Supplementary Material

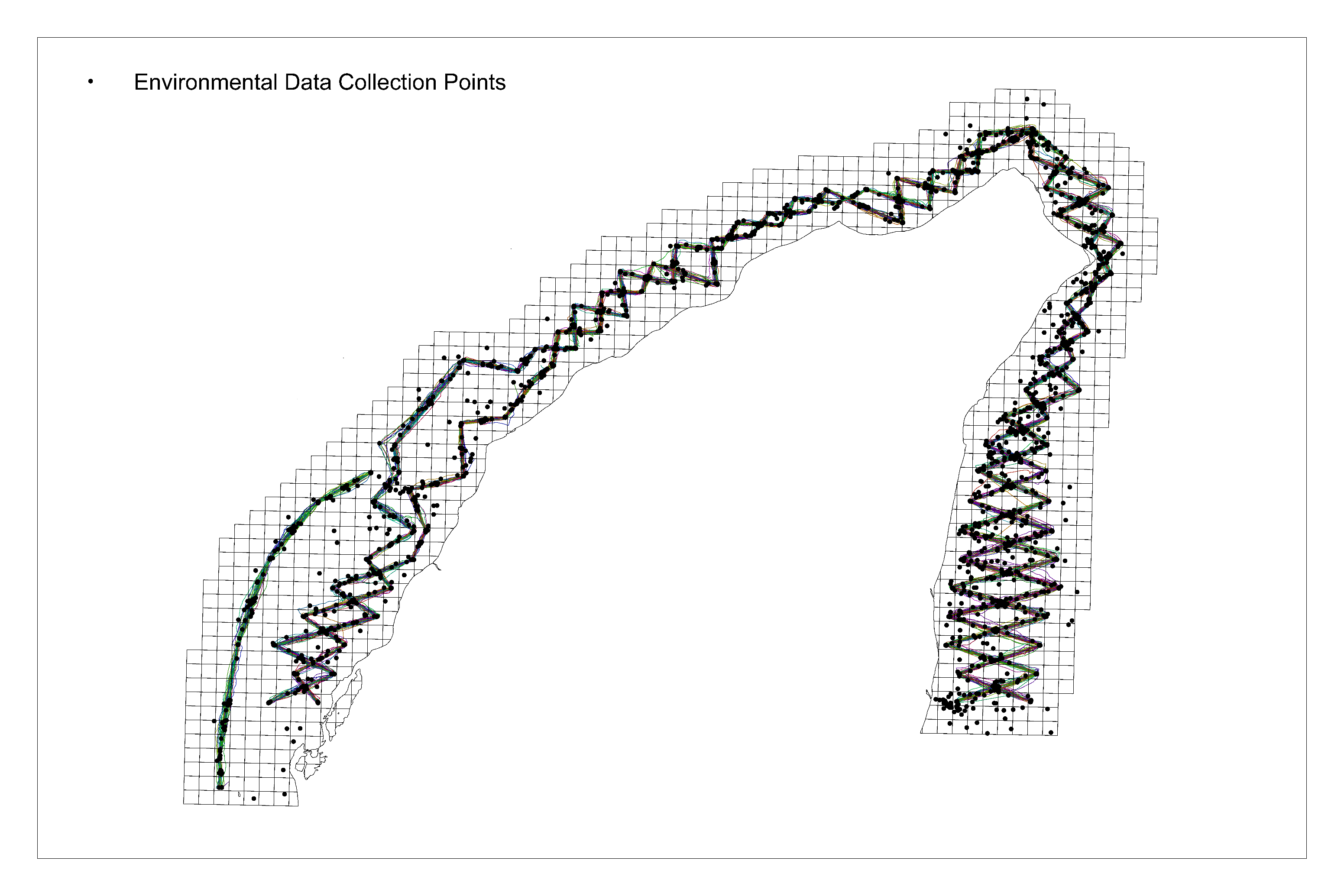
**Supplementary Material 1: Raw survey tracks, environmental data collection locations and spatial distribution of survey effort**

Figure S1.1: Raw survey tracks and environmental data collection locations (2013-2015 and 2018-2019) at the North West Cape

Histogram

Description automatically generated

Figure S1.2: Overall spatial distribution of survey effort (2013-2015 and 2018-2019) at the North West Cape

# Supplementary Material 2: ODMAP Protocol

– ODMAP Protocol –

**Overview**

**Authorship**

Modelling the distribution and habitat preferences of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) inhabiting coastal waters with mixed levels of protection.

Haughey, R., Hunt, T., Hanf, D., Passadore, C., Baring, R. and Parra G J.

Contact: [bec.haughey@hotmail.com.au](mailto:bec.haughey@hotmail.com.au)

**Model objective**

Mapping and interpolation

**Target output**

#### To understand the distribution patterns, habitat preferences and to identify areas of high probability of occurrence for Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) at the North West Cape.

**Focal Taxon**

Indo-Pacific bottlenose dolphin (*Tursiops aduncus*)

**Location**

North West Cape, Western Australia

**Scale of Analysis**

**Spatial extent**

Universal Transverse Mercator zone 50 south based on the WGS 1984 datum

Top: 7591027.338288 m

Left: 17672.044127 m

Right: 211672.044127 m

Bottom: 7565027.338288 m

**Spatial resolution**

500m by 500m

**Temporal extent**

5 years (2013, 2014, 2015, 2018 & 2019)

**Boundary**

Rectangle

**Biodiversity data**

**Observation type**

Field survey

**Response data type**

Presence/absence

**Predictors**

Habitat, topographic and environmental

**Hypotheses**

A number of environmental and anthropogenic variables have been shown to influence the distribution of cetaceans, in particular dolphins. We used measures of water depth, seabed slope, sea surface temperature, distance to coast, distance to the nearest boat ramp and marine park zone as predictor variables for Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) at the North West Cape. We modelled temporal changes by year and austral season and the influence of benthic habitat type within the Ningaloo Marine Park.

**Assumptions**

1. Relevant ecological drivers of species distribution are included

2. Detectability does not change across transects

3. Species are at equilibrium with their environment

4. Sampling is adequate and representative (and any biases are accounted for/corrected)

**Algorithms**

**Modelling techniques**

glm, mars, ann, gam, fda, gbm, randomForest

**Model averaging**

We combined model algorithms to form an ensemble model/prediction.

**Workflow**

Only weakly correlated explanatory variables were included in the final models. Ensemble predictions were derived using mean of mean values from model runs that performed well from the single model algorithms.

**Software**

Software: Analyses were conducted in R Studio using the biomod2 package (<https://cran.r-project.org/web/packages/biomod2/index.html>, <https://github.com/biomodhub/biomod2>). Additional packages required were raster, sp and rjava

ArcMap version 10.7 (ESRI) was used to create spatial layers of response (dolphin presence-absence) and predictor variables.

**Code availability**

Code is available on request

**Data availability**

Data are available on request

**Data**

**Biodiversity data**

**Taxon name**

Indo-Pacific bottlenose dolphin (*Tursiops aduncus*)

**Ecological level**

Population/species

**Data sources**

Survey data was collected in the field over a 5 year period. Each field season (total = 5) was intensive and lasted 6 months, spanning over the austral seasons of autumn, winter and spring.

**Sampling design**

Boat based surveys searching for Indo-Pacific bottlenose dolphins were conducted onboard a 5.6 m vessel during the hours of 0700 and 1800 from May to October in 2013, April to October in 2014, May to October in 2015 and April to September in both 2018 and 2019. Sampling periods spanned the Austral seasons of Autumn (April - May), Winter (June - August) and Spring (September – October). Surveys were conducted in favourable weather conditions (i.e. Beaufort Sea State of ≤3 and no rain) and followed a systematic line transect layout (2 x 93 km opposing zig zag lines and 1 x 13 km single line) covering a wide range of habitats, human use areas and environmental variables within the study area.

**Sample size**

Over the five years of study, almost 723 hours of survey effort were completed. Survey effort varied slightly between years and austral seasons due to variability in weather conditions. Overall, the highest survey effort and number of dolphin sightings occurred, during the winter months (June-August). In total, we encountered 323 groups of IP bottlenose dolphins, with 70 seen in autumn, 184 in winter and 69 in spring. Overall, 227 (70%) of these groups were seen inside the NMP and 96 (30%) outside these boundaries.

**Clipping**

All data covered the extent of the study area.

**Scaling**

500m x 500m grid resolution.

**Absence data**

Each 500 m x 500 m grid within the survey area was assigned either a 1 (dolphin presence) or 0 (dolphin absence). In order to reduce false absences (i.e. determining an absent cell when individuals may in fact occur in that area), absence cells were defined based on areas which had the highest survey effort (Phillips et al. 2009). Survey effort was quantified using the total area of ‘on-effort’ survey tracks within each 500 x 500 m grid cell. A 250 m buffer area either side of each transect line was added, which was considered to be the average distance from the vessel that dolphins could reliably be observed under a variety of sea conditions (Zanardo et al 2017, Hunt 2018). Grid cells were then ranked and cells containing no dolphin presence and values of survey effort higher than the mean were considered most likely to represent true absences and therefore were defined as absence cells (Zanardo et al. 2017, Hunt 2018, Passadore et al. 2018).

**Data partitioning**

**Training data**

A random data splitting procedure of 75/25% was used for model calibration and testing. We implemented a 10-fold cross-validation method.

**Validation data**

See training data

**Predictor variables**

**Predictor variables**

Predictor variables used to model Indo-Pacific bottlenose dolphin distribution and habitat use were classified as: abiotic (i.e. water depth, slope, sea surface temperature (SST) and distance to coast), and anthropogenic (i.e. distance to boat ramp and marine park zone).

**Data sources**

Most data were collected *in situ* or calculated using tools in ArcMap. Marine park zoning data was obtained from the Western Australian Government Department of Biodiversity, Conservation and Attractions. Benthic habitat type data was derived from a broad-scale benthic habitat study of the Ningaloo Marine Park (Bancroft and Sheridan 2000) and obtained from Seamap Australia (Lucieer et al. 2017).

**Spatial extent**

Top: 7591027.338288 m

Left: 17672.044127 m

Right: 211672.044127 m

Bottom: 7565027.338288 m

**Spatial resolution**

500m x 500m

**Coordinate reference system**

All spatial layers were projected to Universal Transverse Mercator zone 50 south based on the WGS 1984 datum.

**Temporal extent**

6-month sampling periods repeated over 5 years (2013, 2014, 2015, 2018, 2019).

#### Temporal resolution

Years and austral seasons.

#### Data processing

SST raster layers were created using the Ordinary Kriging tool with a spherical semi variogram model (500m cell size, 12-point variable search radius size). Slope was calculated as the standard deviation of the depth, using the slope tool. Distance to the coast and boat ramps was measured using the Euclidean distance (i.e. shortest straight-line distance) and Cost distance (the shortest distance factoring in land given study area wraps around a peninsula) functions, respectively, using the Spatial Analyst extension in ArcMap. To evaluate the relevance of the marine park zoning for the conservation of Indo-Pacific bottlenose dolphins within the study area, a raster of marine park zones was created with the following zones; General, Recreational, Sanctuary, Special Purpose, Naval Waters and outside the Ningaloo Marine Park.

