estimates were received overall, with a mean of  $39.8 \pm 17.48$  transmissions (range 30-71) for all five *M. alfredi*. The daily mean number of Argos position estimates was  $3.31 \pm 1.79$  (range 1-7).

**Table 3** - Satellite tracking details of *M. alfredi* tagged in BIOT with deployment information: tag number, date of deployment and location coordinates. Argos position estimate data: tracking duration, number of days the tags transmitted, mean number of position estimates per day with  $\pm$  SD and range in brackets. State-space model (hSSM) position estimate data: total track distance estimated via measurement of direct transit between points and mean travelling speed with  $\pm$  SD and range in brackets calculated from the distance travelled between points at 12hrly timesteps.

Manta Tag ID	Deployment		Argos position estimates			hMMS position estimate	
	Deployment date	Location	Track duration (days)	Transmitting days	Mean number of positions per day	Total track distance (km)	Mean speed (m/s)
144472	17-Jan-15	6°37'40.8"S 71°20'42.0"E	10	7	4.71 $\pm$ 1.49 (3-7)	9.57	0.015 $\pm$ 0.012 (0.003-0.04)
144473	17-Jan-15	6°37'40.8"S 71°20'42.0"E	35	11	2.72 $\pm$ 1.79 (1-6)	576.09	0.177 $\pm$ 0.222 (0.003-0.931)
144474	17-Jan-15	6°37'40.8"S 71°20'42.0"E	32	19	3.73 $\pm$ 1.82 (1-7)	112.05	0.052 $\pm$ 0.054 (0.006-0.221)
144475	17-Jan-15	6°37'40.8"S 71°20'42.0"E	21	11	3.01 $\pm$ 1.55 (1-6)	47.58	0.028 $\pm$ 0.029 (0.006-0.164)
144476	17-Jan-15	6°37'40.8"S 71°20'42.0"E	18	12	2.66 $\pm$ 1.72 (1-6)	174.49	0.115 $\pm$ 0.164 (0.005-0.492)

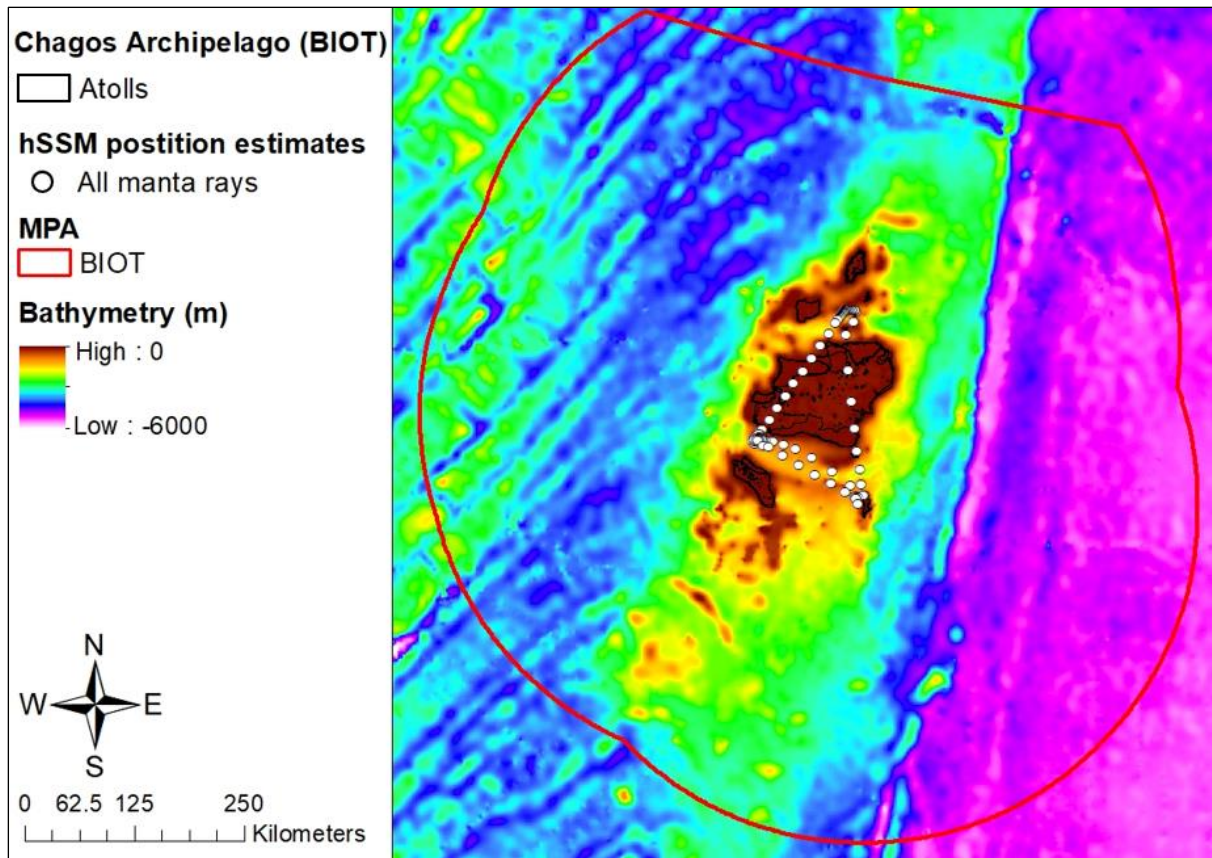
### 3.3. Bayesian joint estimation state-space model (hSSM)

The hSSM estimated a total of 221 positions, ARS was identified at 89% of these locations (Fig. 4) occurring at three atolls (Egmont, Saloman and Diego Garcia). Of the five manta rays, three remained at Egmont Atoll throughout the tracking duration (manta tag ID 144472 = 10 days, 144474 = 32 days and 144475 = 18 days). The other two manta rays travelled to other atolls to forage with manta 144473 visiting all three atolls over the course of 34 days with an estimated total travel distance 576.1km.



**Figure 4** - Position and behavioural state estimates of hSSM. Inset: The hotspot identified at Egmont Atoll from the 10 % KUD, which indicates intensive use by all five manta rays.

All position estimates occurred within the boundary of the regions marine protected area (MPA) in waters with a maximum depth of 1836 m (Fig. 5). The position estimate which occurred closest to MPA limits was located at Salomon Atoll which was 254.5 km inside the MPA boundary.

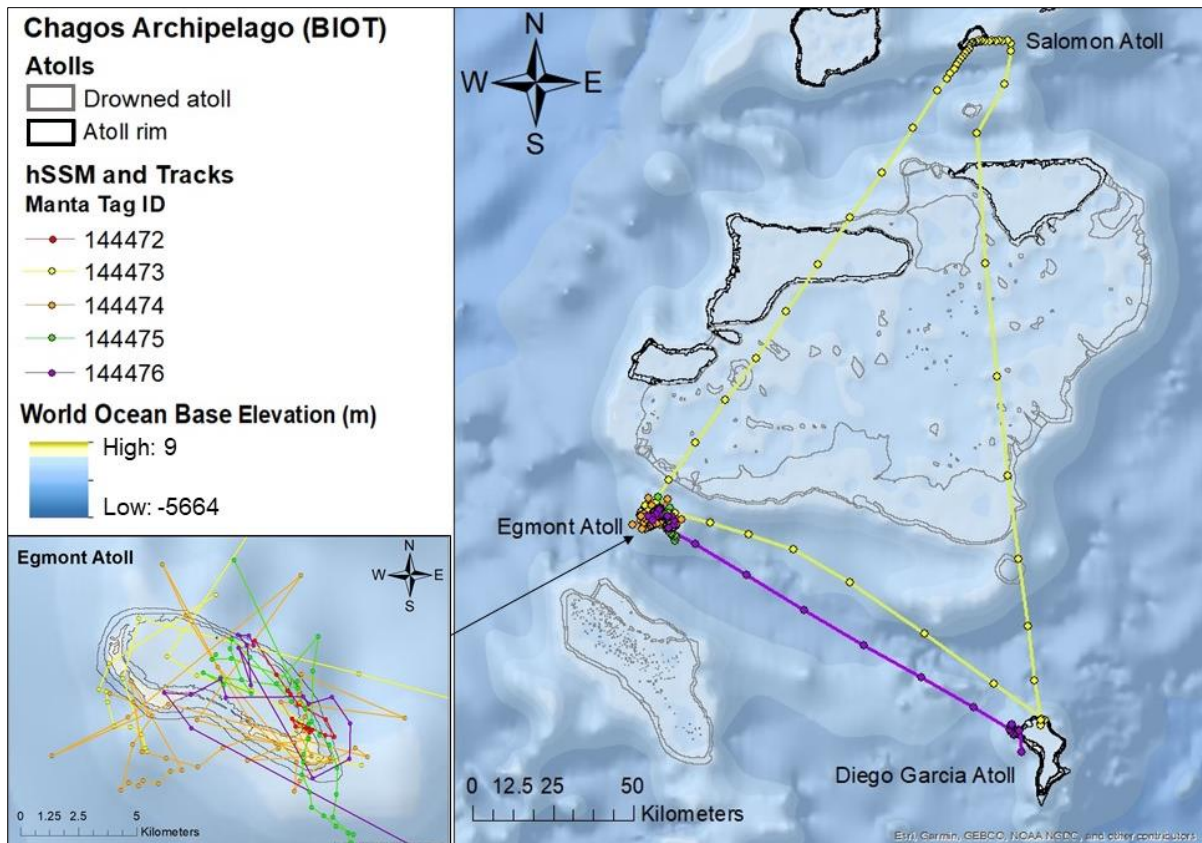


**Figure 5** - Position estimates of hSSM for all five manta rays at BIOT shown with bathymetry and the BIOT MPA. All position estimates occurred within the MPA boundary in waters with a maximum depth of 1836m.

### 3.4. Track

The overall mean distance travelled by the manta rays was  $183.9 \pm 228.1$  km (range 9.57-576.1 km) (Fig. 6 and Table 3). Manta 144473 travelled the furthest and visited three atolls in during 29 of the 34 days the tag was retained covering a straight line distance of 546 km.

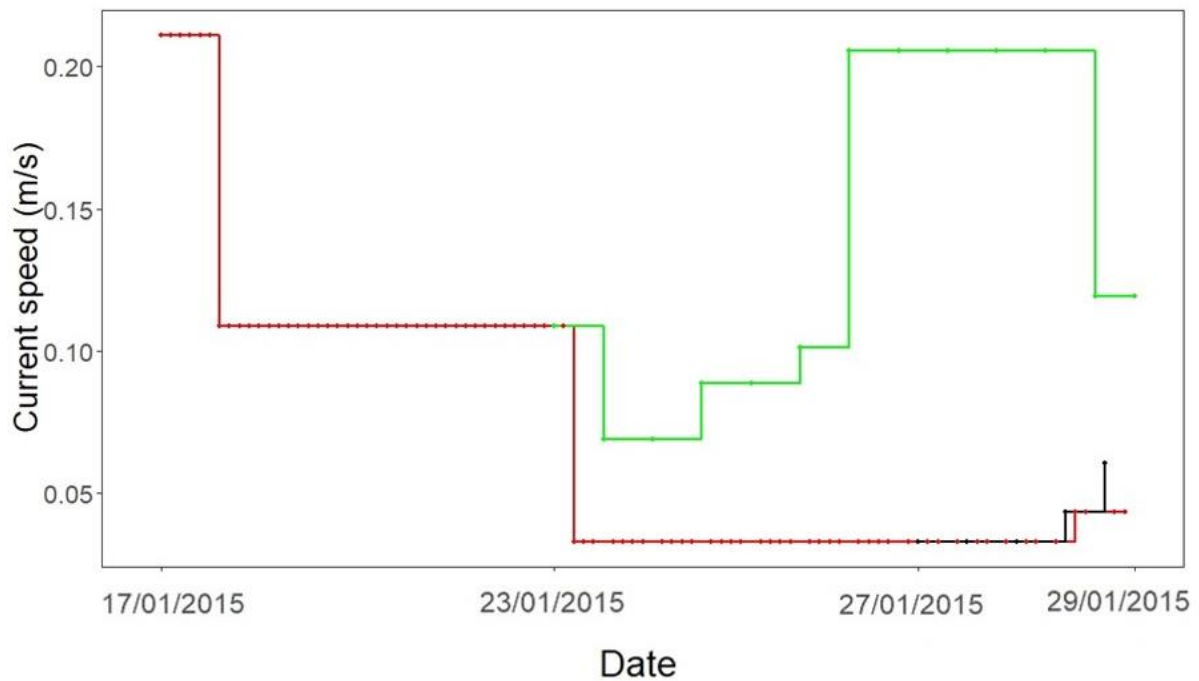
Mean distance travelled between successive hMMS position estimates (12-hourly time-steps) was  $4.98 \pm 7.36$  km with a maximum distance of 40.28 km travelled by manta 144473 during a transiting behavioural state. The mean speed between hMMS position estimates was  $0.12 \pm 0.17$  m/s (range 0.003 – 0.93 m/s). The mean speeds were significantly lower during ARS ( $0.059 \pm 0.105$  m/s) compare to transiting speeds ( $0.48 \pm 0.144$ ; Mann-Whitney U = 1263,  $p < 0.001$ ).



**Figure 6** - Horizontal swim tracks of all five manta rays showing the minimum possible distance travelled between points. Inset: close-up of tracks at Egmont Atoll only.

Manta 144473 departed from Egmont Atoll on 23<sup>rd</sup> January, which coincided with a substantial drop in current speed at the location (Fig. 7). Manta 144476 also migrated away from the location four days later following a similar track to Diego Garcia atoll (Fig. 6). Both manta rays moved to waters with higher current speeds.

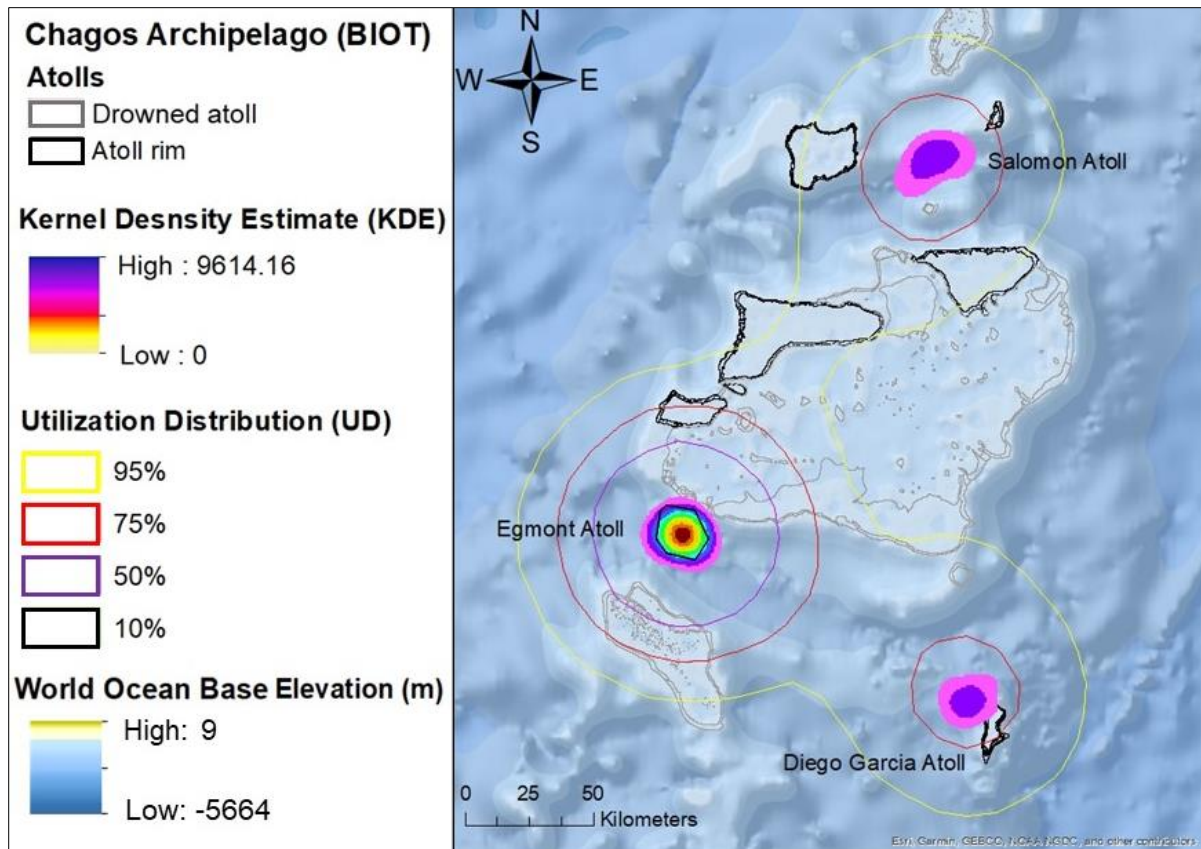




**Figure 7** – Current speed changes recorded for all position estimates within the identified hotspot (red line) from 17/01/2015-29/01/2015. Dates displayed on the x-axis show when the manta rays left the hotspot (144473 = 23/01/2015, 144476 = 27/01/2015). The green and black lines show current speed changes recorded for manta 144473 and 144476 position estimates respectively after departure from the hotspot.

### 3.5. Hotspot

The Kernel density estimate (KDE) indicates the area most intensively utilised by the tagged *M. alfredi* was around Egmont Atoll (Fig. 8). Although relatively lower, the KDE also identifies that Salomon and Diego Garcia Atoll have elevated densities compared to the rest of the region. The Kernel utilisation distribution (KUD) shows that the home range (75% and 95% KUD) of the five manta rays combined encompasses all three of these atolls with a ‘hotspot’ identified by the 10% KUD at Egmont Atoll. Overall, 59% of hSSM position estimates occurred within this hotspot. All manta rays utilised this area, and mantas 144472, 144474 and 144475 were only observed within the hotspot for their entire track duration of 10, 32 and 21 days respectively.



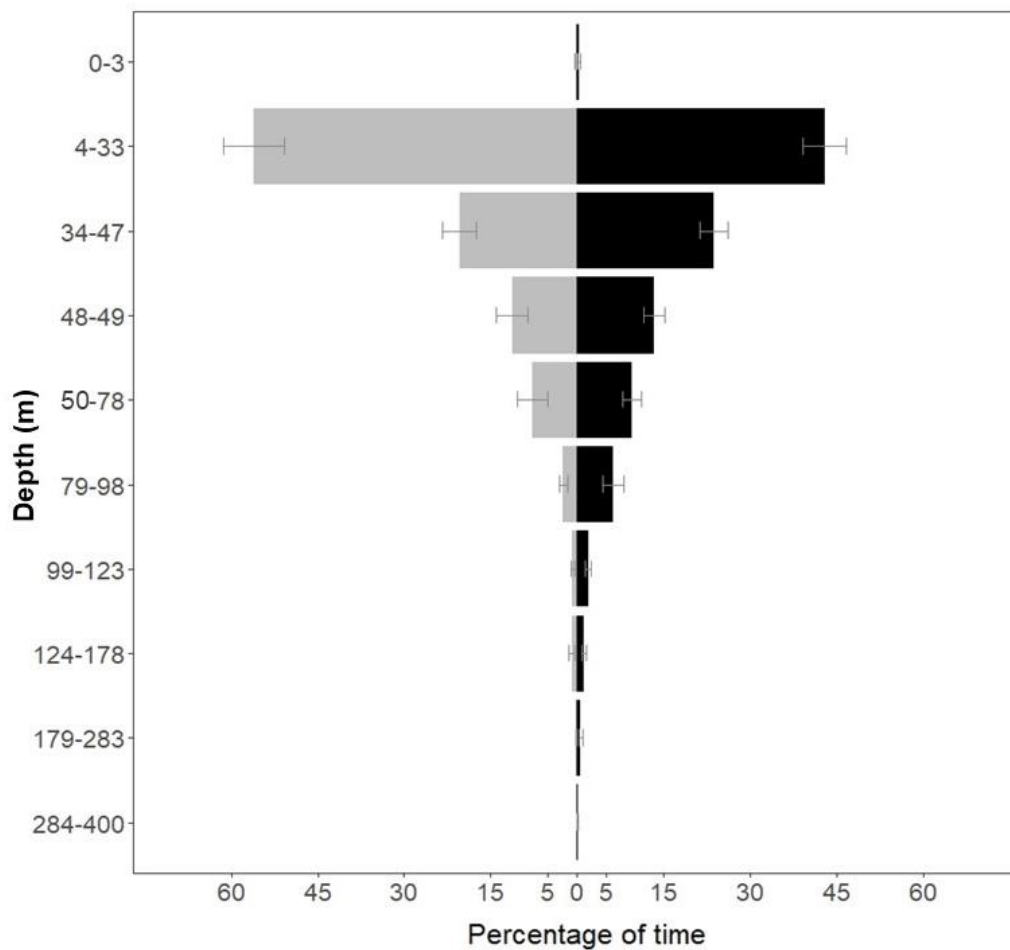
**Figure 8** - Kernel density estimate (KDE) and Kernel utilisation distribution (KUD) showing areas of intensive use and probability density for recapture, respectively. The KDE values are the density of manta rays per km<sup>2</sup>. The highest density area corresponds to the 10% KUD, which was used to define the 'hotspot'.

### 3.6. Time-at-temperature (TAT)

Available time-at-temperature (TAT) data ranged from 43–90 % of tracking days for individual manta rays (mean = 71%) with a total of 177 TAT records over 77 days for all manta rays combined.

The vertical temperature and depth profiles obtained via MMS and CTD in 2015 and 2016 respectively used to estimate the depths of the TAT data (Fig. B1) indicate the manta rays occupied depths from the sea surface down to 400 m. Predominantly shallow waters (4-33 m) were occupied during the day (09.00-15.00) (Fig. 9). At night (21.00-03.00), the highest mean percentage of time was also spent within the 4-33 m depth range. However, diel differences are apparent with a higher mean percentage of time spent at depths >79 m at night compared to during the day. The deepest depths recorded (>179 m) were visited during both daily light and dark with the mean percentage of time of 0.11% and 0.68% respectively.

Depths of up to 400 m were reached during both time periods although for the lowest mean percentage of time for day (0.01%) and night (0.08%).

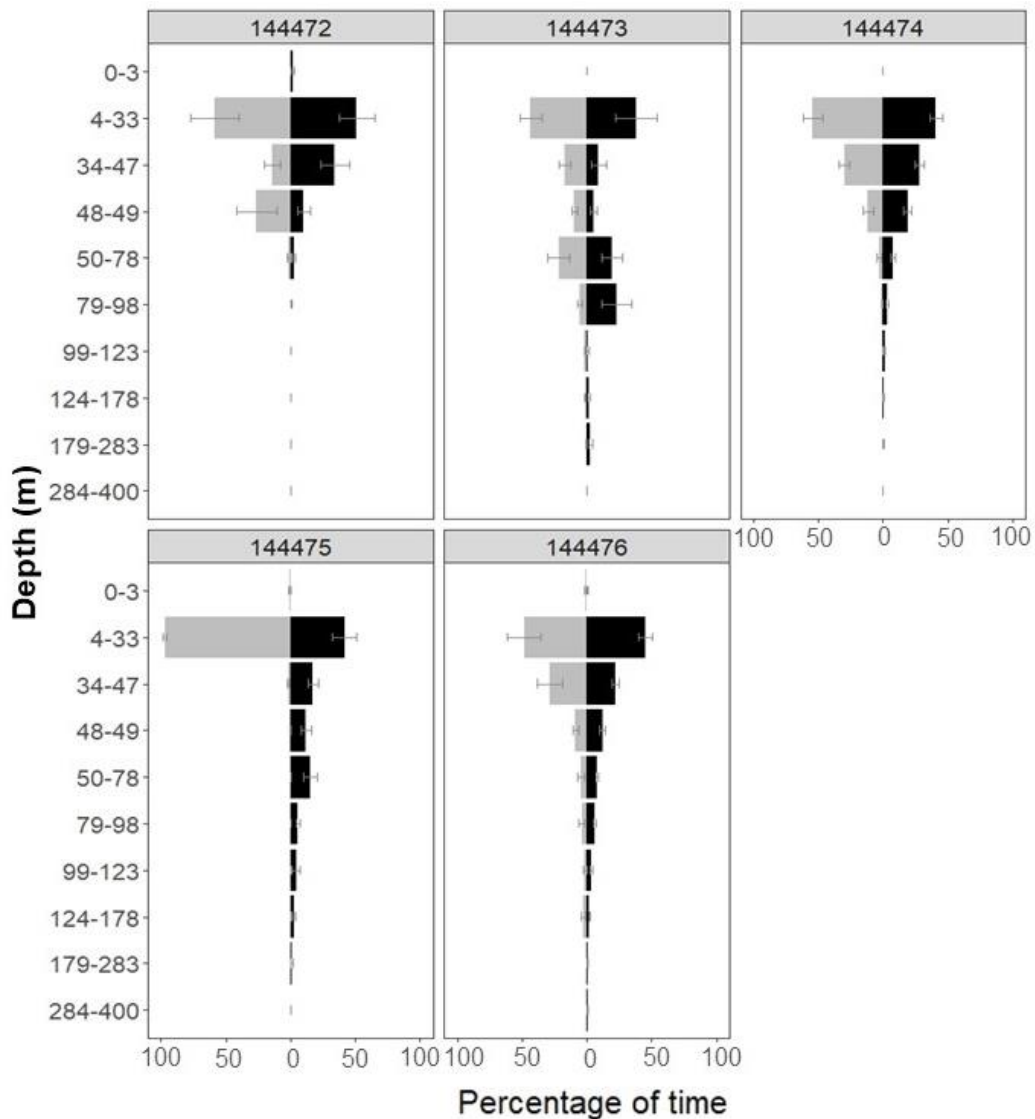


**Figure 9** - Overall mean percentage of time spent at depths during the day (grey) and night (black) ( $\pm$  SE) inferred from the SPOT5 TAT recordings and the CDT/MMS temperature and depth profiles (Appendix B(i)).

