Marine
Management
Organisation

Spatial models of essential fish habitat (South Inshore and Offshore marine plan areas): Technical annex

December 2013

This page has been intentionally left blank.

Spatial models of essential fish habitat (South Inshore and Offshore marine plan areas)

Technical annex

MMO Project No: 1044

Project funded by: The Marine Management Organisation

Project contractor: Institute of Estuarine and Coastal Studies, University of Hull

Disclaimer

This report contributes to the Marine Management Organisation (MMO) evidence base which is a resource developed through a large range of research activity and methods carried out by both the MMO and external experts.

The opinions expressed in this report do not necessarily reflect the views of the MMO nor are they intended to indicate how the MMO will act on a given set of facts or signify any preference for one research activity or method over another. The MMO is not liable for the accuracy or completeness of the information contained nor is it responsible for any use of the content.

This report should be cited as

MMO (2013). Technical annex. Spatial models of Essential Fish Habitat (South Coast Inshore and Offshore Marine Plan Areas). A report produced for the Marine Management Organisation by the Institute of Estuarine and Coastal Studies, 47pp. MMO Project No: 1044. ISBN: 978-1-909452-21-3.

First published November 2013.

Contents

Figures

Tables

List of acronyms

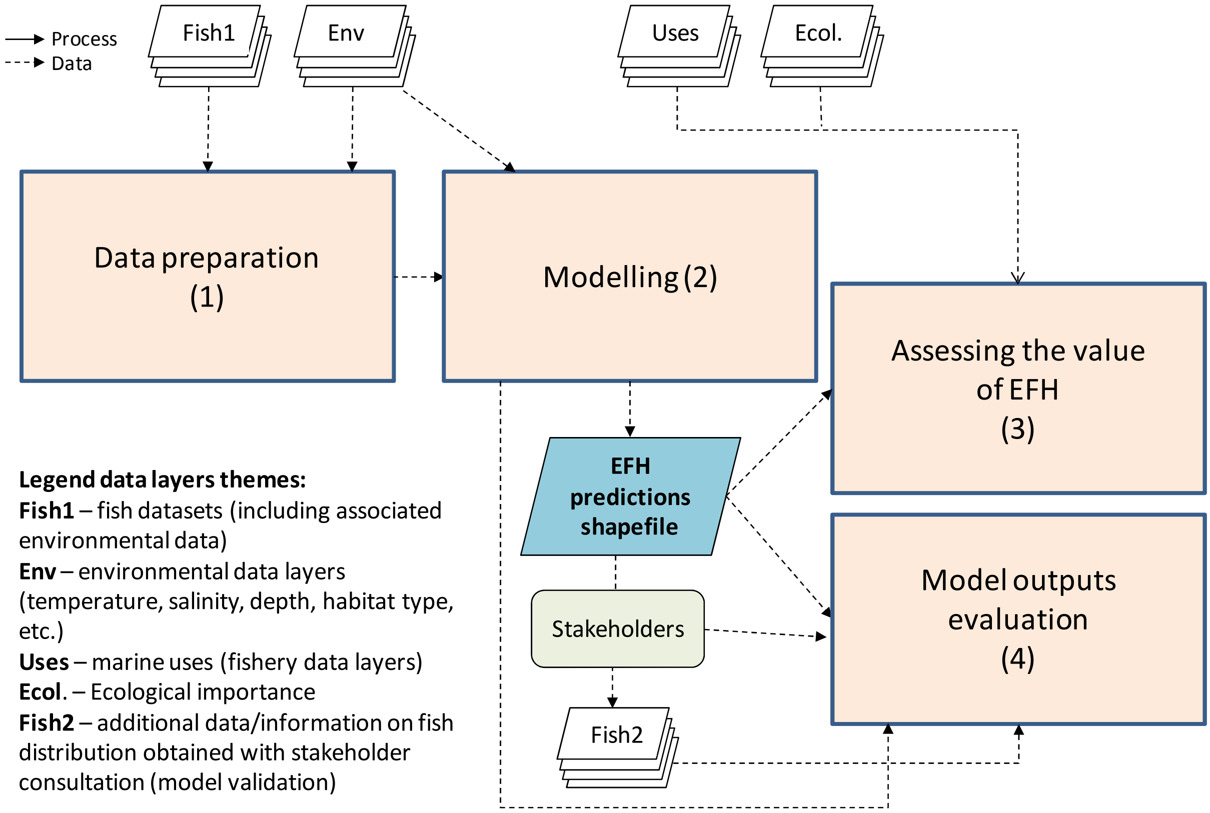
BTS	Beam Trawl Survey
CHARM	Channel Integrated Approach for Marine Resource Management
CPUE	Catch Per Unit Effort
EFH	Essential Fish Habitat
EUNIS	European Nature Information System
ICES	Institute of Estuarine and Costal Studies (University of Hull)
IFCA	Inshore Fisheries and Conservation Authorities
IHLS	International Herring Larval Survey
JNCC	Joint Nature Conservation Council
MCZ	Marine Conservation Zone
MMO	Marine Management Organisation
TidE	Tidal Current Energy
WavE	Wave Energy
WGEGGS	Working Group on North Sea Cod and Plaice Egg Surveys