#### Errors and biases:

Prey availability data is not known, but selected predictor variables have been used as proxies for prey availability/distribution.

### Transfer data

ArcMap version 10.7 (ESRI) was used to create spatial layers of response (dolphin presence-absence) and predictor variables at a 500 x 500 m grid resolution (See Table S2.1, also see *Methods: Predictor variables* section). All spatial layers were projected to Universal Transverse Mercator zone 50 South.

Table S2.1: List of predictor variables used in species distribution modelling of Indo-Pacific bottlenose dolphins in the coastal waters of the North West Cape, Western Australia, their associated data source and how they were derived in ArcMap. \*indicates the variable was only included in Ningaloo Marine Park (NMP) SDMs. Spatial resolution for each variable is 500 x 500 m.

|  |  |  |  |
| --- | --- | --- | --- |
| Type of variable | Predictor variable | Variable abbreviation | Data Source |
| Abiotic | Distance to coast | N/A | Derived using the Euclidean distance tool (Spatial Analyst toolbox). |
| Abiotic | Slope | Slope | Derived using the Slope tool and is measured in decimal degrees (Spatial Analyst toolbox). |
| Abiotic | Sea surface temperature | SST | Derived from *in situ* measurements of SST. Created using the Ordinary Kriging tool with a spherical semi variogram model (500m cell size, 12-point variable search radius size)(Spatial analyst toolbox). |
| Abiotic | Water depth | Depth | Derived from *in situ* measurements of depth. Created using the Ordinary Kriging tool with a spherical semi variogram model (500m cell size, 12-point variable search radius size) (Spatial Analyst toolbox). |
| Abiotic | Water visibility | N/A | Derived from *in situ* measurements of water visibility using a secchi disk and calculated as a proportion of the total depth.  Created using the Ordinary Kriging tool with a spherical semi variogram model (500m cell size, 12-point variable search radius size) (Spatial Analyst Toolbox). |
| Anthropogenic | Distance to boat ramp | N/A | Exmouth, Bundegi and Tantabiddi boat ramps are established vessel launch sites in the study area. Derived using the Cost distance tool (Spatial Analyst toolbox). |
| Anthropogenic | Marine park zone | NMP Zone | A NMP zoning shape file which shows the zone boundaries was obtained from the Western Australian Government Department of Biodiversity, Conservation and Attractions. Each grid cell was assigned a variable according to MP zone using the polygon to raster tool (1=General Use, 2=Recreational, 3=Sanctuary, 4= Special Purpose, 5 = Naval Waters, 6 = outside the NMP) (DPaW. and DoF. 2014) |
| Biotic | Benthic habitat type\* | Habitat | Benthic habitat data only exists for the Ningaloo Marine Park portion of the study area (Figure 1). This data was derived from a broad scale benthic habitat study of the NMP (Bancroft and Sheridan 2000, DPaW. 2006, Lucieer et al. 2017). Habitat type was classified as either 1=mobile sand, 2=mangroves, 3=bare reef (intertidal), 4=coral reef (intertidal), 5=bare reef (subtidal), 6=macro algae (subtidal), 7=coral reef (subtidal), 8=saltmarsh and 10=pelagic (No habitat type associated with a value of 9 (mudflat) is present in this section of the NMP). For habitat type definitions, see Supplementary Material 3, Table S3. Each grid cell was assigned a variable according to habitat type using the polygon to raster tool. |

## Model

### Multicollinearity

Before running the SDM’s, we tested for collinearity between our continuous numerical explanatory variables using stepwise procedures within the usdm package in RStudio (Naimi 2015). Variance inflation factors (VIF) were calculated for all variables. Variable pairs with a maximum linear correlation greater than the threshold (0.7) were identified using ‘vifcor’ and the variable with the highest variance inflation factor (VIF, threshold = 3) was excluded using ‘vifstep’ (Zuur et al. 2010). These procedures were repeated until there was no variable remaining with a correlation coefficient greater than 0.7 and no variables with a VIF greater than the threshold (Naimi et al. 2014).

### Model settings

See Supplementary Material S5 for model settings.

### Model estimates

The importance of explanatory variables was calculated using a 10-permutation run randomisation procedure within BioMod2 (Thuiller et al. 2009). This procedure allows for a direct comparison between model algorithms and calculates the Pearson’s correlation between the standard predictions and predictions where 1 variable has been randomly permutated. High correlation (i.e. little difference between the two predictions) indicates that the variable is not important in the model, and a low correlation indicates that the variable is important. Variables are ranked from 0 to 1 according to the mean correlation coefficient, with the variable with the highest ranking the most influential and the lowest, the least influential (Thuiller et al. 2009).

### Analysis and Correction of non-independence

To ensure independence of data points in the face of bottlenose dolphins exhibiting flexible grouping patterns, termed fission fusion (Wells et al. 1987), only the locational point where the initial group members were first encountered whilst on transect was included in analysis.

### Threshold selection

To compare modelling algorithms and for comparability with other studies, we used the AUC metric which has been standard and widely used in many SDM studies (Fielding and Bell 1997). Values of AUC range from 0 to 1; with values >0.5 indicating that the model predictions perform better than random, whereas values <0.5 indicates that the model predictions are no better than what would be expected by chance. In general, AUC values of 0.5–0.7 are considered low and represent poor model performance, values of 0.7–0.9 are considered moderate to good, and values above 0.9 represent excellent model performance (Peterson et al. 2011).

## Assessment

### Performance statistics

Model performance was assessed based on AUC values using a threshold of 0.7. Final ensembles were generated using AUC values.

### Plausibility check

We referred to the response curves of each algorithm to examine the plausibility of the most important explanatory variables.

## Prediction

### Prediction output

Prediction of probability of occurrence was expressed as 0.00 – 0.40 = low, 0.41 – 0.80 = moderate and 0.81 – 1.00 = high.

### Uncertainty quantification

An ensemble approach was used combining all individual SDM algorithms that had performed well (≥0.7 AUC). The ensembles performed better than all single model algorithms and results from ensembles supported those of the individual algorithms and were able to overcome any discrepancies in most influential explanatory variable between single SDMs, although usually the single SDMs presented one of the top two variables presented by the ensemble.

0.81 – 1.00 = high. Outside the recreational zone boundary is the General Use zone.

**Supplementary Material 3: Habitat type definitions**

Definitions obtained from (Bancroft and Sheridan 2000).

Table S3: Habitat type definitions used in the Ningaloo Marine Park ensemble. Definitions obtained from (Bancroft and Sheridan 2000)

|  |  |
| --- | --- |
| Habitat Type | Definition |
| Bare reef (intertidal) | The bare reef (intertidal) habitat is located in the intertidal zone (between the LAT and HAT) and may be offshore or contiguous to the coast. This habitat includes low cliffs (<5 m high), high cliffs (>5 m high), boulders (>25.6 cm particle size), or pavement of igneous (granite/basalt), metamorphic (gneiss/schists), or sedimentary (limestone/sandstone) substratum. The bare reef (intertidal) is typically unvegetated but may have algal turfs present. This habitat may contain a variety of mollusc species including oysters (eg. *Saccostrea spp*.), abalone (eg. *Haliotis spp*.) nerites (eg. *Nerita spp. Nodolittorina spp., Littoraria spp*.), chitons (eg. *Ischnochiton spp*.) and barnacles (eg. *Tetraclita porosa*). Rock crabs (*F. Grapsidae*) also inhabit this habitat. |
| Bare reef (subtidal) | Bare reef (subtidal) is located in subtidal areas with either sedimentary (eg. limestone, sandstone), igneous (eg. granite, granophyre) or metamorphic (eg. schist, gneiss) substratum, either as pavement or boulder (>25 cm) fields. This habitat typically includes areas covered by mobile sand veneers, and is located in deep water offshore or in subtidal lagoonal areas. Bare reef (subtidal) habitats are typically bare but may have vegetation (eg. *Thalassodendron spp*., *Padina spp*.), or have sparse cover sessile invertebrates such as sponges (eg. *Cymbastella spp., Carteriospongia spp*.), octocorals, soft corals and ascidians. |
| Coral reef (intertidal) | The coral reef (intertidal) habitat is located in the intertidal or shallow regions (<1 m LAT) on a limestone substrate. This habitat includes the reef crest, shallow reef fronts, reef flats and shallow back reef zones (see Veron, 2000). Live coral cover varies greatly and some areas have a high proportion of coral rubble. Macroalgae, sand, reef rubble or pavement also may be present. Hard corals (eg. *Acropora spp*.) and soft corals (eg. *Sinularia spp*.) are typical of the fauna present in these habitats. Parts of this habitat typically support a high diversity and abundance of fish and invertebrate fauna. |
| Coral reef (subtidal) | The coral reef (subtidal) habitat is located in the subtidal zone and often has high live coral cover with macroalgal turf and coralline algae covering areas of reef not occupied by living corals. Sand patches, bare pavement and rubble may also be present. This habitat is used to describe the upper seaward reef slope, sheltered back reef, deep lagoonal reef (Veron, 2000) and bommie clusters. Typically, areas of high coral cover are generally restricted to water depths of less than 15 m depth. Offshore, the coral reef (subtidal) habitats are dominated by the faster growing coral species such as Acropora (eg. *A. hyacinthus)* and Pocillopora (eg. *P. verrucosa*). This habitat typically supports a high diversity and abundance of fish and other coral reef fauna such as crabs (Families Xanthidae and Portunidae) and snapping shrimp (*Alpheus spp*.). |
| Macroalgae | The macroalgae (subtidal) habitat is subtidal areas with sedimentary, igneous or metamorphic substratum of low or high relief. This habitat is found in deep and shallow-waters and also may incorporate mobile sand patches, and scattered isolated hard and soft corals. This habitat generally is covered in large fleshy macroalgae (*eg. Sargassum spp., Cystophora spp., Ecklonia spp*.) or macroalgal turf (thallus height <100mm) comprised of red (eg*. Laurencia spp*.), green (eg. *Enteromorpha spp., Ulva spp. Caulerpa spp*.) and brown (eg*. Padina spp., Turbinaria spp*.) algae. A wide range of invertebrate life such as sponges, ascidians, gastropods, seastars, brittle stars, sea urchins and soft corals, are associated with this habitat. Crustaceans such as the western rock lobster (*Panulirus cygnus*), painted rock lobster (*P. vericolor*), ornate rock lobster (*P. ornatus*) and the southern rock lobster (*Janus edwardsii*) are often found in macroalgae (subtidal) habitats. |
| Mangroves | The mangrove habitat describes areas of mangrove forest greater than 0.05 ha and typically is located in the upper intertidal zone. The substratum of this habitat is typically comprised of mud and silt; however some mangrove species do occur on intertidal rocky shores. In Western Australia, the most common mangrove species are Rhizophora stylosa and Avicennia marina, the latter occurring as far south as Bunbury. Mangrove roots provide a substratum for many gastropods (eg. *Natica sp., Cerithium sp., Strombus spp*.) and other invertebrates, such as the mangrove crab (*Scylla serrata and Scylla olivacea)* and fiddler crab (Uca sp.)are often present. Mangroves are an important habitat for birds such as the mangrove whistler (*Pachycephala melanura*) and brahminy kite (*Haliastur indus*). |
| Mobile sand | The mobile sand (subtidal) habitat is defined as subtidal habitats that have predominantly white carbonate sands (0.1-2 mm grain size) as a substrate, which is constantly being moved by currents or wave action. However, the sand may overlay reef platform or have patches of other habitats present. Mobile sand (subtidal) habitats typically are bare, and may have seasonal vegetation or permanent patches of seagrass or macroalgae. Invertebrate infauna such as scallops (eg. *Pecten spp*.) seastars (eg. *Astropecten spp*.), and sea urchins (eg. *Brissus spp., Echinocardium spp*.), may also be present. |
| Pelagic | The pelagic habitat is defined as habitats with greater than 50 m depth. The pelagic habitat is dominated by the life in the water column, which include pelagic fish, pelagic invertebrates, zooplankton and phytoplankton. Contemporary acoustic mapping techniques have been able to discern hardness (soft and hard) and relief (smooth and rough) which may be used for local scale habitat mapping (Penrose & Siwabessy, 2001; Siwabessy et al., 1999). |
| Saltmarsh | The saltmarsh habitat describes areas of low relief located in the upper intertidal zone of low energy coastlines. The substratum consists of muddy or silty terrigenous sediment. Saltmarsh habitats often occur landward of mangroves, tidal creeks and estuaries, and typically supports vegetation such as the saltwater couch (*Sporobolus virginicus*) and blue-green algal mats (eg. *Microcoleus chthonoplastes, Oscillotoria sp., Phoridium sp*.), but can also occur as unvegetated coastal saline flats. In the tropics, burrowing crabs (*Uca sp*.), soldier crabs (*Mictyris sp*.) and *Cerinthium spp*. gastropods are conspicuous fauna in this habitat. In temperate areas, the glassswort *Sarcocornia quinqueflora* and *Sporobolus virginicus* are conspicuous flora in this habitat. |

