



Centre for Environment  
Fisheries & Aquaculture  
Science



*World Class Science for the Marine and Freshwater Environment*

# Predicting the distribution of seahorse habitat in coastal UK waters

**Cefas reference: C8528**

---

Author(s): Joanna K. Bluemel, Louise Brown, Christopher Taylor and Anna L. Downie

Issue Date: 25/09/2023

## Cefas Document Control

Submitted to:	Natural England
Date submitted:	25/09/2023
Project Manager: Principle Investigator:	Alexander Bird Joanna Bluemel
Report compiled by:	Joanna K. Bluemel, Louise Brown, Chris Taylor, Anna L. Downie
Quality control by:	Elena Couce
Approved by and date:	Elena Couce, 02/06/2023 v1.0; 15/06/2023 v2.0
Version:	V2.3
Recommended citation for this report:	Bluemel, J.K., Brown L., Taylor C. and Downie A.L. (2023). Predicting the distribution of seahorse habitat in coastal UK waters. Cefas Project Report for Natural England, 35 pp.

Version Control History			
Version	Author	Date	Comment
0.1	All	18/01/2023	First draft
1.0	Elena Couce	02/06/2023	Internal QA
1.1	Joanna Bluemel	08/06/2023	Revise comments
2.0	Elena Couce	15/06/2023	Internal QA after revision
2.1	Joanna Bluemel	15/06/2023	Final revision
2.2	Joanna Bluemel	03/07/2023	Address customer comments
2.3	Joanna Bluemel	25/09/2023	Final amendments

## Executive Summary

Effective conservation, marine spatial planning, and sustainable management of anthropogenic activities in our waters requires accurate knowledge on all ecosystem components, including the abundance and distribution of vulnerable Protected, Endangered and Threatened species (PETs) and their associated habitats (Bluemel et al. 2020, Reiss et al. 2021, Weinert et al. 2021, Pierri et al. 2022, Bertelli et al. 2023). In the marine environment it can be difficult to observe and monitor species directly, resulting in a patchy understanding of their distribution (Reiss et al. 2021). This is especially true of very rare and cryptic species, which can be extremely data lacking, and tools such as Species Distribution Models (SDMs) can provide useful information in data-limited settings. Seahorses of the *Hippocampus* genus are renowned for being difficult to study due to their cryptic morphology and behaviour, and there is evidence for widespread declines in some natural populations (Curtis & Vincent 2005, Caldwell & Vincent 2012), highlighting the need for improved assessment, monitoring and management of these species (Woodall et al 2018). The aim of this study is to increase the accuracy and extent of environmental predictor variables in coastal and estuarine environments required to accurately predict coastal *Hippocampus* species habitat preferences.

To provide information on the fine scale environmental variability in shallow water environments, we combined satellite-derived gridded environmental data (offshore) with field-collected water samples from the Water Framework Directive (WFD) statutory monitoring point data (nearshore) using sophisticated kriging methods in ArcGIS software (v10.5). The fine-scale modelled environmental data layers (including distance to seagrass habitat, bathymetry, minimum winter and maximum summer sea surface temperature (SST) and chlorophyll a concentrations) were used to predict the distribution of *Hippocampus* spp. habitat using presence-only species records provided by The Seahorse Trust, Natural England, and additional open-source databases. We tested the commonly used modelling algorithm, maximum entropy (MAXENT), which predicts the probability distribution of a given species using presences and background (pseudo-absence) data (Phillips et al. 2006).

The environmental data generated within this project have enabled the development of finer spatially resolved predictions of the PET *Hippocampus* seahorse species habitats around the British Isles. However, further work is required to deal with sampling bias and spatial autocorrelation that exists in the occurrence data to increase confidence in model predictions, particularly in more northerly, less frequently sampled localities. Further work needed to robustly test the model outcomes with model validation tools appropriate for presence-only input data, test different modelling algorithms to identify consistent predictions, and test different habitat suitability thresholds that are tailored to the intended use of the predicted distributions.

Overall, this study has made important developments in the understanding of these elusive, data deficient species. The habitat partitioning observed between the two species is important to understand and develop the most appropriate conservation and management strategies which will benefit both species. The information provided by this project can be used to support marine spatial planning to reduce the wider impact of anthropogenic activities and enable better decision making to protect these sensitive species and their habitats. Further work to produce finer resolution maps to capture micro-habitat requirements within highly suitable bays and estuaries will provide evidence to better inform conservation strategies at a local scale.

# Table of Contents

1. Introduction .....	5
1.1 Aims and objectives .....	6
2. Methods.....	8
2.1 Environmental data.....	8
2.2 Species data .....	11
2.3 Species distribution modelling.....	13
3. Results.....	14
3.1 Environmental data.....	14
3.2 Model performance .....	16
3.3 Variable importance.....	16
3.4 Habitat distribution.....	17
4. Discussion.....	22
4.1 Environmental variables .....	22
4.2 Modelling limitations and sampling biases.....	22
4.3 Seahorse habitat preferences.....	23
5. Conclusion.....	25
6. Acknowledgements.....	26
7. References .....	26
Appendix 1: Supplementary material .....	30
Appendix 2: Seagrass data sources.....	35

## Tables

Table 1. Table of environmental variable details and source locations. See Table A2.1 for seagrass data sources. See 2.1 for details of the seasonal resolution of interpolated sea surface temperature and chlorophyll a concentration variables..... 10

Table 2. Variable importance (% contribution) for predicting presence of *Hippocampus* spp., *H. hippocampus* and *H. guttulatus* habitats. SST = sea surface temperature..... 17

## Figures

Figure 1. The example shows the marine area around the Isle of Wight, Dorset, to illustrate the available data extents and interpolation process for the chlorophyll a concentration variable. Top: gridded satellite-derived data extent ( $0.0417^{\circ} \times 0.0417^{\circ}$  resolution), WFD field sampling locations (Table 1), the extent of the gridded satellite-derived data used for interpolation. Bottom: the resulting extent of the interpolated data ( $0.0042^{\circ} \times 0.0042^{\circ}$  resolution). ..... 9

Figure 2. Occurrence distribution of A) the *Hippocampus* genus, B) *H. hippocampus* and C) *H. guttulatus* from years 2000 to present, that were used for modelling the species habitat preferences. Occurrences are aggregated for presentation to a  $0.2 \times 0.2$  degrees resolution ( $\sim 10 \times 10$  km) following rules for the presentation of sensitive species data..... 12

Figure 3. Environmental predictor variables used for modelling the distribution of seahorses. A) summer maximum sea surface temperature (SST), B) winter minimum SST, C) summer maximum chlorophyll concentration, D) winter minimum chlorophyll concentration, E) bathymetry, F) distance from seagrass. Locations of Environment Agency Water Framework Directive field sampling locations used in the variable interpolation have been plotted on the maps for the SST and chlorophyll concentrations. .... 15

Figure 4. Maximum entropy (MAXENT) model predictions of habitat suitability for A) *Hippocampus* genus, B) *H. hippocampus* and C) *H. guttulatus*. Values range between zero (low probability) and one (high probability)..... 18

Figure 5. Maximum entropy (MAXENT) model predictions of habitat suitability for A) *Hippocampus* genus, B) *H. hippocampus* and C) *H. guttulatus* for the south of England, Wales and the English Channel. Values range between zero (low probability) and one (high probability)..... 19

## 1. Introduction

The acceleration of man-made global change and increased vulnerability of marine species, especially in coastal areas, means it is critical to establish robust information on the distribution of biodiversity, so that changes in the use patterns of the seas can be managed sustainably. Seahorses (Syngnathidae family) are Protected, Endangered and Threatened species (PETs) that are found in shallow coastal and estuarine waters which are vulnerable to anthropogenic pressures, including coastal development, fisheries bycatch, aquarium trade, traditional medicines and other human activities that cause damage to their habitats (such as bottom trawling, anchoring and dredging etc.) (Vincent et al. 2011, Bluemel et al. 2020). Whilst seahorses are renowned for being difficult to study due to their cryptic morphology and behaviour, there is evidence for widespread declines in some natural populations (Curtis & Vincent 2005, Caldwell & Vincent 2012), highlighting the need for improved assessment, monitoring and management of these species (Woodall et al 2018).

In shallow waters, many marine species favour sheltered and protected habitats, such as kelp forests and seagrass meadows, including commercially valuable juvenile fish species (Bertelli & Unsworth 2013). Seagrasses are marine flowering plants that are widely distributed globally along temperate and tropical coastlines, from the intertidal to subtidal zones. Seagrasses are highly productive, biogenic habitats that have key ecological roles in coastal ecosystems and can form extensive meadows supporting high biodiversity (Short et al. 2007). Through their complex root system (rhizome) seagrasses stabilise sediment, whilst their canopy attenuates wave energy, reducing the impact of coastal erosion as well as facilitating sedimentation decreasing resuspension of particles. Within seagrass canopies, hydrodynamic processes are different compared to nearby unvegetated habitats. Past studies have proposed that higher densities of species from the family Syngnathidae, which includes seahorses, pipefish and seadragons etc., are generally found in and around seagrass beds rather than in neighbouring unvegetated areas of the seafloor due to the food supply and protection they provide (Kendrick & Hyndes 2003, Curtis & Vincent 2005).

The two temperate PET seahorse species found in UK coastal waters (*Hippocampus hippocampus* and *H. guttulatus*) are reportedly concomitant with seagrass habitats, using their cryptic nature for predating on small crustaceans and using the plants stems or algal holdfasts for anchorage (Foster & Vincent 2004). Curtis & Vincent (2005) suggest some habitat partitioning is evident, with *H. guttulatus* preferring complex habitats with more flora and sessile fauna available, whereas *H. hippocampus* is often found using open, less speciose habitats with more exposure to hydrodynamic influences. However, this partitioning does not appear to be consistent across their range (Garrick-Maidment and Jones 2004, Woodall 2009). Both species are known to reside near seagrass beds as well as a wide variety of shallow habitats along the English Channel on the south coast of England (Garrick-Maidment and Jones 2004, Garrick-Maidment 2007, Woodall et al. 2018). When occurring sympatrically, *H. guttulatus* tend to be found in higher densities compared to *H. hippocampus* (Woodall et al. 2018). Previous modelling by Bluemel et al. (2020) of *Hippocampus* spp. identified that phytoplankton concentration was an important environmental driver of their distributions, with optimal habitat predicted in areas with phytoplankton concentrations above  $\sim 10 \text{ mgC/m}^3$ . High plankton biomass and productivity are linked to increased prey availability, which is likely driving habitat preferences in these species (Curtis & Vincent 2005, Woodall et al. 2018).

Both seahorse species are of high conservation concern in UK waters and classified by the following conservation bodies: Convention on International Trade in Endangered Species (CITES) Appendix II (trade restrictions); Berne Convention Appendix II; OSPAR Annex V (List of Threatened and/or Declining Species & Habitats); Wildlife and Countryside Act, 1981 Schedule 5, section 9; UK Biodiversity Action Plan Priority Species (England) and Features of Conservation Importance (England and Wales). Additionally, the International Union for Conservation of Nature (IUCN) Red List of threatened species lists both of these seahorse species as data deficient (DD, Woodall 2017, Pollom 2017), and therefore, more information is needed to conduct robust assessments and establish the status of these species.

In the marine environment it can be difficult to observe and monitor species directly, resulting in a patchy understanding of their distribution (Reiss et al. 2021). Currently there is uncertainty surrounding the distribution, population abundance and specific habitat preferences of both *Hippocampus* species due to their apparent sparse distribution, low density and cryptic nature (Woodall et al. 2018, Curtis & Vincent 2005, Foster & Vincent 2004), giving rise to challenges for monitoring, evaluation and management. Species Distribution Models (SDMs) are a useful tool for data limited situations, to understand species habitat requirements and to help guide effective decision making for biodiversity, conservation and ecosystem management. For example, a suite of SDMs have been used to predict the distribution of rare species, such as *Arctica islandica* in the North Sea (Reiss et al. 2021, Weinert et al. 2021), and in response to projected environmental change (Weinert et al. 2021). Although, the level of performance of SDMs is highly dependent on environmental and species data available, there are recent advances in creating high-resolution datasets to improve modelling outcomes (Bertelli et al. 2023).

Whilst previous attempts to model seahorse distributions around the British Isles have been made, predictions in the shallow coastal and estuarine environments have not been possible due to a lack of available environmental data at a suitable resolution, scale or accuracy. There are inherent difficulties associated with modelling coastal species, including seahorses, due to the fine scale variability over short distances in shallow water environments. Environmental data that are often used to model habitat preference of marine species have limitations nearshore, such as mixed reflectance signals from multiple ground types where grid cells cross land and sea for satellite derived variables (i.e., each raster cell needs to be at least one grid cell from the land for accuracy), and the heavy computational requirements and high resolution input data needed to run ecosystem models of the whole coastline at a relevant resolution. Thus, these datasets currently do not capture the full environmental variability driving the distribution of coastal species.