The vertical movements of individual *M. alfredi* (Fig. 10) were varied. For example, the mean percentage of time manta 144472 occupied depths <49 m was similar during the day and night (approximately 95%). The deepest depth inferred for this manta ray from the tag temperature was c.a. 98 m which was reached only at night and only for a mean of 0.4% of the time. In contrast, manta 144473 spent a mean percentage of c.a. 46% of the night time at depth >49 m with 2.8% of the mean percentage of time spent between 179-283m. Manta 144473 spent a similar mean percentage of time in depths <78 m during the day and night but a greater amount of time was spent below this depth at night. Manta 144475 rarely occupied depths below 33 m during the day with a mean of >97% spent between 4-33 m. At night, the manta ray's tag recorded depths of up to potentially 400 m. Manta 144476 was also recorded



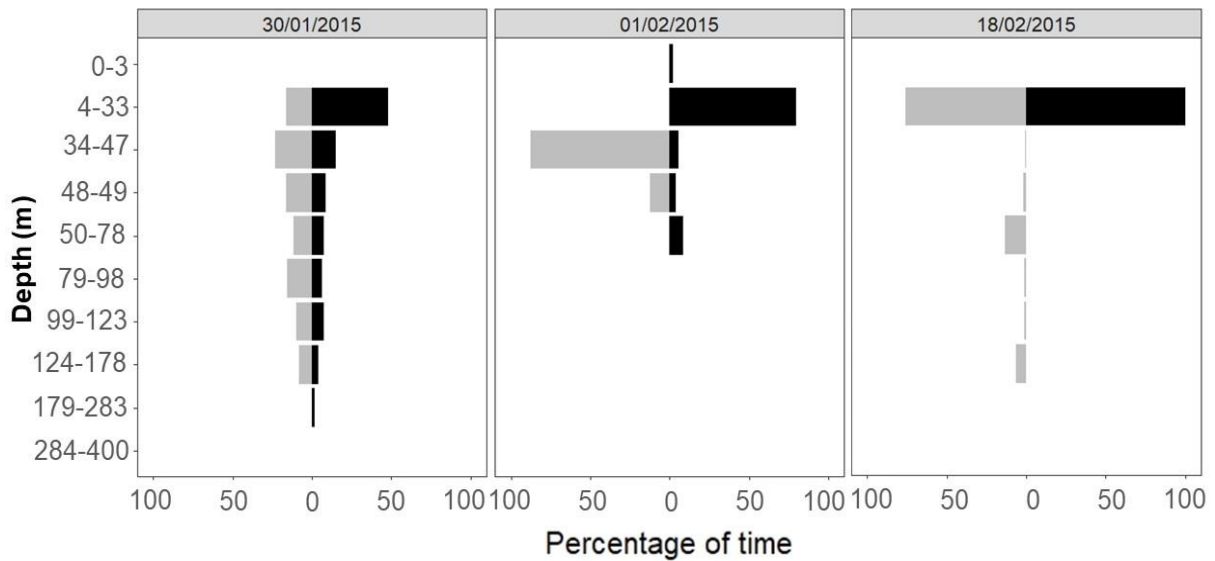
at this depth and was the only manta observed to visit this depth both during the day and at night with a mean percentage of time of 0.06% and 0.37% respectively.



**Figure 10** - Mean percentage of time individual manta rays spent at depths during the day (grey) and night (black) ( $\pm$  SE) inferred from the SPOT5 TAT recordings and the CDT/MMS temperature and depth profiles (Appendix B).

A comparison of the daily total percentage of time each manta ray spent at depths is considered using only days where both day and night data was available (Appendix B(ii), Fig. B2-B6). For each manta ray, the number of days varied with a total of 26 days for all manta rays combined. Fifteen of these days indicate the occupation of deeper waters for higher percentage of time at night than during the day. The opposite diel pattern was observed for four of the five manta rays over a combined total of eleven days (Fig. B2-B6). An example of three of these days is displayed in Figure 11, which shows the manta rays spent a substantially

higher percentage of time in surface waters at night. Most notable is manta 144473 (Fig. 11, left plot) where 100% of the night was spent between 4-33 m while deeper depths were occupied during the day.



**Figure 11** - Example of the total percentage on time spent at depths during the day (grey bars) and night (black bars) using manta 144476 (left and middle plots) and 144473 (right plot).

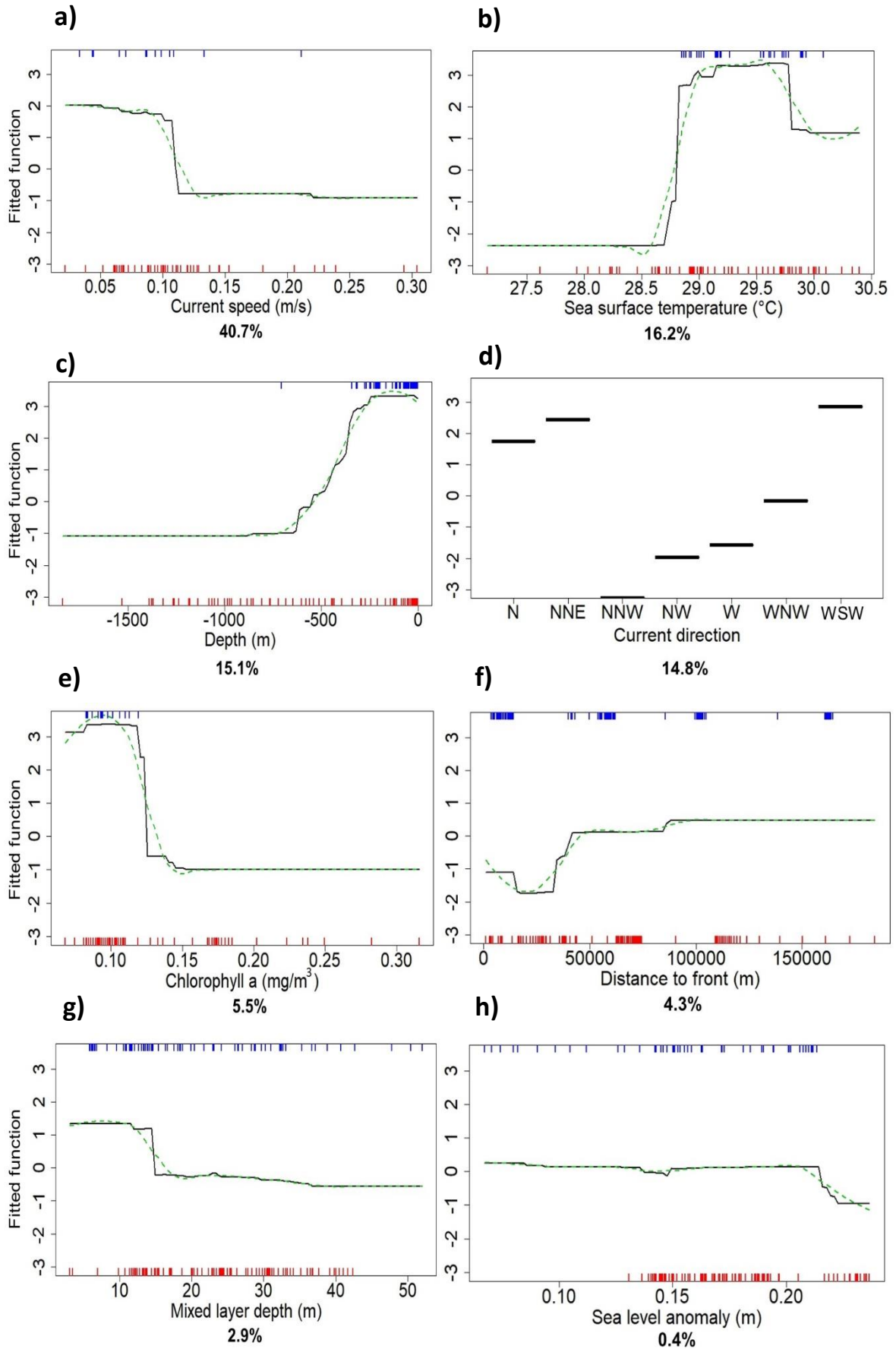
### 3.7. Environmental influences: Boosted regression trees

For the hotspot model, the `gbm.simplify()` output indicated that there was one non-informative predictor variable (slope, SLOPE) (Fig. D2). The exclusion of SLOPE substantially improved model performance thus the final hotspot model included eight predictor variables: 1) chlorophyll-*a* (CHL-A), 2) sea surface temperature (SST), 3) sea level anomalies (SLA), 4) distance to front (FRONT), 5) mixed layer depth (MLD), 6) current speed (C-SPD), 7) current direction (C-DIR), 8) depth (DEPTH). Model performance evaluation suggested the final model had outstanding predictive performance on for both the training (TAUC = 1) and cross-validated data (CVAUC = 0.999) without overfitting ( $\Delta$ AUC = 0.001) (Table 4). The estimated percentage of deviance for the null model explained by the fitted model ( $D^2$ ) suggests that 93% of the deviance is explained.

**Table 4** - Performance of the final hotspot model fitted with eight predictor variables with the tree complexity (TC) and learning rate (LR) and the optimal number of trees (NT) and performance results: mean training AUC (TAUC), cross-validated AUC (CVAUC) and their difference ( $\Delta$ AUC). The total deviance explained ( $D^2$ ) corresponds to the percentage of deviance for the null model explained by the fitted model. The mean deviance for the null model is 1.283.

TC	LR	NT	TAUC	CVAUC	$\Delta$ AUC	CV residue deviance (SE)	$D^2$
2	0.005	3550	1	0.999	0.001	0.095 (0.03)	0.93

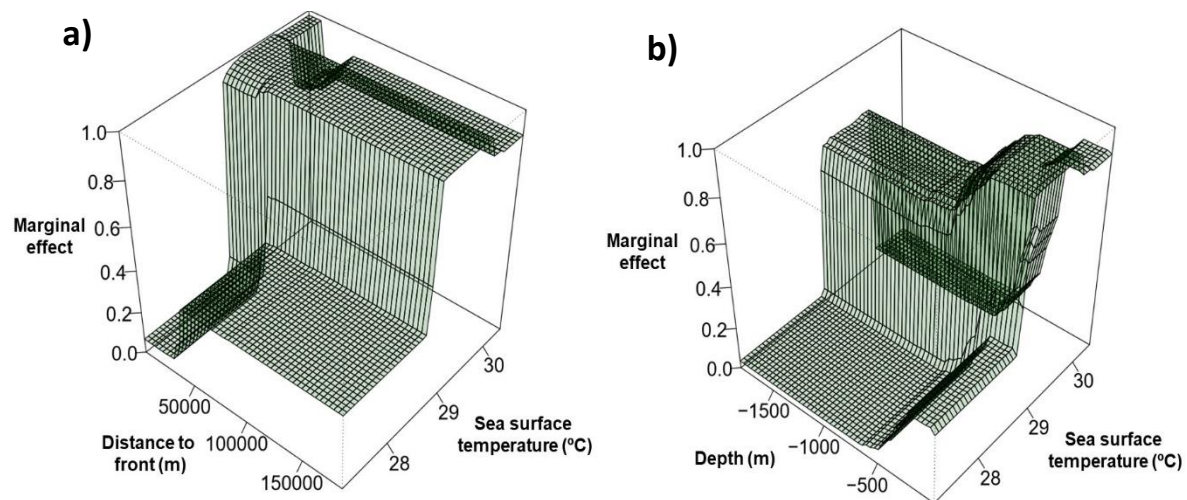
The relative influence of the predictor variables (Fig. 12 and D3) indicate that current speed (C-SPD) at the hotspot location has the greatest influence (40.7%). The variable showed a positive effect at a low C-SPD (0.01-0.11 m/s) (Fig. 12 a) with approximately 75% of presence points occurring where C-SPD was within this range. The influence of sea surface temperature (SST) (16.2%), depth (15.2%) and current direction (14.8%) was substantially lower with their combined percentage contribution being similar to C-SPD alone (Fig. 12 a-d). All presence data occurred at SST between approximately 28.8 and 30.1°C (Fig. 12 b) while manta rays that were absent from the hotspot were observed in SST as low as 27.2°C. Approximately 40% of presence points occurred at depths <30m (Fig. 12 c). The partial dependency plots for C-DIR (Fig. 12 d) do not display percentile rugs; however, the plot indicates that a current direction of N, NNE and WSW influence the presence of *M. alfredi* at the hotspot. Analysis of the raw data shows that 69% of observations within the hotspot occurred when the C-DIR was in one of these directions. Chlorophyll *a* (CHL-A), distance to front (FRONT), mixed layer depth (MLD) and sea level anomalies (SLA) had relatively low influence of 5.5%, 4.3%, 2.9% and 0.4% respectively. For CHL-A, concentrations between 0.08 and 0.11 mg/m<sup>3</sup> appear to have a positive influence with all position estimates occurring within the hotspot when CHL-A values were within this range (Fig. 12 e). The results for FRONT and MLD are ambiguous due to uneven clustering of both presence and absence data (Fig. 12 f-g) however, the plot suggests that being further from a FRONT and a shallower MLD (<30m) influence presence of manta rays at the hotspot. For MLD, 82% of points occurred at the hotspot when MLD <30m. For SLA, up to 0.23m appears to have a marginal positive influence while above this value has a relatively high negative influence.



**Figure 12** - Partial dependency plots of the eight predictor variables included in the final hotspot model; a) current speed (C-SPD), b) sea surface temperature (SST), c) depth (DEPTH), d) current direction (C-DIR), e) chlorophyll-*a* (CHL-A), f) distance to front (FRONT), g) mixed layer depth (MLD), h) sea level anomalies (SLA). The broken green line shows smoothed partial dependency. Rugs display the distribution of data in percentiles for presence (top, blue) and absence (bottom, red).

The most important interactions identified were distance to front and SST, depth and SST (Fig. 13), and current direction and distance to front, as the latter contains a categorical variable, the interaction cannot be plotted. The 3D plot displays a piecewise-constant approximation of the joint partial dependence of their effect [101]. Each split in the model is represented by rectangular regions that divide the predictor variable [107]. A marginal effect of zero indicates that no interactions were fitted [101].

For distance to front and SST (Fig. 13 a), the greatest interaction effect occurs between 29-30°C at all distances with a lesser interaction between 27-29°C at distances over 50,000 m. For depth and SST (Fig. 13 b), the highest interaction occurred between 29-30°C at depths of <500 m and no interaction was fitted between these variables when depth was >500 m and SST was <29°C.



**Figure 13** - Interaction plots for two of the most important interactions identified for the final hotspot model; a) distance to front (FRONT) and sea surface temperature (SST), b) depth (DEPTH) and SST.

For the ARS model, the `gbm.simplify()` output indicated there were two non-informative predictor variables (SLA and C-SPD) (Fig. D5). However, the exclusion of SLA and C-SPD did not markedly alter model performance, therefore, following the concepts of parsimony [109], the

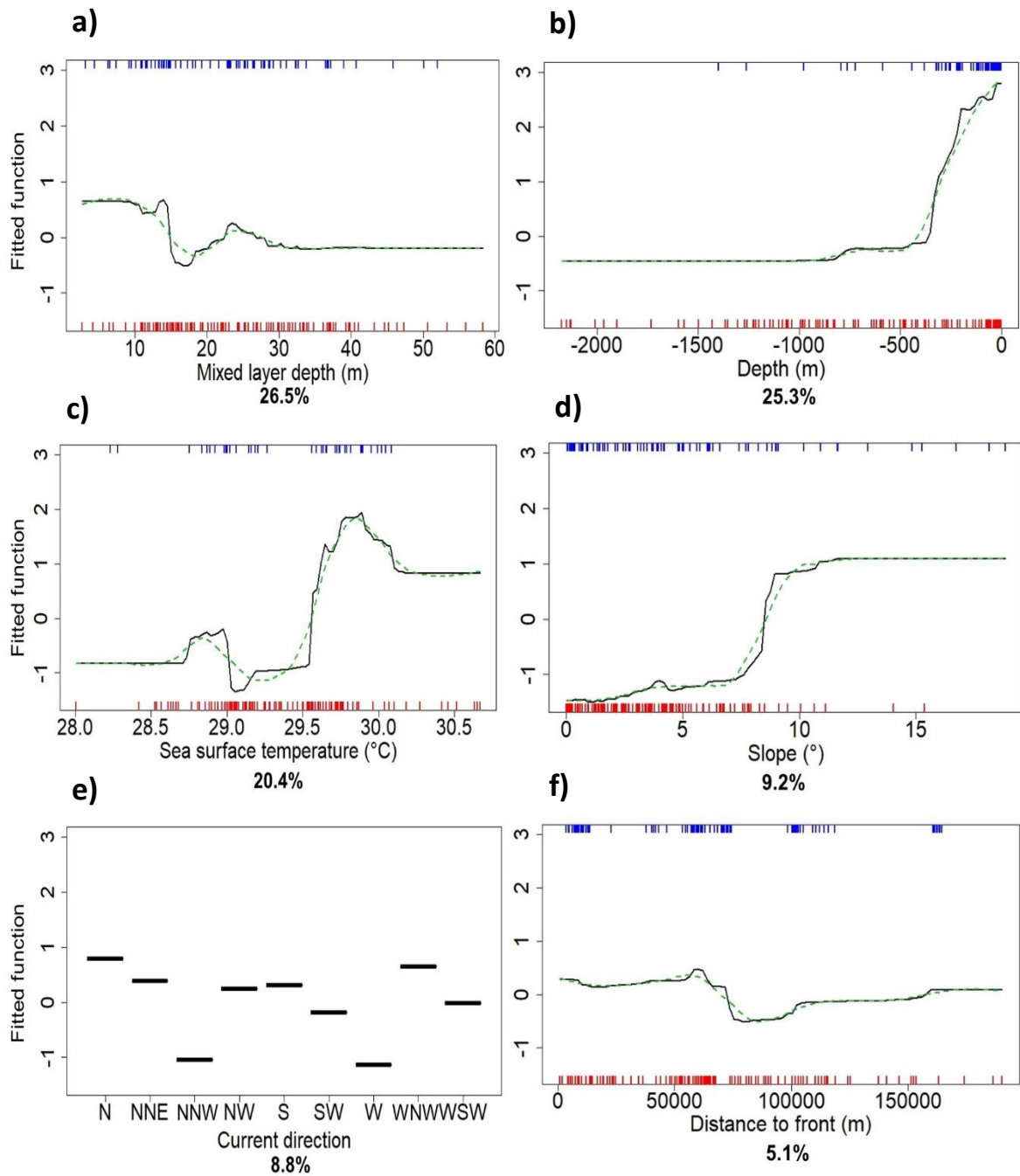
final ARS model was fitted with seven predictor variables: 1) CHL-A, 2) SST, 3) SLA 4), FRONT, 5) MLD, 6) C-DIR, 7) DEP. Model performance evaluation suggested the final model had outstanding predictive performance for both the training (TAUC = 0.999) and cross-validated data (CVAUC = 0.972) with minimal evidence of overfitting ( $\Delta$ AUC = 0.027). The estimated  $D^2$  suggests that 63% of the deviance is explained.

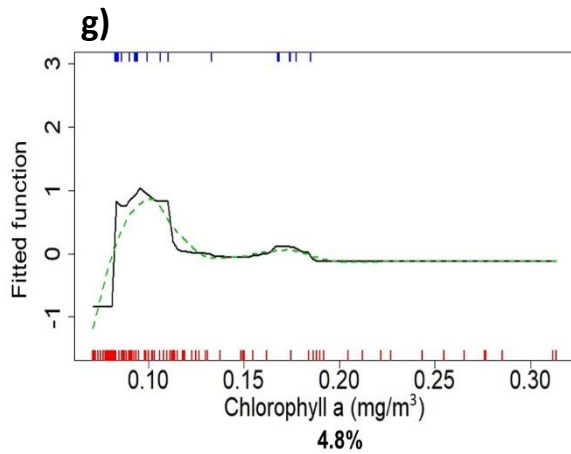
**Table 5** - Performance of the final ARS model fitted with seven predictor variables with the tree complexity (TC) and learning rate (LR) and the optimal number of trees (NT) and performance results: mean training AUC (TAUC), cross-validated AUC (CVAUC) and their difference ( $\Delta$ AUC). The total deviance explained ( $D^2$ ) corresponds to the percentage of deviance for the null model explained by the fitted model. The mean deviance for the ARS null model is 1.386.

TC	LR	NT	TAUC	CVAUC	$\Delta$ AUC	CV residule deviance (SE)	$D^2$
5	0.001	8000	0.999	0.972	0.027	0.447 (0.08)	0.68

Results of the final ARS model suggest that ARS behaviour may be predominantly influenced by three of the seven predictor variables which all had similar scores: MLD (26.5%), DEPTH (25.3%) and SST (20.4%) (Fig. 14 and D6). The MLD partial dependency plot (Fig. 14 a) suggests that ARS behaviour is more likely to occur when the MLD is shallower (<15m). For non-ARS points, 78% occurred where the MLD was >15m compared to 54% at ARS position estimates. The partial dependency plots also indicate that ARS occurs at shallower depths (Fig. 14 b) with 38% of ARS points at <30m compared to just 12% of non-ARS. At depths >100m 76% of non-ARS occurred compared to just 33% of ARS. Sea surface temperature ranged from 27.16-30.6°C and appeared to have the greatest influence on ARS at just under 30°C (Fig. 14 c). Although both ARS and non-ARS occupied a similar mean temperature of 29.4±0.4°C (range 28.2-30.1°C) and 29.5±0.51°C (range 28-30.7°C) respectively, 71% of ARS positions were found at temperatures >29.6°C compared to 27% of non-ARS. Slope (9.2%), C-DIR (8.8%), FRONT (5.1%), and CHL-A (4.8%) (Fig. 14 d-g) had relatively lower influence than the top three predictor variables. Both ARS and non-ARS occurred around relatively gentle slopes (<5°) however, Figure 14 d shows that a slope >9° has a positive influence on ARS whereby 14% of ARS occurred compared to 6% of non-ARS. ARS is suggested to be positively influenced by C-DIR of N and WNW direction (Fig. 14 e) with 48% of ARS occurring in these directions compared to 26% of non-ARS. A C-DIR of NNW and W is indicated to have a negative influence;

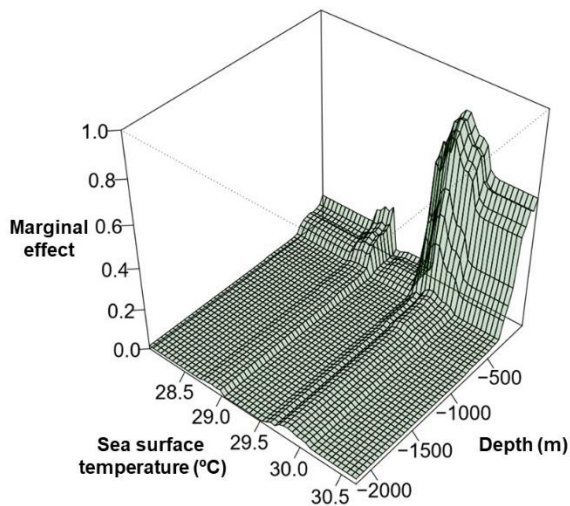
the raw data shows that 30% of non-ARS occurred in these directions compared to zero ARS points. Distance to front (Fig. 14 f) appears to have little influence with the partial dependency showing a fitted function around zero. For CHL-A (Fig. 14 g), all ARS are in concentrations  $<0.2 \text{ mg/m}^3$  with 84% occurring between  $0.08\text{-}0.1 \text{ mg/m}^3$  compared with only 34% of non-ARS. For non-ARS, 45% occurred at concentrations  $>0.1 \text{ mg/m}^3$  with 11% between  $0.2\text{-}0.31 \text{ mg/m}^3$ .





**Figure 14** - Partial dependency plots of the seven predictor variables included in the final ARS model; a) mixed layer depth (MLD), b) depth (DEPTH), c) sea surface temperature (SST), d) slope (SLOPE), e) current direction (C-DIR), f) distance to front (FRONT), g) chlorophyll-*a* (CHL-A). The broken green line shows the smoothed partial dependency. Rugs display the distribution of data in percentiles for presence (top, blue) and absence (bottom, red).

The most important interactions identified for the ARS model were depth and SST (Fig. 15) and current direction and SST. The interactions of depth as SST was highest at SST between 29.5-30°C and depths <500m.



**Figure 15** - Interaction plots for sea surface temperature (SST) and depth (DEPTH), which was one of the most important interactions identified for the final ARS model.



## 4. Discussion

### 4.1. Current understanding of the Chagos BIOT *Mobula alfredi* subpopulation

There were 123 observations of 108 photographically identified *Mobula alfredi* around the Chagos Archipelago; British Indian Ocean Territory (BIOT). The majority were juvenile, which is similar to other studies in Indonesia [110] and the Maldives [111]. Eighteen of the manta rays have been sighted more than once of which three were documented in different years. All re-sightings occurred at the same atoll either in the same location or locations <5km apart indicating some degree of site fidelity for these individuals. The small number of re-sightings may be a product of the infrequency and short duration of surveys due to the logistics of conducting fieldwork in such a remote location.

The tagged manta rays appeared to demonstrate varying degrees of site fidelity. Three remained at Egmont Atoll throughout the tracking period while two were observed to make a relatively long migration away to Diego Garcia Atoll (>130 km). High site fidelity is a regularly reported characteristic of *M.alfredi* which has been observed via passive acoustic [15] and satellite telemetry at Lady Elliot Island (LEI) on the Great Barrier Reef [112] and in observational studies in the Maldives [111] and Hawaii [12]. Although an overall trend, other observational studies have noted that while some individuals displayed high site fidelity others frequently interchange between locations [45,110]. Differences in site residency can be associated with demographic segregation as significant associations between habitat use, and the age and sex of the manta rays have been observed [110]. Adult *M.alfredi* have also been observed to frequently embarked on long migrations while juveniles displayed greater site residency times [45]. Here, one tagged manta ray (144473) visited three atolls, starting and ending at Egmont Atoll. The round trip took 29 days and covered a straight-line distance of 546 km. Similar to one satellite-tagged adult female *M.alfredi* off LEI that travelled 2441 km in 118 days [112], the distance travelled by manta 144473 appears to be one of longest reported distances travelled in such a short space of time. Due to restraints in the field, the demographics of the tagged manta rays were not recorded here; thus, whether migration occurred due to differences in sex or age class of the manta rays cannot be postulated. However, photo-ID surveys observed more females, than males, which were predominantly juvenile and adult respectively at Egmont Atoll, while at Diego Garcia Atoll more males (subadult and adult) than females (one adult) have been sighted. Potentially this indicates

that juveniles remain at Egmont Atoll and that Diego Garcia Atoll is less frequently utilised by female and juvenile manta rays. For various ray species, greater residency times of juveniles have been suggested to be a predator avoidance strategy [45,113,114]. Alternatively, due to the smaller body size, it may not be energetically efficient for juvenile manta rays to make long migrations [45,115] or they may simply lack foraging experience [45].

Foraging (area-restricted search, ARS) was the predominant behaviour identified by the Bayesian joint estimation state-space model (89% of position estimates). Foraging was also the predominant estimated behaviour in a study which conducted similar analysis of satellite tag data from *M. alfredi* in the Red Sea [11]. Satellite tags can only transmit when they are dry for a sufficient period; therefore, they are required to be above the surface of the water. There are several feeding strategies and behaviours where breaching of the surface occurs, including surface feeding as well as direct breaching, which has been shown to be significantly more likely during feeding [111]. It is then plausible that manta rays are more likely to be located while feeding than at any other time hence the high percentage of ARS position estimates. The ARS behaviour was concentrated around three atolls: Egmont, Diego Garcia and Salomon Atoll, which is consistent with photo-ID survey observations at these locations where feeding was consistently recorded as the primary behaviour.