**Supplementary Material 4: Preliminary ensemble species distribution modelling**

**Collinearity testing**

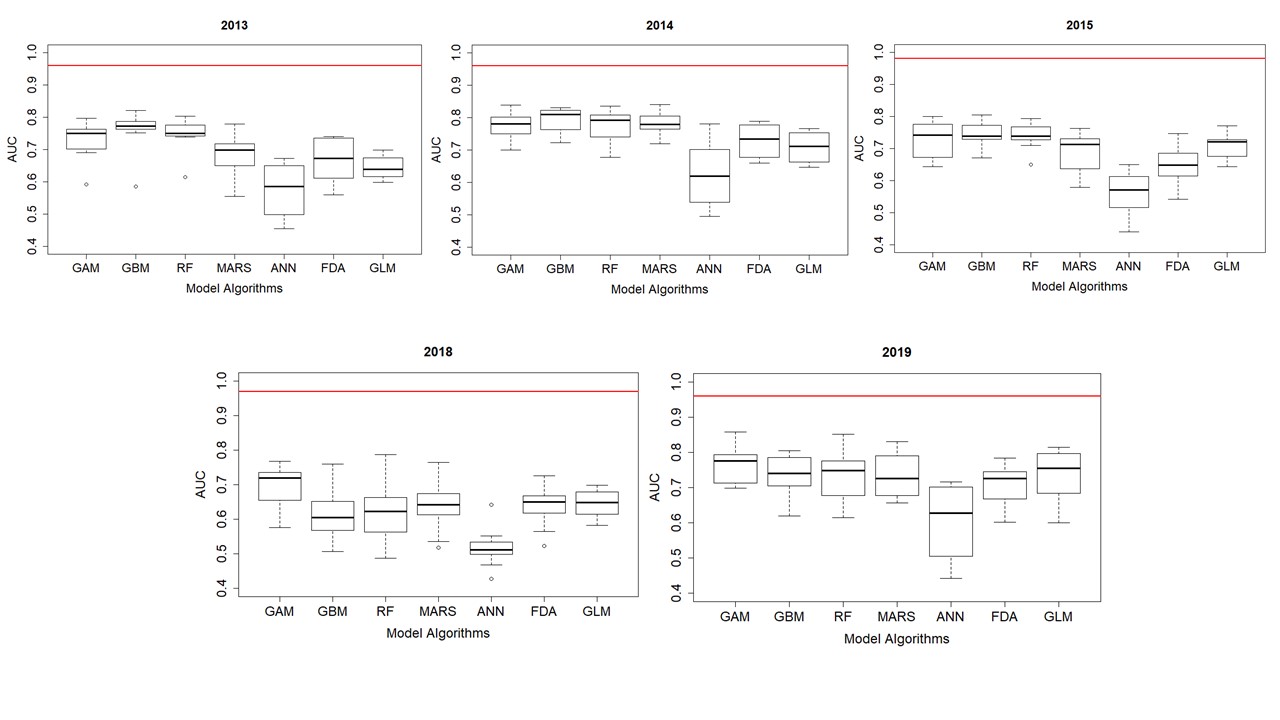
Correlation testing revealed collinearity between water visibility and water depth in the years 2013, 2018 and 2019 (2013; r = 0.73, 2018; r = 0.79, 2019; r = 0.74). In 2018, a collinear relationship was also found between SST and water visibility (r = 0.79). Testing also revealed a VIF of >3 for water depth in all years of study except 2014 and 2015 (2013; VIF = 5.27, 2018; VIF = 5.11, 2019; VIF = 4.64). In 2014 and 2015, a VIF >3 was found for water visibility (2014; VIF = 3.52 and 2015; VIF = 10.60) and additionally in 2015, distance to boat ramp (VIF = 3.63).

Due to the ecological and biological importance of water depth reported in other coastal bottlenose dolphin habitat use studies (Heithaus and Dill 2002, 2006, Zanardo et al. 2017, Passadore et al. 2018, Sprogis et al. 2018a, Vargas-Fonesca et al. 2018), water visibility was dropped from the yearly SDMs as an explanatory variable. Additional correlation testing (after the removal of water visibility), revealed no further collinearity issues except for SST and distance to boast ramp in 2013 (r = 0.72) nor VIFs >3, except for distance to boat ramp in 2015. As these relationships were not replicated in any of the other models, they were treated as outliers and both SST and distance to boat ramp were retained in the set of potential explanatory variables.

***Results of yearly ensemble models of IP bottlenose dolphin distribution***

**Model performance**

Most of the single yearly SDMs performed well (AUC > than 0.7), with the exception of several runs primarily from the ANN and GLM algorithms (AUC < 0.7) (Figure S4.1). All poor performing runs were excluded from the final ensemble models. Ensemble models outperformed all single SDMs with AUC values above 0.9 for all years indicating excellent model performance (2013; range = 0.46-0.82, median = 0.70, ensemble AUC = 0.96, 2014; range = 0.50 to 0.84, median = 0.76, ensemble AUC = 0.96, 2015: range = 0.44 to 0.81, median = 0.72, ensemble AUC = 0.98, 2018; range = 0.43 to 0.79, median = 0.63, ensemble AUC = 0.97 and 2019; range = 0.44 to 0.86, median = 0.73 and ensemble AUC = 0.96 (Figure S4.1).



**Figure S4.1:** Performance of yearly species distribution models of Indo-Pacific bottlenose dolphins at the North West Cape, Western Australia. Boxplots for the AUC (area under the curve of the receiver-operating characteristic) of the 10-cross validation runs of each modelling algorithm (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model). The red line shows the AUC of the ensemble model for each year. Values of AUC ≥0.7 indicates that the model predictive performance is moderate to excellent.

**Variable importance and response curves**

Most yearly SDMs, except for 2015, identified distance to coast as one of the most important variables influencing IP bottlenose dolphin distribution in the study area (Table S4.1). Response curves across these different yearly models indicated that the probability of Indo-Pacific bottlenose dolphin occurrence was highest up to 2000 m from the coast. Other important variables influencing dolphin occurrence varied by year and included water depth (2013), marine park zone (2014 and 2015); seabed slope (2015), SST (2018) and distance to boat ramp (2019). Response curves for these different variables indicated that the probability of dolphin occurrence also tended to be higher in : 1) water depths of 8 – 13 m in 2013, 2) in Sanctuary zones and Naval Waters in 2014; 3) in areas where the seabed slope was steeper in 2015, in water temperatures of ~ 22 – 23.5°C in 2018; and in waters up to 10 000 m from the nearest boat ramp in 2019. For response curves see Figures S4.2 – S4.6.

Table S4.1: Importance of predictor variables used in the yearly species distribution models (SDMs) of Indo-Pacific bottlenose dolphins at the North West Cape, Western Australia. Seven SDM algorithms were used: artificial neural network (ANN), flexible discriminant analysis (FDA), generalised additive model (GAM), generalised boosted model (GBM), generalised linear model (GLM), multivariate adaptive regression splines (MARS) and random forest (RF). Variable importance is presented as the mean value over the 10 runs of each algorithm, the mean of means amongst them and as the ensemble value calculated using only the runs that met the AUC (area under the curve of the receiver operating characteristic) evaluation criteria of ≥0.7. The number of runs of each algorithm that was included in the ensemble is indicated in subscript. Variables of greatest influence are highlighted in bold.