1. Annex Contents

This annex provides technical information regarding the data processing and analysis for the production of EFH maps in the South Inshore and Offshore Marine Plan Areas. The information provided in this annex complements/integrates the information on methods provided in the Final Report.

The overall methodological approach to the project is summarised by the flow chart in Figure 1.

Figure 1: Flow diagram of the main project phases.



2. Fish Survey Data Selection

The collation of fish data was focused on scientific fish surveys. Relevant fish survey data were identified based on the information obtained from recent projects (Defra project MB5301, Cefas 2010, Ellis *et al*., 2012; CHARM project) and enquiries made with relevant data providers (ICES, Cefas, Ifremer, Environment Agency). Multiple datasets were explored, as no single scientific ship-based survey is currently carried out to cover consistently both the Eastern and Western English Channel.

Not all the data collated were used in the project. Due to the short duration of the project a selection of most relevant data was made. This selection was mainly driven by completeness and consistency of the data (in terms of survey design and methods), although, in some cases, time constrains also occurred (when data were received late). The full list of the data collated is provided in Table 1, together with the indication of which datasets have been used in the analysis and the reasons why the other datasets were not used.

Although data for the same species/life stage could be derived from different survey datasets, data from different surveys were not combined into a single dataset (hence preventing a single analysis) for a certain species as the use of different sampling gear and strategies made these data not comparable. Therefore, the best dataset for a species/life stage was selected from among those available based on the dataset size (number of observations) and on considerations about the confidence on the species catch data related to the sampling method employed and its sampling selectivity and efficiency with respect to the species/life stage under consideration.

Table 1: Fish survey data identified with coverage of the study area.

Fish data	Source	Survey/data information	Used for model calibration
UK Eastern English Channel Beam Trawl Survey (BTS)	ICES, online fish trawl surveys database (DATRAS) (public access)	Survey series starting in 1989 and ongoing, carried out by Cefas. Fishing during July/August (Quarter 3) over an allocated area of the Southern North Sea and Eastern English Channel using a standard grid. Station, catch, length (all species) and biological data (selected species) for each of the annual surveys covering the Southern North Sea and Eastern English Channel using research vessels and 4m beam trawl in support of EU data regulations and as part of a research program coordinated by ICES. The primary aim was to assess the relative abundance of pre-recruit plaice and sole in ICES Division VIId (with extension to southern North Sea in 1995); consequently most of the sampling is concentrated in areas that are nursery grounds for these species. Additional aims include collection of water temperature and salinity and acoustic data. (Data 2000-2012 within English Channel: N=852)	Yes