## 1.1 Aims and objectives

This project aims to develop advanced habitat suitability models for two sensitive species of seahorse (*H. hippocampus* and *H. guttulatus*) by predicting habitat suitability in shallow, nearshore waters (coastal and estuarine sites) that are considered important for both species. These shallow areas have previously been omitted from SDMs (e.g., Bluemel et al. 2020) due to the limited extent of the gridded environmental data currently available, that typically does not cover shallow waters. Here, we aim to combine Environment Agency field-collected water samples extracted from the Water Framework Directive (WFD) statutory monitoring point data, with gridded satellite-derived environmental data to

facilitate coastal habitat predictions at a finer resolution and closer to the coastline than previously possible. This project has been prioritised by Natural England's Seahorse Working Group. The information provided by this project will be used to support marine spatial planning to reduce the wider impact of anthropogenic activities on seahorses and enable better decision making to protect these sensitive species and their habitats.



## 2. Methods

To improve the resolution and extent of suitable seahorse habitat predictions in coastal and estuarine environments, satellite-derived gridded environmental data (offshore) were combined with WFD field-collected water sample point data (nearshore) using sophisticated kriging methods to produce an environmental dataset for use in predictive modelling. The fine-scale modelled environmental data layers were tested for use in predicting the distribution of seahorse habitat using presence-only species records. The methods are described in detail below.

### 2.1 Environmental data

The selection of environmental variables to model seahorse habitat included variables identified as important on previous studies (e.g., minimum annual phytoplankton concentration and minimum annual sea surface temperature (SST) from Bluemel et al. 2020) and additional potentially relevant variables identified through discussions with Natural England (i.e., distance to highly productive seagrass habitats). Phytoplankton form the base of the marine food web, converting solar energy into organic matter (primary production). Phytoplankton is an important driver of zooplankton and ichthyoplankton dynamics, which many small planktivorous invertebrates and fish are dependent on, including many prey species of seahorses. Therefore, food availability is likely to be driving this habitat preference, as suggested in other seahorse studies (Curtis and Vincent, 2005; Woodall *et al.*, 2018). Inspection of the WFD field-collected monitoring data revealed no available data for phytoplankton concentration. The only other potentially suitable variable available, with an ample volume of data records was chlorophyll a concentration (Table 1). Sea surface temperature (SST) also had ample WFD coverage (Table 1). Additional variables were calculated, including the distance to the coastline, distance to seagrass habitats and bathymetry. The methods used to calculate or interpolate the selected environmental variables are described below.

#### *Chlorophyll a concentration and sea surface temperature*

A combination of satellite derived data (MODIS-Aqua) and field-collected point data extracted from the WFD archive (Table 1) were produced, and continuous data layers were generated by interpolating between point observations (see [Figure 1](#) example for the chlorophyll a concentration variable), using ordinary kriging within the “Geostatistical Analyst” tool in ArcGIS (v10.5). Kriging (Oliver, 1990) is an advanced geostatistical interpolation method based on the spatial autocorrelation between the known values and uses a statistical formula based on multiple values at multiple distances, incorporating detectable spatial gradients.

Raster cells immediately surrounding land masses were removed from the satellite-derived data to reduce potential inaccuracies from mixed reflectance signals from multiple ground types in grid cells that cross land and sea (see [Figure 1](#) example). WFD archive data were extracted from each source between the years 2008 and 2017 and contained multiple observations at each coastal sampling location, representing different days, months and years. In addition, not every sampling location was sampled consistently over time. Therefore, for each year, data were averaged by season at each location (winter = December – February; spring = March – May; summer = June – August and autumn

= September – November) to minimise potential sampling biases. The minimum winter and maximum summer SSTs and chlorophyll concentrations were determined for each location using R (v4.2.1). A similar process was used to calculate the minimum winter and summer maximum values from the satellite derived data layers in the geospatial information system (GIS) software, ArcGIS (v10.5).

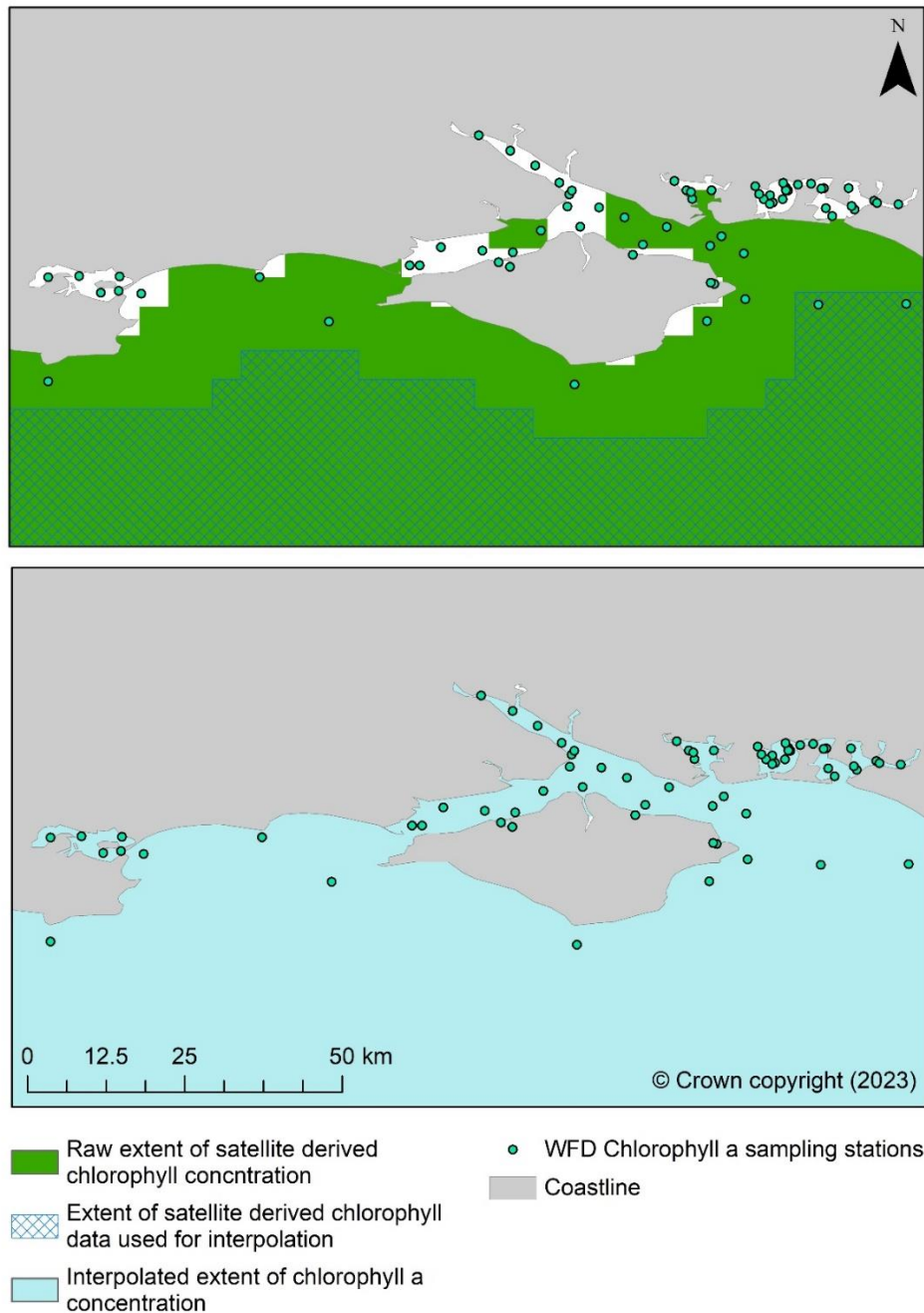


Figure 1. The example shows the marine area around the Isle of Wight, Dorset, to illustrate the available data extents and interpolation process for the chlorophyll a concentration variable. Top: gridded satellite-derived data extent ( $0.0417^{\circ} \times 0.0417^{\circ}$  resolution), WFD field sampling locations (Table 1), the extent of the gridded satellite-derived data used for interpolation. Bottom: the resulting extent of the interpolated data ( $0.0042^{\circ} \times 0.0042^{\circ}$  resolution).

Table 1. Table of environmental variable details and source locations. See Table A2.1 for seagrass data sources. See 2.1 for details of the seasonal resolution of interpolated sea surface temperature and chlorophyll a concentration variables.

Variable	Description	Source(s) and date accessed
Bathymetry (m)	Gridded (0.0042° x 0.0042° resolution), GEBCO (2022) continuous terrain model of the global oceans and land	GEBCO Compilation Group (2022) GEBCO 2022 Grid (doi:10.5285/e0f0bb80-ab44-2739-e053-6c86abc0289c) [Downloaded January 2023]
Sea surface temperature (SST) (°C)	Gridded (0.0417° x 0.0417° resolution), MODIS-Aqua Level-3 Mapped SST (2008 – 2017)	<a href="https://oceancolor.gsfc.nasa.gov/l3/">https://oceancolor.gsfc.nasa.gov/l3/</a> [Downloaded May 2018]
	Point, water quality monitoring SST data (2008 – 2017)	Environment Agency water quality data from the Water Quality Archive [Accessed January 2023]
Chlorophyll a (mg m <sup>-3</sup> )	Gridded (0.0417° x 0.0417° resolution), MODIS-Aqua Level-3 Mapped concentration of chlorophyll a (2008 - 2017)	<a href="https://oceancolor.gsfc.nasa.gov/l3/">https://oceancolor.gsfc.nasa.gov/l3/</a> [Downloaded May 2018]
	Point, water quality monitoring chlorophyll data	Environment Agency water quality data from the Water Quality Archive [Accessed January 2023]
Distance to coast	Gridded (0.0042° x 0.0042° resolution), calculated within ArcGIS (V10.5)	-

### *Bathymetry*

Bathymetry data were downloaded from GEBCO (GEBCO 2022) continuous terrain model of the global oceans and land. In general, the bathymetry data extended close to the shoreline. However, there were numerous occasions where the cell values were > 0m for grid cells that cover above and below the coastline. Therefore, to increase the accuracy of the bathymetry layer, values greater than zero were reassigned because seahorses would not naturally encounter land values > 0m in reality. To reassign these > 0m values, the Natural Neighbour interpolation method in ArcGIS 10.5 was used. Natural Neighbour (Sibson 1981) is a simple interpolation method that uses the immediate surrounding values without considering gradients at a longer distance. It produces smooth values in the prediction area only within the range of the surrounding values. Kriging methods were not appropriate for use here because this would result again in values above 0m so the Natural Neighbour interpolation was deemed appropriate for the purpose of infilling to provide more realistic values that seahorses would naturally encounter i.e., bounded by nearby values < 0m.

### *Distance from the coastline*

The distance from the coastline to the centre of each grid cell in the study area was calculated in ArcGIS 10.5 using the Euclidian distance tool. This calculates the closest distance to a source feature, which in this case was a polygon shapefile representing the coastline.

### *Distance to seagrass habitat*

To calculate the distance to seagrass meadows, all available polygon and point data of seagrass habitats across the shelf-seas were collated from numerous sources (Appendix 2, Table A2.1). For NBN Atlas Wales (2021), seagrass records were obtained from multiple datasets which are listed in Appendix 2.

As the UK is home to numerous species of seagrass, all three species that are most common around UK shores, *Zostera marina*, *Z. noltii* and *Ruppia* spp., were included. All data were uploaded to ArcGIS 10.7 and checked for duplicates. All open access polygon data was added to ArcMap and merged to create one single polygon layer. For the point data records, only those points outside of the polygons were retained, and any records outside of the years 2000 to present were removed. In the absence of seagrass bed extents for the point data, these were converted to polygons by creating a small buffer around them and merged with the existing polygon data to ensure all possible seagrass locations were included in the calculation. A distance from seagrass layer was calculated in ArcGIS 10.5 using the "Euclidean Distance" tool.

### *Predictor variable selection*

The environmental data layers were computed for the study area covering the British Isles, Ireland and English Channel (Figure 1) on a consistent grid with a cell resolution of 0.0042 x 0.0042 degrees. To reduce collinearity among model predictors (Dormann et al. 2013), only variables with a Pearson's correlation < 0.7 were retained for analysis (Appendix 1, Figure A1.1). Distance to seagrass bed and distance to coastline were found to be highly correlated (Pearson's correlation = 0.8) and therefore distance to coastline was not used in the analysis. The final list of six variables used for modelling were distance to seagrass habitat, bathymetry, minimum winter and maximum summer SST and chlorophyll concentrations.