| **Year** | **Model** | **Explanatory variables** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Distance to boat ramp** | **Distance to coast** | **MP zone** | **Seabed slope** | **SST** | **Water depth** |
| 2013 | ANN0 | **0.80** | **0.80** | 0.04 | 0.00 | 0.02 | 0.06 |
| FDA3 | 0.03 | 0.20 | 0.00 | 0.17 | 0.26 | **0.64** |
| GAM8 | 0.26 | **0.57** | 0.26 | 0.04 | 0.33 | 0.19 |
| GBM9 | 0.11 | **0.40** | 0.01 | 0.33 | 0.25 | 0.12 |
| GLM0 | 0.19 | **0.59** | 0.17 | 0.08 | 0.00 | 0.02 |
| MARS5 | 0.14 | **0.52** | 0.00 | 0.10 | 0.26 | 0.40 |
| RF9 | 0.16 | **0.27** | 0.03 | 0.26 | 0.24 | 0.14 |
| Mean of means | 0.24 | **0.48** | 0.07 | 0.14 | 0.19 | **0.22** |
| Ensemble | 0.16 | **0.42** | 0.07 | 0.18 | **0.30** | 0.25 |
| 2014 | ANN3 | 0.83 | **0.86** | 0.09 | 0.00 | 0.01 | 0.14 |
| FDA6 | 0.04 | 0.32 | 0.40 | 0.23 | **0.49** | 0.06 |
| GAM10 | 0.04 | 0.36 | **0.44** | 0.25 | 0.21 | 0.05 |
| GBM10 | 0.09 | **0.39** | 0.17 | 0.15 | 0.37 | 0.08 |
| GLM6 | 0.02 | 0.03 | **0.71** | 0.02 | 0.32 | 0.03 |
| MARS10 | 0.00 | **0.52** | 0.26 | 0.28 | 0.25 | 0.05 |
| RF9 | 0.15 | **0.27** | 0.13 | 0.13 | **0.27** | 0.13 |
| Mean of means | 0.17 | **0.40** | **0.31** | 0.15 | 0.27 | 0.17 |
| Ensemble | 0.14 | **0.39** | **0.36** | 0.14 | 0.27 | 0.06 |
| 2015 | ANN0 | **0.85** | 0.80 | 0.07 | 0.00 | 0.01 | 0.12 |
| FDA2 | 0.05 | 0.09 | **0.63** | 0.31 | 0.13 | 0.06 |
| GAM6 | 0.17 | 0.26 | **0.39** | 0.34 | 0.29 | 0.18 |
| GBM8 | 0.20 | 0.12 | 0.12 | **0.53** | 0.43 | 0.15 |
| GLM7 | 0.06 | 0.08 | **0.48** | 0.43 | 0.35 | 0.06 |
| MARS6 | 0.09 | 0.37 | 0.12 | **0.38** | 0.22 | 0.22 |
| RF9 | 0.22 | 0.11 | 0.11 | **0.36** | 0.33 | 0.13 |
| Mean of means | 0.24 | 0.26 | **0.27** | **0.34** | 0.25 | 0.13 |
| Ensemble | 0.10 | 0.19 | **0.36** | **0.31** | 0.27 | 0.13 |
| 2018 | ANN0 | **0.81** | 0.80 | 0.00 | 0.00 | 0.02 | 0.07 |
| FDA1 | 0.02 | 0.26 | 0.12 | 0.15 | **0.55** | 0.34 |
| GAM6 | 0.08 | 0.42 | 0.36 | 0.08 | **0.43** | 0.12 |
| GBM2 | 0.11 | 0.09 | 0.10 | 0.38 | **0.58** | 0.17 |
| GLM0 | 0.00 | 0.08 | 0.10 | 0.17 | **0.70** | 0.10 |
| MARS2 | 0.07 | 0.43 | 0.03 | 0.13 | **0.52** | 0.22 |
| RF2 | 0.12 | 0.11 | 0.06 | 0.24 | **0.36** | 0.16 |
| Mean of means | 0.17 | **0.31** | 0.11 | 0.16 | **0.45** | 0.17 |
| Ensemble | 0.10 | **0.18** | 0.10 | 0.16 | **0.60** | 0.06 |
| 2019 | ANN4 | **0.85** | 0.80 | 0.09 | 0.08 | 0.03 | 0.37 |
| FDA6 | 0.25 | 0.22 | 0.00 | **0.44** | 0.41 | 0.07 |
| GAM9 | **0.52** | **0.52** | 0.26 | 0.17 | 0.06 | 0.20 |
| GBM8 | 0.32 | 0.23 | 0.01 | **0.47** | 0.28 | 0.10 |
| GLM7 | **0.46** | 0.34 | 0.00 | 0.38 | 0.06 | 0.00 |
| MARS6 | 0.35 | **0.47** | 0.00 | 0.31 | 0.12 | 0.07 |
| RF6 | 0.28 | 0.19 | 0.04 | **0.31** | 0.23 | 0.12 |
| Mean of means | **0.43** | **0.40** | 0.06 | 0.31 | 0.17 | 0.13 |
| Ensemble | **0.42** | **0.39** | 0.06 | 0.30 | 0.19 | 0.13 |

**2013 response curves**

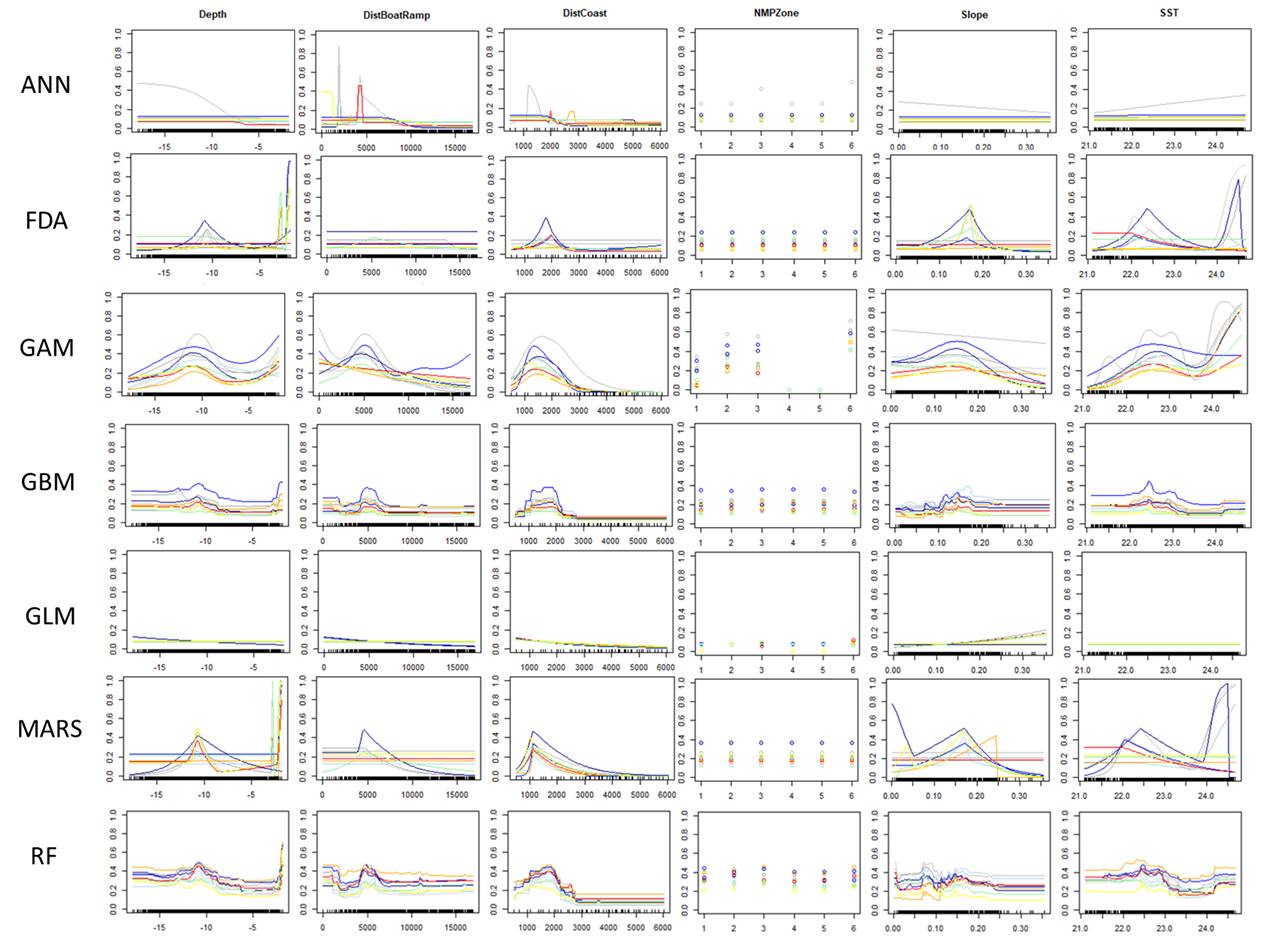


Figure S4.2: Response curves of the seven modelling algorithms used in the ensemble to model 2013 Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**2014 response curves**

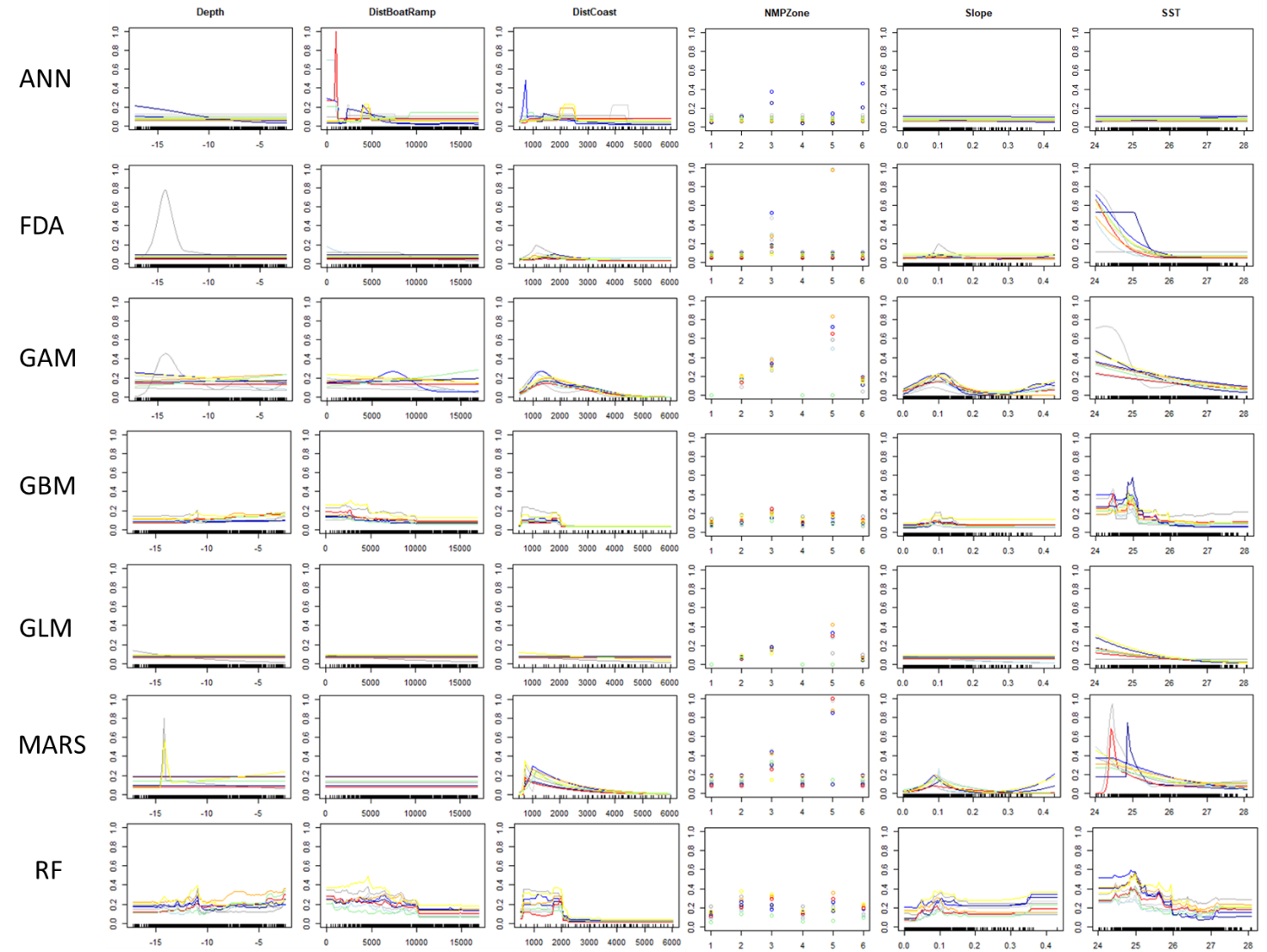


Figure S4.3: Response curves of the seven modelling algorithms used in the ensemble to model 2014 Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