Kernel utilisation estimates identified Egmont Atoll as a hotspot for *M. alfredi*. As a species that are reliant on zooplankton, utilisation of the area is likely to be associated with feeding opportunities [6]. Hydrodynamics, controlled largely by bathymetric complexity [116], play a key role in the accumulation of zooplankton in coral reef ecosystems [117]. These systems encompass a multitude of hydrodynamically connected sub-systems, including the reef, lagoons and the ocean [116]. The spatial and temporal dynamics of these sub-systems can influence the abundance of zooplankton [116,117]. For example, coral reefs act as a barrier to ocean currents which diverge around the reef creating a pocket of relatively motionless water immediately upstream, which can trap nutrients and plankton [117]. Acceleration of the diverging currents induces longshore currents parallel to the reef sides where abrupt changes in bathymetric features can induce flow separation, which result in the formation of plankton-rich eddies. [117]. Eddies on the lee of the reef can create a zone of reduced flushing which retains zooplankton, for example, that may have emerged at night, a phenomenon noted in a study of zooplankton community structure on the Great Barrier Reef [117].

Plankton can also be advected up the reef slope when tidal currents become entrapped by bathymetric features [117]. Boosted regression trees (BRT) modelling indicated current speed was the most influential variable for hotspot utilisation. There was a strong positive influence at low current speeds plausibly because the manta rays were taking advantage of zooplankton accumulation induced by bathymetric features [117]. Moreover, in areas where the current speeds are higher, bottom and vertical shear is increased, which can disperse zooplankton [118]. Such dynamics have been observed in marine coastal systems where the accumulation of zooplankton was characterised by low current velocities and a Richardson number  $>0.25$ , indicating a stable water column [119]. Increased wind speeds, which can be concomitant with ocean surface currents, have also been linked to reduced *M. alfredi* presence and foraging behaviour [73]. Conversely, the departure of the two *M. alfredi* from Egmont Atoll appears to correspond to a substantial decrease in current speed. Lower current speeds reduce zooplankton transport and dispersal [118] thus potentially the dense aggregations which are necessary for energetically efficient foraging may have become depleted, which might have led to the migration of the manta rays. However, the spatial and temporal resolution of the current speed and other environmental data hinders the identification of any fine-scale changes that may also have coincided with the migration.

Sea surface temperature (SST) was one of the most influential variables for both the hotspot utilisation and ARS behaviour models. The full range of SST was 27.2-30.6°C, and both models indicate a positive influence at a higher temperature niche; 28.8-30.1°C and 28.2-30.1°C for the hotspot and ARS model, respectively. A similar association between foraging and temperature was found to occur for satellite-tracked giant manta rays (*Mobula birostris*) in the southern Gulf of Mexico where SST ranged from 25.5-30°C and >80% of foraging position estimates occurred in temperatures between 27-30.0°C [36]. A warmer temperature niche was also apparent at *M. alfredi* aggregation sites Komodo Marine Park, Indonesia where the SST ranged from 24–29°C and the species predominantly visited (91%) during periods it was 26 to 28°C [13]. These results are in contrast to an observational study at LEI, which showed that sightings of foraging *M. alfredi* occurred between 21–23°C with a decrease in numbers at higher temperatures [73]. Recently presented acoustic tagging data in the region concurs, whereby hourly visitation patterns of *M. alfredi* consistently increased with decreasing temperatures [15]. These different thermal niches likely reflect oceanographic processes that

are enhancing local zooplankton production, which may be species or regionally-specific [120,121]. For example, upwelling of cool nutrient-rich water can enhance phytoplankton growth [122,123] and subsequent production of cool water zooplankton species [120,121]. In the LEI studies, each concluded that cooler SST and higher chl-*a* indicated that such upwelling might be increasing zooplankton biomass in the region [15,73,75]. These conclusions are supported by modelling of zooplankton dynamics and local environmental variables in the area, which found significantly higher biomass at lower SST, which coincided with *M. alfredi* feeding events [8]. Sufficient upwelling conditions to support enhanced primary production at LEI are also evident from other oceanographic data collected in the region [124]. Very little is known about the local oceanographic processes at BIOT; however, if upwelling intensity is variable, other factors such as temperature can become more influential [120].

For both phytoplankton and zooplankton, increased temperature enhances metabolic rate [125], thus accelerates physiological processes [126]. In temperature-controlled experiments, phytoplankton biomass was observed to decline by as much as 50% with a 2°C increase [126], and reduced phytoplankton size and changes in community composition were also evident [126]. These results were suggested to be consistent with increased zooplankton grazing outpacing phytoplankton production at higher temperatures [126,127] as heterotrophic processes are more sensitive to temperature increases [128]. *In situ* sampling of phytoplankton and zooplankton has also revealed a similar phenomenon whereby zooplankton biomass increased, but phytoplankton biomass remained steady likely due to accelerated zooplankton grazing rate [123]. Under these conditions, chl-*a* concentrations, both satellite-derived and sampled *in situ* can remain low and lack any clear association with zooplankton production [123]. Here, chl-*a* was only minimally influential in both models which appear to identify that there may be specific chl-*a* niche associated with feeding behaviour. Potentially, this reflects the level to which phytoplankton is reduced by grazing where zooplankton biomass becomes sufficient for manta rays to feed. This is in contrast to an upwelling region where a direct linear relationship can be observed for chl-*a* and the presence and feeding behaviour of *M. alfredi* [73].

The elevated SST, relatively slow current speeds and chl-*a* at the position estimates are not consistent with upwelling conditions [129]. However, due to the coarse resolution of the data considered here, upwelling in the region cannot be precluded. Potentially, it could be

intermittent providing ideal conditions for maximising zooplankton populations [129] as intense periods of upwelling can disperse zooplankton; thus, peak densities could occur when upwelling is reduced [123]. Alternatively, the location of the biological response to upwelling may have shifted due to advection [123]. Currents interacting with abruptly changing bathymetry could then be acting to aggregate zooplankton at in areas around Egmont Atoll [130]. The positive influence of several different current directions in the hotspot model may reflect such interaction with bathymetric features [129]. The model also indicated that there were important interactions between shallow bathymetry (depth) and SST, which could be working to concentrate and increase the production of zooplankton [117,127]. Similarly, various current directions had a positive influence in the ARS model, and there were important interactions between depth and SST. These similarities likely occurred because the majority of ARS position estimates occurred around atolls which may have similar bathymetric features to Egmont Atoll.

Overall, manta rays tended to remain in the upper epipelagic layer (<49 m) during both the day and the night but a greater mean percentage of time was spent at depths below 49 m at night. In a study conducted in the Red Sea, fine-scale dive profiles (depth recorded every 10-15 seconds) displayed a similar reverse diel vertical migration pattern [10]. Similar nocturnal habits have also been inferred from acoustic tagging data and observational studies where manta rays left the shallow study areas at night presumably to forage in deeper waters [13,15,73]. As there have been no direct observations of *M. alfredi* foraging at great depths [76], interpretation of dive data in this, as in studies of other mobulid species, is related mainly to known associations such as their reliance on zooplankton [76]. A prevalent theory is that the behaviour is in response to the diel vertical migration of zooplankton, whereby manta rays dive to depths to intercept their prey as they emerge at night [10,13,131]. The theory is supported by stable isotope (SI) analysis whereby a large proportion of manta ray's diet has been found to be made up of species which matched the SI of demersal zooplankton [131]. In the Red Sea, the rate of ascent and descent during dives were observed to be faster and slower, respectively, and it was suggested that this might be evidence of glide diving [10]. It has been proposed that glide diving is an energetically efficient strategy used by planktivores to locate prey that has an inconsistent spatial and temporal distribution [132]. Conversely, the theory that manta rays exploit prey in deep waters at night is contested by evidence that

dive profiles did not seem to indicate any period of levelling out behaviour which was presumed would occur when the manta ray had successfully located food so would remain at a specific depth for an extended period [10]. Contrastingly, in the present study, there were several records where the manta rays remained in a deep, narrow depth range for the entire six-hour recording period, which may be evidence of successful foraging activities. Although the overall pattern appeared to indicate a reverse diel vertical migration, diel vertical migration occurred on 36% of the days where both day and night data was available. Manta rays may remain at depths during the day when utilising cleaning stations [10] at which they have been shown to spend as much as five hours [133]. However, this is unlikely to be the case at depths below 40m, which is the maximum depth of occurrence for cleaner fish in BIOT [134].

Manta rays were also recorded moving through almost the full range of depths during the day and night without spending a notably higher percentage of time within a specific depth range. The time-at-temperature (TAT) histograms of the current study are constructed by recording the percentage of time within a six-hour period that a manta spent within a temperature (depth) range. Hence, it is not possible to determine in what order the depth ranges were occupied for example, whether the manta ray remained at a certain depth for several hours or spent short periods at the depth but on multiple occasions during the six hours. However, possibly manta rays could be feeding continuously while travelling from the surface to depths. One of many feeding strategies employed by the species [6,111] is cyclone feeding, which is categorised by decent from the surface in a circular motion [6] which works to concentrate zooplankton [111]. Potentially use of this technique could lead to the histograms being distributed throughout the entire depth of the water column. However, observations of this feeding strategy suggest the cyclones only last for a limited amount of time (<30 minutes) [111], therefore, it appears unlikely that this would explain vertical movement patterns observed in the current study.

Changes in the diving behaviour of *M.birostris* have been found to be associated with variations in the mixed layer depth (MLD) [76]. Zooplankton often aggregate at the base of the MLD; the upper limit of the thermocline [135,136] and data recorded by tags attached to *M.birostris* indicated they spent up to 22% of their time at the base of the MLD during the day and 21% of their time during the night [76]. As TAT histograms in the current study are not

accompanied by position estimates, associations between diving behaviour and MLD cannot be inferred; however, MLD was the most significant variable in the ARS model. The results indicated that ARS was most likely to occur where the MLD was shallower (<15m). As well as aggregating at the thermocline, zooplankton biomass has been found to be highest in the MLD [135,136] becoming more concentrated as MLD decreases [137] thus, plausibly foraging within a shallow MLD is energetically efficient. A shallower MLD may also improve the likelihood of satellite transmission as the manta ray will be foraging more closely to the surface thus are more likely to expose the tag to the air.

To continue to develop the current understanding of the BIOT *M. alfredi* subpopulation so to construct a natural baseline for the species, future research would benefit from the use of satellite and acoustic telemetry to further elucidate the subpopulations movements and identify other key habitats [114]. The subpopulations reproductive success and overall fitness should also be assessed by demographically defining habitat use, which will help identify essential requirements for the species. Modelling methods, such as ecological niche factor analysis, could then be applied to identify areas throughout the species range that might be essential to survivorship.

#### 4.2. Effectiveness of the Chagos BIOT MPA

Quantifying the impact of fisheries on *M. alfredi* subpopulations in the Indian Ocean is challenging. On a local scale, many target fisheries are largely unregulated [21]. Outside of the countries exclusive economic zones (EEZs) fisheries regulations are managed by the regional fisheries management organization (RFMO); the Indian Ocean Tuna Commission (IOTC) [22], which encompasses BIOT within their remit [22,23]. Although fisheries statistics are collected by the IOTC, calculating accurate catch values is hindered by systemic underreporting and inconsistent reporting methods [24]. For example, a general heading of 'Mobulidae' [138] is often used. In 2018, available IOTC records indicated a total of 4,018 metric tons of *Mobula spp.* were caught [138]. However, a review of publicly available data including scientific reports from sampling programmes and peer-reviewed literature, estimates the true catch rate to be substantially greater with an annual catch (2014-2016) of around 10,500 metric tons of mantas and devil rays [139]. Mobulid ray protection such as a retention ban and mandatory release requirement have been repeatedly proposed and rejected under the IOTC [140]. As Mobulid ray catch rate continues to increase in the Indian

Ocean [23], protection for *M.alfredi* is essential. Although MPAs exist in the Indian Ocean [16,19], as a pelagic species shown to make long migrations, the limited size of these areas does not effectively protect subpopulations from anthropogenic pressures [16,19]. For example, there are several manta ray sanctuaries in Indonesia which range from 200-11,655 km<sup>2</sup> [16]. The waters which divide these sanctuaries are heavily fished with many fisheries which directly target manta rays [16]. Photo-ID surveys provided evidence that *M. alfredi* regularly migrates between these sanctuaries, making them vulnerable to target and bycatch fisheries as well as boat strikes [16]. Due to its extensive size, the BIOT no-take MPA has been recognised as an essential conservation strategy which may be spatially sufficient to protect such a highly mobile species [23]. In the current study, spatially sufficient protection is evident as all of the tagged individuals remained within the boundary of the MPA throughout the study period. As a species capable of making long migrations, it is theoretically possible that the BIOT *M.alfredi* subpopulation has some degree of connectivity with other subpopulations, for example, in the Seychelles. If so, they are potentially migrating through regions where they are vulnerable to fisheries [16]. However, the evidence presented here indicates that the tagged manta rays home range is centred around the atolls and it has been suggested that where an *M.alfredi* subpopulation has sufficient resources to sustain it, migration away from the region is unlikely [12]. Furthermore, to date, no clear evidence of connectivity between the highly fragmented *M.alfredi* subpopulations has been presented [6,9,12,141]. Further evidence gleaned from long-term research via active and passive tracking is required to investigate *M.alfredi* use of the BIOT MPA more thoroughly. Genetic analysis of this and other populations is also essential so to ascertain whether connectivity between neighbouring populations exists [9,12]. As well as identifying whether manta rays are migrating, genetic analysis may also be used to assess the population's genetic diversity [142]. Genetic diversity is associated with a species ability to adapt to environmental changes whereby a loss of diversity with increased levels of inbreeding can hinder survivorship through the accumulation of genetic mutations [142].

The BIOT MPA also provides *M.alfredi* with protection from some of the other most prominent threats which have been highlighted as a significant concern for *M. alfredi* subpopulations [9,16,143]. These threats include disturbance from unregulated tourism which has been shown to drive *M. alfredi* away from productive feeding areas [32,33]



potentially reducing the fitness and survivorship of the population [19,32]. Concomitant to touristic pressure is habitat degradation and an increased vulnerability to boat strikes and propeller injuries [9,16,143] which are much less likely to occur within the BIOT MPA. Although further research is required to determine whether the *M.alfredi* move outside of the MPA boundary, current evidence suggests the region's subpopulation do not thus may be the most highly protected in the world. This protection status is essential for this species as almost all the worlds subpopulations face future hardship induced by anthropogenic stressors and climate change [19] which will further decrease the already diminished population. Using the present study as a baseline, future research should continue to assess the effectiveness of the protection offered by the BIOT MPA, which can help to inform management planning of other MPAs in the Indian Ocean. For example, if the BIOT MPA is spatially sufficient, the expansion of other existing MPAs which only encompass limited areas within a subpopulation's home range should be considered so to reduce the impact of anthropogenic stressors (Stevens and Froman, 2018).

## 5. Conclusion

The present study utilised all of the currently available data to provide a summation of the current understanding of the BIOT *M.alfredi* subpopulation. To date, only 123 *M.alfredi* have been individually identified via photo-ID surveys, but the data provides evidence of productive feeding locations at three atolls: Egmont, Diego Garcia and Salomon Atoll. There may be some degree of demographic segregation of these habitats and the substantially higher number of sightings at Egmont Atoll indicates it is an aggregation hotspot. Analysis of tag data supports these observations whereby most position estimates occurred at this location although some individuals made relatively long migrations away from the area. Low current speeds potentially influenced by abrupt changes in bathymetric features and high SST around Egmont Atoll are plausibly increasing zooplankton biomass which supports *M.alfredi* aggregations. When foraging, MLD is also important as zooplankton can become more concentrated as it narrows. Generally, dive behaviour was consistent with reverse diel vertical migration, although this was not persistent and varied between individuals. The tagged manta rays did not appear to leave the BIOT MPA during this study, indicating that the MPA is providing spatially sufficient protection for this subpopulation. In a region where mobulids are under increasing threat from fisheries [20], an MPA which encompasses the entire range

of an *M.alfredi* subpopulation is essential for the protection of the whole population of the species. With aggregation hotspots and important environmental influences identified, the present study provides an essential baseline for future research into the BIOT *M.alfredi* subpopulation. The baseline can now be developed and may assist in identifying crucial habitats throughout the species range that might be essential for *M. alfredi* survivorship.

## **Declarations**

### *Acknowledgements*

Acknowledgements to the Foreign and Commonwealth Office (FCO) and the British Indian Ocean Territory Administration (BIOTA) who granted their supported for the research conducted in this study. Thank you to Guy Stevens for sharing the Chagos Manta Ray Project data and for all the support throughout, and to Josh Stewart and all the Manta Trust team who have been of assistance. I would also like to thank my supervisor, Clare Embling, for unwavering support and enthusiasm. I am also grateful to Phil Hosegood for assistance with the collection and analysis of essential oceanographic data and Jill Schwarz for the introduction to the chosen modelling techniques. Finally, thank you to Robert Pethick and Marie and Tony Lang whose support and generosity has made this research possible.

### *Ethics approval*

All tagging techniques implemented in this project were approved by the Foreign and Commonwealth Office (FCO) BIOT Administration (BIOTA) and are covered under the University of California, San Diego (UCSD) animal care protocol: UCSD IACUC Protocol #S12116.

### *Authors' contributions*

All authors contributed equally in the preparation of this manuscript. All authors read and approved the final manuscript.

### *Funding*

The expedition was funded through a grant from the Bertarelli Foundation.

### *Availability of data and materials*

The data that support the findings of this study are available from the Manta Trust, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Manta Trust

### *Competing interests*

The authors declare that they have no competing interests.

## References

1. White WT, Corrigan S, Yang L, Henderson AC, Bazinet AL, Swofford DL, *et al.* Phylogeny of the manta and devilrays (Chondrichthyes: mobulidae), with an updated taxonomic arrangement for the family. *Zool J Linn Soc.* 2017;182(1):50–75.
2. Ward-Paige CA, Davis B, Worm B. Global population trends and human use patterns of Manta and Mobula rays. *PLoS One.* 2013;8(9):e74835.
3. Kitchen-Wheeler AM, Ari C, Edwards AJ. Population estimates of Alfred mantas (*Manta alfredi*) in central Maldives atolls: North Male, Ari and Baa. *Environ Biol Fishes.* 2012;93(4):557–575.
4. Marshall AD, Bennett MB. Reproductive ecology of the reef manta ray *Manta alfredi* in southern Mozambique. *J Fish Biol.* 2010 Jul;77(1):169–90.
5. Kashiwagi T, Marshall AD, Bennett MB, Ovenden JR. Habitat segregation and mosaic sympatry of the two species of manta ray in the Indian and Pacific Oceans: *Manta alfredi* and *M. birostris*. *Mar Biodivers Rec.* 2011;4:1–8.
6. Couturier LIE, Marshall AD, Jaine FRA, Kashiwagi T, Pierce SJ, Townsend KA, *et al.* Biology, ecology and conservation of the Mobulidae. *J Fish Biol.* 2012;80:1075–119.
7. McCauley DJ, DeSalles PA, Young HS, Papastamatiou YP, Caselle JE, Deakos MH, *et al.* Reliance of mobile species on sensitive habitats: A case study of manta rays (*Manta alfredi*) and lagoons. *Mar Biol.* 2014;161(9):1987–98.
8. Armstrong AO, Armstrong AJ, Jaine FRA, Couturier LIE, Fiora K, Uribe-Palomino J, *et al.* Prey Density Threshold and Tidal Influence on Reef Manta Ray Foraging at an Aggregation Site on the Great Barrier Reef. *PLoS One.* 2016;11(5):e0153393.
9. Stewart JD, Jaine FRA, Armstrong AJ, Armstrong AO, Bennett MB, Burgess KB, *et al.* Research Priorities to Support Effective Manta and Devil Ray Conservation. *Front Mar Sci.* 2018;5:1–27.
10. Braun CD, Skomal GB, Thorrold SR, Berumen ML. Diving behavior of the reef manta ray links coral reefs with adjacent deep pelagic habitats. *PLoS One.* 2014;9(2):e88170.
11. Kessel ST, Elamin NA, Yurkowski DJ, Chekchak T, Walter RP, Klaus R, *et al.*

- Conservation of reef manta rays (*Manta alfredi*) in a UNESCO World Heritage Site: Large-scale island development or sustainable tourism? PLoS One. 2017;12(10): e0185419.
12. Deakos MH, Baker JD, Bejder L. Characteristics of a manta ray *Manta alfredi* population off Maui, Hawaii, and implications for management. Mar Ecol Prog Ser. 2011;429:245–60.
  13. Dewar H, Mous P, Domeier M, Muljadi A, Pet J, Whitty J. Movements and site fidelity of the giant manta ray, *Manta birostris*, in the Komodo Marine Park, Indonesia. Mar Biol. 2008;155(2):121–33.
  14. Setyawan E, Sianipar AB, Erdmann M V., Fischer AM, Haddy JA, Beale CS, *et al.* Site fidelity and movement patterns of reef manta rays (*Mobula alfredi*: Mobulidae) using passive acoustic telemetry in northern Raja Ampat, Indonesia. Nat Conserv Res. 2018;3(4).
  15. Couturier LIE, Newman P, Jaine FRA, Bennett MB, Venables WN, Cagua EF, *et al.* Variation in occupancy and habitat use of *Mobula alfredi* at a major aggregation site. Mar Ecol Prog Ser. 2018;599:125–45.
  16. Germanov ES, Marshall AD. Running the Gauntlet: Regional movement patterns of *Manta alfredi* through a complex of parks and fisheries. PLoS One. 2014;9(10): e110071.
  17. Couturier LIE, Jaine FRA, Townsend KA, Weeks SJ, Richardson AJ, Bennett MB. Distribution, site affinity and regional movements of the manta ray, *Manta alfredi* (Krefft, 1868), along the east coast of Australia. Marine and Freshwater Research. 2011;62:628–637.
  18. Rohner CA, Pierce SJ, Marshall AD, Weeks SJ, Bennett MB, Richardson AJ. Trends in sightings and environmental influences on a coastal aggregation of manta rays and whale sharks. Mar Ecol Prog Ser. 2013;482:153–68.
  19. Harris JL, McGregor P, Oates Y, Stevens G. Gone with the wind – seasonal movement ecology, habitat use, and conservation management of the reef manta ray (*Mobula alfredi*) in the Maldives. Submitted to Aquat Conserv Mar Freshw Ecosyst. 2019.

20. Lawson JM, Fordham S V., O'Malley MP, Davidson LNK, Walls RHL, Heupel MR, *et al.* Sympathy for the devil: a conservation strategy for devil and manta rays. *PeerJ*. 2017;5:e3027.
21. Dulvy NK, Pardo SA, Simpfendorfer CA, Carlson JK. Diagnosing the dangerous demography of manta rays using life history theory. *PeerJ*. 2014;2:e400.
22. Sheppard, Ateweberhan M, Bowen BW, Carr P, Chen CA, Clubbe C, *et al.* Reefs and islands of the Chagos Archipelago, Indian Ocean: Why it is the world's largest no-take marine protected area. *Aquat Conserv Mar Freshw Ecosyst*. 2012;22(2):232–61.
23. Koldewey HJ, Curnick D, Harding S, Harrison LR, Gollock M. Potential benefits to fisheries and biodiversity of the Chagos Archipelago / British Indian Ocean Territory as a no-take marine reserve. *Mar Pollut Bull*. 2010;60(11):1906–15.
24. Temple AJ, Kiszka JJ, Stead SM, Wambiji N, Brito A, Poonian CNS, *et al.* Marine megafauna interactions with small-scale fisheries in the southwestern Indian Ocean: a review of status and challenges for research and management. *Rev Fish Biol Fish*. 2017;28(1):89-115.
25. Cortés E, Papastamatiou Y, Carlson J, Ferry-Graham L, Wetherbee B. An Overview of the Feeding Ecology and Physiology of Elasmobranch Fishes. JEP C, D B, BG K, editors. *Feeding and Digestive Functions in Fishes*. Boca Raton, FL: Science Publishers; 2014. p. 393–443.
26. Croll DA, Dewar H, Dulvy NK, Fernando D, Francis MP, Galván-Magaña F, *et al.* Vulnerabilities and fisheries impacts: the uncertain future of manta and devil rays. *Aquat Conserv Mar Freshw Ecosyst*. 2016;26(3):562–75.
27. Marshall A, Kashiwagi T, Bennett MB, Deakos M, Stevens G, McGregor F, *et al.* Reef Manta Ray (*Manta alfredi*). The IUCN Red List of Threatened Species 2016. 2011. Available from: <http://www.iucnredlist.org/details/195459/0>. Accessed 10 Oct 2017.
28. CMS (Convention on the Conservation of Migratory Species of Wild Animals). Proceedings of the Tenth Meeting of the Conference of the Parties. Bergen, Norway; 2011.

29. O'Malley MP, Lee-Brooks K, Medd HB. The Global Economic Impact of Manta Ray Watching Tourism. *PLoS One*. 2013;8(5):e65051.
30. Dharmadi, Fahmi, Satria F. Fisheries management and conservation of sharks in Indonesia. *African J Mar Sci*. 2015;37(2):249–58.
31. Dulvy NK, Pardo SA, Simpfendorfer CA, Carlson JK. Diagnosing the dangerous demography of manta rays using life history theory. *PeerJ*. 2014;2:e400.
32. Murray A, Garrud E, Ender I, Lee-Brooks K, Atkins R, Lynam R, *et al*. Protecting the million-dollar mantas; creating an evidence-based code of conduct for manta ray tourism interactions. *In Press J Ecotourism*. 2019;
33. Anderson RC, Adam MS, Kitchen-Wheeler A-M, Stevens G. Extent and Economic Value of Manta Ray Watching in Maldives. *Tour Mar Environ*. 2010;7(1):15–27.
34. Singleton RL, Roberts CM. The contribution of very large marine protected areas to marine conservation: Giant leaps or smoke and mirrors? *Mar Pollut Bull*. 2014;87(1):7–10.
35. Toonen RJ, Wilhelm TA, Maxwell SM, Wagner D, Bowen BW, Sheppard CRC, *et al*. One size does not fit all: The emerging frontier in large-scale marine conservation. *Mar Pollut Bull*. 2013;77(1–2):7–10.
36. Graham RT, Witt MJ, Castellanos DW, Remolina F, Maxwell S, Godley BJ, *et al*. Satellite tracking of manta rays highlights challenges to their conservation. *PLoS One*. 2012;7(5):e36834.
37. Brooks K. Investigating tourism at Hanifaru Bay marine protected area, Maldives. MSc Thesis. York. University of York. 2010.
38. Steneck RS, Mumby PJ, MacDonald C, Rasher DB, Stoyle G. Attenuating effects of ecosystem management on coral reefs. *Sci Adv*. 2018;4(5):1–12.
39. Braun CD, Skomal GB, Thorrold SR, Berumen ML. Movements of the reef manta ray (*Manta alfredi*) in the Red Sea using satellite and acoustic telemetry. *Mar Biol*. 2015;162(12):2351–62.
40. Readman JW, Deluna F, Ebinghaus R, Guzman A, Price ARG, Emily E, *et al*.