## 2.2 Species data

Occurrence records for modelling the ecological niche of *Hippocampus* spp. were compiled from Bluemel et al. (2020) from fisheries monitoring (trawl) surveys and other open-source databases reported, with additional data from The Seahorse Trust (National Seahorse Database) and Natural England's database. All occurrence records of seahorses that coincided with the study area limits (Figure 1) were retained and filtered to remove records that were duplicated, or where species identification was dubious. Prior to modelling, records were reduced to one point per grid cell to reduce sampling biases. All occurrence records from the year 2000 to present were used for model training and testing (*Hippocampus* genus n = 165, *H. hippocampus* = 144, *H. guttulatus* = 45) (Figure 2). Older occurrence records (pre-2000) were excluded from the analysis to ensure that the occurrences roughly coincide with the environmental data time frame used to model the distribution (2008 – 2017) and that model predictions represent the species' current habitat distribution.

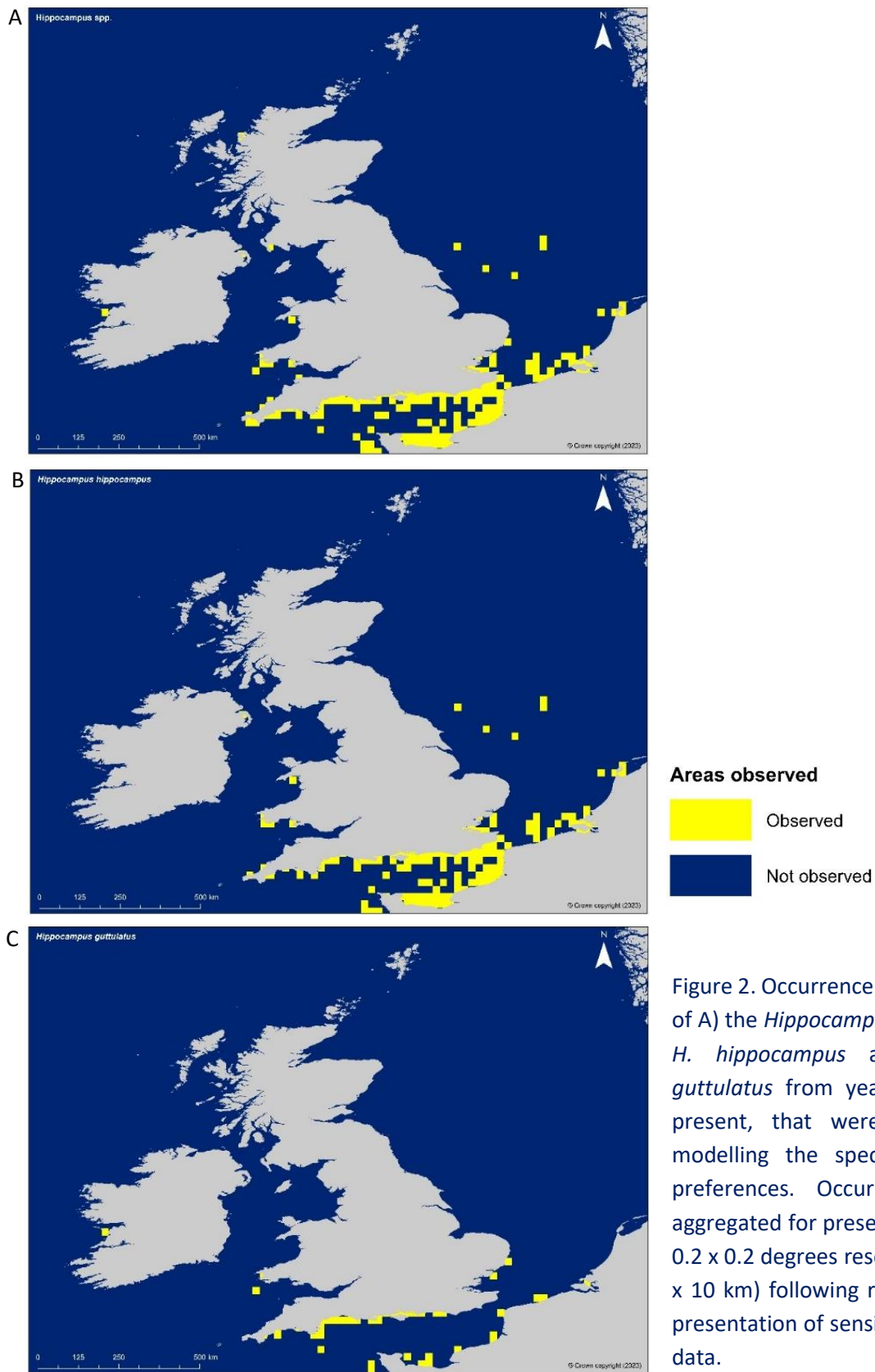


Figure 2. Occurrence distribution of A) the *Hippocampus* genus, B) *H. hippocampus* and C) *H. guttulatus* from years 2000 to present, that were used for modelling the species habitat preferences. Occurrences are aggregated for presentation to a 0.2 x 0.2 degrees resolution (~10 x 10 km) following rules for the presentation of sensitive species data.

## 2.3 Species distribution modelling

Maximum entropy (MAXENT) is a commonly used modelling algorithm which predicts the probability distribution of a given species using presences and background (pseudo-absence) data (Phillips et al. 2006). The ecological niche (*sensu* Hutchinson, 1957) of a species is modelled in relation to several environmental variables. The model minimises the relative entropy between the probability densities calculated from the presence data and the background landscape in covariate space (Elith et al. 2011). This modelling method has been considered in many studies and has been found to have a good performance when comparing the effectiveness of different species distribution modelling (SDM) techniques within the marine environment (Ready et al. 2010, Reiss et al. 2011, Jones et al. 2012, Bučas et al. 2013). Whilst MAXENT has been identified as a useful tool for describing ecological relationships when only species presences are available, it is crucial to deal with sampling biases that may exist in the occurrence data (Elith et al. 2011), which is beyond the scope of this project due to time constraints.

Here we run MAXENT to test the useability of the environmental variables calculated in the nearshore environment. Additionally, we make recommendations for future work modelling seahorse habitat that would improve confidence in the predictions (i.e., testing multiple algorithms, habitat suitability thresholds, validation approaches suited to presence-only situations and dealing with sampling bias and spatial autocorrelation).

### *Model calibration*

MAXENT version 3.4.3 (Phillips et al. 2020) was used to predict the distribution of habitats suitable for supporting *H. hippocampus* and *H. guttulatus* and the combined genus. The default convergence threshold of  $10^6$  and a maximum of 5,000 iterations were used. A maximum of 10,000 background samples were randomly selected from the study area and the jackknife option was selected to determine the importance of each predictor variable in the final model. Within the model settings, the random test percentage was set to 25% and random seed was selected (i.e., a different random subset of presences were used for each model run). Ten replicates were performed and the replicate run type was set to subsample. All other MAXENT settings were default.

### *Model evaluation and variable importance*

Model performance was examined by computing the Area Under the Curve (AUC) score of the receiver operating characteristic (ROC) curve for each species, which is calculated by the MAXENT software as part of the modelling process and is a widely used measure of model performance (e.g., Reiss et al. 2011), but see Lobo et al. (2008) for a review of its performance and questioned reliability. The AUC represents the relationship between sensitivity and specificity and varies between 0 and 1 (Reiss et al., 2011). However, true absences are required to calculate AUC scores, and MAXENT calculates AUC differently by comparing presences with background points for presence-only models (Phillips et al. 2006). In general, values  $>0.9$  indicate an excellent model fit, values between 0.7 and 0.9 represent a good fit, anything  $<0.7$  represents a poor model fit and values around 0.5 represents nothing better than random (Reiss et al., 2011). The importance of environmental variables was established based on the relative percentage contribution.



## 3. Results

### 3.1 Environmental data

#### *Sea surface temperature (SST)*

The SST layers (Figure 3A,B) follow expected trends. In summer months the SST is highest in the south of the study area (16.5-23.3°C), particularly around the southwest coast of England, through the channel to East Anglia and the coasts of France, Belgium and the Netherlands. Similar temperatures are also experienced along the coast of Wales and north-east England, with cooler temperatures found around Scotland and Northern Ireland (11.0 – 15.9°C).

In winter months the warmest surface temperatures are observed offshore extending in from the Atlantic (8.3 – 11.4°C) around the island of Ireland, to north of Scotland, and through the south-west approaches into the channel. The North Sea appears to be cooler with temperatures between 2.4°C and 7.3°C, and temperatures < 6°C are found from the Thames up to Bridlington on the northeast coast, in the Solent, the Severn, around the Wirral Peninsula and up to the Scottish border.

#### *Chlorophyll concentration*

Chlorophyll concentrations (Figure 3C,D) are higher in the summer months compared to the winter months. In the summer months there is a large range in chlorophyll concentrations (0.4 mg m<sup>-3</sup> – 252.5 mg m<sup>-3</sup>). The highest values observed can be found at the south-west tip of England at Porthleven, with values ranging from 60.5 mg m<sup>-3</sup> to 252.5 mg m<sup>-3</sup>. Concentrations within the rest of the study area are generally less than 20 mg m<sup>-3</sup>. There is a higher concentration of chlorophyll found nearer the coastline, with the higher concentrations (6 - 20 mg m<sup>-3</sup>) being found around the Gower peninsula, Bangor to Morecombe Bay, the Firth of Clyde, the Firth of Forth, the Humber to the wash, and the Thames. These concentrations can also be observed within the Normandy region of France and along the coastlines of Belgium and The Netherlands.

The chlorophyll concentration within the winter months is much lower (0.1 mg m<sup>-3</sup> to 13.9 mg m<sup>-3</sup>) in comparison to the summer. The highest concentrations are found in the Firth of Clyde in Scotland (7.1 mg m<sup>-3</sup> to 13.9 mg m<sup>-3</sup>). Coastal areas generally have concentrations between 1.2 mg m<sup>-3</sup> and 7.1 mg m<sup>-3</sup>). There are lower concentrations around the north coast of mainland Scotland and the Scottish islands, and the south coast of Cornwall, which generally have concentration of <1 mg m<sup>-3</sup>.

#### *Distance to seagrass beds*

Seagrass beds are generally found in shallow waters close to the coastline (Figure 3F), which explains the high correlation (Pearson's  $r = 0.8$ ) between distance to coastline (not shown) and distance to seagrass bed layers.

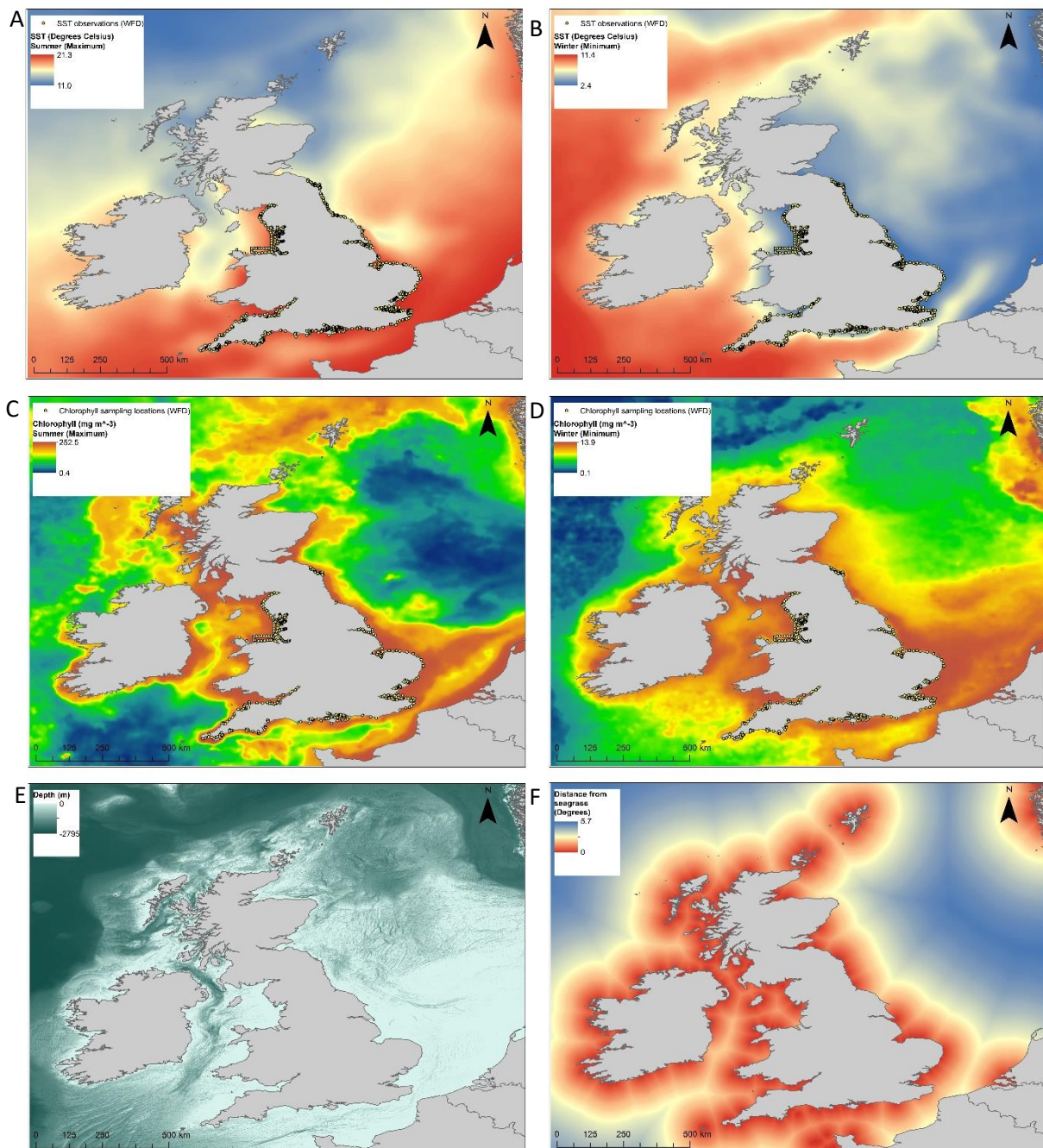


Figure 3. Environmental predictor variables used for modelling the distribution of seahorses. A) summer maximum sea surface temperature (SST), B) winter minimum SST, C) summer maximum chlorophyll concentration, D) winter minimum chlorophyll concentration, E) bathymetry, F) distance from seagrass. Locations of Environment Agency Water Framework Directive field sampling locations used in the variable interpolation have been plotted on the maps for the SST and chlorophyll concentrations.