Fish data	Source	Survey/data information	Used for model calibration
ICES International Herring Larval Survey (IHLS)	ICES, online fish eggs and larvae database (public access)	Survey series starting in 1967 and ongoing, with combined effort of different countries (UK, France, Germany, Netherlands), as part of a research programme coordinated by ICES. Surveys carried out in specific periods and areas, following autumn and winter spawning activity of herring from north to south (December/January in the English Channel), with double oblique hauls of high-speed plankton sampler deployed on a fixed stations grid from research vessels. Data on herring larvae CPUE (individuals per square meter) per haul per length class (small, medium, large larvae), sampling methods (e.g., gear type, hauling duration) and environmental conditions measured during sampling (e.g., depth, water temperature, salinity). The main purpose of the international herring larval surveys (IHLS) programme is to provide quantitative estimates of herring larval abundance, which are used as a relative index of changes of the herring spawning-stock biomass in the assessment. (Data 2000-2011 within English Channel: N=1503)	Yes
ICES North Sea Cod and Plaice Egg Surveys in the North Sea (WGEGGS)	ICES, online fish eggs and larvae database (public access)	Survey series conducted in winter (December/January) 2003/04 and 2008/09, with combined effort of different countries (France, Germany, Netherlands), as part of a research programme coordinated by ICES. Use of different sampling strategies (e.g., double oblique hauls of high-speed plankton sampler, surface sampling with continuous underway fish egg sampler) Station, egg abundance (eggs per haul per species), egg stage (all species) and length (selected species) data for each of the annual surveys covering the North Sea, down to Eastern English Channel using research vessels and different sampling gears. The database contains also the haul information data, position, time, duration, filtered water volume, depth, temperature and salinity. The surveys were originally directed at cod and plaice, but also supply data of other winter spawning North Sea fish. (Data 2003/4 and 2008/09 within English Channel: N=172 with high-speed plankton sampler 280um mesh; N=93 (Jan 2009 only) with continuous underway fish egg sampler)	Yes
French groundfish survey in the Eastern English Channel (FR_CGFS)	Ifremer	Survey series starting in 1989 and ongoing (October, Quarter 4), carried out by Ifremer using GOV trawler. Surveys as part of the ICES programme of International Bottom Trawl Surveys in the Western and Southern Areas (WS-IBTS). These surveys aim to provide consistent and standardized data for examining spatial and temporal changes in the distribution and relative abundance of fish and fish assemblages and of the biological parameters of commercial fish species for stock assessment purposes. Fish CPUE per haul per species per length class. (Data 2000-2010 within English Channel: N=1111)	No (incomplete data obtained)
Cefas Southern North Sea and English Channel Sole Egg Survey	Cefas	Four cruises were undertaken in 1991 (Spring) collecting 70-80 samples to estimate the spawning stock biomass of the sole (*Solea solea*) in the English Channel and southern North Sea. Abundance / density of fish eggs and fish larvae from plankton tows. Eggs from sole assigned to developmental stages. Associated environmental data (temperature salinity).	No (no data after 2000)

Fish data	Source	Survey/data information	Used for model calibration
National Fish Population Dataset (inshore fish data)	Environment Agency	Collation of data obtained by the EA between 2004 and 2012 from different fish surveys of inshore/estuarine water bodies (Adur, Arun, Cuckmere, Dart, Exe, Lime Bay West, Pool Harbour, Rother, Southampton Water) for WFD assessment purposes. Surveys combine different methods (e.g., beam trawls, fyke nets, otter trawls, seine nets) and sampling months (March to December). Station, catch (counts), length for each survey. Additional information on sampling event (gear used, date, effort, the latter not recorded for all data and with inconsistencies) (Data 2000-2010 within English Channel: N=730)	No (not directly comparable data (multiple methods/ strategies); missing/ inconsistent data on sampling effort; need further processing/analysis before it can be used; data received late)
UK South West Beam Trawl Survey (Q1SW)	Cefas	Survey series starting in 2006 and ongoing, carried out by Cefas. Fishing during March (Quarter 1) over an allocated area (with random-stratified design) covering the ICES Division VII e-h (including Western English Channel) using two 4m beam trawls (with different mesh size). Station, catch, length and biological data for each of the annual surveys in support of EU data regulations and as part of a research program coordinated by ICES. (Data 2006-2013 within English Channel: N=1037)	No (non comparable catch data with those from BTS in Eastern English Channel (different methods and season); data received late)
Cefas Young Fish Surveys in South Coast areas	Cefas	Survey series carried out between 1981 and 2006 by Cefas. Fishing inshore with 2m scientific beam trawl (with 4mm mesh liner) in September each year. Surveys aim to provide indices of abundance of small demersal fish, in particular juvenile 0-group and 1-group plaice and sole, prior to their recruitment to the fishery. The data is in support of the EU Data Collection Regulation. Station, catch, length data for each of the annual surveys. (Data 2000-2006 within English Channel: N=496)	No (non comparable catch data with those from BTS in Eastern English Channel (different methods and season); data received late)
Cefas Small Pelagic Fish Western Channel and Celtic Sea plankton survey (PELTIC11)	Cefas	Cefas surveys in the Western Channel and the Celtic Sea targeting small pelagic fish. Surveys in May/June 2011, using multiple methods (sandeel trawl, otter trawl, rosette sampler, drop nets, high speed manta trawl, sounders). Station, catch, length and biological data, as well as associated oceanographic data. (Data 2011 within English Channel: N=56)	No (time constraints)