- Contaminants, Pollution and Potential Anthropogenic Impacts in Chagos/BIOT. In: Sheppard C, editor. Coral reefs of the United Kingdom overseas territories. 1st ed. Dordrecht: Springer Netherlands; 2013. p. 283-295.
41. Graham NAJ, Purkis SJ, Harris A. Diurnal, land-based predation on shore crabs by moray eels in the Chagos Archipelago. *Coral Reefs*. 2009;28(2):397.
  42. Hosegood PJ, Nimmo-Smith WAM, Proud R, Adams K, Brierley AS. Internal lee waves and baroclinic bores over a tropical seamount shark 'hot-spot.' *Prog Oceanogr*. 2019;172:34–50.
  43. Sheppard CRC, Bowen BW, Chen AC, Craig MT, Eble J, Fitzsimmons N, *et al*. *Coral Reefs of the United Kingdom Overseas Territories*. Dordrecht: Springer Netherlands. 2013.
  44. Manta Trust. Chagos Manta Ray Project. <https://www.mantatrust.org/chagos> (2019). Accessed 25 June 2019.
  45. Peel L, Stevens G, Daly R, Daly C, Lea J, Clarke C, *et al*. The island life: movement and residency patterns of reef manta rays *Mobula alfredi* in the Amirante Islands, Seychelles. *Mar Ecol Prog Ser*. 2019;621:169–84.
  46. Kitchen-Wheeler A. The behaviour and ecology of Alfred mantas ( *Manta alfredi* ) in the Maldives. Newcastle: Newcastle University; 2013.
  47. Rohner CA, Richardson AJ, Jaine FRA, Bennett MB, Weeks SJ, Cliff G, *et al*. Satellite tagging highlights the importance of productive Mozambican coastal waters to the ecology and conservation of whale sharks. *PeerJ*. 2018;6:e4161.
  48. Hearn AR, Green JR, Espinoza E, Peñaherrera C, Acuña D, Klimley AP. Simple criteria to determine detachment point of towed satellite tags provide first evidence of return migrations of whale sharks (*Rhincodon typus*) at the Galapagos Islands, Ecuador. *Anim Biotelemetry*. 2013;1(1):1–10.
  49. George Karleskint, Richard Turner JS. *Introduction to Marine Biology*. 3rd ed. Belmont: Cengage Learning; 2009.
  50. Create T, Data E, Using V, Description G. *Create Elegant Data Visualisations Using the*



- Grammar of Graphics Package 'ggplot2.' 2019. <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>. Accessed 21 June 2019.
51. Jonsen I. Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Sci Rep.* 2016;6:1–9.
  52. Jonsen ID, Flemming JM, Myers RA. Robust state-space modeling of animal movement data. *Ecology.* 2005;86(11):2874–80.
  53. Jonsen ID, Myers RA, James MC. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Mar Ecol Prog Ser.* 2007;337:255–64.
  54. Breed GA, Jonsen ID, Myers RA, Bowen WD, Marty L, Breed GA, *et al.* Sex-Specific , Seasonal Foraging Tactics of Adult Grey Seals (*Halichoerus grypus*) Revealed by State-Space Analysis. *Ecology.* 2009;90(11):3209–21.
  55. Jonsen ID, Wotherspoon SJ, Bravington MV, Thomson R, Bestley S, Patterson TA, *et al.* State-space models for bio-loggers: A methodological road map. *Deep Sea Res Part II Top Stud Oceanogr.* 2013:34–46.
  56. Jonsen I, Bestley S, Wotherspoon S, Sumner M, Mills J. Bayesian State-Space Models for Animal Movement (Package bsam). 2017. <https://cran.r-project.org/web/packages/bsam/bsam.pdf> Accessed 12 Nov 2018.
  57. Plummer MM. Bayesian Graphical Models using MCMC (Package rjags). 2018. <https://cran.r-project.org/web/packages/rjags/rjags.pdf>. Accessed 12 Jan 2019.
  58. Plummer M. JAGS: Just Another Gibbs Sampler. Version 4.3.0. 2019. <http://mcmc-jags.sourceforge.net> Accessed 12 Jan 2019.
  59. Reisinger RR, Keith M, Andrews RD, de Bruyn PJN. Movement and diving of killer whales (*Orcinus orca*) at a Southern Ocean archipelago. *J Exp Mar Bio Ecol.* 2015;473:90–102.
  60. Prieto R, Silva MA, Waring GT, Gonçalves JMA. Sei whale movements and behaviour in the North Atlantic inferred from satellite telemetry. *Endanger Species Res.* 2014;26(2):103–13.

61. Robert CP, Casella G. *Introducing Monte Carlo Methods with R*. New York: Springer Science & Business Media; 2010.
62. Hijmans RJ, Williams E, Vennes C. Spherical Trigonometry Version Package "geosphere". 2017. <https://cran.r-project.org/web/packages/geosphere/geosphere.pdf>. Accessed 19 Feb 2019.
63. Henry L. A Grammar of Data Manipulation Package "dplyr". 2019. <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>. Accessed 19 July 2019.
64. Lichti NI, Swihart RK. Estimating utilization distributions with kernel versus local convex hull methods. *J Wildl Manage*. 2011;75(2):413–22.
65. Calenge C, Fortmann-Roe S. Home Range Estimation Package "adehabitatHR". 2017. <https://cran.r-project.org/web/packages/adehabitatHR/adehabitatHR.pdf>. Accessed 9 March 2019.
66. Deng H, Wickham H. Density Estimation In R. 2011. <https://vita.had.co.nz/papers/density-estimation.pdf>. Accessed 9 March 2019.
67. Acuña-Marrero D, Jiménez J, Smith F, Doherty PF, Hearn A, Green JR, *et al*. Whale shark (*Rhincodon typus*) seasonal presence, residence time and habitat use at darwin island, galapagos marine reserve. *PLoS One*. 2014;9(12):e115946.
68. Worton BJ. Using Monte Carlo Simulation to Evaluate Kernel-Based Home Range Estimators. *J Wildl Manage*. 1995;59(4):794–800.
69. Hammerschlag N, Luo J, Irschick DJ, Ault JS. A Comparison of Spatial and Movement Patterns between Sympatric Predators: Bull Sharks (*Carcharhinus leucas*) and Atlantic Tarpon (*Megalops atlanticus*). *PLoS One*. 2012;7(9).
70. Thums M, Waayers D, Huang Z, Pattiaratchi C, Bernus J, Meekan M. Environmental predictors of foraging and transit behaviour in flatback turtles *Natator depressus*. *Endanger Species Res*. 2017;32(1):333–49.
71. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: How, where and how many? *Methods Ecol Evol*. 2012;3(2):327–38.

72. Lobo JM, Jiménez-valverde A, Real R. AUC: A misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr.* 2008;17(2):145–51.
73. Jaine FRA, Couturier LIE, Weeks SJ, Townsend, Kathy A. Bennett MB, Fiora K, Richardson AJ. When Giants Turn Up: Sighting Trends, Environmental Influences and Habitat Use of the Manta Ray *Manta alfredi* at a Coral Reef. *PLoS One.* 2012;7(10):e46170.
74. Cotté C, D'Ovidio F, Chaigneau A, Lévy M, Taupier-Letage I, Mate B, *et al.* Scale-dependent interactions of Mediterranean whales with marine dynamics. *Limnol Oceanogr.* 2011;56(1):219–32.
75. Weeks SJ, Magno-Canto MM, Jaine FRA, Brodie J, Richardson AJ. Unique sequence of events triggers manta ray feeding frenzy in the Southern Great Barrier Reef, Australia. *Remote Sens.* 2015;7(3):3138–52.
76. Stewart JD, Hoyos-Padilla EM, Kumli KR, Rubin RD. Deep-water feeding and behavioral plasticity in *Manta birostris* revealed by archival tags and submersible observations. *Zoology.* 2016;119(5):406–13.
77. Sleeman JC, Meekan MG, Wilson SG, Polovina JJ, Stevens JD, Boggs GS, *et al.* To go or not to go with the flow: Environmental influences on whale shark movement patterns. *J Exp Mar Bio Ecol.* 2010;390(2):84–98.
78. McKinney JA, Hoffmayer ER, Wu W, Fulford R, Hendon JM. Feeding habitat of the whale shark *Rhincodon typus* in the northern Gulf of Mexico determined using species distribution modelling. *Mar Ecol Prog Ser.* 2012;458(July):199–211.
79. Roberts JJ, Best BD, Dunn DC, Treml EA, Halpin PN. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environ Model Softw.* 2010;25(10):1197–207.
80. Welch H, Pressey RL, Reside AE. Using temporally explicit habitat suitability models to assess threats to mobile species and evaluate the effectiveness of marine protected areas. *J Nat Conserv.* 2018;41:106–15.
81. Welch H, Pressey RL, Heron SF, Ceccarelli DM, Hobday AJ. Regimes of chlorophyll-a in

- the Coral Sea: Implications for evaluating adequacy of marine protected areas. *Ecography (Cop)*. 2016;39(3):289–304.
82. D’Errico JR. Understanding Gridfit The mechanical and philosophical underpinnings. *Methodology*. 2006;1(1):1–6.
  83. Chassignet EP, Halliwell GR, Hurlburt HE, Hogan PJ, Wallcraft AJ, Smedstad OM, *et al*. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *J Mar Syst*. 2007;65:60–83.
  84. Kara AB, Rochford PA, Hurlburt HE. An optimal definition for ocean mixed layer depth. *J Geophys Res Ocean*. 2000;105(C7):16803–21.
  85. Keerthi MG, Lengaigne M, Vialard J, de Boyer Montégut C, Muraleedharan PM. Interannual variability of the Tropical Indian Ocean mixed layer depth. *Clim Dyn*. 2013;40(3–4):743–59.
  86. de Boyer Montégut C, Madec G, Fischer AS, Lazar A, Iudicone D. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J Geophys Res C Ocean*. 2004;109(12):1–20.
  87. PO.DAAC. Physical Oceanography Distributed Active Archive Center [Internet]. NASA EOSDIS PO.DAAC. 2015 [cited 2019 Mar 3]. <https://podaac.jpl.nasa.gov/>. Accessed 20 March 2019.
  88. Cayula J-F, Cornillon P. Edge Detection Algorithm for SST Images. *Am Meteorol Soc*. 1992;9(1):67–80.
  89. British Oceanographic Data Centre (BODC). General Bathymetric Chart of the Oceans (GEBCO). 2014. <https://www.bodc.ac.uk/>. Accessed 20 Dec 2018.
  90. NASA. Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra satellites. 2015. [oceancolor.gsfc.nasa.gov](http://oceancolor.gsfc.nasa.gov). 1 Nov 2018.
  91. Copernicus Marine Environment Monitoring Service (CMEMS). Archiving, Validation and Interpretation of Satellite Oceanographic Data (AVISO). 2015. [marine.copernicus.eu](http://marine.copernicus.eu). Accessed 20 Feb 2019.
  92. Duke University Marine Geospatial Ecology Laboratory. Marine Geospatial Ecology

- Tools (MGET). 2010. <http://mgel.env.duke.edu/tools>. Accessed 5 Jan 2019.
93. NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC). Group for High-Resolution Sea Surface Temperature (GHRSSST). 2015. [podaac.jpl.nasa.gov](http://podaac.jpl.nasa.gov). Accessed 5 March 2019.
  94. Bonjean F, Lagerloef GSE. Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean. *J Phys Oceanogr*. 2002;32(10):2938–54.
  95. NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC). Ocean Surface Current Analyses Real-time (OSCAR). 2015. [podaac.jpl.nasa.gov](http://podaac.jpl.nasa.gov). Accessed 3 March 2019.
  96. Metzger EJ, States U, Smedstad OM, States U, Chassignet EP. Predicting “ocean weather” using the HYbrid Coordinate Ocean Model (HYCOM). *Ocean Sci Technol*. 2009:1–8.
  97. HYCOM. HYbrid Coordinate Ocean Model. 2015. [www.hycom.org](http://www.hycom.org). Accessed 1 March 2019.
  98. Hijmans, Phillips S, Leathwick JR, Elith J. Species Distribution Modeling Version Dismo package for R, version 1.1-4. 2017. <http://cran.r-project.org/package=dismo>. Accessed 3 Jan 2019.
  99. Leathwick JR, Elith J, Francis MP. Variation in demersal fish species richness in the oceans surrounding New Zealand. *Mar Ecol Prog Ser*. 2006;321:267–81.
  100. Froeschke J, Stunz GW, Wildhaber ML. Environmental influences on the occurrence of coastal sharks in estuarine waters. *Mar Ecol Prog Ser*. 2010;407:279–92.
  101. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. 2008;77(4):802–13.
  102. Dedman S, Officer R, Brophy D, Clarke M, Reid DG. Advanced Spatial Modeling to Inform Management of Data-Poor Juvenile and Adult Female Rays. *Fishes*. 2017;2(3):12.
  103. Torres L, Baker C, Derville S, Constantine R, Oremus M. Environmental correlates of nearshore habitat distribution by the Critically Endangered Maūi dolphin. *Mar Ecol*

- Prog Ser. 2016;551:261–75.
104. Fielding AH, Bell JF, Fielding AH, Bell JF. A review of methods for the assessment of prediction errors in conservation presence / absence models. *Environ Conserv.* 1997;24(1):38–49.
  105. Elith J, Leathwick J. Boosted Regression Trees for ecological modeling. 2017. <https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf>. Accessed 20 Nov 2018.
  106. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. Toronto: John Wiley & Sons Inc.; 2000.
  107. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning The Elements of Statistical Learning Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer-Verlag; 2009.
  108. Greenwell B, Boehmke B, Cunningham J. *Generalized Boosted Regression ModelsPackage ' gbm '*. 2019. <https://cran.r-project.org/web/packages/gbm/gbm.pdf>. Accessed 20 April 2019.
  109. Burnham K., Anderson D. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Vol. 172. New York: Springer; 2002. p. 488.
  110. Germanov ES, Bejder L, Chabanne DBH, Dharmadi D, Hendrawan IG, Marshall AD, *et al.* Contrasting Habitat Use and Population Dynamics of Reef Manta Rays Within the Nusa Penida Marine Protected Area, Indonesia. *Front Mar Sci.* 2019. doi: 10.3389/fmars.2019.00215.
  111. Stevens GMW. *Conservation and Population Ecology of Manta Rays in the Maldives*. York: University of York; 2016.
  112. Jaine FRA, Rohner CA, Weeks SJ, Couturier LIE, Bennett MB, Townsend KA, *et al.* Movements and habitat use of reef manta rays off eastern Australia: Offshore excursions, deep diving and eddy affinity revealed by satellite telemetry. *Mar Ecol Prog Ser.* 2014;510:73–86.
  113. Cerutti-Pereyra F, Thums M, Austin CM, Bradshaw CJA, Stevens JD, Babcock RC, *et al.*

- Restricted movements of juvenile rays in the lagoon of Ningaloo Reef, Western Australia - evidence for the existence of a nursery. *Environ Biol Fishes*. 2014;97(4):371–83.
114. Stewart JD, Nuttall M, Hickerson EL, Johnston MA. Important juvenile manta ray habitat at Flower Garden Banks National Marine Sanctuary in the northwestern Gulf of Mexico. *Mar Biol* [Internet]. 2018; doi:10.1007/s00227-018-3364-5.
  115. Nøttestad L, Giske J, Holst JC, Huse G. A length-based hypothesis for feeding migrations in pelagic fish. *Can J Fish Aquat Sci*. 1999;56(S1):26–34.
  116. Hearn CJ. Hydrodynamics of coral reef systems. In: Hopley D, editor. *Encyclopedia of Modern Coral Reefs: Structure, Form and Process*. Dordrecht: Springer Netherlands; 2011.
  117. Carleton JH, Brinkman R, Doherty PJ. Zooplankton community structure and water flow in the lee of Helix Reef (Great Barrier Reef, Australia). *Mar Biol*. 2001;139(4):705–17.
  118. McManus MA, Woodson CB. Plankton distribution and ocean dispersal. *J Exp Biol*. 2012;215(6):1008–16.
  119. McManus M, Cheriton O, Greenlaw C, Donaghay P, Storlazzi C, Holliday D, *et al*. Effects of physical processes on structure and transport of thin zooplankton layers in the coastal ocean. *Mar Ecol Prog Ser*. 2005;301:199–215.
  120. Bode A, Alvarez-Ossorio MT, Miranda A, López-Urrutia A, Valdés L. Comparing copepod time-series in the north of Spain: Spatial autocorrelation of community composition. *Prog Oceanogr*. 2012;97–100:108–19.
  121. Cedras RB. Patterns in diversity, abundance, distribution and community structure of epi-pelagic copepods in the south-western Indian Ocean. PhD Thesis. Western Cape: University of the Western Cape; 2016.
  122. González-Gil R, Taboada FG, Höfer J, Anadón R. Winter mixing and coastal upwelling drive long-term changes in zooplankton in the Bay of Biscay (1993-2010). *J Plankton Res*. 2015;37(2):337–51.

123. Voynova YG, Oliver MJ, Sharp JH. Wind to zooplankton: Ecosystem-wide influence of seasonal wind-driven upwelling in and around the Delaware Bay. *J Geophys Res Ocean*. 2013;118(12):6437–50.
124. Choukroun S, Ridd P V., Brinkman R, McKinna LIW. On the surface circulation in the western Coral Sea and residence times in the Great Barrier Reef. *J Geophys Res Ocean*. 2010;115(6):1–13.
125. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. Toward a Metabolic Theory of Ecology. *Ecology*. 2004;85(7):1771–89.
126. Sommer U, Lengfellner K. Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom. *Glob Chang Biol*. 2008;14(6):1199–208.
127. Olonscheck D, Hofmann M, Worm B, Schellnhuber HJ. Decomposing the effects of ocean warming on chlorophyll a concentrations into physically and biologically driven contributions. *Environ Res Lett*. 2013;8(1): 014043.
128. Lewandowska AM, Boyce DG, Hofmann M, Matthiessen B, Sommer U, Worm B. Effects of sea surface warming on marine plankton. *Ecol Lett*. 2014;17(5):614–23.
129. Kämpf J, Chapman P. The Functioning of Coastal Upwelling Systems. *Upwelling Systems of the World: A Scientific Journey to the Most Productive Marine Ecosystems*. 2016. p. 1–433.
130. Genin A, Jaffe JS, Reef R, Richter C, Franks PJS. Swimming against the flow: A mechanism of zooplankton aggregation. *Science*. 2005;308(5723):860–2.
131. Couturier LIE, Rohner CA, Richardson AJ, Marshall AD, Jaine FRA, Bennett MB, *et al*. Stable Isotope and Signature Fatty Acid Analyses Suggest Reef Manta Rays Feed on Demersal Zooplankton. *PLoS One*. 2013;8(10):e77152.
132. Gleiss AC, Norman B, Wilson RP. Moved by that sinking feeling: Variable diving geometry underlies movement strategies in whale sharks. *Funct Ecol*. 2011;25(3):595–607.
133. Shea ORO, Kingsford MJ, Seymour J. Tide-related periodicity of manta rays and sharks



- to cleaning stations on a coral reef. *Mar Freshw Res.* 2010;61:65–73.
134. Kuitert RH. *Fishes of the Maldives – Indian Ocean: Applicable to Central and Western Indian Ocean: East Africa, Seychelles, Mauritius, Reunion, Madagascar, Sri Lanka, Chagos, Laccadives, Cocos Keeling Islands.* Cairns, Australia: Atoll Editions; 2014.
  135. Saltzman J, Wishner KF. Zooplankton ecology in the eastern tropical pacific oxygen minimum zone above a seamount: 1. General trends. *Deep Res Part I Oceanogr Res Pap.* 1997;44(6):907–30.
  136. Madhuratap M, Gopalakrishnan TC, Haridas P, Nair KKC, Aravindakshan PN, Padmavati G, *et al.* Lack of seasonal and geographic variation in mesozooplankton biomass in the Arabian Sea and its structure in the mixed layer. *Curr Sci.* 1996;71(11):863–8.
  137. Litchman E, Miller JR, Schofield OM, Klausmeier CA, Falkowski PG. Multi-nutrient, multi-group model of present and future oceanic phytoplankton communities. *Biogeosciences Discuss.* 2010;3(3):607–63.
  138. IOTC. Indian Ocean Tuna Commission Database: Nominal catch by species and gear, by vessel flag reporting country. 2019. <https://www.iotc.org/data/datasets/latest/NC>. Accessed 3 July 2019.
  139. Garcia A, Herrera M. Assessing the Contribution of Purse Seine Fisheries to Overall Levels of Bycatch in the Indian Ocean. 2018. <https://www.iotc.org/documents/WPDCS/14/26-PSBycatch>. Accessed 3 July 2019.
  140. Lawson JM, Fordham S V. *Sharks Ahead: Realizing the Potential of the Convention on Migratory Species to Conserve Elasmobranchs.* Washington, DC; 2018. [https://docs.wixstatic.com/ugd/5c8a4b\\_c7b7dd1abd1e4c2d9ea26f88a87a0edd.pdf](https://docs.wixstatic.com/ugd/5c8a4b_c7b7dd1abd1e4c2d9ea26f88a87a0edd.pdf). Accessed 20 July 2019.
  141. Couturier LIE, Dudgeon CL, Pollock KH, Jaine FRA, Bennett MB, Townsend KA, *et al.* Population dynamics of the reef manta ray *Manta alfredi* in eastern Australia. *Coral Reefs.* 2014;33(2):329–42.
  142. Heyden S Von Der. *Coral Reefs.* Heidelberg: Springer Berlin Heidelberg; 2016.

143. Stevens GM., Froman N. The Maldives Archipelago. In: Sheppard C, editor. World Seas: An Environmental Evaluation: Volume II: The Indian Ocean to the Pacific. 2nd ed. London: Academic Press; 2018. p. 211–36.

## Appendices

### Appendix A – Argos satellite data processing

#### (i) Location quality class

Argos position estimates are calculated by measuring the Doppler Effect on transmission frequency from a transmitter [1]. The estimated accuracy of a position is classed based on location process (e.g. number of transmissions received during a satellite pass), and estimated errors established by a positioning algorithm (least squares analysis or Kalman filtering). Each location point is then assigned one of seven location quality classes (LCs) [2]. Higher quality locations that are derived from four or more transmissions fall into the top four LCs (Argos classes 0-3) and are accompanied by an estimated radius of error [2] which is the radius of the ellipse estimated to contain the true location [1]. Commonly, low-quality LCs (Argos classes A, B and Z) for which estimated radius of error could not be established dominate animal tracking data [3]. Rather than discard potentially highly erroneous location estimates Argos derived location errors may be corrected by fitting a state-space model (SSM) to the transmission data [4]. Moreover, an SSM can be used regularise Argos position estimates (see Appendix C(i)) that are generally temporally irregular [4] which is a particular problem in the case of a pelagic species [5] such as the manta ray.

#### (ii) Removal of data received from a detached tag

Before Argos data can be utilised, the point of tag detachment needs to be established. Tag detachment can be determined through the assessment of the following three factors:

##### (1) *The temporal consistency of transmissions to Argos.*

The satellite tags can only transmit via Argos when the tag is dry [1] therefore, tags deployed on a pelagic species such as *M. alfredi* will only transmit when the animal surfaces [1,6]. While attached, daily transmissions will vary in their temporal distribution [6]. After detachment, they are more likely to occur in the first few hours after midnight [6]. Table A1 shows an example of the point of detachment (red box) of the tag deployed on manta 144475 ascertained from reduced temporal variability in the daily transmission after this point.

(2) *The quality of transmissions*

Argos position estimates are assigned one of seven location quality classes (LCs) (Appendix A(i)). When a tag is floating on the surface, the LCs are generally of consistently higher quality [6] as >4 transmissions are likely to be achieved [1]. After the red box in Table A1, LC is of consistently higher quality indicating the detachment of the tag deployed on manta 144475.

**Table A1** - Argos position estimates for a manta 144475. The red box indicates the last transmission before the tag detached. Above the red box, position estimates are temporally variable, and the location quality class (LC) is generally low (LC = A, B or Z). Below the red box, the daily transmissions are received in the first few hours after midnight and are of a consistently higher LC (LC = 0-3).

DeployID	Date	LC	Latitude	Longitude
144475	25/01/2015 20:00	B	-6.675	71.391
144475	25/01/2015 21:39	B	-6.676	71.39
144475	26/01/2015 13:39	B	-6.675	71.391
144475	26/01/2015 16:19	B	-6.676	71.391
144475	26/01/2015 17:51	B	-6.678	71.394
144475	26/01/2015 19:52	A	-6.673	71.399
144475	30/01/2015 13:46	B	-7.286	72.245
144475	06/02/2015 11:33	1	-6.674	71.391
144475	06/02/2015 12:55	A	-6.642	71.387
144475	06/02/2015 13:18	B	-6.641	71.396
144475	07/02/2015 00:03	2	-6.561	71.35
144475	07/02/2015 00:17	3	-6.575	71.24
144475	07/02/2015 01:39	2	-6.567	71.237
144475	08/02/2015 00:06	2	-6.408	71.098
144475	08/02/2015 01:13	2	-6.544	71.047
144475	09/02/2015 00:02	1	-6.591	70.824
144475	09/02/2015 00:45	2	-6.59	70.918
144475	10/02/2015 00:28	1	-6.576	70.999
144475	10/02/2015 02:01	2	-6.579	71.034
144475	11/02/2015 00:02	3	-6.664	71.473
144475	11/02/2015 01:17	2	-6.487	71.191
144475	11/02/2015 01:28	2	-6.483	71.186

(3) *The range of time-at-temperature (TAT) bins recorded by the tag*

The satellite tags were programmed to record the ambient temperature. Recordings result in a time-at-temperature (TAT) histogram which averages the temperature measurements at six-hourly intervals into 12 predefined temperature ranges (bins) and gives the percentage of time within that six hours that the manta spent within each bin. When a tag is attached, there

will be 100% spread over a range of bins [6]. When the tag begins to report its time as 100% in a single bin or split between the same two bins for multiple consecutive histogram entries, particularly the warmest bin, the tag has detached and is drifting on the surface [6] (Table A2, red box).