2015 Response curves

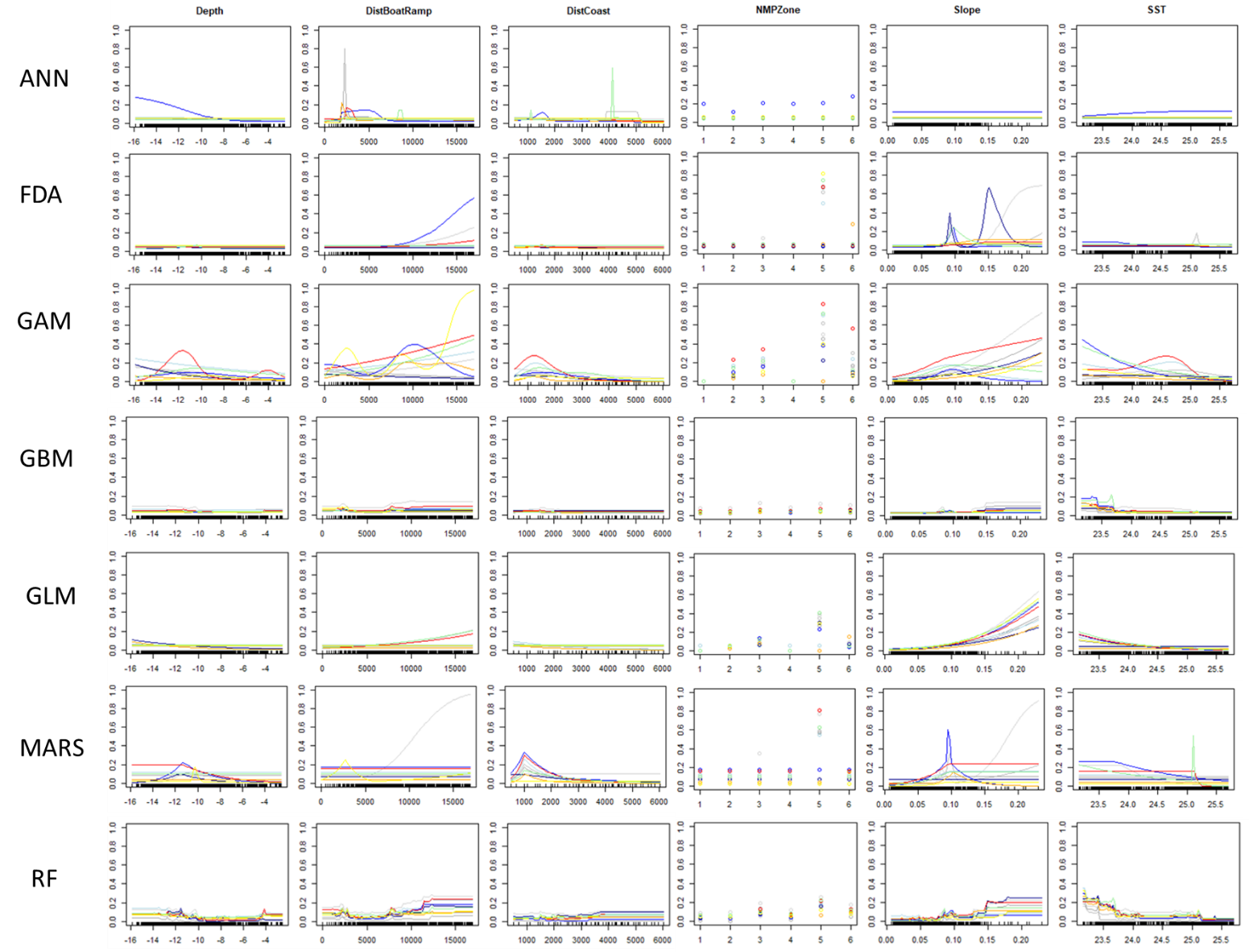


Figure S4.4: Response curves of the seven modelling algorithms used in the ensemble to model overall Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**2018 Response curves**

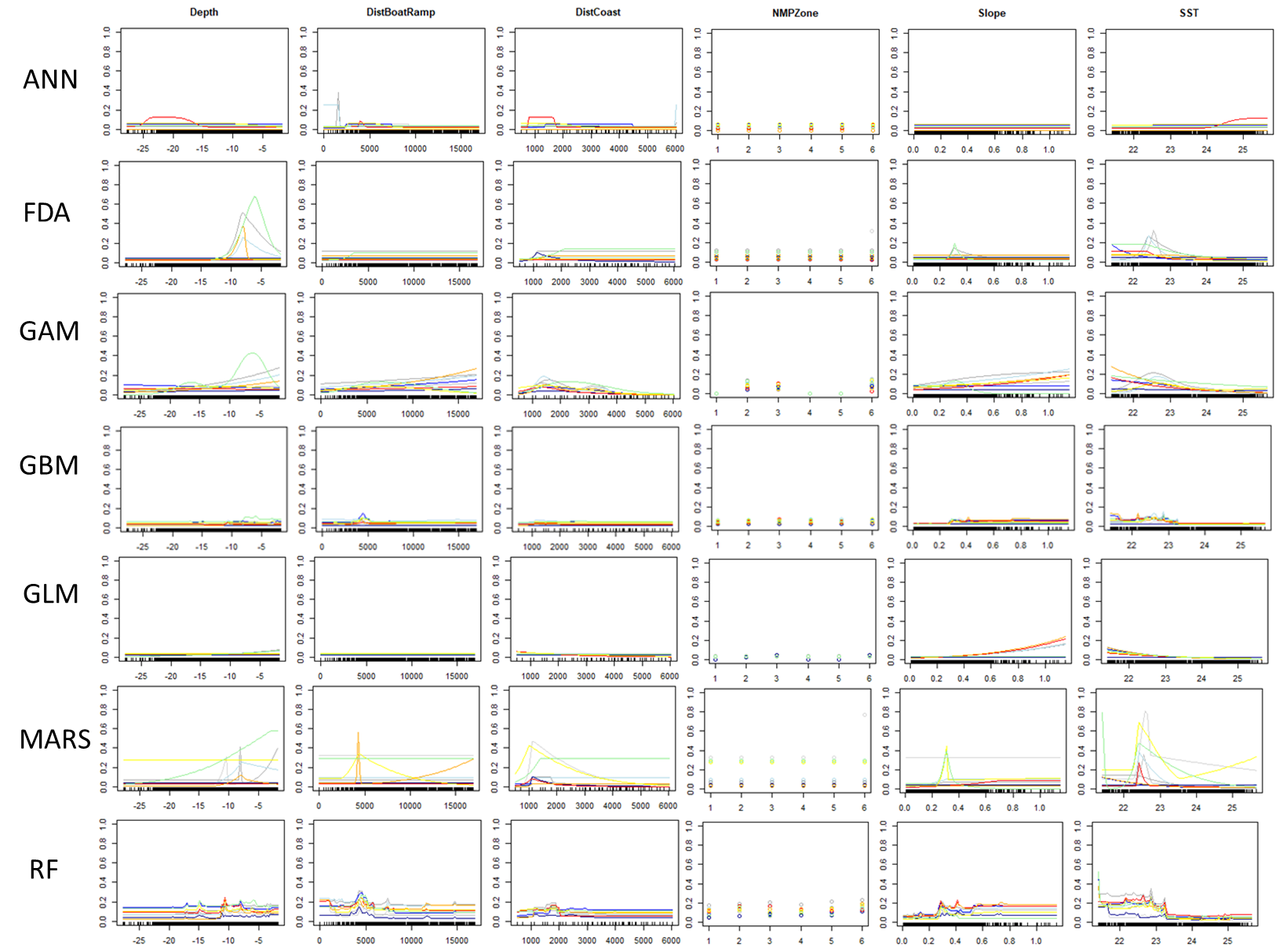


Figure S4.5: Response curves of the seven modelling algorithms used in the ensemble to model overall Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**2019 Response curves**