It is acknowledged that Table 1 does not represent an exhaustive list of all the fish data available for the study area. For example, a wide range of fishery-dependent data is available for the study area. This type of data (in particular, logbook data on landings) were not considered suitable for the purpose of this project, due to their bias towards commercially relevant species and fish size, possible issues associated with the taxonomic identification of the catches, and also due to the low resolution (ICES rectangle) at which these data are available as well as inaccuracy as the data relate to port at which they are taken rather than the precise place of capture (Ellis *et al*., 2010a). It is acknowledged that other types of fishery-dependent data exist (e.g., self sampling or skippers logs) that could be used, although issues on the quality of these data (e.g., associated to taxonomic standards) might make difficult their integration with data from fish surveys. An additional potential source of data has been identified in Cefas at-sea observer programme, whereby Cefas scientific observers are placed on a sample of UK-registered commercial fishing vessels in each quarter of the year to estimate the quantities and sizes of each species discarded at sea (the size composition of the retained catch is also recorded). The Cefas at-sea observer programme covers a wide range of gears, areas and times, and since 2002, the observer programme has covered most areas in ICES Areas IV (North Sea) and VII (English Channel and western waters). These data were not considered in this project as their existence was not known at the time of the data collation[1]. however this type of data is potentially valuable for the integration of fish data in the Channel area (e.g., for the validation of the EFH models), therefore they should be considered in future work.

2.1 Life stages identification and fish species selection

The BTS dataset was analysed in order to distinguish catches of juveniles (indicative of nursery grounds) and adults (indicative of the general habitat used by the species), based on the size (length) of individuals. Size thresholds for identifying juveniles of a species have been derived from available literature (Stephens *et al*., 2010; Ellis *et al*., 2010b, 2012; Lauria *et al*., 2011; Froese and Pauly 2013). Different size thresholds can be identified for a species, based on the juvenile stages taken into account (from 0-group fish only to the wider identification of immature individuals). Whenever possible, the most restrictive threshold was considered to increase the confidence in the ability of the distribution of these life stages to represent primary nursery habitats. When published thresholds were too restrictive (i.e. with too few data available for juveniles), length frequency histograms of fish data in the analysed dataset were examined to identify the length threshold for the first cohort(s). With regard to datasets for larvae (IHLS) and eggs (WGEGGS), the data for the earlier larval stages and egg stages (EG1) were considered to increase the confidence in the ability of the distribution of these life stages to represent spawning habitats.

Based on the initial list of species identified as being relevant to the project (Table 1 in Final Report), fish species were then prioritised and selected for the analysis based on the:

- relevance of the species for commercial exploitation (within the English Channel, as per results of projects MMO1011 and CHARM, and in general, as per FishBase) and for conservation (Ellis *et al*., 2010a) (see Table 1 in Final Report)
- confidence in the fish data, based on:
    - confidence in the survey design/method to be able to capture abundance/occurrence of a species/life stage depending on gear selectivity (as informed by Elliott and Hemingway, 2002; Ellis *et al*., 2010a, 2010b) and on season of sampling (in relation to the seasonality of occurrence of life cycles/spawning period, as informed by Ellis *et al*., 2012; Froese and Pauly, 2013)

---

[1] Data enquiries were sent to Cefas, asking for provision of specific datasets and also for suggestions on other available data possibly useful to the project, but these data were not identified during data collation.

    - confidence in the survey design to be able to capture abundance/occurrence of a species/life stage depending on the spatial coverage of its habitats (as informed by Ellis *et al*., 2010a)
    - known limitations in the taxonomic standards applied to the data (as informed by Ellis *et al*., 2010a)
    - confidence in the ability of the identified life stages to represent the distribution of EFH (depending on how life stage has been identified; e.g. early vs. later egg stages; 0-group vs. immature individuals)
- data availability (frequency of occurrence of different life stages in the dataset).