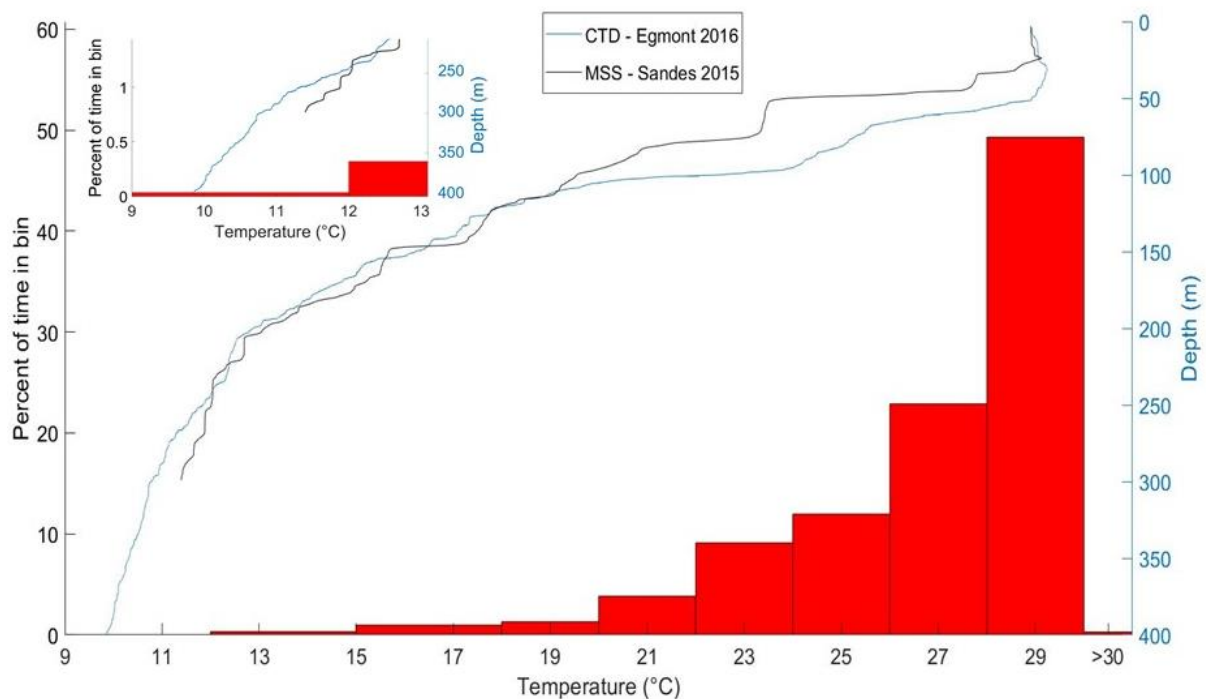
**Table A2** - Time-at-temperature (TAT) histogram data recorded at six-hourly intervals and displayed as the proportion (%) of time the manta (manta 144475) spent within each predefined temperature range (bin). The red box indicates the last recording prior to the detachment of the tag. Subsequent histogram entries are only recorded in one or two of the warmest bins for multiple consecutive entries.

Tag ID	Date	Predefined Temperature Bins/Percentage of Time											
		Bin1	Bin2	Bin3	Bin4	Bin5	Bin6	Bin7	Bin8	Bin9	Bin10	Bin11	Bin12
144475	03/02/2015 21:00	0	0	0	0	0	0	2.8	22.6	44.2	16.3	14.2	0
144475	04/02/2015 03:00	0	0	0	3.3	6.1	3.1	5.8	24.7	23.1	20	13.9	0
144475	04/02/2015 09:00	0	0	0	0.2	1.5	1.7	3.1	10.2	19	37.3	27	0
144475	04/02/2015 21:00	0	0	0	0	0	0	1.5	2	19	44.5	32.9	0
144475	05/02/2015 09:00	0	0	0	0	0	0	2.8	6.1	10	18.1	63.1	0
144475	05/02/2015 15:00	0	0	0	0	0	0	1.6	9	8.8	28.6	52	0
144475	05/02/2015 21:00	0	0	0	0	0	0	0	7.7	2.6	4.4	85.2	0
144475	06/02/2015 03:00	0	0	0	0	0	0	5.7	44.6	31.1	12.2	6.3	0
144475	06/02/2015 09:00	0	0	0	0	0	2.3	2.6	13.6	1.9	2.3	77.3	0
144475	06/02/2015 15:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	06/02/2015 21:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	07/02/2015 09:00	0	0	0	0	0	0	0	0	0	0	59.4	40.6
144475	07/02/2015 21:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	08/02/2015 03:00	0	0	0	0	0	0	0	0	0	0	86.7	13.3
144475	08/02/2015 09:00	0	0	0	0	0	0	0	0	0	0	75	25
144475	08/02/2015 15:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	08/02/2015 21:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	09/02/2015 09:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	10/02/2015 03:00	0	0	0	0	0	0	0	0	0	0	100	0
144475	10/02/2015 09:00	0	0	0	0	0	0	0	0	0	0	100	0

## Appendix B - Time at temperature (TAT)

### (i) Method for establishing depth

To establish depth from temperature range, vertical temperature profiles along with the mean percentage of time the manta rays spent within each temperature bin were plotted in MATLAB using the code provided in Appendix B(ii). The resulting plot (Fig. B1) shows that temperature profiles converge at depths, and there is  $<0.5^{\circ}\text{C}$  difference at 300m. The range of each temperature bin (displayed as the width of the red bars) was used to infer the approximate depth range, which is indicated by the vertical profiles. The Sandes (2015) profile was used for depths up to 300m, and the Egmont (2016) profile was used for depths  $>300\text{m}$ .

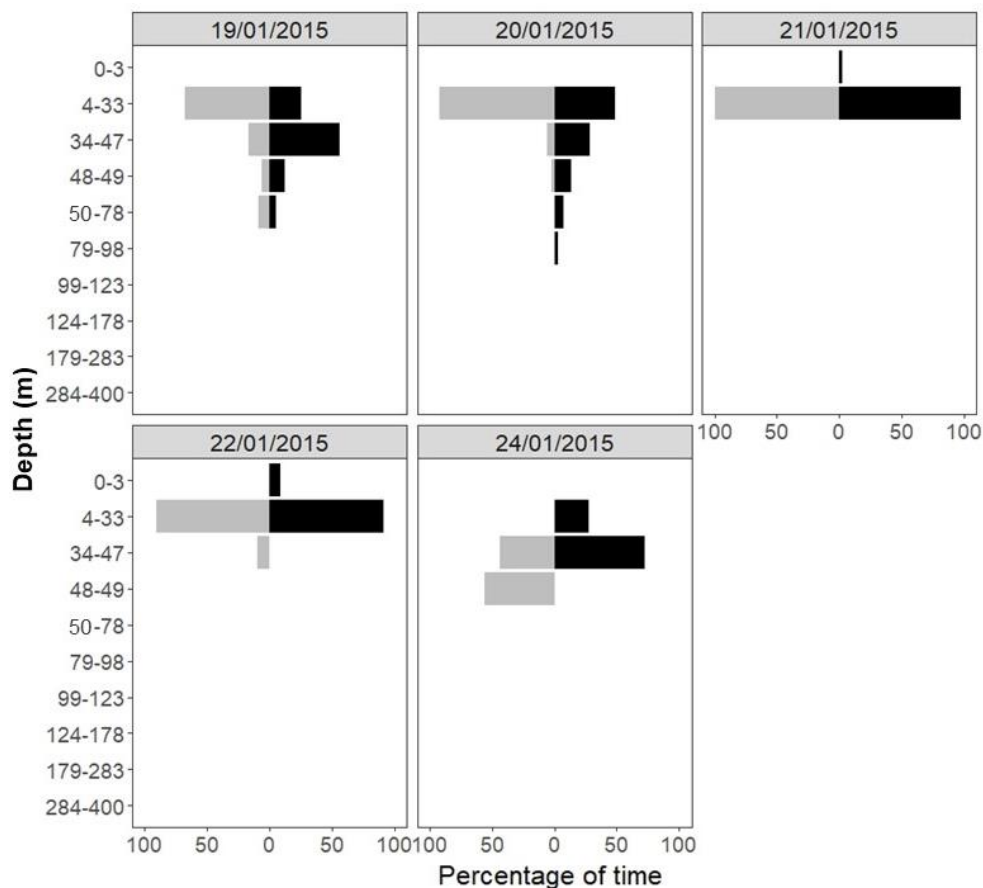


**Figure B1** - Vertical temperature and depth profiles constructed from data obtained at two locations (Fig. 1). ISW Microstructure Sensing System (MSS) equipped with a temperature sensor cast to 300m (black line) and compact multi-parameter conductivity, temperature, depth (CTD) system with a cast to 400m (blue line) in 2015 and 2016 respectively. Red bars indicate the overall percentage of time manta rays occupied each temperature bin whereby the width indicates the temperature range of the bin. Inset: lowest temperature bin is displayed from 200-400m showing a small percentage of time was spent in the 9-12°C bin.

## (ii) Daily total TAT data for manta rays

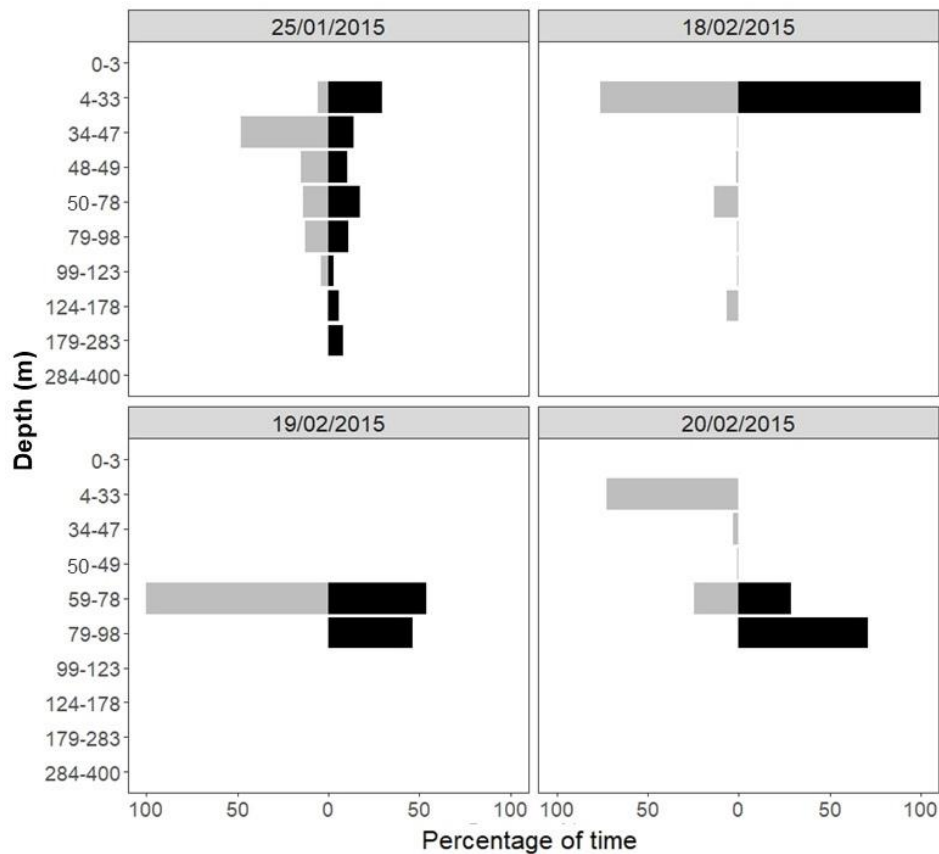
Time-at-temperature (TAT) data is only transmitted if measurements recorded precede a satellite transmission, thus large gaps in the data can occur if the manta ray does not surface regularly [7].

When considering the daily total percentage of time manta rays spent at depths where both day and night data is available (n=30), all manta rays were observed to occupy deeper water during the day for a combined total of eleven days (37%). Manta 144472 followed this diel pattern for three out of five days (Fig. B2). For two of these days (21/01/2015 and 22/01/2015) 100% of the night was spent at depths <33 m. During the day on 24/01/2015 this manta ray remained between 34-49 m and only occupied <33 m during the night. The remaining two days (19/01/2015 and 20/01/2015) are more typical of a reverse diel vertical migration pattern.



**Figure B2** – Total percentage of time spent at depths for the five days where both day and night were recorded for manta 144472. Three days (21/01/2015, 22/01/2015 and 24/01/2015) show the manta ray predominantly occupied shallower water at night and deeper waters during the day.

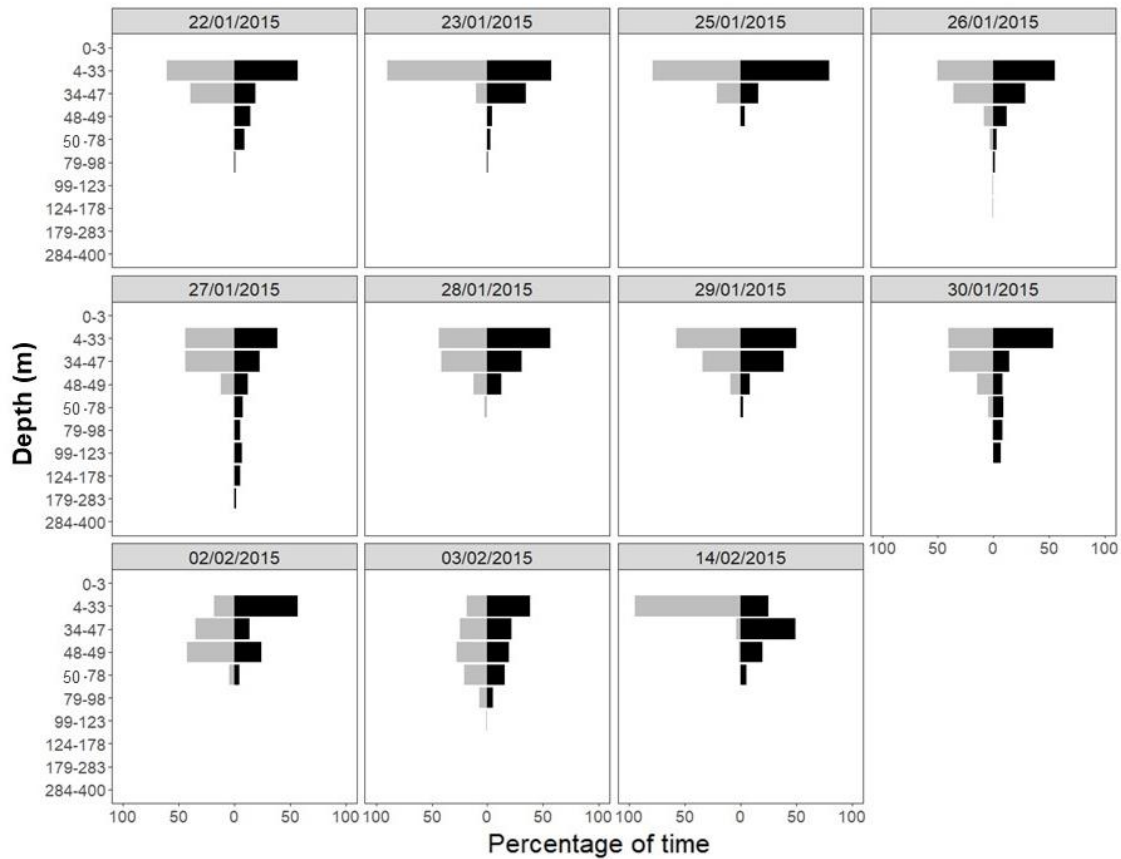
For manta 144473 (Fig. B3), diel vertical migration is apparent most notably on 18/02/2015 where 100% of the night was spent in shallow water (<33 m) and depths of up to 178m are occupied during the day. On the 25/01/2015 the manta occupied shallower water (<33m) for a higher percentage of the night than during the day and on 19/02/2015 the manta remained between 59-78m for 100% of the day. On 20/02/2015 reverse diel vertical migration is evident.



**Figure B3** – Total percentage of time spent at depths for the four days where both day and night were recorded for manta 144473. Two days (25/01/2015 and 18/02/2015) show the manta ray predominantly occupied shallower water (<33m) at night and deeper waters during the day. On the 19/02/2015 the manta ray spent 12 hours continuously at depths between 59-98 m with the longest period of any of the manta rays to remain in deep waters during the day.

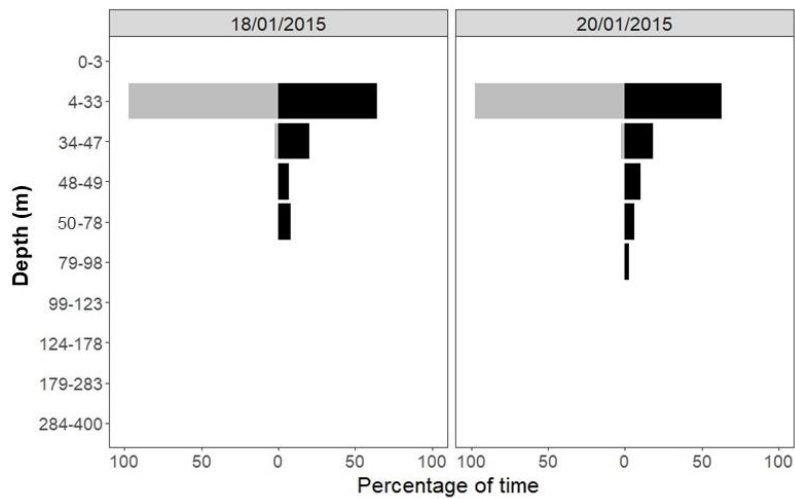
Manta 144474 had the highest number of days where both day and night data were retrieved (Fig. B4). A total of three days (28/01/2015, 02/02/2015 and 03/02/2015) provide evidence of diel vertical movement and a fourth (26/01/2015) indicates that the deepest depths occupied occurred during the day. The remaining seven days follow a reverse diel vertical migration pattern.





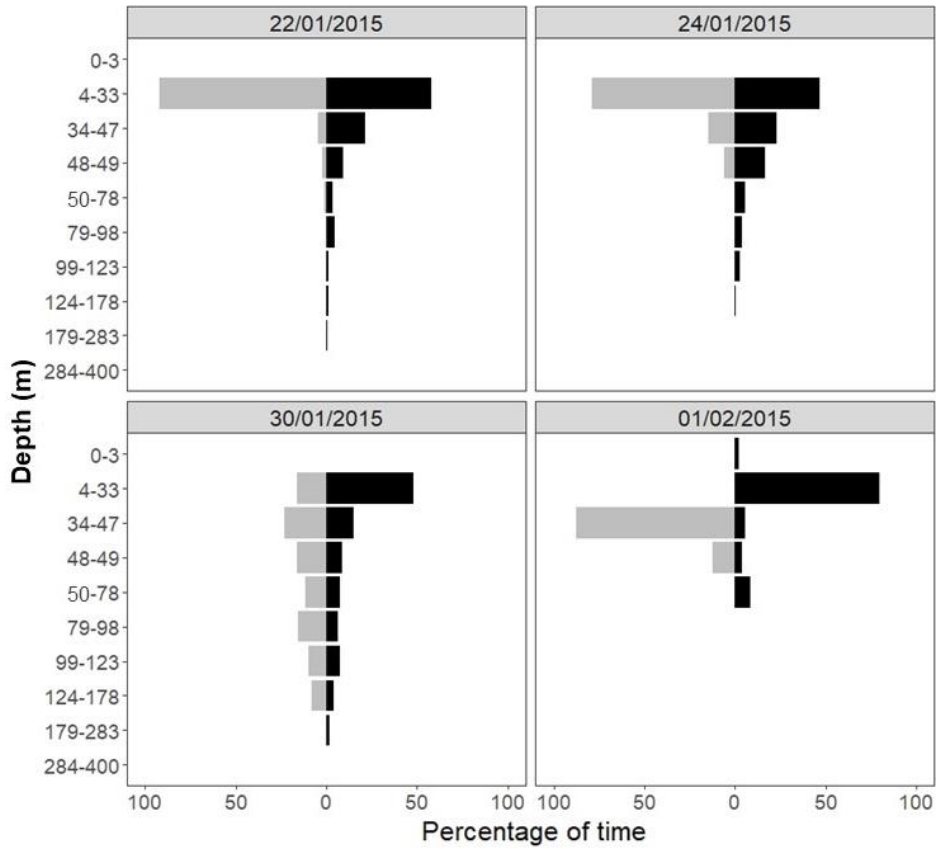
**Figure B4** – Total percentage of time spent at depths for the 11 days where both day and night were recorded for manta 144474. Three days (28/01/2015, 02/02/2015 and 03/02/2015) show the manta ray predominantly occupied shallower water at night and deeper waters during the day. On the 26/01/2015 the manta ray spent an almost even amount of time at depths <49 m and >49 m however, the deepest depths recorded were during the day.

For manta 144475 there were only two days where both day and night data were available. Both days follow a reverse diel vertical migration pattern where 100% of the day was spent in shallow waters (<33m) and deeper waters were only occupied at night.



**Figure B5** – Total percentage of time spent at depths recorded for the two days where both day and night were recorded for manta 144475. Both days are consistent with a reversed diel vertical migration pattern.

For manta 144476, there were four days when data for both day and night was received (Fig. B6). Of these days two (30/01/2015 and 01/02/2015) show evidence of diel vertical migration. On the 30/01/2015 the manta ray spent considerably more time in deeper waters (>48m) than on any other day. On the remaining two days reverse diel vertical migration is apparent.



**Figure B6** – Total percentage of time spent at depths for the four days where both day and night were recoded for manta 144476. Two days (30/01/2015 and 01/02/2015) show the manta ray predominantly occupied shallower water (<33m) at night and deeper waters during the day. On the 30/01/2015 the manta ray spent an considerably more time in deeper waters than on any other day.

## Appendix C - Bayesian joint estimation state-space model (hSSM)

### (i) Modelling method

A state-space switching model (SSSM) incorporates a process model which models the movement process to predict unobserved locations and an observation model which relates the predicted locations to the observed data [8]. The model used in the current study is a process model which indexes movement parameters by behavioural mode [5] known as a Bayesian state-space model of the first-difference correlated random walk with switching (DCRWS) [4] which assumes a correlated random walk on the difference in successive locations e.g. changes in their direction and speed [8]. Two states are derived from the difference: transiting which is characterised by faster more direct movements and area-restricted search (ARS) which is assumed when movements are slower with frequent change in direction [8]. Behavioural state at time  $t$  is denoted  $b_t$  where  $b = 1$  is a transiting state and  $b = 2$  is ARS [8]. The states are defined via mean turn angle  $\theta_{bt}$  and movement persistence  $\gamma_{bt}$  (the autocorrelation in both direction and speed [5]). Jonsen *et al.* (2007) depict the model using the following form:

$$d_t \sim N_2(\gamma_{bt} \mathbf{T}(\theta_{bt})d_{t-1}, \mathbf{\Sigma}),$$

where  $d_t$  is the difference between locations  $x_t$  and  $x_{t-1}$ .  $x_t$  and  $x_{t-1}$  are the true unobserved locations of the animal at times  $t$  and  $t - 1$  ( $x_t$  is a coordinate vector thus  $d_t$  is also a vector [4]) and  $d_{t-1}$  is the difference between locations  $x_{t-1}$  and  $x_{t-2}$ . Describing  $\theta_{bt}$  is a transition matrix,  $\mathbf{T}$ :

$$\mathbf{T}(\theta) = \begin{pmatrix} \cos \theta_{bt} & -\sin \theta_{bt} \\ \sin \theta_{bt} & \cos \theta_{bt} \end{pmatrix}$$

The randomness in the manta ray's behaviour is represented by  $N_2$  which is modelled by a bivariate Gaussian distribution with mean zero [9] and covariance matrix  $\mathbf{\Sigma}$ :

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{lon}^2 & \rho\sigma_{lon}\sigma_{lat} \\ \rho\sigma_{lon}\sigma_{lat} & \sigma_{lat}^2 \end{pmatrix}$$

where  $\sigma_{lon}^2$  and  $\sigma_{lat}^2$  is the process variance in longitude and latitude, respectively and  $\rho$  is the correlation coefficient [4].

The process described uses random sampling so the outcome can be strongly influenced by the starting point [10]. To reduce this influence, a ‘burn in’ phase is imposed which discards a specified number of samples from the start of the chain.

Two Markov Chain Monte Carlo (MCMC) describe the evolution of the manta rays behaviour through time with the possibility of two states at time  $t$  which are denoted foraging ( $F_t$ ) or not-foraging ( $NF_t$ ) [11]. The probability of switching states from  $F$  at time  $t$  to  $NF$  at time  $t + 1$  specifies the Markov chain which is presented as a transition matrix [11]:

$$\mathbf{P} = \begin{pmatrix} \Pr(F_{t+1}|F_t) & \Pr(NF_{t+1}|F_t) \\ \Pr(F_{t+1}|NF_t) & \Pr(NF_{t+1}|NF_t) \end{pmatrix}$$

The probable state of the manta ray at any time ( $\Pr(F_t)$  or  $\Pr(NF_t)$ ) is denoted  $S_t$  which is updated through time by  $S_{t+1} = S_t \cdot \mathbf{P}$  [11].

Behavioural state ( $b_t$ ) is a discrete parameter however, the MCMC samples provide means of these variables which are continuous from 1 to 2 [5]. These means can be used to identify behaviour switches whereby those closer to 1 represent transiting while those closer to 2 indicate ARS [8].

To regularise Argos position estimates and account for location quality classes (LCs) (see Appendix A(i)), errors are directly incorporated as a  $t$ -distributed variable to the model [4,12]:

$$\mathbf{y}_t = \mathbf{X}_t + t(0, \boldsymbol{\tau}_q, \mathbf{v}_q)$$

where  $\boldsymbol{\tau}_q$  are the scale parameters of longitude and latitude for the LC ( $q$ ), and  $\mathbf{v}_q$  are the associated degrees of freedom [12].

The longitude and latitude components of the estimation error allows the best estimate for each of the observed locations  $y_{t,i}$ . For simplification, the assumption is made that the manta rays travel in a straight line between  $x_{t-1}$  and  $x_t$  [4]:

$$y_{t,i} = (1 - j_i)\mathbf{X}_{t-1} + j_i\mathbf{X}_t + \boldsymbol{\varepsilon}_t$$

where  $i$  is an index for locations (Argos position estimates) between time  $t$  and  $t + 1$  and  $j_i$  is the proportion of the regular time interval between  $x_{t-1}$  and  $x_t$  at which the  $i$ th observation is made ( $0 < j_i < 1$ ) and  $\boldsymbol{\varepsilon}_t$  is the estimation error [4].  $j_i$  is calculated from the

time of day recorded with Argos position estimates. Where no observation exists,  $i = 1$  and  $j_i = 0.5$ .