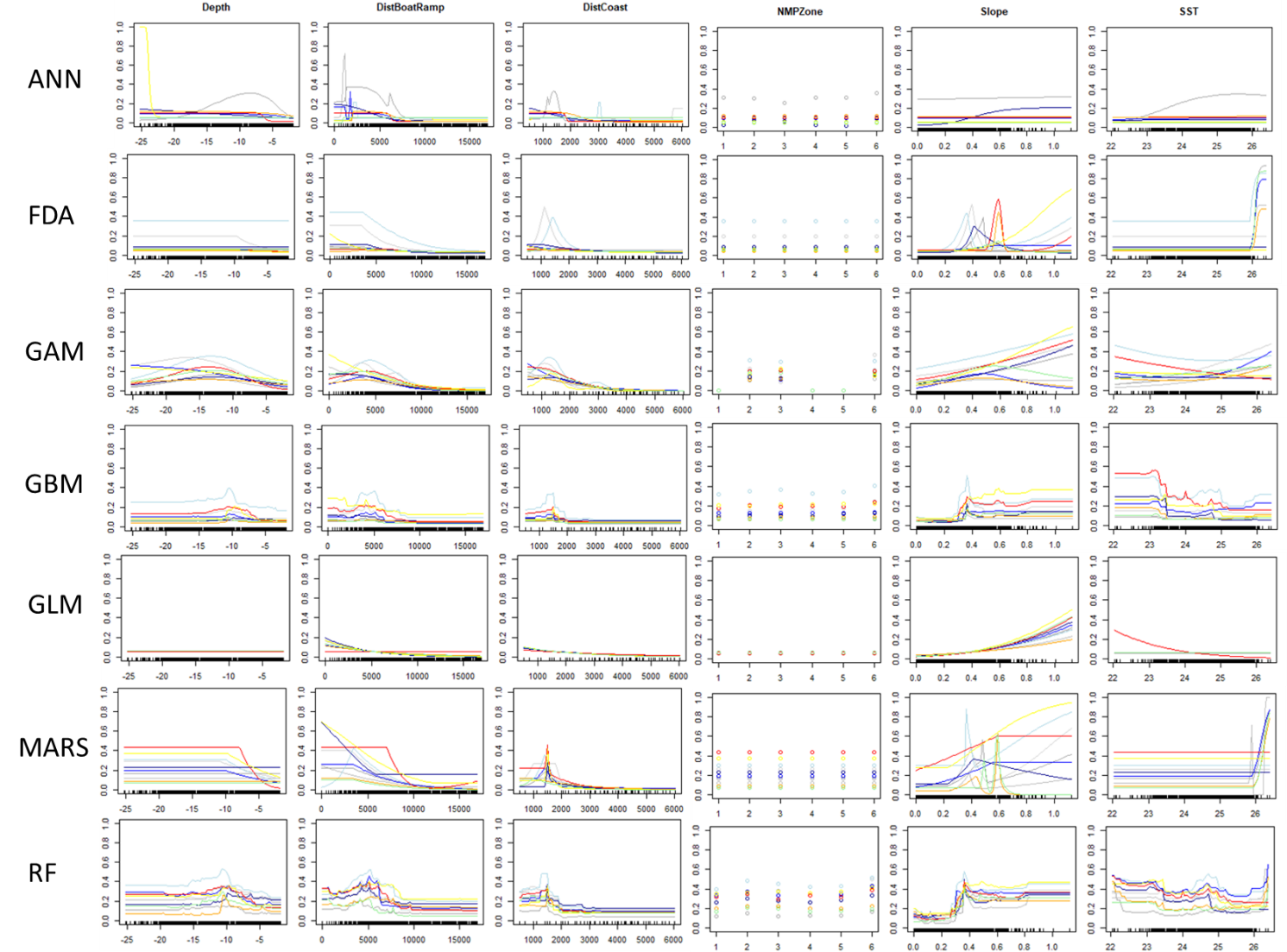


Figure S4.6: Response curves of the seven modelling algorithms used in the ensemble to model overall Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**Distribution and space use maps**

In all five years, the coastal lagoons on the west side, the tip and eastern side of the Cape had a moderate probability of occurrence (Figure S4.7). Typically, the highest concentration of moderate to high probability cells were located at the tip and in the waters of Exmouth Gulf to the east of the Cape (Figure S4.7).

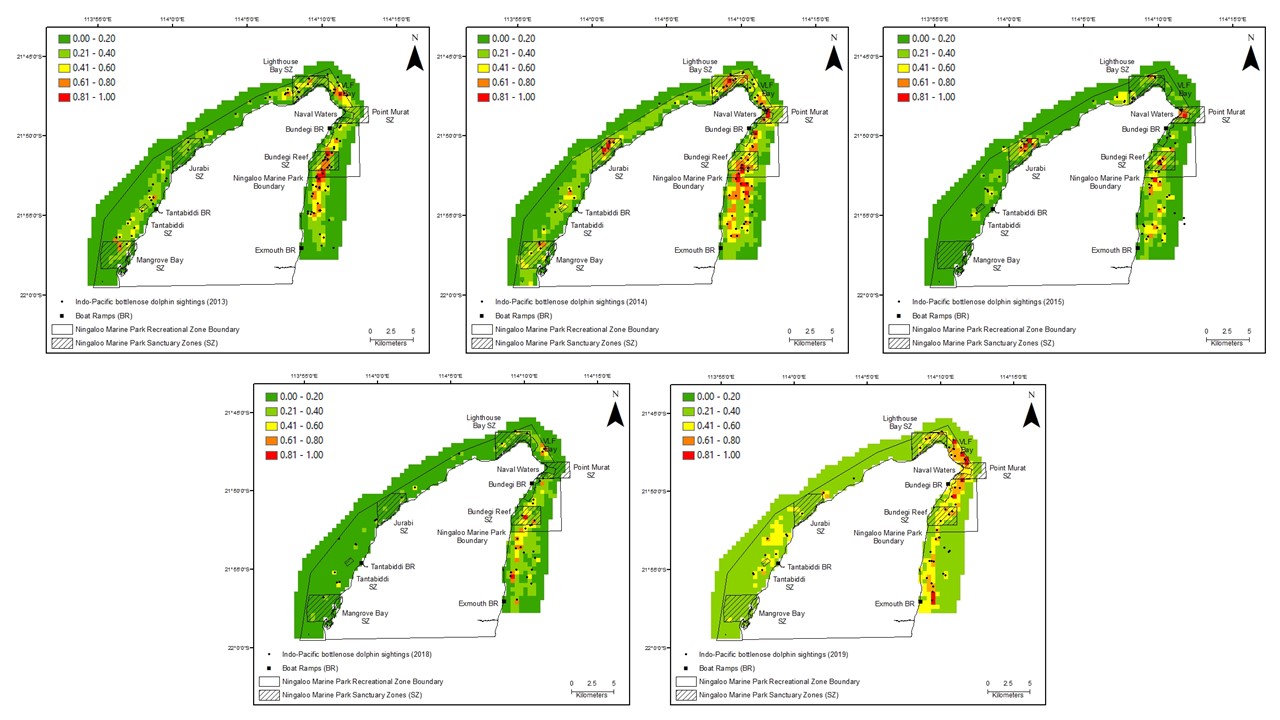


Figure S4.7: Ensemble models of Indo-Pacific bottlenose dolphin probability of occurrence at the North West Cape for each survey year (2013, 2014, 2015, 2018 and 2019). Colours as shown in the legend indicate the probability of occurrence 0.00 – 0.40 = low, 0.41 – 0.80 = moderate and 0.81 – 1.00 = high. Outside the recreational zone boundary is the General Use zone.

***Results of Ningaloo Marine Park ensemble species distribution models***

After establishing that Indo-Pacific bottlenose dolphin distribution was not influenced by temporal changes, we examined the Ningaloo Marine Park portion of the study area to determine if benthic habitat type (only available for this section of the study area), was influencing space use and distribution.

**Model performance**

Most of the single algorithms performed well (AUC > than 0.7, range = 0.43-0.86 and median = 0.73), with the exception of several runs from the ANN algorithm (AUC < 0.7) (Figure S4.8). All poor performing runs were excluded from the final ensemble models. The ensemble model outperformed all single SDMs with an AUC value of 0.95 indicating excellent model performance (Figure S4.8).

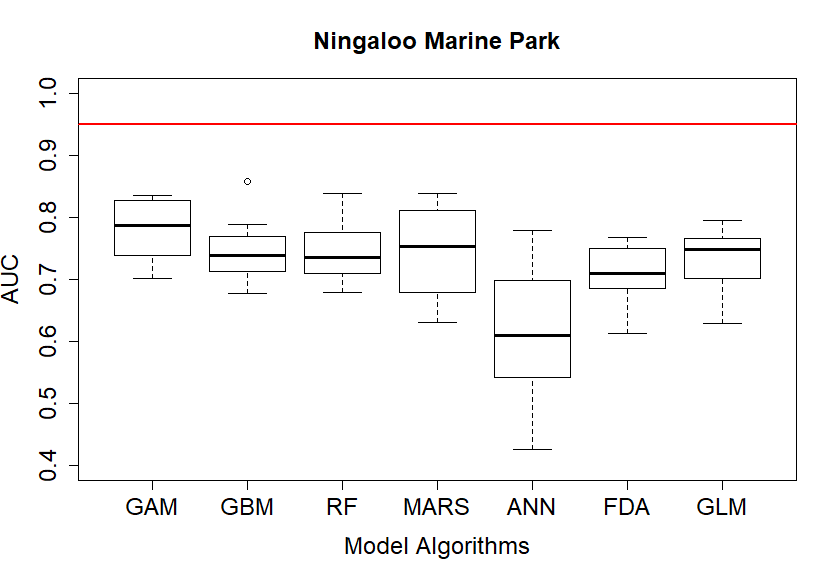


Figure S4.8: Performance of the Ningaloo Marine Park species distribution models of Indo-Pacific bottlenose dolphins at the North West Cape, Western Australia. Boxplots for the AUC (area under the curve of the receiver-operating characteristic) of the 10-cross validation runs of each modelling algorithm (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model). The red line shows the AUC of the ensemble model for each year. Values of AUC ≥0.7 indicates that the model predictive performance is moderate to excellent.

**Variable importance and response curves**

Distance to boat ramp (0.47) and distance to coast (0.38) were the most influential variables explaining IP bottlenose dolphin distribution within the Ningaloo Marine Park. Habitat type ranked number five out of the seven explanatory variables and on that basis, it was concluded that habitat type was not strongly influencing dolphin distribution (Table S4.4). Response curves indicated that within the NMP, IP bottlenose dolphins were most likely to occur up to 5000 m from the nearest boat ramp and 1000 – 2000 m from the coast (Figure S4.9).

Table S4.4: Importance of predictor variables used in the Ningaloo Marine Park species distribution models (SDMs) of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) at the North West Cape, Western Australia. Eight SDM algorithms were used: artificial neural network (ANN), flexible discriminant analysis (FDA), generalised additive model (GAM), generalised boosted model (GBM), generalised linear model (GLM), maximum entropy (MaxEnt), multivariate adaptive regression splines (MARS) and random forest (RF). Variable importance is presented as the mean value over the 10 runs of each algorithm, the mean of means amongst them and as the ensemble value calculated using only the runs that met the AUC (area under the curve of the receiver operating characteristic) evaluation criteria of ≥0.7. The number of runs of each algorithm that was included in the ensemble is indicated in subscript. Variables of greatest influence are highlighted in bold.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Explanatory variables** | | | | | | |
| **Distance to boat ramp** | **Distance to coast** | **MP zone** | **Seabed slope** | **SST** | **Water depth** | **Habitat** |
| ANN2 | **0.88** | 0.79 | 0.00 | 0.00 | 0.01 | 0.23 | 0.04 |
| FDA7 | 0.27 | 0.21 | 0.39 | **0.40** | 0.01 | 0.08 | 0.07 |
| GAM10 | 0.34 | **0.52** | 0.36 | 0.38 | 0.04 | 0.13 | 0.39 |
| GBM8 | **0.53** | 0.14 | 0.02 | 0.31 | 0.09 | 0.03 | 0.16 |
| GLM8 | **0.40** | 0.11 | 0.41 | 0.19 | 0.03 | 0.02 | 0.07 |
| MARS6 | 0.42 | **0.54** | 0.06 | 0.37 | 0.03 | 0.05 | 0.01 |
| RF8 | **0.29** | 0.13 | 0.04 | 0.20 | 0.10 | 0.07 | 0.11 |
| Mean of means | **0.45** | **0.35** | 0.18 | 0.27 | 0.05 | 0.09 | 0.12 |
| Ensemble | **0.47** | **0.38** | 0.16 | 0.26 | 0.04 | 0.10 | 0.12 |