### 3.2 Model performance

The mean AUC for all three models indicates an excellent performance ( $AUC > 0.9$ ). The model for *H. guttulatus* had the highest AUC, which had a mean value of  $0.988 \pm 0.001$  across the 10 runs, followed by *H. hippocampus* ( $0.972 \pm 0.001$ ) and finally the model predicting the distribution for the genus ( $0.967 \pm 0.001$ ). Note, the AUC metric provided by MAXENT evaluates presences against background absences only (i.e., the fraction of the total study area predicted present as opposed to the fraction of absences predicted present when absence data is available) and are likely to be elevated for species with narrow ranges relative to the study area, which is an artifact of the AUC statistic (Phillips et al. 2006).

### 3.3 Variable importance

#### *Hippocampus spp.* & *H. hippocampus*

Of the six environmental variables included within model runs, the bathymetry and summer maximum SST had the highest percent contribution for predicting the distribution of the genus *Hippocampus* (78.8% cumulative importance) and *H. hippocampus* (76.3% cumulative importance; Table 2). These results are supported by the Jackknife test of variable importance (Figure A1.5). Habitat suitability for the genus increased with increasing temperature, with an optimum suitability at  $\sim 19^{\circ}\text{C}$ , and then decreased with temperatures  $> 19^{\circ}\text{C}$  (Figure A1.2). In comparison, habitat suitability for the species *H. hippocampus* increased with increasing temperature, with the most suitable habitat found at temperatures above  $19^{\circ}\text{C}$  and did not appear to decrease (Figure A1.3). Habitat suitability was also highest for both the genus and *H. hippocampus* in areas which were shallow ( $\sim 18\text{m}$ ).

#### *Hippocampus guttulatus*

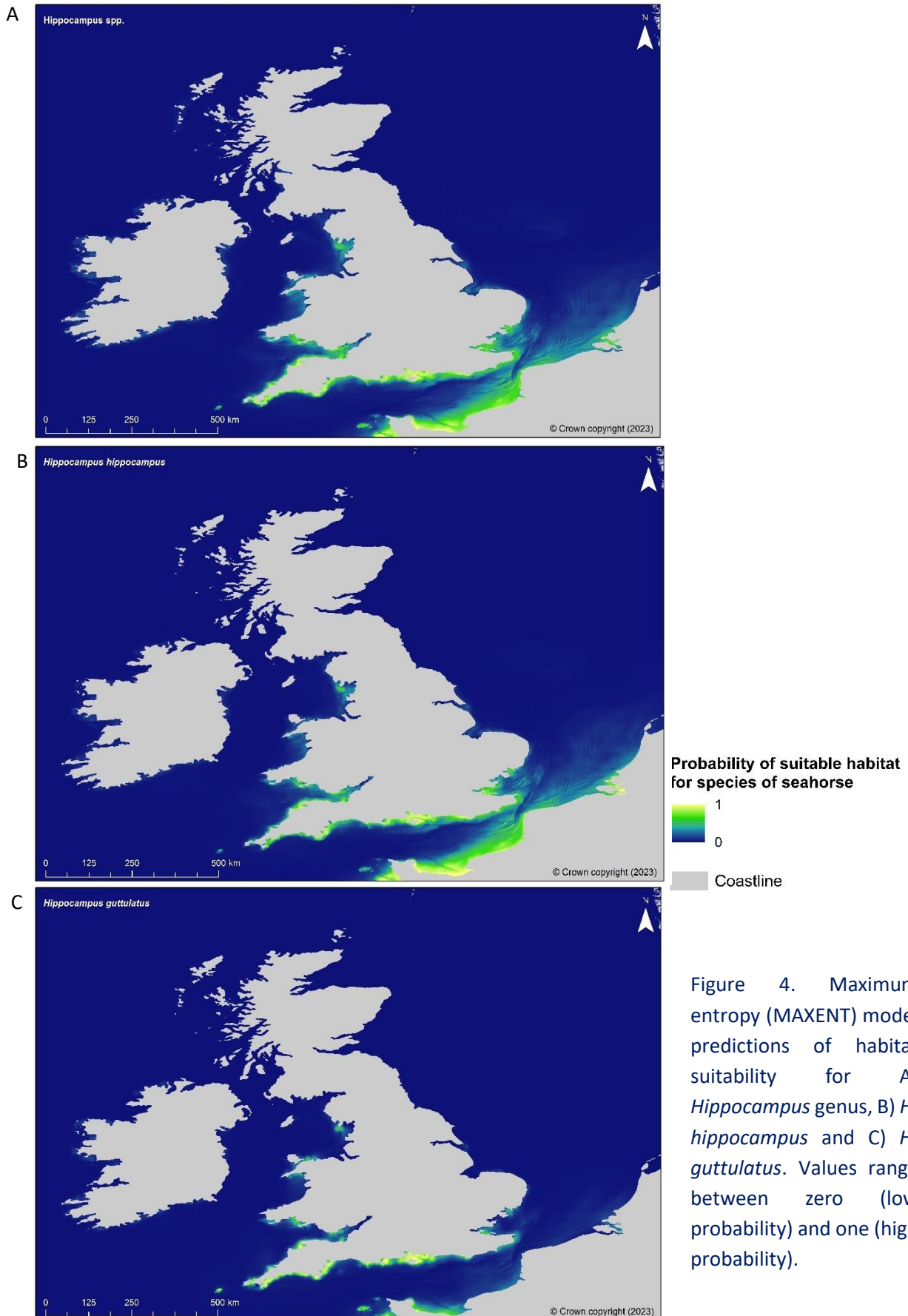
The variable importance for predicting the presence of *H. guttulatus* habitat is marginally different to that of the genus and *H. hippocampus* models (Table 2). The most important variables for predicting *H. guttulatus* habitat were the bathymetry, distance from seagrass beds and the summer maximum SST (93.4% cumulative importance). This was also supported by the jackknife test for variable importance (Figure A1.5). Habitat suitability for *H. guttulatus*, based on maximum summer SST and bathymetry, demonstrate similar trends to that of the genus. Habitat suitability increases with increasing summer temperatures to  $\sim 19^{\circ}\text{C}$  and then decreases with temperatures  $> 19^{\circ}\text{C}$  (Figure A1.4). The most suitable habitat is also found in shallow waters. However, unlike *H. hippocampus*, *H. guttulatus* appears to be more dependent on the presence of seagrass beds. The habitat suitability of *H. guttulatus* is highest when in close proximity to seagrass beds and decreases as distance from seagrass increases.

Table 2. Variable importance (% contribution) for predicting presence of *Hippocampus* spp., *H. hippocampus* and *H. guttulatus* habitats. SST = sea surface temperature.

Environmental predictor	<i>Hippocampus</i> spp.	<i>H. hippocampus</i>	<i>H. guttulatus</i>
Bathymetry	54.4	43.8	38.8
Summer maximum SST	24.4	32.5	21.9
Summer maximum chlorophyll concentration	8.9	11.0	1.0
Winter minimum SST	5.4	6.2	4.9
Distance from seagrass beds	3.5	2.6	32.7
Winter minimum chlorophyll concentration	3.2	3.9	1.2

### 3.4 Habitat distribution

The raw MAXENT predicted probabilities of the occurrence of suitable seahorse habitat across the whole study area are presented in Figure 4 and for the southwest of England and the English Channel in Figure 5. Values range from zero, indicating low probability of suitable habitat being present, and one, indicating high probability. The highest probabilities (i.e., greater than 0.6) are observed predominantly in southern coastal areas (bays and estuaries) of the British Isles and the English Channel, as well as shallow waters of the southern North Sea (France, Belgium and Netherlands coastlines and the Thames estuary), northern Cardigan Bay and the Menai Strait (northern Wales) and the Bristol Channel. The distributions for the two species follow a broadly similar pattern, with higher probabilities over larger extents observed for *H. hippocampus* in most of the regions, except for the southwestern tip of England (Cornwall and Devon coastlines), the Isles of Scilly, and around Portland and Weymouth in Dorset, where the *H. guttulatus* model tended to show higher probabilities. Both species show particularly high probabilities around the Isle of Wight and Solent, Poole Harbour, west Torbay (Exmouth to Brixham), and north Devon (River Taw and Torridge estuary) (Figure 5).



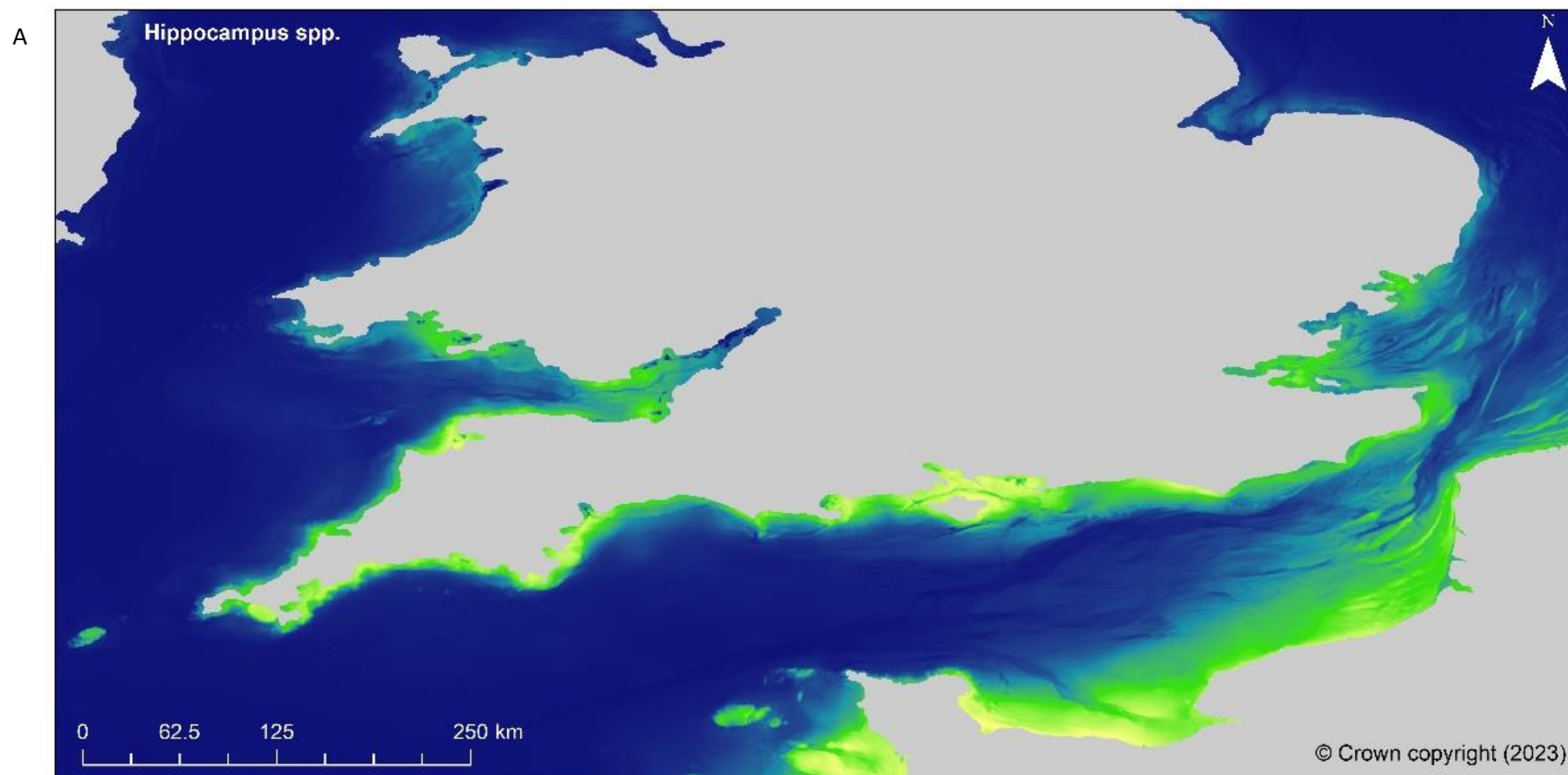


Figure 5. Maximum entropy (MAXENT) model predictions of habitat suitability for A) *Hippocampus* genus, B) *H. hippocampus* and C) *H. guttulatus* for the south of England, Wales and the English Channel. Values range between zero (low probability) and one (high probability).