It is unlikely that the DCRWS will be able to categorise behaviour effectively for all the manta rays individually due to the limited number of Argos position estimates for several of the animals [9]. The problem is addressed here by estimating behavioural state jointly [12] across all the manta rays in the hierarchical version of DCRWS (hDCRWS) [12] referred to as a Bayesian joint estimation state-space model (hSSM) [12].

As a Bayesian approach, prior probabilities for all unknown parameters are incorporated into analysis [11]. Prior probabilities weight the likelihood values to obtain posterior probabilities (predictions of ARS and transiting behaviour) [13]. In the absence of prior information, a uniform prior is used; thus, all possible values are equally as likely [13].

#### (ii) *hSSM model diagnostics*

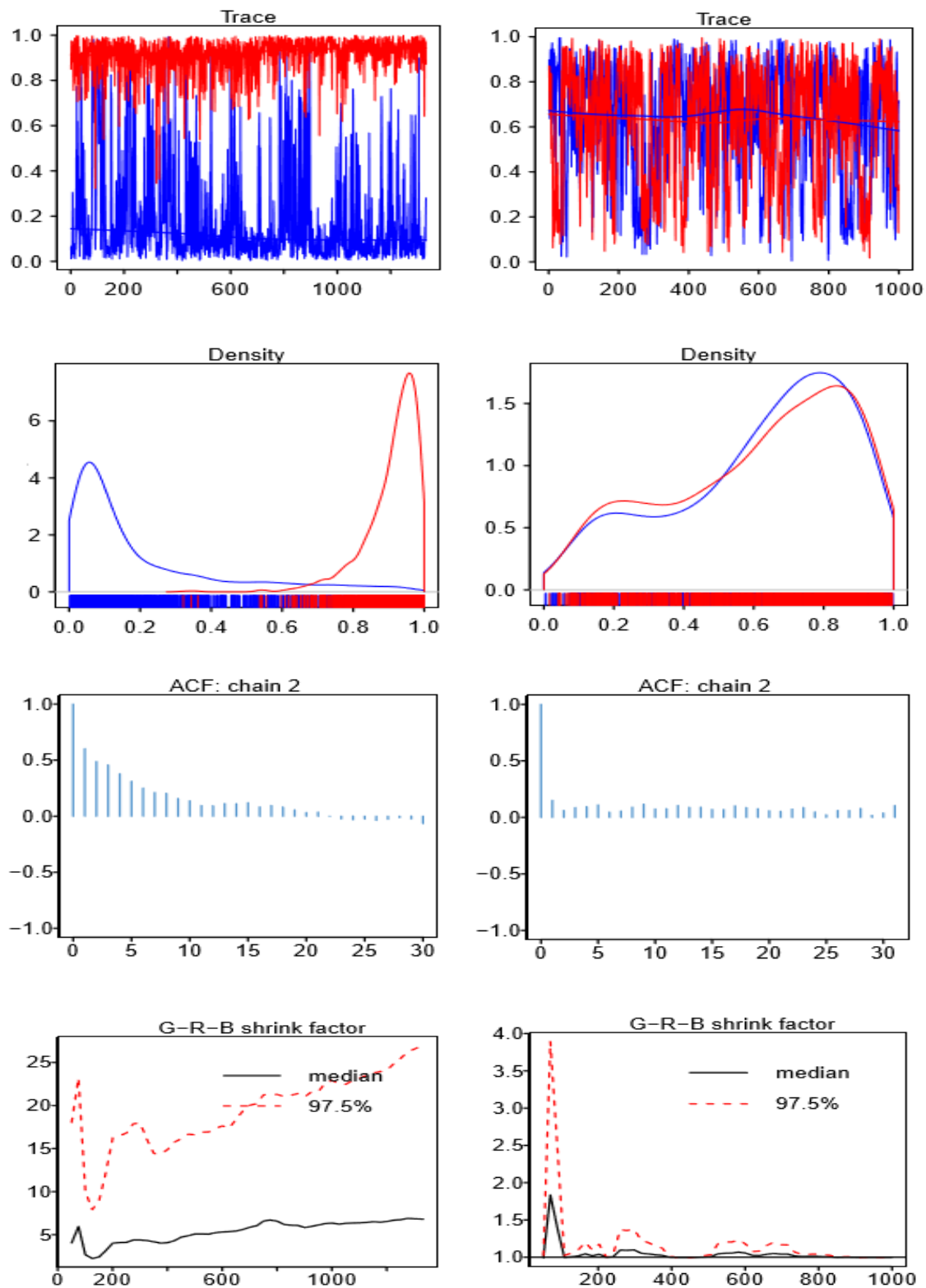
There is no set criterion for hSSM model diagnostics [8, 10, 12] however, certain aspects of the model should be checked to ensure the standard of the model is satisfactory [11]. These aspects include model convergence, which is assessed by the mixing of the two MCMC chains, posterior sample stationarity, in-chain autocorrelation and Gelman-Rubin-Brooks potential scale reduction factor ( $\hat{r}$ ) [9, 11]. The  $\hat{r}$  estimates the factor by which the scale of the posterior distribution may be reduced if simulations were increased which declines to 1 as they reach infinity [15]. Values closer to 1 are consistent with convergence with  $\hat{r} < 1.1$  being the acceptable rule of thumb [10]. All of these parameters can be plotted using the `diag_ssm()` function of the `bsam` R package [16] so to be visually assessed.

The results for model diagnostics indicate how the model can be improved. For example, if mixing or  $\hat{r}$  are not satisfactory, the number of iterations should be increased to give the MCMC chains time to converge [9]. Where autocorrelation is an issue, the gap between the samples to retain (thinning factor) should be increased [12]. Further to these diagnostics, the model output should also be checked to ensure it is plausible as satisfactory diagnostic results do not guarantee a suitable model [9]. For example, all behavioural state estimates may be of an ARS state, in this case, the result may have been biased by the initial sample of the random sampling and an increase in burn-in should be applied [9].

For the current study suitable parameter settings were established by first constructing a model with a low number of samples, a short burn-in phase and low thinning factor. Improvement was then achieved by varying these parameters to balance diagnostic results with the model output.

Figure C1 gives an example of four plots produced by the `diag_ssm()` function for the movement persistence ( $\gamma_1$ ) parameter that show a poor (left plots) and satisfactory (right plots) diagnostic result. The poor diagnostic results were obtained from one of the rejected models during preliminary analysis. For this model two MCMC chains of 16,000 samples were run with the first 2,000 from each chain discarded as burn-in then retaining every 10<sup>th</sup> sample of the remaining 14,000 resulting in a total of 1,400 iterations. The satisfactory model is the one used for analysis for the current study which consisted of two MCMC chains of 40,000 samples with the first 10,000 from each chain discarded as burn-in then retaining every 30<sup>th</sup> sample of the remaining 30,000 resulting in a total of 1,000 iterations.

The two MCMC chains are shown as red and blue in the trace plots (Fig. C1). In the satisfactory model, colours are well mixed while the poor model shows almost no mixing of MCMC chains. Stationarity is displayed when the two chains cluster around the same mean [9] shown as the solid blue and red lines running through the trace plots. The satisfactory plot shows these lines remain at similar values throughout the 1,000 iterations. The density plot also shows the two MCMC chains which should be unimodal [9] indicating convergence, which is not apparent in the poor model. Autocorrelation, shown by the autocorrelation factor (ACF) plots is acceptable for both models however, it is reduced in the final model. The  $\hat{r}$  (G-R-B shrink factor) are above the acceptable level of 1.1 [10] in the poor model and substantially reduced in the model which was chosen for analysis.



**Figure C1** - Example of poor hSSM model diagnostics (left plots) where MCMC chains are not well mixed (trace plot), the density plot is bimodal and the Gelman-Rubin-Brooks potential scale reduction factor ( $\hat{\rho}$ ) (G-R-B shrink factor plot) is  $>1.1$ . On the right is the model chosen for analysis where MCMC chain mixing (trace plot) is satisfactory, the density plots are close to unimodal and  $\hat{\rho} < 1.1$  (G-R-B shrink factor plot). Autocorrelation (ACF plots) for both models is acceptable but show improvement in the chosen model (right plot).



## **Appendix D - Boosted Regression Tree Modelling**

### **(i) Model building and assessment**

Boosted regression trees (BRT) use a model averaging (ensemble) method that allows for both explanation and prediction [17]. The technique fits many models which are combined for prediction using two algorithms: regression trees models and boosting, which builds and combines the models [17]. The regression tree is constructed from predictor variables which are passed from root node to leaves based on a series of binary splits [18]. Splits occur based on the homogeneity of the predictor variables relationship to the response variable [19]. Multiple possible splits will be tested by the algorithm and partitioning will occur when the greatest improvement of homogeneity is found [19].

The BRT modelling technique can accommodate both numerical and categorical predictors [17]. It is also able to accommodate missing values [15, 18] via the construction of surrogates [16, 19]. Briefly, where a value for a variable is not available (NA), an alternative variable with a similar splitting property is used [21]. The alternative is selected by first applying the initial split using only the observations where the value of that predictor is not missing (primary split) [18]. Once the split that provides the greatest improvement is found, a potential list of surrogates and split points are developed [18]. The predictor and split point that best mimics the primary split becomes the surrogate [18]. A predictors ability to mimic another is categorised based on correlations, whereby a higher correlation reduces the loss of information due to the missing value [18].

Another advantage of BRT is its ability to fit complex, non-linear relationships and model interactions between response variables [17]. When building the model, interactions can be controlled for using the tree complexity (TC) (number of nodes in a tree) parameter where the value provided specifies the number of interactions, for example, TC = 2 fits a model with up to two-way interactions [17]. Another controlled parameter is learning rate (LR) which regulates the contribution of each tree to the growing model [17] whereby the lower the LR, the higher the number of trees fitted but the lower the individual influence on their overall result [22]. While these controls can enhance predictive performance [23], it is noted that BRT is prone to overfitting [22]. As a general rule of thumb, an increase in TC should be

accompanied by a decrease in LR, but variations should be tested to find the optimal setting for the data [17,24].

The next parameter to be set is bag fraction, which controls stochasticity by randomly selecting (without replacement) a specified subset of the data at each iteration [17]. The optimal number of trees (NT) is also important, but this can be determined by the model itself [17] using k-fold cross-validation (CV) [17]. The technique works by using part of the available data to fit the model and a different part to test it [18]. The current study applied ten-fold CV, whereby the data is divided into ten subsets, nine which are used to build and one which is used to test the model [17]. During the building of the model, the ten-fold CV processes is tightly interlinked with the bag fraction thus may best be explained with the following example which is based on the hypothetical dataset which contains 1000 observations of presence ( $n = 500$ ) and absence ( $n=500$ ).

To start, the ten-fold CV first splits the data into ten subsets [17], each containing 100 observations. Ten unique combinations of nine subsets (900 observations) are then used to construct ten BRT models with a selected number of trees (i.e. 50) [17]. With a bag fraction of 0.5, the first tree for each of the models will be built using 450 randomly selected observations (50% of the data) [17]. Predictions are then made on all 900 observations using this tree. The following tree is then built to minimise the residuals (450 randomly selected residuals) of this tree and added to the model [17]. The next tree is then fitted to 450 residuals that are selected at random from the two trees, and the process is continually repeated. Once a model of 50 trees has been constructed the performance of each of the ten models is tested on their ability to predict their respective 10<sup>th</sup> (hold-out) subset of data (100 observations) [15, 16]. Average residual deviance and standard error are recorded [17]. The number of trees is then increased, and the process is repeated [17]. Predictive performance is measured after ten steps and compares the last five iterations with the five previous to those [17]. A reduction in average residual deviance indicates model improvement [15, 16] and when deviance begins to increase the optimal number of trees has been passed [17].

As described in the main text, preliminary analysis to find the suitable TC and LR included fitting models with progressive TC (2,3,4 and 5) and decreased LR (0.05, 0.01 and 0.005) [18, 20] (Table D1 and Table D3). The performance of each model was then assessed using 1) the area under the receiver operating characteristic curve (AUC) [25], and 2) proportion (%) of

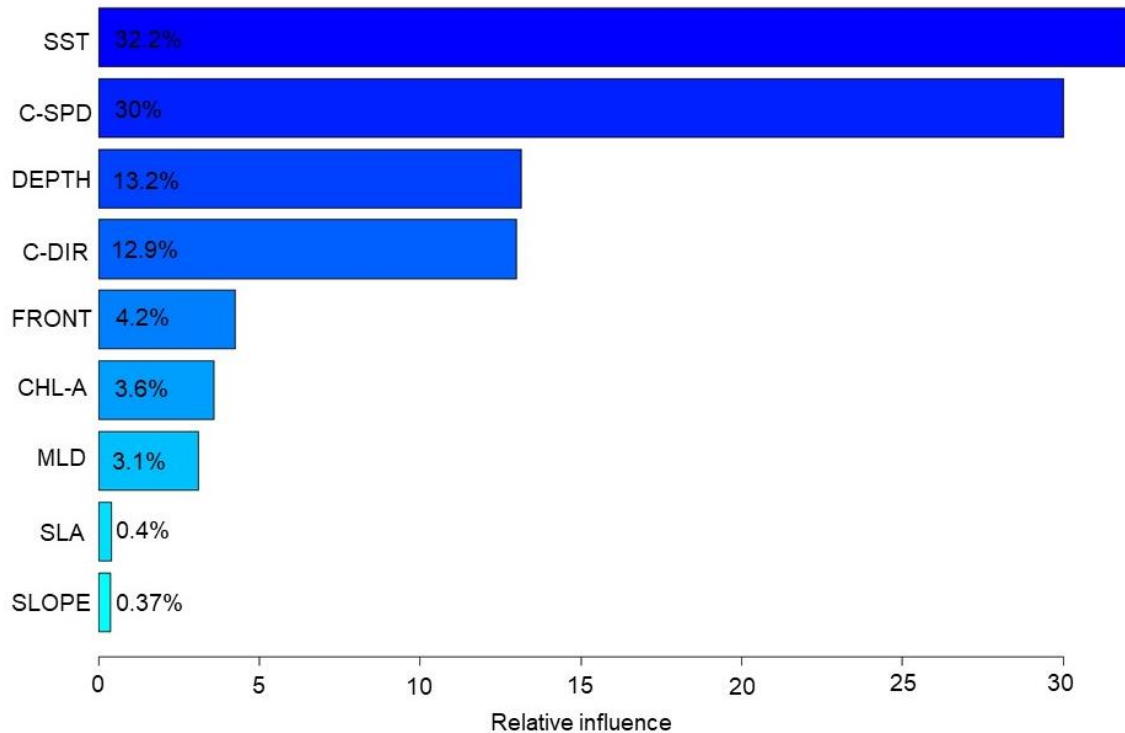
total deviance explained by the model ( $D^2$ ) [22]. During the current study it was found that generally, models that perform better based on the AUC values will also have a higher  $D^2$  however, during a pilot study there were some exceptions thus these methods were used in combination specifically to deal with occasions where two models had the same AUC values.

**Table D1** - Performance of the hotspot models with progressively increased tree complexity (TC) and reduced learning rate (LR). The optimal number of tress (NT) is ascertained automatically by the `gbm.fit()` function. The mean training AUC (TAUC) values indicate how well the model fit the training data and the mean cross-validated AUC (CVAUC) shows how well that model was able to classify the data in the hold-out sample during the ten-fold cross-validation procedure. The difference between these values ( $\Delta$ AUC) is a measure of overfitting whereby a greater difference indicates the model is overfitting the training data. The highest performing model is the one shown in bold, which has the lowest  $\Delta$ AUC. The cross-validated proportion of the total deviance explained ( $D^2$ ) corresponds to the percentage of deviance for the null model explained by the fitted model. The mean deviance for the null model is 1.283.

TC	LR	NT	TAUC	CVAUC	$\Delta$ AUC	CV residule deviance (SE)	$D^2$
2	0.01	1200	1	0.987	0.013	0.243 (0.048)	0.81
<b>2</b>	<b>0.005</b>	<b>3250</b>	<b>1</b>	<b>0.998</b>	<b>0.002</b>	<b>0.174 (0.058)</b>	<b>0.86</b>
2	0.001	12850	1	0.997	0.003	0.175 (0.042)	0.86
2	0.0001	54580	1	0.995	0.005	0.27 (0.038)	0.79
3	0.01	900	1	0.992	0.008	0.225 (0.078)	0.82
3	0.005	1550	1	0.996	0.004	0.242 (0.067)	0.81
3	0.001	8050	1	0.996	0.004	0.206 (0.067)	0.84
3	0.0001	47700	1	0.994	0.006	0.242 (0.039)	0.81
4	0.01	800	1	0.997	0.003	0.232 (0.056)	0.82
4	0.005	1550	1	0.992	0.008	0.265 (0.061)	0.79
4	0.001	7250	1	0.995	0.005	0.217 (0.049)	0.83
4	0.0001	45750	1	0.993	0.007	0.249 (0.029)	0.81
5	0.01	600	1	0.997	0.003	0.204 (0.053)	0.84
5	0.005	1350	1	0.996	0.004	0.22 (0.062)	0.83
5	0.001	6350	1	0.996	0.004	0.238 (0.071)	0.81
5	0.0001	44450	1	0.996	0.004	0.264 (0.04)	0.79

The results of the preliminary analysis for the hotspot model are shown in Table D1 with the best performing model in bold. The results of this model (Fig. D1) reflect the relative

importance of predictor variables which is scaled to 100 across all the variables [17]. The predictor variables with the highest numbers indicate a higher influence on the response variable [17].



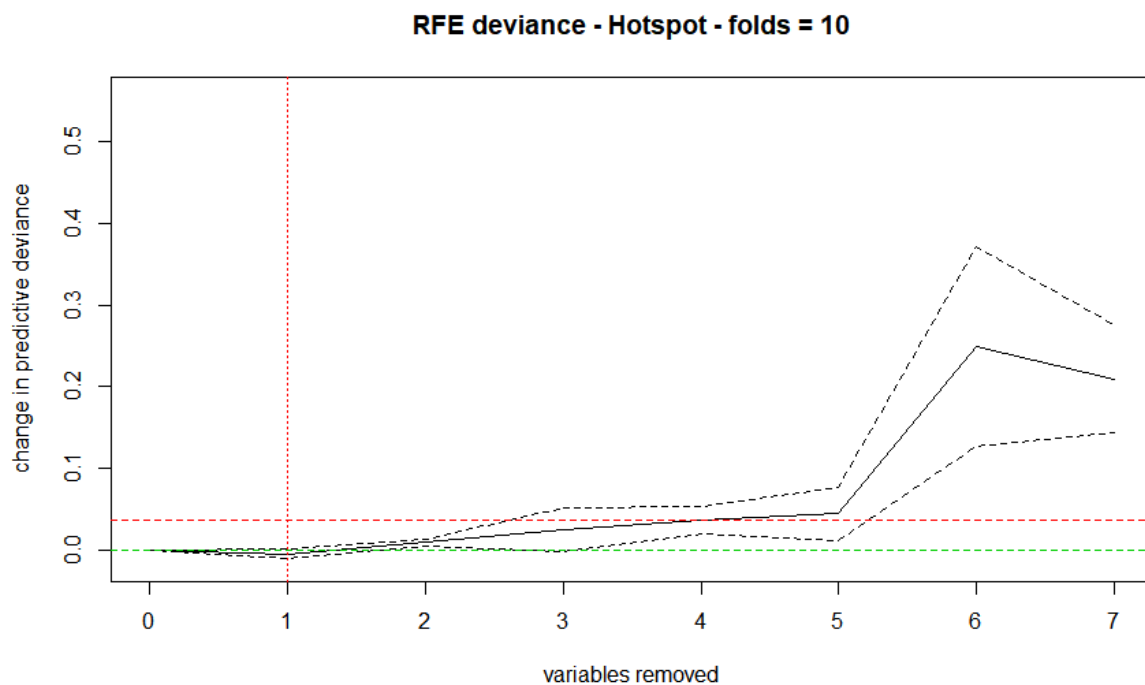
**Figure D1** - Results of the hotspot model highlighted in bold in Table A\*\* showing the percentage contribution of each of the nine predictor variables: sea surface temperature (SST), current speed (C-SPD), depth (DEPTH), current direction (C-DIR), distance to front (FRONT), chlorophyll-*a* (CHL-A), mixed layer depth (MLD), sea level anomalies (SLA) and seabed slope (SLOPE)

These results suggest that sea bed slope (SLOPE) and sea level anomalies (SLA) have the lowest influence in the model thus may be considered non-informative predictors [17]. Non-informative predictors have the potential to hinder model performance by increasing variance [17]. The dismo R package includes the function `gbm.simplify()`, which tests the effect of the elimination of predictor variables, excluding up to a specified maximum [17]. The function starts with the least important predictor, re-fitting the model using the ten-fold CV procedure using the mean CV error to evaluate the effect on predictive performance [17]. The function output provides a list of the least informative predictor variables in ascending order along with a plot which indicates how many of these variables should be removed [17]. For the hotspot model the list was presented in the following order:

```
# sample of gbm.simplify() output for the hotspot model
```

- 1-SLOPE
- 2-SLA
- 3-MLD
- 4-CHL.A
- 5-FRONT
- 6-DEPTH
- 7-C\_DIR

The plot generated by the function indicates that one predictor variable should be dropped (red dotted vertical line) (Fig. D2)



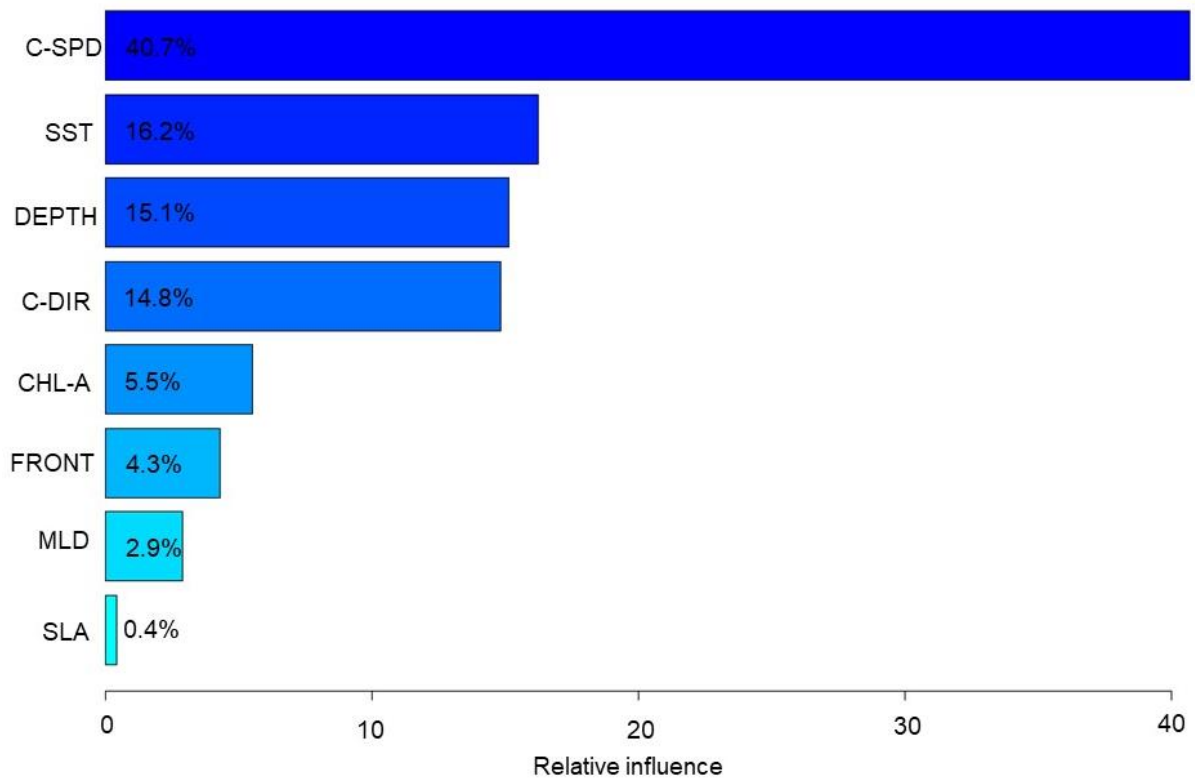
**Figure D2-** Output of the `gbm.simplify()` for the hotspot model showing the change in the predictive deviance when up to seven explanatory variables are removed. The vertical red dotted line indicates the optimal number of variables to remove.

A new BRT model was then constructed whereby the output list and the number of predictors to exclude are provided to the response variable argument of the gbm.step script (shown in Appendix E(vi))

The results of the final hotspot model are shown in Figure D3. Note that the contribution of the predictor variables has changed from the initial model (Fig. D1). The model performance results from the new hotspot model indicate substantial improvement in both AUC values and  $D^2$  from the initial model (Table D2) which indicates that SLOPE was hindering model performance by increasing variance [17].

**Table D2** - Performance results of the final hotspot model (Fig. D3) indicating that the removal of a non-informative predictor variable has substantially improved values of AUC and  $D^2$ .