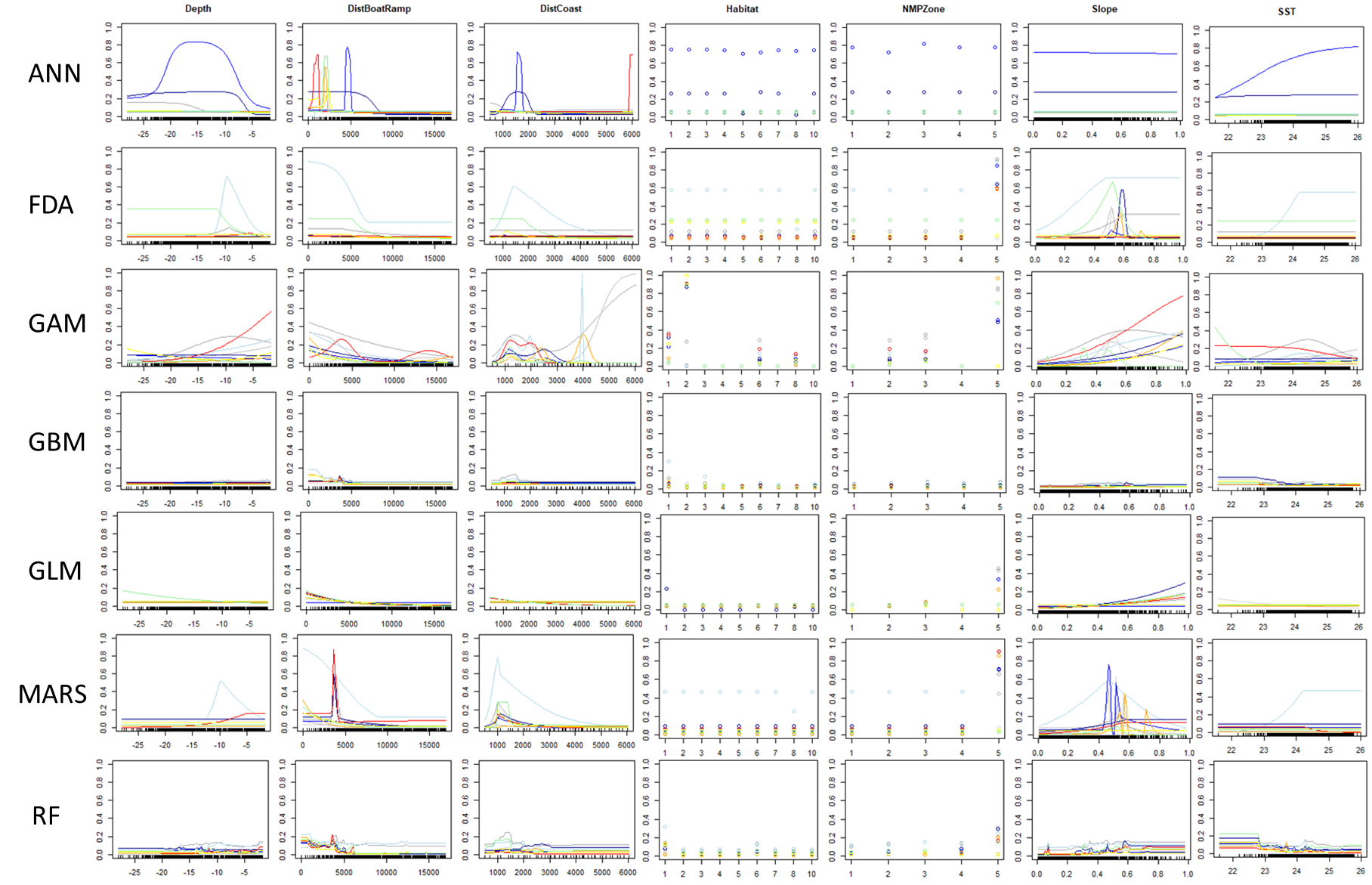


Figure S4.9: Response curves of the seven modelling algorithms used in the ensemble to model Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences within the Ningaloo Marine Park at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**Distribution and space use maps**

Within the Ningaloo Marine Park, moderate to high probability of occurrence cells were located in coastal lagoons to the west of the Cape, at the tip and waters to the east of the Cape, at the top of the Exmouth Gulf (Figure S4.10). One high probability of occurrence cell was modelled, located in the Naval waters.

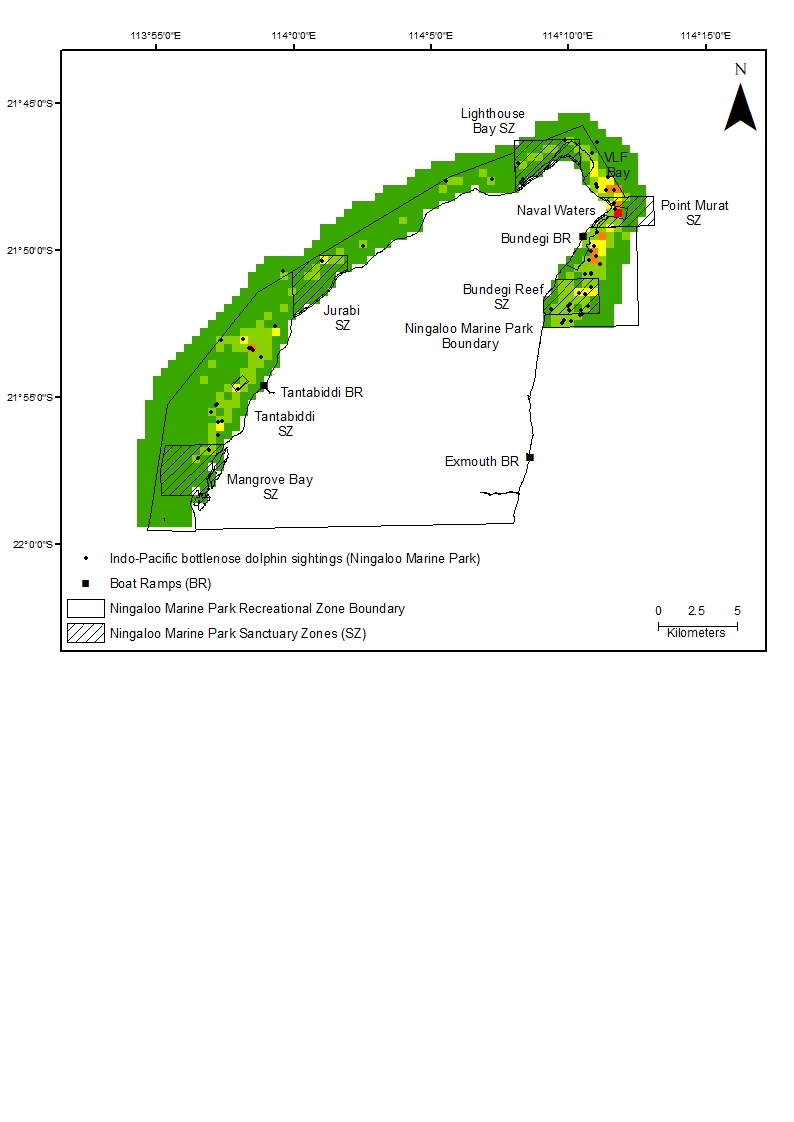


Figure S4.10: Ensemble models of Indo-Pacific bottlenose dolphin probability of occurrence at the North West Cape for the Ningaloo Marine Park ensemble. Colours as shown in the legend indicate the probability of occurrence 0.00 – 0.40 = low, 0.41 – 0.80 = moderate and 0.81 – 1.00 = high.

**Supplementary Material 5: Biomod2 algorithm default settings**

Artificial Neural Network (ANN) = NbCV = 5, size = NULL,decay = NULL, rang = 0.1, maxit = 200

Flexible Discriminant Analysis (FDA) = method = 'mars', add\_args = NULL

Generalised Additive Modelling (GAM) = algo = 'GAM\_mgcv',type = 's\_smoother', k = -1, interaction.level = 0, myFormula = NULL, family = binomial(link = 'logit'), method = 'GCV.Cp',optimizer = c('outer','newton'), select = FALSE, knots = NULL,paraPen = NULL, control = list(nthreads = 1, irls.reg = 0, epsilon = 1e-07, maxit = 200, trace = FALSE, mgcv.tol = 1e-07, mgcv.half = 15, rank.tol = 1.49011611938477e-08, nlm = list(ndigit=7, gradtol=1e-06, stepmax=2, steptol=1e-04, iterlim=200, check.analyticals=0), optim = list(factr=1e+07), newton = list(conv.tol=1e-06, maxNstep=5, maxSstep=2, maxHalf=30, use.svd=0), outerPIsteps = 0, idLinksBases = TRUE, scalePenalty = TRUE, efs.lspmax = 15, efs.tol = 0.1, keepData = FALSE, scale.est = fletcher, edge.correct = FALSE)

Generalised Boosted Modelling (GBM) = distribution = 'bernoulli',n.trees = 2500,interaction.depth = 7,n.minobsinnode = 5,shrinkage = 0.001,bag.fraction = 0.5,train.fraction = 1,cv.folds = 3,keep.data = FALSE,verbose = FALSE, perf.method = 'cv',n.cores = 1)

Generalised Linear Modelling (GLM) = type = 'quadratic',interaction.level = 0, myFormula = NULL,test = 'AIC', family = binomial(link = 'logit'), mustart = 0.5, control = glm.control(epsilon = 1e-08, maxit = 50, trace = FALSE) )

Multivariate Adaptive Regression Splines (MARS) = type = 'simple', interaction.level = 0, myFormula = NULL, nk = NULL, penalty = 2, thresh = 0.001, nprune = NULL, pmethod = 'backward')

Random Forest (RF) = do.classif = TRUE, ntree = 500, mtry = 'default', nodesize = 5, maxnodes = NULL)

**Supplementary Material 6: Response curves – overall**

A picture containing line chart

Description automatically generated

**Figure S6:** Response curves of the seven modelling algorithms used in the ensemble to model overall Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**Supplementary Material 7: Response curves – seasons**