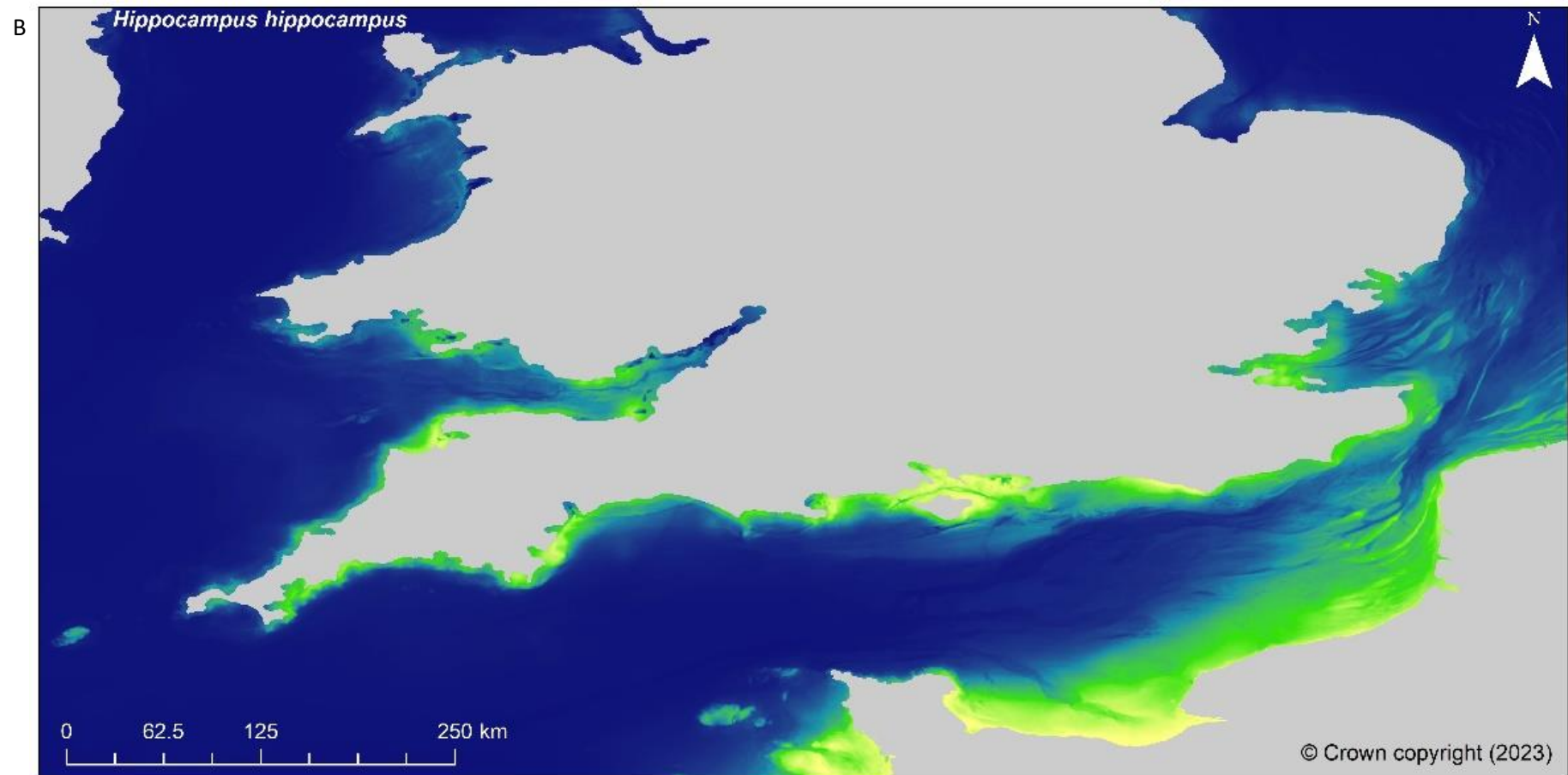


Figure 5 (continued)

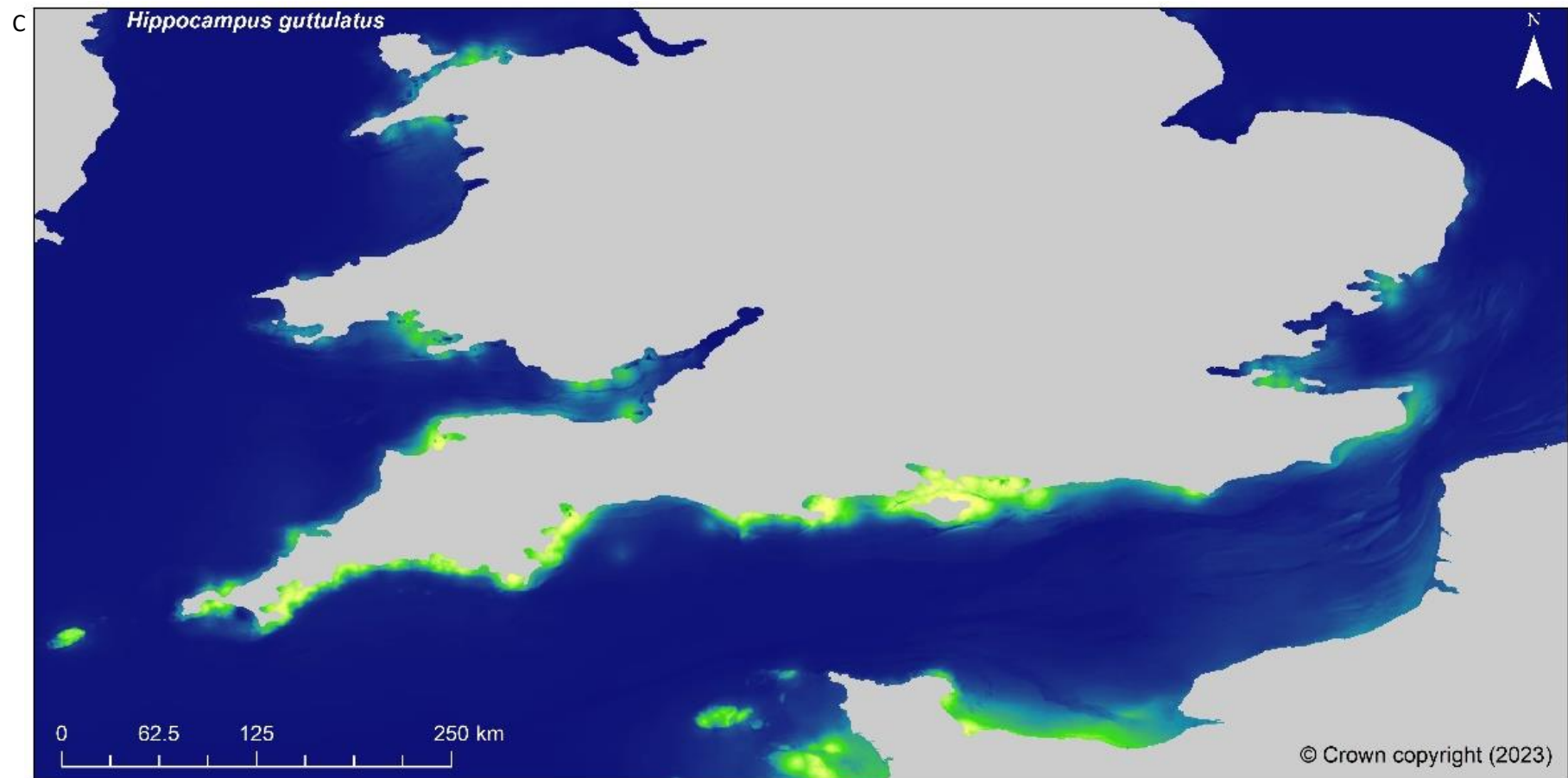


Figure 5 (continued)



## 4. Discussion

Effective conservation, marine spatial planning, and sustainable management of anthropogenic activities in our waters requires accurate knowledge on all ecosystem components, including the abundance and distribution of vulnerable PET species and their associated habitats (Bluemel et al. 2020, Reiss et al. 2021, Weinert et al. 2021, Pierri et al. 2022, Bertelli et al. 2023). In the marine environment it is difficult to observe and monitor species directly, resulting in a patchy understanding of their distribution (Reiss et al. 2021). This is especially true for rare, cryptic and data-limited species such as seahorses. Coastal species are also inherently difficult to model due to a lack of available environmental data at suitable resolutions, scales or accuracies in these highly variable marine environments. This project has successfully developed methods to improve the available environmental data that is required to effectively model and predict this distribution of species habitats in coastal, shallow water environments.

### 4.1 Environmental variables

A total of six environmental raster layers (distance to seagrass habitat, bathymetry, minimum winter and maximum summer SST and chlorophyll concentrations) have been successfully produced under this project by combining multiple open-source databases to create higher resolution environmental predictor layers that extend into shallow, coastal waters. These data provide information on the fine scale variability over short distances in shallow water environments required to accurately predict coastal species' habitat preferences.

To further improve our capability in predicting the distribution of coastal species habitats, it would be beneficial to extend the methods to additional environmental variables, such as phytoplankton, zooplankton, primary production, wave energy, current velocity, oxygen concentration, nutrients, substrate type, where possible. The inclusion of additional environmental variables in SDMs will provide further environmental context, which is potentially not captured by the variables used in this study. Additionally, the distance to seagrass bed layer could be further improved using modelled seagrass distributions, such as those produced by Berlotelli et al. (2023), to predict the area of individual seagrass beds in areas where only point observations are available.

### 4.2 Modelling limitations and sampling biases

We run MAXENT here to test the useability of the environmental variables calculated in the nearshore environment for modelling seahorse habitat preferences. While MAXENT has been identified as a useful tool for describing ecological relationships when only species presences are available, it is however crucial to deal with sampling biases that may exist in the occurrence data (Elith et al. 2011). MAXENT is sensitive to biased data distributions, and there is a strong assumption of random sampling distribution for presences when random background data are used. Known sampling biases exist in the seahorse occurrence data used in this study and therefore the results should be taken with caution. Biases include higher sampling effort in the southwest and English Channel, due to high

recreational SCUBA diving, boating and snorkelling intensity in this area, resulting in uneven sampling distributions across the study area. Recorder biases exist due to known locations of seahorses being regularly visited and targeted by snorkellers / divers, and records are often provided of the same seahorse that has moved to a nearby locality (Garrick-Maidment pers. comm.). The increased resolution of the survey grid used here will further exacerbate these sampling biases. Also, occurrences were restricted to the year 2000 onwards in this study, which may have further exasperated the sampling bias effect by removing some of the historic records in less frequently sampled areas.

To improve confidence in the habitat predictions for seahorses, further work should include testing different modelling algorithms (e.g., Bayesian Additive Regression Trees (BART), a machine learning technique that allows for estimation of prediction error (Chipman et al., 2010)). Variations in model predictions were observed by Bluemel et al. (2020), highlighting the need to compare outputs and identify consistent robust predictions, especially in data-limited, presence-only settings. Additionally, testing different habitat suitability thresholds that are tailored to the intended use of the predicted distributions (Freeman and Moisen 2008), testing validation approaches more suited to presence-only situations, because using AUC in isolation can be uninformative when species prevalence is low (e.g., the Continuous Boyce Index (CBI, Hirzel et al. 2006, after Boyce et al. 2002), which is also useful for defining discrete habitat suitability classes). Furthermore, sampling bias and spatial autocorrelation in the occurrence data requires further work. Methods to deal with spatial autocorrelation and sampling biases could include defining a bias grid to include in the modelling process (i.e., known dive sites with higher sampling activity thus a higher likelihood of observation), using other suitable point process models that account for these biases. Additionally, assessing model performance with a block cross validation approach (Roberts et al. 2017)) could better capture the model's ability to predict to novel areas.