TC	LR	NT	TAUC	CVAUC	$\Delta$ AUC	CV residue deviance (SE)	$D^2$
2	0.0050	3550	1	0.999	0.001	0.095 (0.03)	0.93



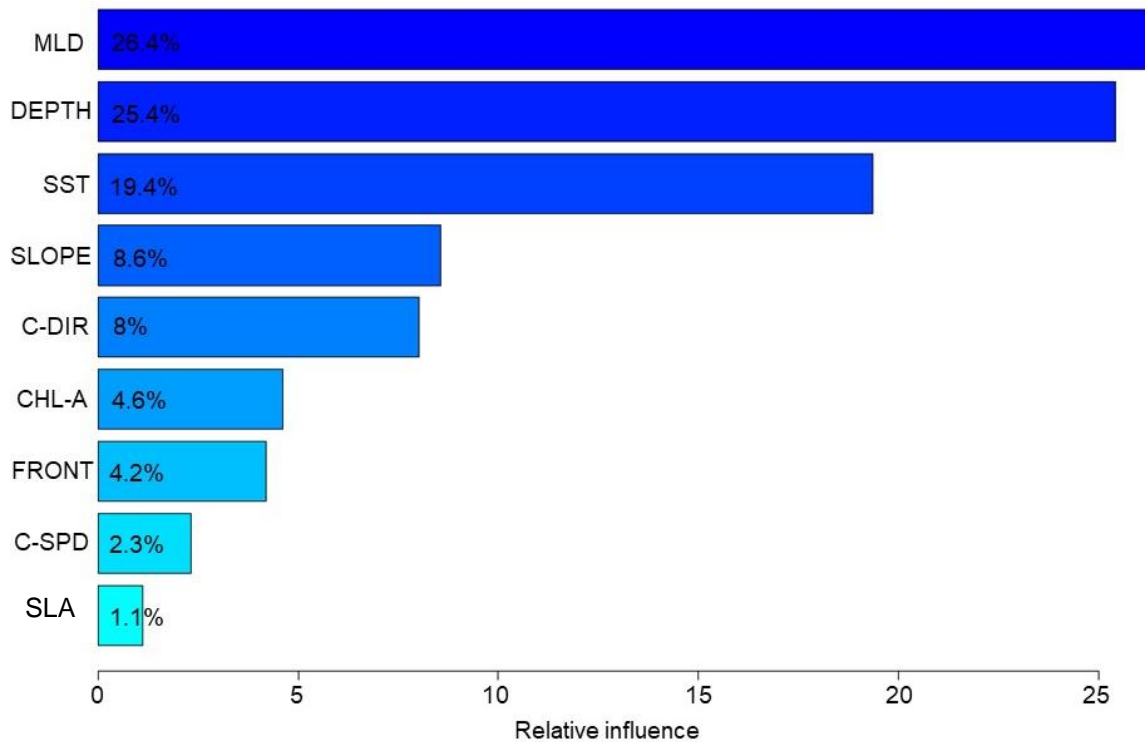
**Figure D3** - Results of the final hotspot model showing the percentage contribution of each of the eight predictor variables that were retained after the model was simplified using the `gbm.simplify()` function: current speed (C-SPD), sea surface temperature (SST), depth (DEPTH), current direction (C-DIR), chlorophyll-*a* (CHL-A), distance to front (FRONT), mixed layer depth (MLD) and sea level anomalies (SLA).

The results of this preliminary analysis for the ARS model are shown in Table D3 with the highest performing model highlighted in bold. The contribution of each of the variables from this model are displayed in Figure D4.

**Table D3** - Performance of the ARS models with progressively increased tree complexity (TC) and reduced learning rate (LR). The optimal number of trees (NT) is ascertained automatically by the `gbm.fit` function. The mean training AUC (TAUC) values indicate how well the model fit the training data and the mean cross-validated AUC (CVAUC) shows how well that model was able to classify the data in the hold-out sample during the k-fold cross-validation procedure. The difference between these values ( $\Delta$ AUC) is a measure of overfitting whereby a greater difference indicates the model is overfitting the training data. The highest performing model is the one with the shown in bold, which has the lowest  $\Delta$ AUC. The cross-validated proportion of the total deviance explained ( $D^2$ ) corresponds to the percentage of deviance for the null model explained by the fitted model. The mean deviance for the null model is 1.386.

TC	LR	NT	TAUC	CVAUC	$\Delta$ AUC	CV residule deviance (SE)	$D^2$
2	0.01	1500	0.999	0.972	0.027	0.441 (0.042)	0.68
2	0.005	2900	0.998	0.967	0.031	0.464 (0.06)	0.67
2	0.001	12700	0.998	0.970	0.028	0.449 (0.049)	0.68
2	0.0001	49600	0.986	0.960	0.026	0.528 (0.052)	0.62
3	0.01	1100	0.999	0.972	0.027	0.428 (0.05)	0.69
3	0.005	2250	0.999	0.967	0.032	0.455 (0.08)	0.67
3	0.001	10000	0.999	0.969	0.030	0.454 (0.06)	0.67
3	0.0001	44350	0.991	0.960	0.031	0.518 (0.06)	0.63
4	0.01	850	0.999	0.967	0.032	0.47 (0.062)	0.66
4	0.005	1850	0.999	0.971	0.028	0.448 (0.064)	0.68
4	0.001	9000	0.999	0.970	0.029	0.436 (0.051)	0.69
4	0.0001	38300	0.993	0.961	0.032	0.533 (0.044)	0.62
5	0.01	800	0.999	0.971	0.028	0.449 (0.063)	0.68
5	0.005	1450	0.999	0.966	0.033	0.469 (0.036)	0.66
<b>5</b>	<b>0.001</b>	<b>7050</b>	<b>0.999</b>	<b>0.976</b>	<b>0.023</b>	<b>0.456 (0.04)</b>	<b>0.67</b>
5	0.0001	43000	0.996	0.972	0.024	0.488 (0.037)	0.65





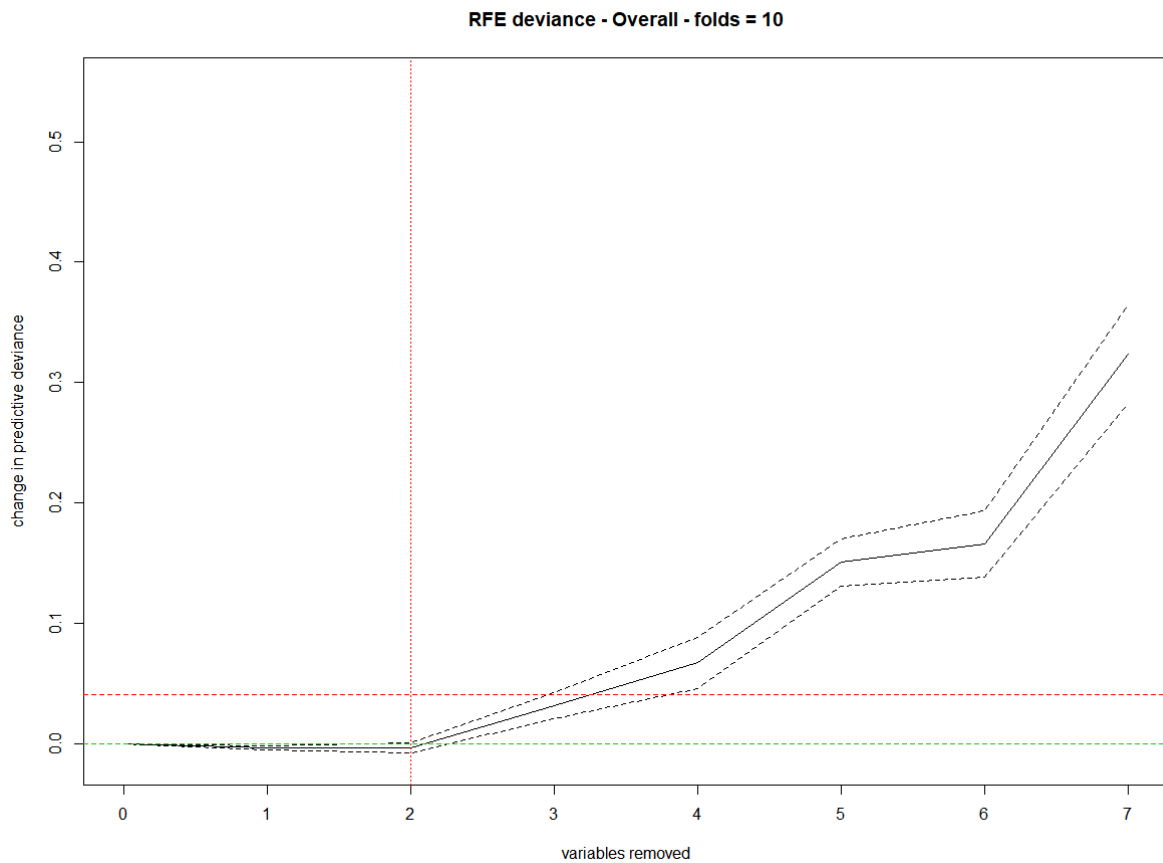
**Figure D4** - Results of the ARS model highlighted in bold in Table D3, showing the percentage contribution of each of the nine predictor variables: mixed layer depth (MLD), depth (DEPTH), surface temperature (SST), seabed slope (SLOPE), sea current direction (C-DIR), chlorophyll-*a* (CHL-A), distance to front (FRONT), current speed (C-SPD) and sea level anomalies (SLA).

The `gbm.simplify()` function was applied to tests the effect of the elimination of up to seven predictor variables from the ARS model (R output below).

```
# sample of gbm.simplify() output for the hotspot model
```

```
1-SLA
2-C_SPD
3-CHL.A
4-FRONT
5-C_DIR
6-SLOPE
7-MLD
```

The plot generated by the function indicates that two predictor variables should be excluded (red dotted vertical line) (Fig. D5).



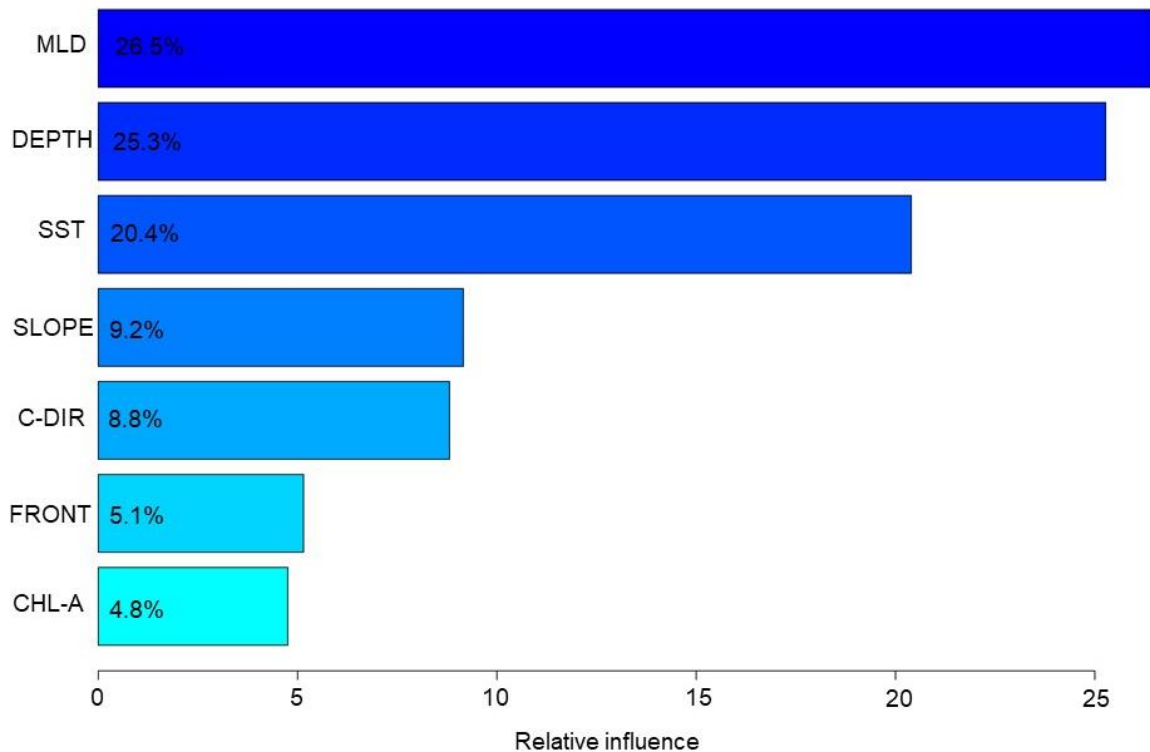
**Figure D5** - Output of the `gbm.simplify()` for the ARS model showing the change in the predictive deviance when up to seven explanatory variables are removed. The vertical red dotted line indicates the optimal number of variables to remove.

A new ARS model was then constructed with the exclusion of the SLA and current speed (C-SPD), which did not substantially alter the performance of the model (Table D3 and Table D4). Therefore, the simpler model was retained; thus, the final ARS model included seven predictor variables.

**Table D4** - Performance results of the ARS model with seven predictor variables showing there was little change in model performance compared to the model with nine predictor variables (Table D3).

TC	LR	NT	TAUC	CVAUC	$\Delta$ AUC	CV residule deviance (SE)	$D^2$
5	0.001	8000	0.999	0.972	0.027	0.447 (0.08)	0.68

The results of the final ARS model are shown in Figure D6. Note that the relative contribution of the predictor variables has not changed considerably from the initial model (Fig. D4) although the order of variables has altered.



**Figure D6** - Results of the final ARS model showing the percentage contribution of each of the seven predictor variables that were retained after the model was simplified using the `gbm.simplify()` function: mixed layer depth (MLD), depth (DEPTH), surface temperature (SST), seabed slope (SLOPE), sea current direction (C-DIR), chlorophyll- $\alpha$  (CHL-A) and distance to front (FRONT).

The BRT also reports the marginal effect of each of the predictor variables as 2D partial dependency plots (Fig. 12 and Fig. 14) [17]. The x-axis of the plot is the predictor variable, and the y-axis is the fitted function (an estimate of log-odds) [18]. The plot represents the effect of the explanatory variable after accounting for the mean effects of all other explanatory variables [18]. The effect can be positive or negative, with zero being no effect [17]. The rugs show the distribution of the presence and absence data in percentiles.

The interactions modelled by BRT can also be visualised by forming predictions on a linear scale for a grid of paired variable with all other variables held to their respective mean [17].

The resulting 3D plot (Fig. 13 and Fig. 14) displays a piecewise-constant approximation of the joint partial dependence of their effect [17]. Each split in the model is represented by rectangular regions that divide the predictor variable [18]. A marginal effect of zero indicates that no interactions were fitted [17]. The `gbm.interactions()` function creates a data frame which includes the most important interactions between variables. These interactions were plotted for the current study (Fig. 13 and Fig. 14) with the exception of interactions that included current direction (C-DIR) as a categorical variable cannot be visualised in this format.

## Appendix E – R scripts and MATLAB code

### (i) MATLAB code: Time-at-temperature (TAT) and vertical profiles

```
% set divisions of the bars based on temperature bins
% recorded by tags

midbin = [9 12 15 18 20 22 24 26 28 30 31]

% set the mean percentage of time spent in each bin

percent = [0.033522727 0.319318182 0.965340909 1.300568182 3.836363636
9.113068182 11.96704545 22.87954545 49.29545455 0.293181818 0]

%change bar width (load midbin and percent first

y = zeros(length(midbin),1);
dx = diff([midbin 1.8]); % width of bar

p = figure, hold on
for ii=1:length(midbin)
    rectangle('position',[midbin(ii) y(ii) dx(ii) percent(ii)])
end

hold on

%plotting both line and bar
yyaxis left
b = bar(midbin,percent); %bar plot
yyaxis right
p = plot(CTD6.T,CTD6.depth); %add CTD data to plot
set(gca, 'YDir','reverse')
xlim([8 30])

hold on

yyaxis right
plot(MSS.TEMP,MSS.PRESS,'.-k') %add MMS data to plot

%label axis

yyaxis left
ylabel('Percent of time in bin')

yyaxis right
ylabel('Depth (m)')

xlabel('Temperature (°C)')

%axis labels for big plot

set(gca, 'XLim',[9 31], 'XTick',[9 11 13 15 17 19 21 23 25 27 29 30.5])
xticklabels({'9','11','13','15','17','19','21','23','25','27','29','>30'})

% get tick marks on top of the bars
set(gca, 'Layer','top')

set(gca, 'XLim',[9 31], 'XTick',[9:2:31])
```

```

%axis labels for inset only

set(gca, 'XLim', [9 31], 'XTick', [9 10 11 12 13 14 15 17 19 21 23 25 27 29
30.5])
xticklabels({'9', '10', '11', '12', '13', '14', '15', '17', '19', '21', '23', '25', '27',
', '29', '>30'})

% edit plot aesthetics using the graphical user interface

```

## (ii) Plotting Time-at-temperature (TAT) results in R

```

# covert time at temperature data into a stacked dataset

data<-read.csv("Compiled Temp Histos with date and time.csv",header=T,sep=",")

dat2a <- data.frame(data[1:4], stack(data[5:ncol(data)]))

write.table(dat2a,file="tempstackwithDATEANDTIMEBIN_2.csv",sep="," ,row.names=F,col
.names=T)

# load data

dat<-read.csv("tempstackwithDATEANDTIMEBIN_2.csv",header=T,sep=",")

# raw data reports percentage of time spent within each bin

# all daytime percentages were converted to a negative in Excel to make plotting
# easier

# the following script converts these bins to depth ranges calculates the mean
# and standard deviation of the mean (SE)

library(Rmisc) # for calculating SE

All <- summarySE(dat, measurevar="percent", groupvars=c("bin","day"))

# change the temperature (depth) bin to a factor rather than numeric

All2 <- All
All2$bin <- factor(All2$bin)

# plot overall percentage of time in each bin (all mantas combined)

library(ggplot2) #for plotting

n<-ggplot(All2, aes(x=bin, y=percent, fill=day)) +
  geom_bar(stat="identity") +
  geom_errorbar(aes(ymin=percent-se, ymax=percent+se),width=.2, colour="azure4")
+ scale_fill_manual("legend", values = c("Day" = "grey", "Night" = "black"))+
  theme_bw() + scale_x_discrete(name="Depth (m)",
  breaks=c(2,3,4,5,6,7,8,9,10,11,12), labels=c(">400","284-400", "179-283","124-
178", "99-123", "79-98", "59-78", "48-49", "34-47", "4-33", "0-3"))+

```

```

  theme(axis.text.y = element_text(size = 20))

n + scale_y_continuous(name="Percentage of time",breaks=c(-60,-45,-30,-15,-
  5,0,5,15,30,45,60),labels=c("60","45","30","15","5","0","5","15","30","45",
  "60"), limits = c(-70, 70)) + theme(legend.position='none') +theme(axis.text.x
  = element_text(size = 20)) +theme(axis.text.y = element_text(size = 20)) +
  coord_flip() + theme(axis.title.y = element_text(size = 16)) +
  theme(axis.title.x = element_text(size = 20)) + theme(panel.grid =
  element_line("white")) +theme(axis.title.y = element_text(size = 20))

# calculate SE including tags separately

tag <- summarySE(dat, measurevar="percent", groupvars=c("bin","day", "tag"))

# change bin to factor rather than numeric
tag2 <- tag
tag2$bin <- factor(tag2$bin)

n<-ggplot(tag2, aes(x=bin, y=percent, fill=day)) +
  geom_bar(stat="identity") + geom_errorbar(aes(ymin=percent-se,ymax=percent+se),
  width=.2, colour="azure4") + scale_fill_manual("legend", values = c("Day" =
  "grey", "Night" = "black"))+ theme_bw() + scale_x_discrete(name="Depth (m)",
  breaks=c(2,3,4,5,6,7,8,9,10,11,12),labels=c(">400", "284-400", "179-283", "124-
  178", "99-123", "79-98", "59-78", "48-49", "34-47", "4-33", "0-3"))+
  theme(axis.text.y = element_text(size = 16))

n + scale_y_continuous(name="Percentage of time",breaks=c(-60,-45,-30,-15,
  -5,0,5,15,30,45,60), labels=c("60","45","30","15","5","0","5","15","30","45",
  "60"), limits = c(-70, 70)) + theme(legend.position='none') +theme(axis.text.x
  = element_text(size = 16)) +theme(axis.text.y = element_text(size = 16)) +
  coord_flip() + theme(axis.title.y = element_text(size = 16)) +
  theme(axis.title.x = element_text(size = 16)) + theme(panel.grid =
  element_line("white"))

n + facet_wrap( ~ tag) +coord_flip()+theme(panel.grid = element_line("white")) +
  theme(strip.text.x = element_text(size = 16))+ theme(panel.spacing = unit(0.5,
  "lines")) + theme(axis.text.x = element_text(size = 12)) +
  theme(legend.position='none')+ scale_y_continuous(name="Percentage of time",
  limits = c(-100, 100), labels=c("100","50","0","50","100"))

```

### (iii) Bayesian joint estimation state-space model (hSSM) in R

```

# Joint estimation state-space model (hSSM)

library(bsam)
library(rjags)
library(chron)

argos = read.csv("ArgosData.csv")

# change data to correct format (required to run model)

argos$date = as.POSIXct(as.character(argos$date), "%Y-%m-%d %H:%M:%S", tz="GMT")

# the following is the final model but >30 were constructed to establish

```

```

# satisfactory diagnostics. span = smoothing. The package help files recommended
# 0.2 for small data sets

hSSM.Final<-fit_ssm(argos, model="hDCRWS", tstep=0.5, adapt=10000, samples=30000,
thin=30, span = 0.4)

# map the data in the R plot
map_ssm(hSSM.Final)

# run diagnostics
diag_ssm(hSSM.Final)

# plot as 1-D time-series individual mantas estimated locations and behavioural
states (posterior means).
plot_fit(hSSM.Final)

# create a summary of the data
result.hs <- get_summary(hSSM.Final)

# write the data to a .csv file
write.table(speed, file="hSSM.Final.csv",sep="," ,row.names=F,col.names=T)

```

#### **(iv) Track distance and travel speed in R**

```

# calculate distance between successive longitude/latitude points
data<-read.csv("MantasforDistance.csv",header=T,sep="," )

library(geosphere)

# split records into sections by manta tag ID so that each manta track starts at 0
myList <- setNames(split(data[,c(3,4)], seq_len(nrow(data))), data$ID)

# get the distance between points (default in metres)
distMat <- outer(myList, myList, Vectorize(distVincentyEllipsoid))

# write to csv
write.table(distMat,file="distance.csv",sep="," ,row.names=F,col.names=T)

# calculate speed between successive points grouped by manta ID to ensure
# tracks of individual mantas are calculated

# the following script requires that the .csv has a 'time' column which is
# the time elapsed between observations. This was added in Excel before running
# the script

# the tidyverse R package which will automatically access all the packages

```



```

# required for the subsequent script e.g. dplyr R package

library(tidyverse)

# create function to calculate the Euclidean distance (straight line) between
# points

euclidean_speed <- function(lat2, lat1, long2, long1, time2, time1) {
  latdiff <- lat2 - lat1
  longdiff <- long2 - long1
  distance <- sqrt(latdiff^2 + longdiff^2)
  timediff <- time2 - time1
  return(distance / timediff)
}

# apply to data sorting by ID and calling the dplyr::lag() function

speed<-data %>%
  group_by(ID) %>%
  arrange(ID, time) %>%
  mutate(speed = euclidean_speed(lat, lag(lat), long, lag(long), time, lag(time)))

# write to a new csv file.

write.table(speed, file="distance.csv",sep="," ,row.names=F,col.names=T)

```

### (v) Kernel utilisation distribution (KUD) in R

```

# Kernel utilisation distribution calculation and write to shapefile

# calculate KUD

library(sp)
library(adehabitatHR)

hotspot<-read.csv("Hotspotdata.csv")

# first make a spatial points data frame which binds the longitude and latitude
# columns

manta<-SpatialPoints(cbind(hotspot$lon,hotspot$lat))

# calculate the kernel density utilization (KUD) using the kernelUD function.

kud <- kernelUD(manta, h=0.2) # KUD (h=smoothing factor)

#convert into vector object, use 95,70 etc. for quartic kernel

homerange95 <- getverticeshr(kud, percent=95)
homerange75 <- getverticeshr(kud, percent=75)
homerange50 <- getverticeshr(kud, percent=50)
homerange10 <- getverticeshr(kud, percent=10)

# plot the data in R to check - if the plots are not satisfactory the smoothing
# factor may not be high/low enough which should be adjusted with the

```

```

# h=smoothing factor argument (above)

plot(homerange95,col="#FF6A6A55",add=T)
plot(homerange75,col="#FF6A6A50",add=T)
plot(homerange50,col="#EE636350",add=T)
plot(homerange10,col="#8B3A3A50",add=T)

# project in the correct coordinates system

WGScoor<- homerange95
proj4string(WGScoor)<- CRS("+proj=longlat +datum=WGS84")

# write to shapefile that can be projected in ArcGIS

raster::shapefile(WGScoor, "homerange95.shp")

WGScoor<- homerange75
proj4string(WGScoor)<- CRS("+proj=longlat +datum=WGS84")
raster::shapefile(WGScoor, "homerange75.shp")

WGScoor<- homerange50
proj4string(WGScoor)<- CRS("+proj=longlat +datum=WGS84")
raster::shapefile(WGScoor, "homerange50.shp")

WGScoor<- homerange10
proj4string(WGScoor)<- CRS("+proj=longlat +datum=WGS84")
raster::shapefile(WGScoor, "homerange10.shp")

```

## (vi) Boosted regression trees (BRT) in R

```

# R script for boosted regression trees (BRT)

# R packages required for model building and plotting

library(dismo)
library(gbm)

# build initial hotspot BRT models for preliminary analysis to determine tree
complexity (TC) and learning rate (LR)

# the following script was repeated 16 times for all combinations of TC and LR

obs.data.hotspot<-read.csv("myhotspotdata.csv")

hotspot_TC_LR0.01<-gbm.step(data=obs.data.hotspot, gbm.x=22:30, gbm.y=20,
                           family="bernoulli", tree.complexity = 2,
                           learning.rate = 0.01, bag.fraction = 0.5,
                           max.trees = 100000)

# chosen models based on AUC and D2

hotspot_TC2_LR0.05<-gbm.step(data=obs.data.hotspot, gbm.x=22:30, gbm.y=20,
                              family="bernoulli", tree.complexity = 2,
                              learning.rate = 0.005, bag.fraction = 0.5,
                              max.trees = 100000)

```

```

# now to test model simplification

# testing all combinations of drops up to a maximum of 7 variables

hotspot_TC2_LR0.05_SIMP<-gbm.simplify(hotspot_TC2_LR0.05, n.drops = 7)

summary(hotspot_TC2_LR0.05_SIMP)

# summary and resulting plot indicates that one predictor variable should be
dropped

# construct model again with the prediction list passed to the response argument
and the number 1
# to drop the least informative variable

hotspot_TC2_LR0.05_SIMP_drop1<-gbm.step(obs.data.hotspot,
                                       gbm.x=hotspot_TC2_LR0.05_SIMP$pred.list[[1]],
                                       gbm.y=20, tree.complexity = 2,
                                       learning.rate = 0.005, max.trees = 100000)

# get the results of the model

summary(hotspot_TC2_LR0.05_SIMP_drop1) # variable contributions scaled to 100

# plot partial dependency plots

gbm.plot(hotspot_TC2_LR0.05_SIMP_drop1, variable.no=1, smooth=TRUE,
         common.scale=TRUE, write.title=FALSE, y.label="Fitted function",
         x.label=expression(paste("Sea surface temperature (°C)")),
         show.contrib=FALSE, plot.layout=c(1, 1),cex.lab=1.5, cex.axis=1.5,
         lty = 1, lwd = 2)

# edit plot via source code

trace(gbm.plot, edit = T) # calls editable source code

# sections of source code edited
# set to display rugs in quartiles: probs = seq(0, 1, 0.25)
# move rugs: side =, colour rugs: col =
# Line thickness: lwd =

# set the absence rugs by defining only to use rows where Hotspot == 0
# set placement of rugs to the bottom of the plot with red tick marks

    }
    if (rug & is.vector(data[, gbm.call$gbm.x[variable.no]])) {
      rug(quantile(data[gbm.call$dataframe$Hotspot ==
        0, gbm.call$gbm.x[variable.no]], prob = seq(0, 1, 0.01), na.rm = TRUE,
        type = 7), side = 1, col = 2, lwd = 2)

# set the presence rugs by defining only to use rows where Hotspot == 1
# set placement of rugs to the bottom of the plot with blue tick marks

    }
    if (rug & is.vector(data[, gbm.call$gbm.x[variable.no]])) {
      rug(quantile(data[gbm.call$dataframe$Hotspot ==
        1, gbm.call$gbm.x[variable.no]], prob = seq(0, 1, 0.01), na.rm = TRUE,