**7.1 Autumn**

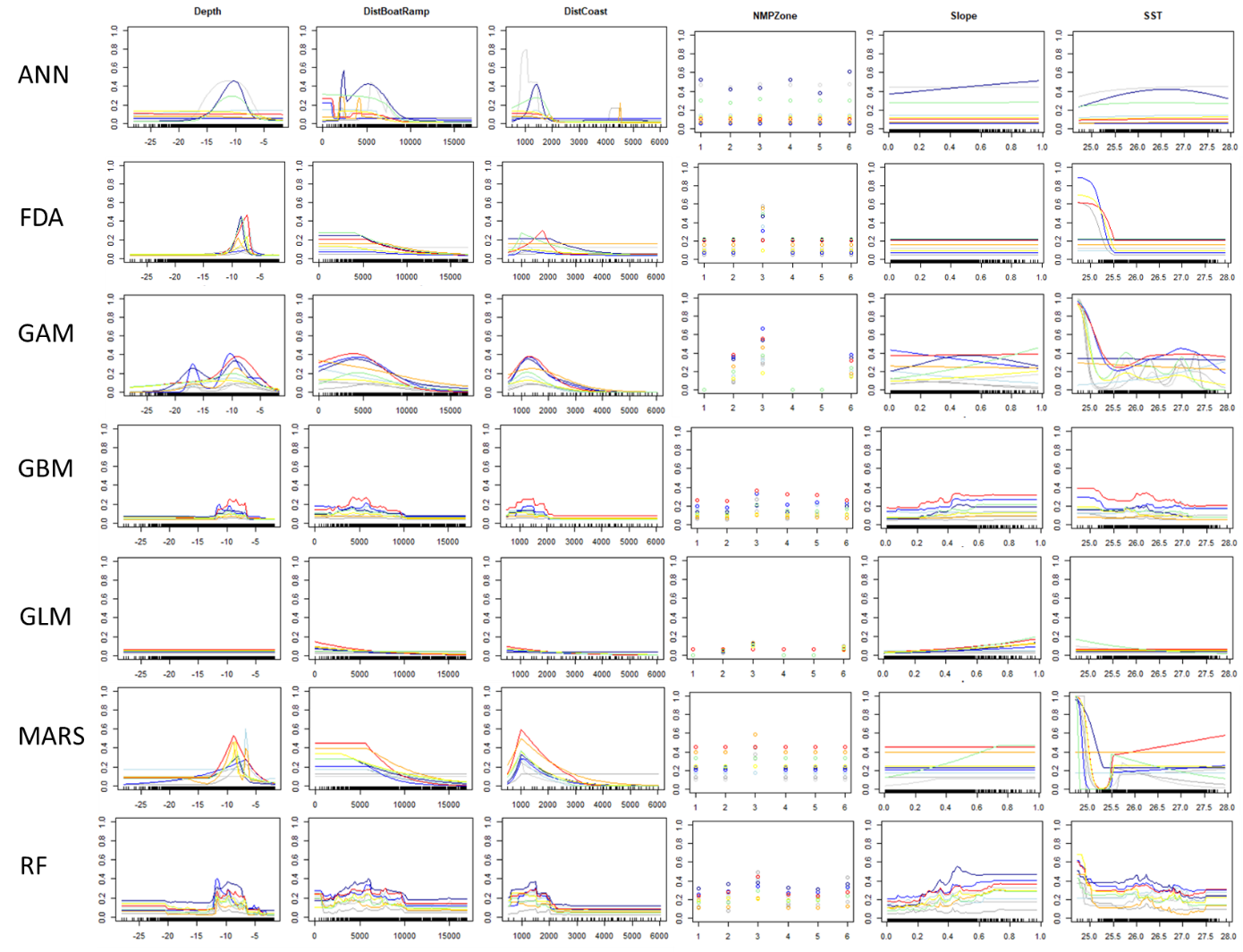


Figure S7.1: Response curves of the seven modelling algorithms used in the ensemble to model autumn Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**7.2 Winter**

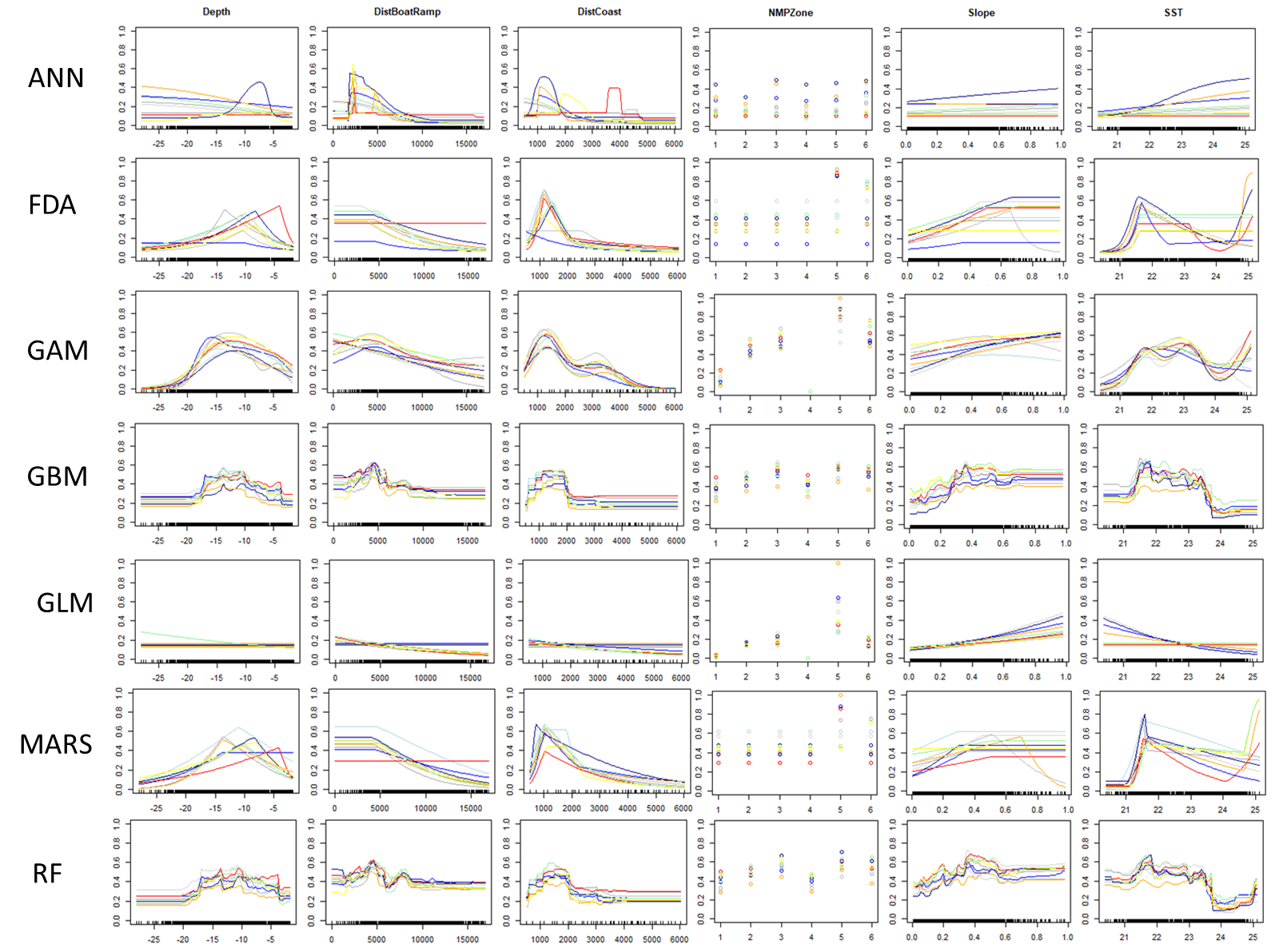


Figure S7.2: Response curves of the seven modelling algorithms used in the ensemble to model winter Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).

**7.3 Spring**

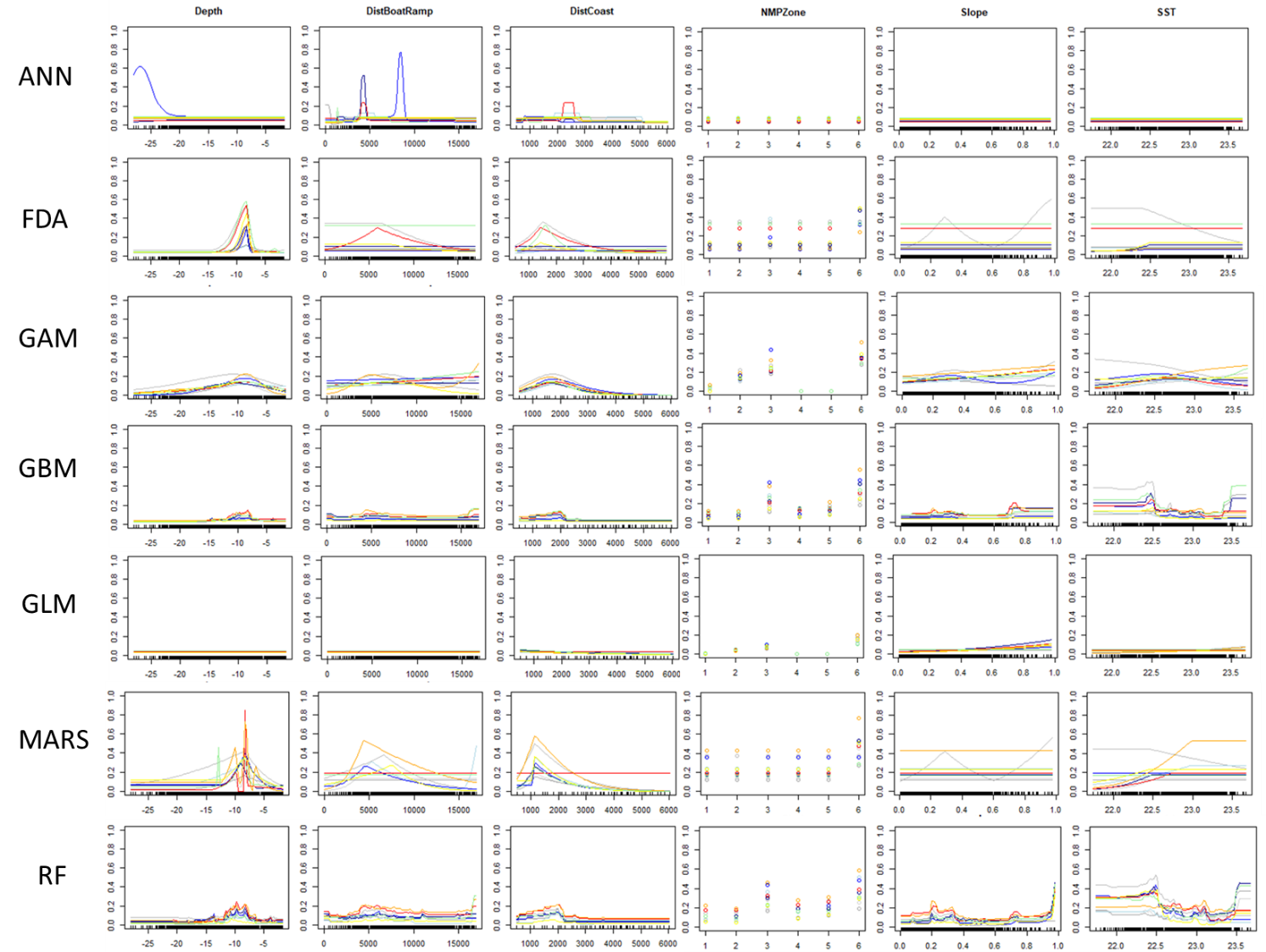


Figure S7.3: Response curves of the seven modelling algorithms used in the ensemble to model spring Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) distribution and habitat preferences at the North West Cape, Western Australia. (GAM: generalised additive model, GBM: generalised boosted model, RF: random forest, MARS; multivariate adaptive regression splines, ANN: artificial neural network, FDA: flexible discriminant analysis and GLM: generalised linear model).