### 4.3 Seahorse habitat preferences

We obtained strong environmental signals from our models, with a similar spatial distribution of predictions to those presented by Bluemel et al. (2020). High predicted probabilities of suitable habitat for both *H. hippocampus* and *H. guttulatus* are concentrated in coastal locations around the English Channel, southern England, Wales and northwest England. However, the underlying model assumptions are invalidated by the sampling bias discussed in 4.2, which could be restricting predictions to this southerly locality where sampling biases (increased effort and recorder bias) are known to exist. There are some occurrences reported in coastal sites around western Scotland (genus only) and western Ireland (*H. guttulatus* only) and offshore in the southern North Sea (*H. hippocampus* only) since 2000 (Figure 1), suggesting these sites do potentially provide suitable habitat for these species for at least part of the year (i.e. during winter migrations offshore). It is possible that due to the comparatively low sampling effort in Scottish, Irish and northern English waters compared to southern England, these samples did not significantly contribute to the model.

Our study provides further evidence of niche partitioning between the two *Hippocampus* species with *H. guttulatus* showing a higher preference for habitats that are in close proximity to seagrass meadows



compared to *H. hippocampus*, also reported by Curtis & Vincent (2005). As seagrass beds are highly productive, complex habitats found close to shore, this suggests *H. guttulatus* predominantly prefers complex habitats in shallow waters, compared to *H. hippocampus*, as previously reported in Woodall et al. (2018) and Bluemel et al. (2020). This has also been shown by Correia et al. (2018) in South Portugal, where there were higher abundances of *H. guttulatus* in locations with a high availability of holdfasts between 3m and 6m, whereas higher abundances of *H. hippocampus* were observed in locations with lower availability of holdfasts. Pierri et al. (2022) undertook a systematic review looking at the large-scale distribution of *Hippocampus* spp. in Europe. The authors noted that both species of seahorse can be found between 0m and 6m within confined locations (lagoons, estuaries etc.). However, *H. guttulatus* is more frequently found in shallower waters (0-1m) than *H. hippocampus*. In shelf locations, *H. guttulatus* was most frequently located in areas with a depth between 5m and 10m, with decreasing observations up to a depth of 20m. In comparison, *H. hippocampus* could be found up to a depth of 30m (Pierri et al. 2022). This supports the outputs of the models produced in this study, which indicate that the distribution of *H. guttulatus* is more restricted to coastal areas compared to *H. hippocampus*. Although, the models presented here do not predict suitable habitat further offshore for *H. hippocampus*, i.e., Dogger bank in the North Sea (Pinnegar et al., 2008) or in deeper parts of the English Channel, where they are regularly reported in scientific fisheries surveys (Bluemel et al. 2020) particularly during the winter months when they undertake small-scale migrations to avoid harsher conditions (Boisseau, 1967; Garrick-Maidment and Jones, 2004; Garrick-Maidment, 2007). It would therefore be beneficial to run seasonal models to gain a better understanding of temporal variation in species distribution.

Both species predicted distributions were similarly influenced by the maximum summer SST, preferring sites with temperatures around 18°C or higher. Temperature has been shown to play an important role in fish and can influence aspects including survival, reproduction, and growth (Planas et al. 2012). The breeding season for *H. guttulatus* can extend from April to October (Boisseau 1967, Planas et al. 2012), and the temperatures experienced during this period are region dependent (Planas et al. 2012). It is possible that temperature could be a limiting factor in the distribution of seahorse species, especially *H. guttulatus*, as the most northerly extent of this species is around the coast of the UK and Ireland (Planas et al. 2012). Temperature has also been demonstrated to influence egg production, meaning the total number of eggs produced, clutch size and number of clutches per female (Planas et al. 2013). The experiment undertaken by Planas et al. (2013) manipulated both temperature and light exposure whilst examining both *H. hippocampus* and *H. guttulatus*. The study indicated that colder temperatures and shorter periods of light exposure, resulted in a lower egg production. The authors also noted that egg clutches were primarily released when temperatures reached 16°C and at increased light exposure. Examining the summer maximum temperatures interpolated around the British Isles (Figure 3) indicate there are more northerly areas along the west coast of England and Scotland and around Ireland that are in a suitable temperature range for the species. As mentioned previously, dealing with sampling biases and spatial autocorrelation in the data would provide greater model confidence in areas where the sampling effort is lower.

It was not possible to include some variables in the modelling process that influence seahorse distributions, i.e., *Hippocampus* spp. preference for winter minimum phytoplankton concentrations above ~10 mgC/m<sup>3</sup> (Bluemel et al. 2020), due to the unavailability of WFD field-collected point data. High phytoplankton concentrations are associated with areas of high production and thus high prey

food availability is likely driving this habitat preference (Curtis and Vincent 2005, Woodall et al. 2018). A comparison of the winter minimum chlorophyll to the winter minimum phytoplankton variable from Bluemel et al. (2020) in the offshore environment identified a relatively high correlation (Pearsons correlation coefficient = 0.7). However, when the variable was included in the models computed here (including the additional coastal areas) it was not found to have particularly high importance, possibly due to differences in the shallow coastal zone. Other variables included in the model were likely to be explaining part of variation 'missing' from not including phytoplankton directly in the model (i.e., summer maximum SST was identified as an important predictor that also had a relatively high correlation with phytoplankton (0.6)).

## 5. Conclusion

The aim of this study was to increase the accuracy and extent of environmental predictor variables into the coastal and estuarine environments required to accurately predict coastal species habitat preferences. A total of six environmental raster layers have been successfully developed that provide information on the fine scale variability over short distances in shallow water environments. These data have enabled the development of finer spatially resolved predictions of the PET *Hippocampus* seahorse species habitats around the British Isles. However, further work is required to deal with sampling bias and spatial autocorrelation that exists in the occurrence data and to increase confidence in model predictions, particularly in more northerly, less frequently sampled localities. Future assessments of cumulative human impacts on these species will benefit from more accurate predictions in both inshore and offshore habitats.

Overall, this study has made important developments in the understanding of these elusive, data deficient species. The habitat partitioning observed between the two species is important to understand and develop the most appropriate conservation and management strategies which will benefit both species. The information provided by this project can be used to support marine spatial planning to reduce the wider impact of anthropogenic activities and enable better decision making to protect these sensitive species and their habitats. Further work to produce finer resolution maps to capture micro-habitat requirements within highly suitable bays and estuaries will provide evidence to better inform conservation strategies at a local scale e.g., see recent advances in creating high-resolution datasets to improve modelling outcomes (Bertelli et al. 2023).

## 6. Acknowledgements

We would like to acknowledge The Seahorse Trust for providing the seahorse occurrence records from their National Seahorse Database that were used to inform the species distribution models, as well as Neil Garrick-Maidment for providing valuable information on the ecology of seahorses in the UK. Acknowledgements go to Natural England for providing seahorse records for the same purpose. Thanks to Dr Elena Couce for providing quality assurance and to Alexander Bird for project management and support. This work was funded by Natural England under Cefas contract code C8528.

## 7. References

- Bertelli, C.M. and Unsworth, R.K. (2014). Protecting the hand that feeds us: Seagrass (*Zostera marina*) serves as commercial juvenile fish habitat. *Marine pollution bulletin*, 83(2): 425-429.
- Bertelli, C.M. et al. (2023). High-resolution wave data for improving marine habitat suitability models. *Frontiers in Marine Science*, 9, p.2728. Available at: <https://doi.org/10.3389/fmars.2022.1004829>.
- Bluemel, J.K., Lynam, C. and Ellis, J. (2020). Fish biodiversity: state and pressure indicators. Cefas Project Report for Defra, 67 pp.
- Boisseau, J. 1967. Les régulations hormonales de l'incubation chez un Vertébré mâle: recherches sur la reproduction de l'Hippocampe. PhD thèses, L' Université de Bordeaux, France. 379 pp.
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157: 281–300.
- Bučas, M., Bergström, U., Downie, A.L., Sundblad, G., Gullström, M., Von Numers, M., Šiaulys, A. and Lindegarth, M. (2013). Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: evaluating the scope for predictive mapping using different modelling approaches. *ICES Journal of Marine Science*, 70(6): 1233-1243.
- Caldwell, I. R. and Vincent, A. C. J. (2012). Revisiting two sympatric European seahorse species: apparent decline in the absence of exploitation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 22(4): 427-435.
- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. (2010). "BART: Bayesian Additive Regression Trees." *Annals of Applied Statistics* 4(1): 266–98.
- Correia, M., Koldewey, H.J., Andrade, J.P., Esteves, E. and Palma, J. (2018). Identifying key environmental variables of two seahorse species (*Hippocampus guttulatus* and *Hippocampus hippocampus*) in the Ria Formosa lagoon, South Portugal. *Environmental Biology of Fishes*, 101: 1357-1367.

- Curtis, J. M. R. and Vincent, A. C. J. (2005). Distribution of sympatric seahorse species along a gradient of habitat complexity in a seagrass- dominated community. *Marine Ecology Progress Series*, 291: 81–91.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carr, G., Garc, J. R., *et al.* (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1): 27-46.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17: 43–57.
- EMODnet. (2021). Seagrass cover (Essential Ocean Variable) in Europe - points (2021) and polygons (2019). e166016f-e6b7-41e0-af48-fb8b51ac1a9c Available at: <https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/metadata/39746d9c-4220-425c-bc26-7cb3056c36a5> (Accessed: March 1, 2023).
- Foster, S. J. & Vincent, A. C. J. (2004). Life history and ecology of seahorses: implications for conservation and management. *Journal of Fish Biology*, 65: 1–61. <http://onlinelibrary.wiley.com/doi/10.1111/j.0022-1112.2004.00429.x/full>.
- Freeman, E.A. and Moisen, G.G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological modelling*, 217(1-2): 48-58.
- Garrick-Maidment, N., and Jones, L. (2004). British seahorse survey 2004. pp.1-83
- Garrick-Maidment, N. (2007). British seahorse survey 2007. pp.1-26
- GEBCO Compilation Group. (2022). GEBCO 2022 Grid (doi:10.5285/e0f0bb80-ab44-2739-e053-6c86abc0289c) [Downloaded January 2023]
- Hirzel, A. H., Randin, C., and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences, 9: 142–152.
- Hutchinson, G. E. 1957. Concluding remarks: The demographic symposium as a heterogeneous unstable population. *Foundations of Ecology: Classic Papers with Commentaries*, 22: 415–427. <http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1957.022.01.039>.
- Jenkin, A., Sturgeon S and Trundle, C. (2022a). Survey field report for the 2022 acoustic survey of the seagrass bed within Gerrans Bay. Cornwall Inshore Fisheries and Conservation Authority (Cornwall IFCA), Hayle, on behalf of Cornwall Wildlife Trust and Natural England as part of the G7 Legacy Project for Nature Recovery (G7LPNR).
- Jenkin, A., Sturgeon S and Trundle, C. (2022b). Survey field report for the 2022 acoustic survey of the seagrass bed within St. Austell Bay. Cornwall Inshore Fisheries and Conservation Authority (Cornwall IFCA), Hayle, on behalf of Cornwall Wildlife Trust and Natural England as part of the G7 Legacy Project for Nature Recovery (G7LPNR).