```

```

    type = 7), side = 3, col = 4, lwd = 2)
  }

# find predictor variable interactions fitted within the model

find.int<-gbm.interactions(hotspot_TC2_LR0.05_SIMP_drop1)

find.int$interaction # gives a matrix of interactions

find.int$rank.list # ranks important interactions between predictor variables

# get a 3D plot of interactions
# the numbers specified indicate the number of the variable assigned in the rank
list
# these numbers do not reflect rank they just ID the variable for plotting
# example using SST and depth

gbm.perspec(hotspot_TC2_LR0.05_SIMP_drop1, 2, 1, pred.means = NULL, x.label =
  "Depth (m)", x.range = NULL, y.label = " Sea surface temperature (°C)",
  z.label = " Marginal effect", y.range = NULL, z.range = NULL,
  leg.coords = NULL, ticktype = "detailed", theta = 40, phi = 40,
  smooth = "none", mask = FALSE, perspective = TRUE,
  col=rgb(0,100,0,50,maxColorValue=255), cex.lab=1.5, cex.axis=1.5)

# build initial ARS BRT models for preliminary analysis to determine tree
complexity (TC) and learning rate (LR)

# the following script was repeated 16 times for all combinations of TC and LR

obs.data.ars<-read.csv("myARSdata.csv")

ARS_TC2_LR0.01<-gbm.step(data=obs.data.ars, gbm.x=21:29, gbm.y=20,
  family="bernoulli", tree.complexity = 2,
  learning.rate = 0.01, bag.fraction = 0.5,
  max.trees = 1000000)

# chosen models based on AUC and D2

ARS_TC5_LR0.001<-gbm.step(data=obs.data.ars, gbm.x=21:29, gbm.y=20,
  family="bernoulli", tree.complexity = 5,
  learning.rate = 0.001, bag.fraction = 0.5,
  max.trees = 1000000)

#now to test model simplification

# testing all combinations of drops up to a maximum of 7 variables

ARS_TC5_LR0.001_SIMP<-gbm.simplify(ARS_TC5_LR0.001, n.drops = 7)

summary(ARS_TC5_LR0.001_SIMP)

# summary and resulting plot indicates that two predictor variable should be
dropped

# construct model again with the prediction list passed to the response argument
and the number 2
# to drop the two least informative variables

```

```

ARS_TC5_LR0.001_SIMP_drop2<-gbm.step(data=obs.data.ars,
                                     gbm.x=ARS_TC5_LR0.001_SIMP$pred.list[[2]], gbm.y=20,
                                     family="bernoulli", tree.complexity = 5,
                                     learning.rate = 0.001, bag.fraction = 0.5,
                                     max.trees = 1000000)

# get the results of the model

summary(ARS_TC5_LR0.001_SIMP_drop2) # variable contributions scaled to 100

# plot partial dependency plots

gbm.plot(ARS_TC5_LR0.001_SIMP_drop2, variable.no=1, smooth=TRUE,
         common.scale=TRUE, write.title=FALSE, y.label="Fitted function",
         x.label=expression(paste("Sea surface temperature (°C)")),
         show.contrib=FALSE, plot.layout=c(1, 1), cex.lab=1.5, cex.axis=1.5)

# edit plot via source code

trace(gbm.plot, edit = T) # calls editable source code

# sections of source code edited
# set to display rugs in quartiles: probs = seq(0, 1, 0.25)
# move rugs: side =, colour rugs: col =
# Line thickness: lwd =

# set the absence rugs by defining only to use rows where Overall == 0
# set placement of rugs to the bottom of the plot with red tick marks

    }
    if (rug & is.vector(data[, gbm.call$gbm.x[variable.no]])) {
      rug(quantile(data[gbm.call$dataframe$Overall ==
        0, gbm.call$gbm.x[variable.no]], prob = seq(0, 1, 0.01), na.rm = TRUE,
        type = 7), side = 1, col = 2, lwd = 2)
    }

# set the presence rugs by defining only to use rows where Overall == 1
# set placement of rugs to the bottom of the plot with blue tick marks

    }
    if (rug & is.vector(data[, gbm.call$gbm.x[variable.no]])) {
      rug(quantile(data[gbm.call$dataframe$Overall ==
        1, gbm.call$gbm.x[variable.no]], prob = seq(0, 1, 0.01), na.rm = TRUE,
        type = 7), side = 3, col = 4, lwd = 2)
    }
  }

# find predictor variable interactions fitted within the model

find.int<-gbm.interactions(ARS_TC5_LR0.001_SIMP_drop2)

find.int$interaction # gives a matrix of interactions

find.int$rank.list # ranks important interactions between predictor variables

# get a 3D plot of interactions
# the numbers specified indicate the number of the variable assigned in the rank
list

```

```
# these numbers do not reflect rank they just ID the variable for plotting
# example using SST and depth where the depth has been restricted as interaction
# is most substantial
# at shallower depths

gbm.perspec(ARS_TC5_LR0.001_SIMP_drop2, 1, 5,y.range=c(-1000, 0.99),
            z.label = "Marginal effct")
```

## Preparing your manuscript

This section provides general style and formatting information only. Formatting guidelines for specific article types can be found below.

- [Marine Record](#)
- [Research](#)
- [Review](#)

### General formatting guidelines

- [Preparing main manuscript text](#)
- [Preparing illustrations and figures](#)
- [Preparing tables](#)
- [Preparing additional files](#)

### Preparing main manuscript text

#### [Back to top](#)

Quick points:

- Use double line spacing
- Include line and page numbering
- Use SI units: Please ensure that all special characters used are embedded in the text, otherwise they will be lost during conversion to PDF
- Do not use page breaks in your manuscript

### File formats

The following word processor file formats are acceptable for the main manuscript document:

- Microsoft word (DOC, DOCX)
- Rich text format (RTF)
- TeX/LaTeX (use BioMed Central's TeX template)

**Please note:** editable files are required for processing in production. If your manuscript contains any non-editable files (such as PDFs) you will be required to re-submit an editable file when you submit your revised manuscript, or after editorial acceptance in case no revision is necessary.

### Additional information for TeX/LaTeX users

Please use BioMed Central's TeX template and BibTeX stylefile if you use TeX format. Submit your references using either a bib or bbl file. When submitting TeX submissions, please submit both your TeX file and your bib/bbl file as manuscript files. Please also convert your TeX file into a PDF (please do not use a DIV file) and submit this PDF as a supplementary file with the name 'Reference PDF'. This PDF will be used by our production team as a reference point to check the layout of the article as the author intended.

The Editorial Manager system checks for any errors in the TeX files. If an error is present then the system PDF will display LaTeX code and highlight and explain the error in a section beginning with an exclamation mark (!).

All relevant editable source files must be uploaded during the submission process. Failing to submit these source files will cause unnecessary delays in the production process.

## **Research [26]**

### **Criteria**

Research articles should report on original primary research, but may report on systematic reviews of published research provided they adhere to the appropriate reporting guidelines which are detailed in our [editorial policies](#). Please note that non-commissioned pooled analyses of selected published research will not be considered.

### **Preparing your manuscript**

The information below details the section headings that you should include in your manuscript and what information should be within each section.

Please note that your manuscript must include a 'Declarations' section including all of the subheadings (please see below for more information).

### **Title page**

The title page should:

- present a title that includes, if appropriate, the study design
- list the full names and institutional addresses for all authors
  - if a collaboration group should be listed as an author, please list the Group name as an author. If you would like the names of the individual members of the Group to be searchable through their individual PubMed records, please include this information in the "Acknowledgements" section in accordance with the instructions below



- indicate the corresponding author

## **Abstract**

The Abstract should not exceed 350 words. Please minimize the use of abbreviations and do not cite references in the abstract. The abstract must include the following separate sections:

- **Background:** the context and purpose of the study
- **Results:** the main findings
- **Conclusions:** a brief summary and potential implications

## **Keywords**

Three to ten keywords representing the main content of the article.

## **Background**

The Background section should explain the background to the study, its aims, a summary of the existing literature and why this study was necessary.

## **Results**

This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

## **Discussion**

For research articles this section should discuss the implications of the findings in context of existing research and highlight limitations of the study. For study protocols and methodology manuscripts this section should include a discussion of any practical or operational issues involved in performing the study and any issues not covered in other sections.

## **Conclusions**

This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study to the field.

## **Methods**

The methods section should include:

- the aim, design and setting of the study
- the characteristics of participants or description of materials

- a clear description of all processes, interventions and comparisons. Generic names should generally be used. When proprietary brands are used in research, include the brand names in parentheses
- the type of statistical analysis used, including a power calculation if appropriate

### **List of abbreviations**

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations can be provided.

### **Declarations**

All manuscripts must contain the following sections under the heading 'Declarations':

- Ethics approval and consent to participate
- Consent for publication
- Availability of data and material
- Competing interests
- Funding
- Authors' contributions
- Acknowledgements
- Authors' information (optional)

Please see below for details on the information to be included in these sections.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

### ***Ethics approval and consent to participate***

Manuscripts reporting studies involving human participants, human data or human tissue must:

- include a statement on ethics approval and consent (even where the need for approval was waived)
- include the name of the ethics committee that approved the study and the committee's reference number if appropriate

Studies involving animals must include a statement on ethics approval.

See our [editorial policies](#) for more information.

If your manuscript does not report on or involve the use of any animal or human data or tissue, please state “Not applicable” in this section.

### ***Consent for publication***

If your manuscript contains any individual person’s data in any form (including any individual details, images or videos), consent for publication must be obtained from that person, or in the case of children, their parent or legal guardian. All presentations of case reports must have consent for publication.

You can use your institutional consent form or our [consent form](#) if you prefer. You should not send the form to us on submission, but we may request to see a copy at any stage (including after publication).

See our [editorial policies](#) for more information on consent for publication.

If your manuscript does not contain data from any individual person, please state “Not applicable” in this section.

### ***Availability of data and materials***

All manuscripts must include an ‘Availability of data and materials’ statement. Data availability statements should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the findings reported in the article. We recognise it is not always possible to share research data publicly, for instance when individual privacy could be compromised, and in such instances data availability should still be stated in the manuscript along with any conditions for access.

Data availability statements can take one of the following forms (or a combination of more than one if required for multiple datasets):

- The datasets generated and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS]
- The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
- All data generated or analysed during this study are included in this published article [and its supplementary information files].
- The datasets generated and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.

- Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.
- The data that support the findings of this study are available from [third party name] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [third party name].
- Not applicable. If your manuscript does not contain any data, please state 'Not applicable' in this section.

More examples of template data availability statements, which include examples of openly available and restricted access datasets, are available [here](#).

BioMed Central also requires that authors cite any publicly available data on which the conclusions of the paper rely in the manuscript. Data citations should include a persistent identifier (such as a DOI) and should ideally be included in the reference list. Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style. Dataset identifiers including DOIs should be expressed as full URLs. For example:

Hao Z, AghaKouchak A, Nakhjiri N, Farahmand A. Global integrated drought monitoring and prediction system (GIDMaPS) data sets. figshare. 2014. <http://dx.doi.org/10.6084/m9.figshare.853801>

With the corresponding text in the Availability of data and materials statement:

The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].<sup>[Reference number]</sup>

### ***Competing interests***

All financial and non-financial competing interests must be declared in this section.

See our [editorial policies](#) for a full explanation of competing interests. If you are unsure whether you or any of your co-authors have a competing interest please contact the editorial office.

Please use the authors initials to refer to each authors' competing interests in this section.

If you do not have any competing interests, please state "The authors declare that they have no competing interests" in this section.

### ***Funding***

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared.

### ***Authors' contributions***

The individual contributions of authors to the manuscript should be specified in this section. Guidance and criteria for authorship can be found in our [editorial policies](#).

Please use initials to refer to each author's contribution in this section, for example: "FC analyzed and interpreted the patient data regarding the hematological disease and the transplant. RH performed the histological examination of the kidney, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript."

### ***Acknowledgements***

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials.

Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.

See our [editorial policies](#) for a full explanation of acknowledgements and authorship criteria.

If you do not have anyone to acknowledge, please write "Not applicable" in this section.

Group authorship (for manuscripts involving a collaboration group): if you would like the names of the individual members of a collaboration Group to be searchable through their individual PubMed records, please ensure that the title of the collaboration Group is included on the title page and in the submission system and also include collaborating author names as the last paragraph of the "Acknowledgements" section. Please add authors in the format First Name, Middle initial(s) (optional), Last Name. You can add institution or country information for each author if you wish, but this should be consistent across all authors.

Please note that individual names may not be present in the PubMed record at the time a published article is initially included in PubMed as it takes PubMed additional time to code this information.

### ***Authors' information***

This section is optional.

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to authors using their initials. Note this section should not be used to describe any competing interests.

### **Endnotes**

Endnotes or Footnotes should be designated within the text using a superscript lowercase letter. These should be captured in the Footnotes, alternatively, all notes (along with their corresponding letter) should be included in an Endnotes section, please format this section in a paragraph rather than a list.

### **References**

Examples of the Vancouver reference style are shown below.

See our [editorial policies](#) for author guidance on good citation practice

**Web links and URLs:** All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database. <http://tumor.informatics.jax.org/mtbwi/index.do>. Accessed 20 May 2013. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

### **Example reference style:**

*Article within a journal*

Smith JJ. The world of science. Am J Sci. 1999;36:234-5.

*Article within a journal (no page numbers)*

Rohrmann S, Overvad K, Bueno-de-Mesquita HB, Jakobsen MU, Egeberg R, Tjønneland A, et al. Meat consumption and mortality - results from the European Prospective Investigation into Cancer and Nutrition. *BMC Medicine*. 2013;11:63.

*Article within a journal by DOI*

Slifka MK, Whitton JL. Clinical implications of dysregulated cytokine production. *Dig J Mol Med*. 2000; doi:10.1007/s801090000086.

*Article within a journal supplement*

Frumin AM, Nussbaum J, Esposito M. Functional asplenia: demonstration of splenic activity by bone marrow scan. *Blood* 1979;59 Suppl 1:26-32.

*Book chapter, or an article within a book*

Wyllie AH, Kerr JFR, Currie AR. Cell death: the significance of apoptosis. In: Bourne GH, Danielli JF, Jeon KW, editors. *International review of cytology*. London: Academic; 1980. p. 251-306.

*OnlineFirst chapter in a series (without a volume designation but with a DOI)*

Saito Y, Hyuga H. Rate equation approaches to amplification of enantiomeric excess and chiral symmetry breaking. *Top Curr Chem*. 2007. doi:10.1007/128\_2006\_108.

*Complete book, authored*

Blenkinsopp A, Paxton P. *Symptoms in the pharmacy: a guide to the management of common illness*. 3rd ed. Oxford: Blackwell Science; 1998.

*Online document*

Doe J. Title of subordinate document. In: *The dictionary of substances and their effects*. Royal Society of Chemistry. 1999. [http://www.rsc.org/dose/title of subordinate document](http://www.rsc.org/dose/title%20of%20subordinate%20document). Accessed 15 Jan 1999.

*Online database*

Healthwise Knowledgebase. US Pharmacopeia, Rockville. 1998.  
<http://www.healthwise.org>. Accessed 21 Sept 1998.

#### *Supplementary material/private homepage*

Doe J. Title of supplementary material. 2000. <http://www.privatehomepage.com>.  
Accessed 22 Feb 2000.

#### *University site*

Doe, J: Title of preprint. <http://www.uni-heidelberg.de/mydata.html> (1999). Accessed  
25 Dec 1999.

#### *FTP site*

Doe, J: Trivial HTTP, RFC2169. <ftp://ftp.isi.edu/in-notes/rfc2169.txt> (1999). Accessed  
12 Nov 1999.

#### *Organization site*

ISSN International Centre: The ISSN register. <http://www.issn.org> (2006). Accessed 20  
Feb 2007.

#### *Dataset with persistent identifier*

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, et al. Genome data from sweet  
and grain sorghum (*Sorghum bicolor*). GigaScience Database.  
2011. <http://dx.doi.org/10.5524/100012>.

## **Preparing figures**

### [Back to top](#)

When preparing figures, please follow the formatting instructions below.

- Figures should be numbered in the order they are first mentioned in the text, and uploaded in this order. Multi-panel figures (those with parts a, b, c, d etc.) should be submitted as a single composite file that contains all parts of the figure.
- Figures should be uploaded in the correct orientation.
- Figure titles (max 15 words) and legends (max 300 words) should be provided in the main manuscript, not in the graphic file.



- Figure keys should be incorporated into the graphic, not into the legend of the figure.
- Each figure should be closely cropped to minimize the amount of white space surrounding the illustration. Cropping figures improves accuracy when placing the figure in combination with other elements when the accepted manuscript is prepared for publication on our site. For more information on individual figure file formats, see our detailed instructions.
- Individual figure files should not exceed 10 MB. If a suitable format is chosen, this file size is adequate for extremely high quality figures.
- **Please note that it is the responsibility of the author(s) to obtain permission from the copyright holder to reproduce figures (or tables) that have previously been published elsewhere.** In order for all figures to be open access, authors must have permission from the rights holder if they wish to include images that have been published elsewhere in non open access journals. Permission should be indicated in the figure legend, and the original source included in the reference list.

## Figure file types

We accept the following file formats for figures:

- EPS (suitable for diagrams and/or images)
- PDF (suitable for diagrams and/or images)
- Microsoft Word (suitable for diagrams and/or images, figures must be a single page)
- PowerPoint (suitable for diagrams and/or images, figures must be a single page)
- TIFF (suitable for images)
- JPEG (suitable for photographic images, less suitable for graphical images)
- PNG (suitable for images)
- BMP (suitable for images)
- CDX (ChemDraw - suitable for molecular structures)

For information and suggestions of suitable file formats for specific figure types, please see our [author academy](#).

## Figure size and resolution

Figures are resized during publication of the final full text and PDF versions to conform to the BioMed Central standard dimensions, which are detailed below.

Figures on the web:

- width of 600 pixels (standard), 1200 pixels (high resolution).

Figures in the final PDF version:

- width of 85 mm for half page width figure
- width of 170 mm for full page width figure
- maximum height of 225 mm for figure and legend
- image resolution of approximately 300 dpi (dots per inch) at the final size

Figures should be designed such that all information, including text, is legible at these dimensions. All lines should be wider than 0.25 pt when constrained to standard figure widths. All fonts must be embedded.

### ***Figure file compression***

- Vector figures should if possible be submitted as PDF files, which are usually more compact than EPS files.
- TIFF files should be saved with LZW compression, which is lossless (decreases file size without decreasing quality) in order to minimize upload time.
- JPEG files should be saved at maximum quality.
- Conversion of images between file types (especially lossy formats such as JPEG) should be kept to a minimum to avoid degradation of quality.

If you have any questions or are experiencing a problem with figures, please contact the customer service team at [info@biomedcentral.com](mailto:info@biomedcentral.com).

## **Preparing tables**

[Back to top](#)

When preparing tables, please follow the formatting instructions below.

- Tables should be numbered and cited in the text in sequence using Arabic numerals (i.e. Table 1, Table 2 etc.).
- Tables less than one A4 or Letter page in length can be placed in the appropriate location within the manuscript.
- Tables larger than one A4 or Letter page in length can be placed at the end of the document text file. Please cite and indicate where the table should appear at the relevant location in the text file so that the table can be added in the correct place during production.
- Larger datasets, or tables too wide for A4 or Letter landscape page can be uploaded as additional files. Please see [below] for more information.
- Tabular data provided as additional files can be uploaded as an Excel spreadsheet (.xls ) or comma separated values (.csv). Please use the standard file extensions.

- Table titles (max 15 words) should be included above the table, and legends (max 300 words) should be included underneath the table.
- Tables should not be embedded as figures or spreadsheet files, but should be formatted using 'Table object' function in your word processing program.
- Color and shading may not be used. Parts of the table can be highlighted using superscript, numbering, lettering, symbols or bold text, the meaning of which should be explained in a table legend.
- Commas should not be used to indicate numerical values.

If you have any questions or are experiencing a problem with tables, please contact the customer service team at [info@biomedcentral.com](mailto:info@biomedcentral.com).

## Preparing additional files

### [Back to top](#)

As the length and quantity of data is not restricted for many article types, authors can provide datasets, tables, movies, or other information as additional files.

All Additional files will be published along with the accepted article. Do not include files such as patient consent forms, certificates of language editing, or revised versions of the main manuscript document with tracked changes. Such files, if requested, should be sent by email to the journal's editorial email address, quoting the manuscript reference number. Please do not send completed patient consent forms unless requested.

Results that would otherwise be indicated as "data not shown" should be included as additional files. Since many web links and URLs rapidly become broken, BioMed Central requires that supporting data are included as additional files, or deposited in a recognized repository. Please do not link to data on a personal/departmental website. Do not include any individual participant details. The maximum file size for additional files is 20 MB each, and files will be virus-scanned on submission. Each additional file should be cited in sequence within the main body of text.

If additional material is provided, please list the following information in a separate section of the manuscript text:

- File name (e.g. Additional file 1)
- File format including the correct file extension for example .pdf, .xls, .txt, .pptx (including name and a URL of an appropriate viewer if format is unusual)
- Title of data
- Description of data

Additional files should be named "Additional file 1" and so on and should be referenced explicitly by file name within the body of the article, e.g. 'An additional movie file shows this in more detail [see Additional file 1]'.

For further guidance on how to use Additional files or recommendations on how to present particular types of data or information, please see [How to use additional files](#).

## References

1. CMS Collecte Localisation Satellites. Argos User's Manual © 2007 -2016 CLS Last updated on June 15, 2016. Toulouse; 2016.
2. Douglas DC, Weinzierl R, C. Davidson S, Kays R, Wikelski M, Bohrer G. Moderating Argos location errors in animal tracking data. *Methods Ecol Evol.* 2012;3(6):999–1007.
3. Costa DP, Robinson PW, Arnould JPY, Harrison AL, Simmons SE, Hassrick JL, *et al.* Accuracy of ARGOS locations of pinnipeds at-sea estimated using fastloc GPS. *PLoS One.* 2010;5(1):e8677
4. Jonsen ID, Flemming JM, Myers RA. Robust state-space modeling of animal movement data. *Ecology.* 2005;86(11):2874–80.
5. Reisinger RR, Keith M, Andrews RD, de Bruyn PJN. Movement and diving of killer whales (*Orcinus orca*) at a Southern Ocean archipelago. *J Exp Mar Bio Ecol.* 2015;473:90–102.
6. Hearn AR, Green JR, Espinoza E, Peñaherrera C, Acuña D, Klimley AP. Simple criteria to determine detachment point of towed satellite tags provide first evidence of return migrations of whale sharks (*Rhincodon typus*) at the Galapagos Islands, Ecuador. *Anim Biotelemetry.* 2013;1(1):1–10.
7. Rohner CA, Richardson AJ, Jaine FRA, Bennett MB, Weeks SJ, Cliff G, *et al.* Satellite tagging highlights the importance of productive Mozambican coastal waters to the ecology and conservation of whale sharks. *PeerJ.* 2018;6:e4161.
8. Jonsen ID, Myers RA, James MC. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Mar Ecol Prog Ser.* 2007;337:255–64.
9. Breed GA, Jonsen ID, Myers RA, Bowen WD, Marty L, Breed GA, *et al.* Sex-Specific , Seasonal Foraging Tactics of Adult Grey Seals (*Halichoerus grypus*) Revealed by State-Space Analysis. *Ecology.* 2009;90(11):3209–21.
10. Brooks SP, Gelman A, Brooks SP. General Methods for Monitoring Convergence of Iterative Simulations General Methods for Monitoring Convergence of Iterative

- Simulations. *J Comput Graph Stat.* 1998;7(4):434–55.
11. Jonsen ID, Wotherspoon SJ, Bravington MV, Thomson R, Bestley S, Patterson TA, *et al.* State-space models for bio-loggers: A methodological road map. *Deep Sea Res Part II Top Stud Oceanogr.* 2013;88–89:34–46.
  12. Jonsen I. Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Sci Rep.* 2016;6:1–9.
  13. McCarthy MA. *Bayesian Methods for Ecology.* Cambridge: Cambridge University Press; 2007.
  14. Robert CP, Casella G. *Introducing Monte Carlo Methods with R.* New York: Springer Science & Business Media; 2010.
  15. Rubin DB, Gelman A. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci.* 1992;7(4):457–72.
  16. Jonsen I, Bestley S, Wotherspoon S, Sumner M, Mills J. *Bayesian State-Space Models for Animal Movement (Package bsam).* 2017. Available from: <https://cran.r-project.org/web/packages/bsam/bsam.pdf>. Accessed 3 Feb 2019.
  17. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol.* 2008;77(4):802–13.
  18. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning The Elements of Statistical Learning Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer-Verlag; 2009.
  19. Colin B, Clifford S, Wu P, Rathmanner S, Mengersen K. Using Boosted Regression Trees and Remotely Sensed Data to Drive Decision-Making. *Open J Stat.* 2017;07(05):859–75.
  20. Froeschke J, Stunz GW, Wildhaber ML. Environmental influences on the occurrence of coastal sharks in estuarine waters. *Mar Ecol Prog Ser.* 2010;407:279–92.
  21. Tierney NJ, Harden FA, Harden MJ, Mengersen KL. Using decision trees to understand structure in missing data. *BMJ Open.* 2015;5(6):1–11.

22. Leathwick JR, Elith J, Francis MP. Variation in demersal fish species richness in the oceans surrounding New Zealand. *Mar Ecol Prog Ser*. 2006;321:267–81.
23. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat*. 2001;29(5):1189–232.
24. Elith J, Leathwick J. Boosted Regression Trees for ecological modeling. 2017. <https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf>. Accessed 3 Feb 2019. Accessed 3 Nov 2018.
25. Fielding AH, Bell JF, Fielding AH, Bell JF. A review of methods for the assessment of prediction errors in conservation presence / absence models. *Environ Conserv*. 1997;24(1):38–49.
26. Marine Biodiversity Records. Submission Guidelines: Research. 2019. <https://mbr.biomedcentral.com/submission-guidelines/preparing-your-manuscript#preparing+main+manuscript+text>. Accessed 10 March 2019.