- Jones, M.C., Dye, S.R., Pinnegar, J.K., Warren, R. and Cheung, W.W. (2012). Modelling commercial fish distributions: Prediction and assessment using different approaches. *Ecological Modelling*, 225:133-145.
- Kendrick, A.J. and Hyndes, G.A. (2003). Patterns in the abundance and size-distribution of syngnathid fishes among habitats in a seagrass-dominated marine environment, Estuarine, Coastal and Shelf Science, 57(4):631–640. Available at: [https://doi.org/10.1016/s0272-7714\(02\)00402-x](https://doi.org/10.1016/s0272-7714(02)00402-x).
- Lobo, J.M., Jiménez-Valverde, A. and Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography*, 17(2): 145-151.
- Magic - gov.uk. (2022). National Seagrass Layer. Natural England and the Environment Agency. Available at: <https://magic.defra.gov.uk/> (Accessed: February 14, 2023).
- NBN Atlas. (2021). Seagrass occurrence records obtained via NBN Atlas, search for records. (2021). Available at: <https://nbnatlas.org/> (Accessed: March 1, 2023).
- NBN Atlas Wales. (2021). Seagrass occurrence records obtained via NBN Atlas Wales, search for records. Available at: <https://wales-records.nbnatlas.org/> (Accessed: March 1, 2023).
- Natural Resource Wales. (2019). DataMapWales: Environment (Wales) Act Section 7 and OSPAR: Marine Habitats. Natural Resource Wales. Available at: <https://datamap.gov.wales/layergroups/inspire-nrw:MarineBAPOSPARHabitats> (Accessed: March 1, 2023).
- Oliver, M. A. (1990). Kriging: A Method of Interpolation for Geographical Information Systems. *International Journal of Geographic Information Systems* 4: 313–332.
- Phillips, S.J., Dudik, M., Schapire, R.E. (2020). [Internet] Maxent software for modelling species niches and distributions (Version 3.4.3). Available from url: [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/). (Accessed: 24/03/23).
- Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3-4): 231-259.
- Pierri, C., Lazic, T., Gristina, M., Corriero, G. and Sinopoli, M. (2022). Large-scale distribution of the European seahorses (*Hippocampus Rafinesque*, 1810): a systematic review. *Biology*, 11(2), p.325.
- Pinnegar, J.K., Stelzenmüller, V., Van Der Kooij, J., Engelhard, G.H., Garrick Maidment, N. and Righton, D.A. (2008). Occurrence of the short-snouted seahorse *Hippocampus hippocampus* in the central North Sea. *Cybium*, 32(4):343-346.
- Planas, M., Quintas, P. and Chamorro, A. (2013). Maturation of *Hippocampus guttulatus* and *Hippocampus hippocampus* females by manipulation of temperature and photoperiod regimes. *Aquaculture*, 388:147-152.

- Planas, M., Blanco, A., Chamorro, A., Valladares, S. and Pintado, J. (2012). Temperature-induced changes of growth and survival in the early development of the seahorse *Hippocampus guttulatus*. *Journal of Experimental Marine Biology and Ecology*, 438:154-162.
- Project Seagrass. (2023). SeagrassSpotter. A global tool for people to help locate seagrass. The SeagrassSpotter project has been created by Project Seagrass in association with Cardiff University and Swansea University. Available at: <https://seagrassspotter.org/> (Accessed: February 14, 2023).
- Pollom, R. (2017). *Hippocampus guttulatus*. The IUCN Red List of Threatened Species 2017: e.T41006A67617766. <https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T41006A67617766.en>. Accessed on 27 March 2023.
- Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J., Agbayani, E., Kullander, S. and Froese, R. (2010). Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, 221(3): 467-478.
- Reiss, H., Cunze, S., König, K., Neumann, H. and Kröncke, I. (2011). Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecology Progress Series*, 442: 71-86.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., others. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*. 40: 913-929.
- Short, F. et al. (2007). "Global Seagrass Distribution and diversity: A bio regional model," *Journal of Experimental Marine Biology and Ecology*, 350(1-2): 3–20. Available at: <https://doi.org/10.1016/j.jembe.2007.06.012>.
- Sibson, R. (1981). "A Brief Description of Natural Neighbor Interpolation," chapter 2 in *Interpolating Multivariate Data*. New York: John Wiley & Sons. 21–36.
- Vincent, A.C., Foster, S.J. and Koldewey, H.J. (2011). Conservation and management of seahorses and other Syngnathidae. *Journal of fish biology*, 78(6):1681-1724.
- Weinert, M., Mathis, M., Kröncke, I., Pohlmann, T. and Reiss, H. (2021). Climate change effects on marine protected areas: Projected decline of benthic species in the North Sea. *Marine Environmental Research*, 163, p.105230.
- Woodall, L. (2017). *Hippocampus hippocampus*. The IUCN Red List of Threatened Species 2017: e.T10069A67618259. <https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T10069A67618259.en>. Accessed on 27 March 2023.
- Woodall, L. C., Otero, F., Miguel, F., Curtis, J. M. R., Garrick, N., Shaw, P. W., and Koldewey, H. J. (2018). A synthesis of European seahorse taxonomy, population structure, and habitat use as a basis for assessment, monitoring and conservation. *Marine Biology*, 165: 19. Springer Berlin Heidelberg. <https://doi.org/10.1007/s00227-017-3274-y>.

## Appendix 1: Supplementary material

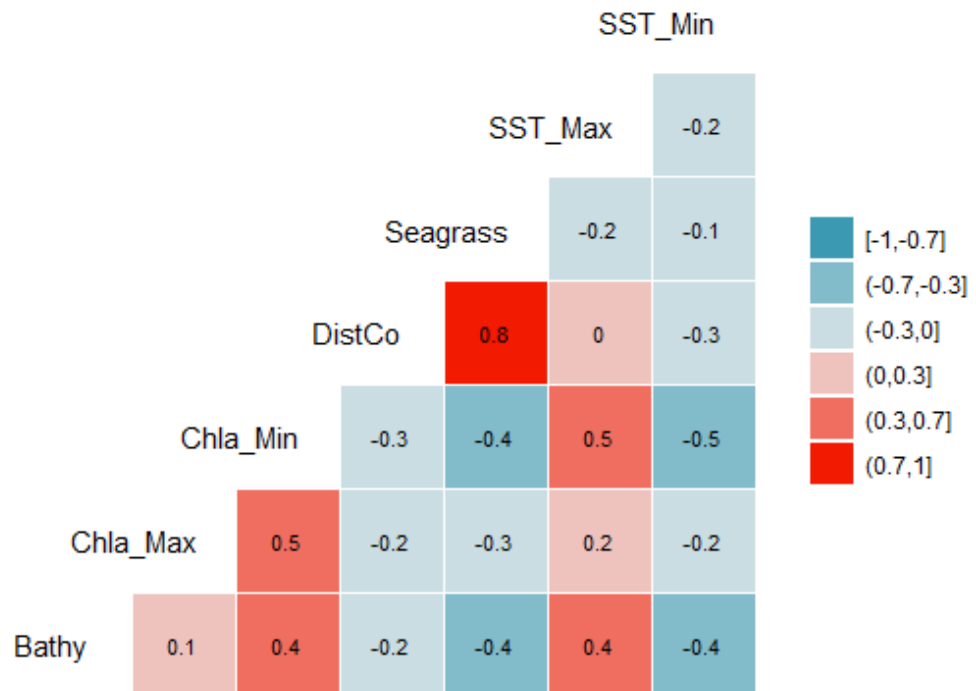


Figure A1.1. Pearson's correlation matrix showing the pairwise correlation scores of the environmental predictor variables used for modelling seahorse habitat distribution. Bathy = bathymetry, Chla\_Max = maximum summer chlorophyll concentration, Chla\_Min = minimum winter chlorophyll concentration, DistCo = distance to coastline, Seagrass = distance to seagrass habitat, SST\_Max = maximum summer sea surface temperature (SST), SST\_Min = minimum winter SST.

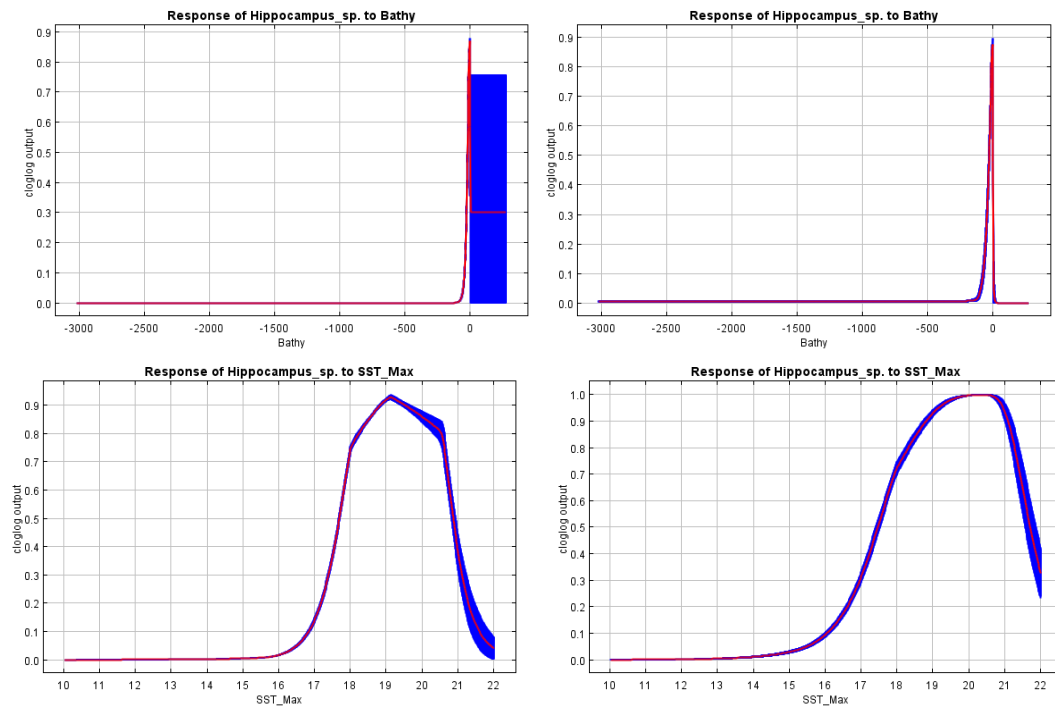


Figure A1.2: Variable response curves averaged over 10 MAXENT model runs for the two environmental predictors with the highest contributions to the MAXENT model for the *Hippocampus* genus. Top to bottom: Bathymetry and summer maximum SST (°C). Left column = how the predicted probability of habitat suitability changes as each environmental predictor is varied, keeping all other environmental predictors at their average sample value. Right column = A MAXENT model created using only the corresponding variable



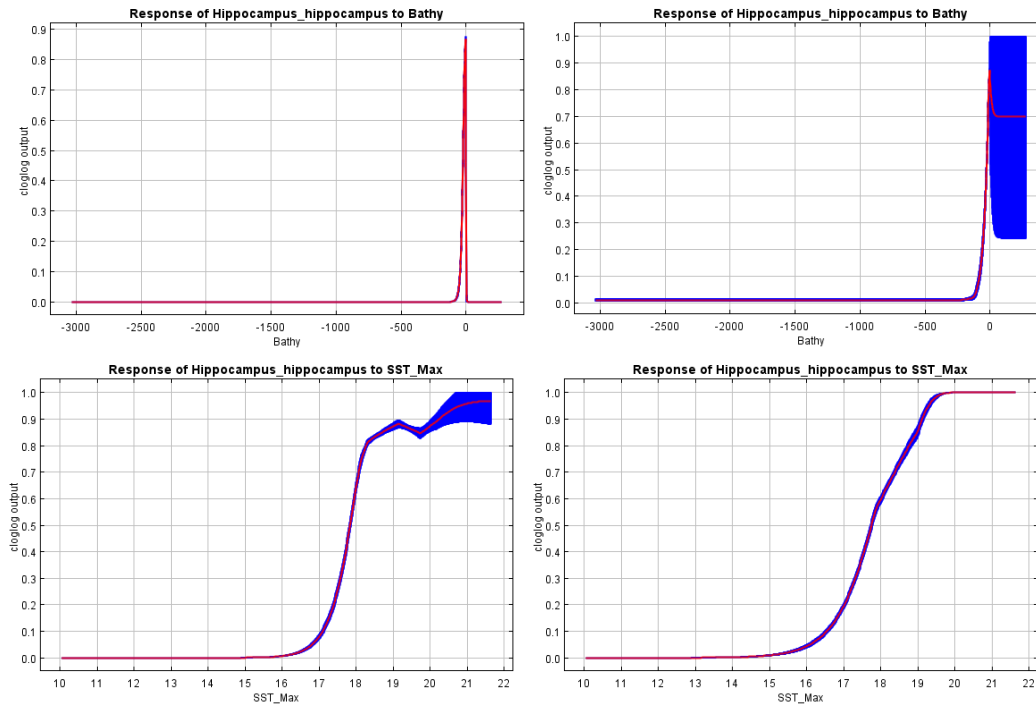


Figure A1.3: Variable response curves averaged over 10 MAXENT model runs for the two environmental predictors with the highest contributions to the MAXENT model for *Hippocampus hippocampus*. Top to bottom: Bathymetry and summer maximum SST (°C). Left column = how the predicted probability of habitat suitability changes as each environmental predictor is varied, keeping all other environmental predictors at their average sample value. Right column = A MAXENT model created using only the corresponding variable

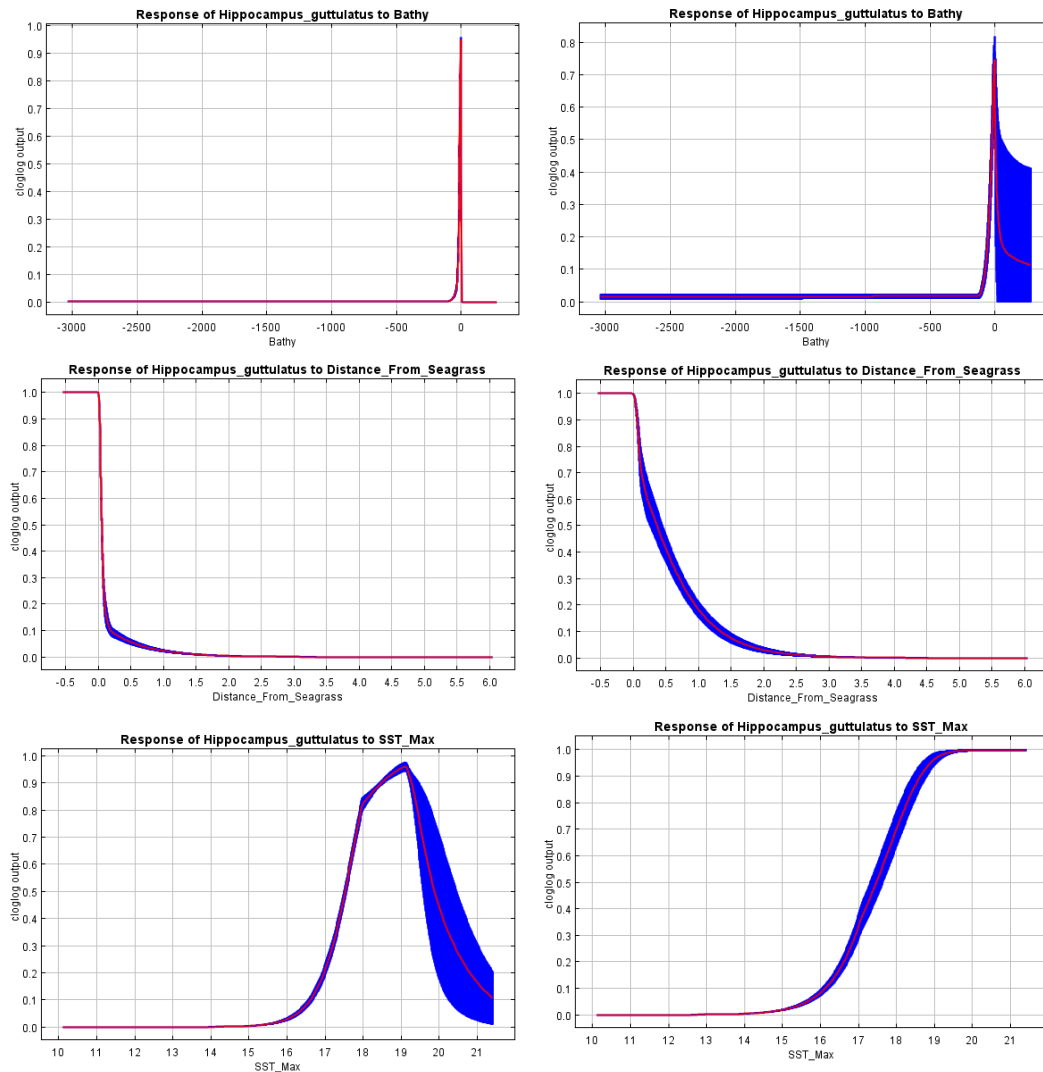


Figure A1.4: Variable response curves averaged over 10 MAXENT model runs for the three environmental predictors with the highest contributions to the MAXENT model for *Hippocampus guttulatus*. Top to bottom: Bathymetry, Distance from seagrass beds (degrees) and summer maximum SST (°C). Left column = how the predicted probability of habitat suitability changes as each environmental predictor is varied, keeping all other environmental predictors at their average sample value. Right column = A MAXENT model created using only the corresponding variable.

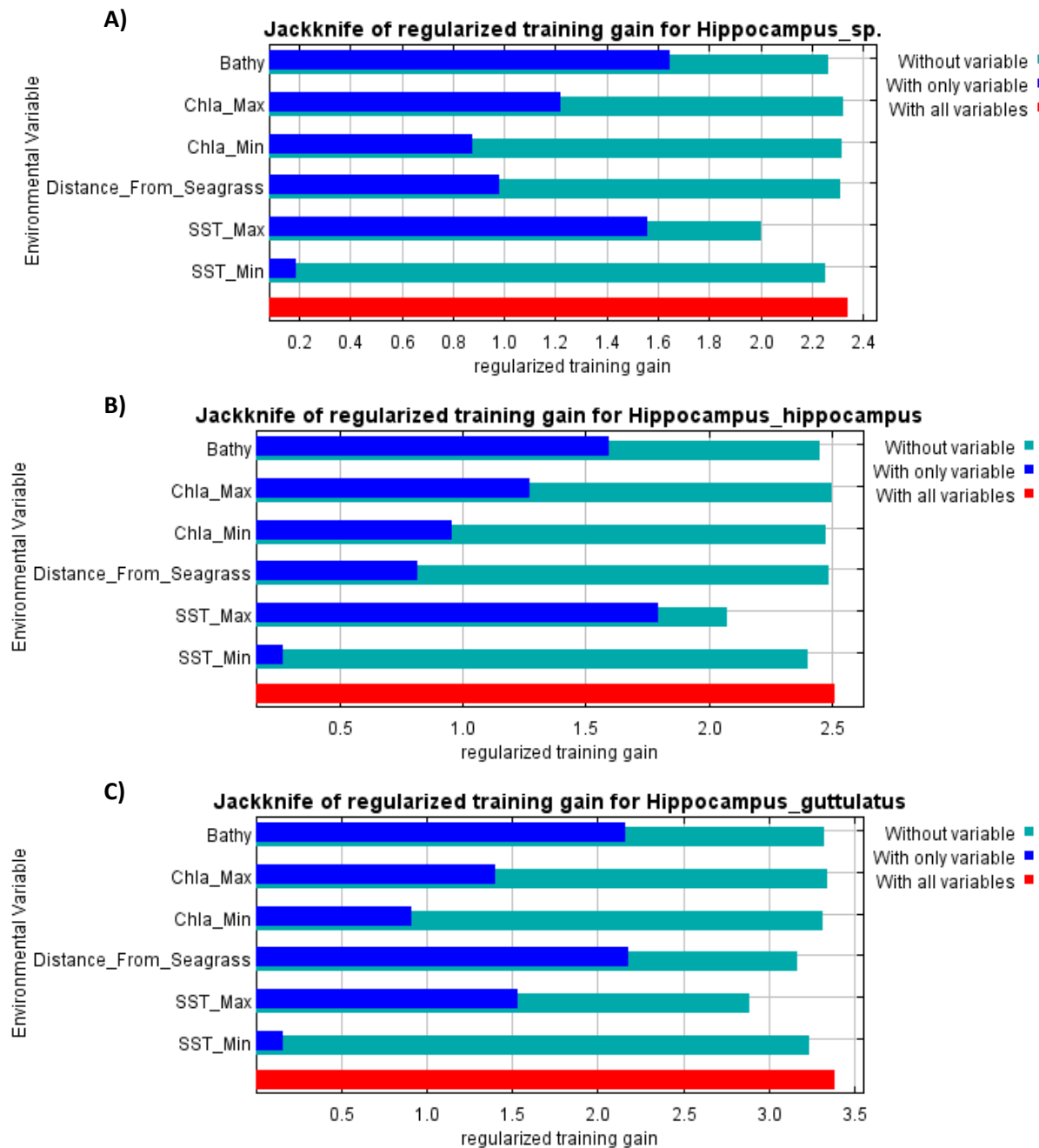


Figure A1.5: Jackknife tests of variable importance for A) the genus, B) *H. hippocampus* and C) *H. guttulatus*.

## Appendix 2: Seagrass data sources

Table A2.1. Seagrass data sources, geographical region, type of data and date of acquisition.

Seagrass data source	Geographical region	Data class	Date
Seagrass cover (Essential Ocean Variable) in Europe - points (2021) and polygons (2019) (EMODnet 2021) <a href="https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/metadata/39746d9c-4220-425c-bc26-7cb3056c36a5">https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/metadata/39746d9c-4220-425c-bc26-7cb3056c36a5</a>	UK/Europe	Point/ Polygon	1/3/2023
National Seagrass Layer (Magic – gov.uk 2022) <a href="https://magic.defra.gov.uk/MagicMap.aspx">https://magic.defra.gov.uk/MagicMap.aspx</a>	England	Polygon	14/2/2023
National Biodiversity Network (NBN) (NBN Atlas 2021, <a href="https://nbnatlas.org/">https://nbnatlas.org/</a> and NBN Atlas Wales 2021, <a href="https://wales-records.nbnatlas.org/">https://wales-records.nbnatlas.org/</a> )	UK/Ireland	Point	1/3/2023
DataMapWales: Environment (Wales) Act Section 7 and OSPAR: Marine Habitats. (Natural Resource Wales 2019) <a href="https://datamap.gov.wales/layergroups/inspire-nrw:MarineBAPOSPARHabitats">https://datamap.gov.wales/layergroups/inspire-nrw:MarineBAPOSPARHabitats</a>	Wales	Point/ Polygon	1/3/2023
Seagrass Spotter (Project Seagrass 2023) <a href="https://seagrassspotter.org/">https://seagrassspotter.org/</a>	Worldwide	Point	23/2/2023
Seagrass bed acoustic surveys (Jenkin et al. 2022a,b)	St Austell Bay & Gerrans Bay	Polygon	9/2/2023

For NBN Atlas Wales (2021), seagrass records were obtained from Porcupine Natural History Society, National Resource Wales (NRW), South East Wales Biodiversity Records Centre (SEWBreC), Seasearch Marine Surveys and the Joint Nature Conservation Committee (JNCC). For NBN Atlas (2021) records were obtained from the following data contributors: Bristol Regional Environmental Records Centre, Botanical Society of Britain & Ireland and Biological Records Centre, Ulster Wildlife, Malcolm Storey personal records, JNCC, Marine Biological Association, NRW, Natural England, Department of Agriculture Environment and Rural Affairs, National Trust, NatureScot, Fife Nature Records Centre, Manx Biological Recording Partnership, Yorkshire Naturalists' Union, Royal Botanic Garden Edinburgh, SEWBreC, Suffolk Biodiversity Information Service, The Wildlife Trusts, The Rock Pool Project database, Scottish Wildlife Trust, Argyll Biological Records Centre, Highland Biological Recording Group, Natural History Museum, Northern Ireland Environment Agency, Centre for Ecology and Conservation, Isle of Wight Local Records Centre, Norfolk Biodiversity Information Service, Environment Agency, Highland Seashore Project Dataset, National Plant Monitoring Scheme, APHOTOMARINE (David Fenwick), National Trust for Scotland and Project Seagrass.



# Centre for Environment Fisheries & Aquaculture Science



## About us

We are the Government's marine and freshwater science experts. We help keep our seas, oceans and rivers healthy and productive and our seafood safe and sustainable by providing data and advice to the UK Government and our overseas partners.

We are passionate about what we do because our work helps tackle the serious global problems of climate change, marine litter, over-fishing and pollution in support of the UK's commitments to a better future (for example the UN Sustainable Development Goals and Defra's 25 year Environment Plan).

We work in partnership with our colleagues in Defra and across UK government, and with international governments, business, maritime and fishing industry, non-governmental organisations, research institutes, universities, civil society and schools to collate and share knowledge.

Together we can understand and value our seas to secure a sustainable blue future for us all, and help create a greater place for living.

### Head office

Pakefield Road  
Lowestoft  
Suffolk  
NR33 0HT  
Tel: +44 (0) 1502 56 2244  
Fax: +44 (0) 1502 51 3865

### Weymouth office

Barrack Road  
The Nothe  
Weymouth  
DT4 8UB

Tel: +44 (0) 1305 206600  
Fax: +44 (0) 1305 206601

Innovative, world-class science is central to our mission. Our scientists use a breadth of surveying, mapping and sampling technologies to collect and analyse data that are reliable and valuable. We use our state-of-the-art Research Vessel Cefas Endeavour, autonomous marine vehicles, remotely piloted aircraft and utilise satellites to monitor and assess the health of our waters.

In our laboratories in Lowestoft and Weymouth we:

- safeguard human and animal health
- enable food security
- support marine economies.

This is supported by monitoring risks and disease in water and seafood; using our data in advanced computer models to advise on how best to manage fish stocks and seafood farming; to reduce the environmental impact of man-made developments; and to respond to serious emergencies such as fish disease outbreaks, and to respond to oil or chemical spills, and radioactivity leaks.

Overseas, our scientists currently work in Commonwealth countries, United Kingdom Overseas Territories, South East Asia and the Middle East.

Our customer base and partnerships are broad, spanning Government, public and private sectors, academia, non-governmental organisations (NGOs), at home and internationally.



[www.cefas.co.uk](http://www.cefas.co.uk)