Based on these criteria, scores (1/low to 3/high) were assigned to the different species/life stages and combined to obtain a total score to identify priority species for the analysis.

Species which could not be separated into different life stages (due to lack of information on size thresholds or very low frequency of occurrence in the data) were excluded from the analysis and the resulting data would not allow the identification of EFH. Also it has been decided not to use data on cod and whiting eggs (in WGEGGS) due to the small size of the dataset available for these stages (N=93), the relatively low frequency of occurrence, and the lack of data for early egg stages (EG1) leading to a lower confidence in the identification of spawning habitats using later or unidentified development stages (due to the increased probability of pelagic eggs being transported away from the spawning areas).

3. Environmental Data Selection

The collation and selection of environmental datasets for the project was based on the following criteria:

- data availability for the main environmental variables relevant to fish species (as described in the Final Report)
- full spatial coverage of the study area, and, if possible, of the wider area in the English Channel where fish survey stations are located
- data layers at a spatial resolution equal or higher than the spatial resolution associated with fish data
- for variables showing a marked seasonal and inter-annual variability (e.g., oceanographic data, like water temperature), data layers available for different seasons and years, covering the temporal extent/resolution of the specific fish survey dataset.

As a result, the following datasets were used:

Data layer (Source)	Description
Bathymetry (EMODnet)	For each maritime region bathymetric survey data and aggregated bathymetry data sets have been collated from public and private organizations. These have been processed and quality controlled and used to produce a regional Digital Terrain Model (DTM) with a grid size of .25 minute * .25 minute. The DTM values have been determined from the combination of bathymetric survey data (high resolution data sets from single and multibeam surveys), composite data sets produced and delivered by a number of external data providers such as Hydrographic Offices derived from their internal bathymetric database and based upon historic surveys, and GEBCO 30” gridded data, used to complete area coverage in case there are no survey data or composite data sets available to the partners.
Seabed substratum type (EMODnet)	The current map is collated from more than 200 separate sea-bed substrate maps provided by different partners (based on sediment sampling, multibeam echosounder, Side Scan Sonar, bathymetric and seismic surveys). Each partner harmonised their available sea-bed substrate data according to a common classification scheme (modified Folk triangle). Data are provided at a 1:1 million scale (the smallest cartographic unit (polygon) on the map being about 4 km^2).
JNCC EuSeaMap North and Celtic Seas Energy data layers (EUSeaMap)	Under a specific contract for the EUSeaMap project, energy layers were produced for the North and Celtic seas. Energy layers are built using data from National Oceanographic Centre (NOC) wave (ProWAM at a resolution of 12.5km) and current models (the CS20, CS3 and NEA models at resolutions of 1.8km, 10km and 35km respectively). These were all processed to populate a 1km resolution grid, with a high (~300m) bespoke resolution DHI Spectral Wave model used to augment the coastal areas where the ProWAM model resolution was inadequate. Data cover the EU Continental Shelf with variable resolution (0.1 to 35 kilometres). Wave and current data were combined to produce the input energy layer for the EUSeaMap model after classification into energy categories. No confidence estimates are available for the original data layers, but uncertainty in the class boundaries was assessed.

Data layer (Source)	Description
Habitats Directive Annex 1 Reefs (JNCC)	This is a collation of all data identifying surveyed Annex I reefs in UK waters out to the edge of the UK continental shelf. Data sources include Natural England, Countryside Council for Wales, Scottish Natural Heritage, Joint Nature Conservation Committee, British Geological Survey and National Oceanography Centre. This dataset shows both potential and known Annex I reefs. Potential reefs include areas where seismic surveys show that there is bedrock up to 0.5m below the seabed (and there is therefore a possibility of exposed bedrock). It should be noted that areas which are dominated by a sand veneer are also classed as 'potential reefs', therefore the mapped occurrence of potential reefs is likely to overestimate actual reef habitats
Marine Water Column Features (JNCC)	This dataset describes aspects of the water column over the UKCS. 4 shapefiles, one for each season (Autumn, Winter, Spring, and Summer) are given. It describes stratification and mixing of water types. Source data for these data layers were obtained from a number of hydrographical data sets were obtained from the Proudman Oceanographic Laboratory (POL) and these datasets were used within the UKSeaMap project (2006).
Global Ocean OSTIA Sea Surface Temperature and Sea Ice analysis REPROCESSED (1985-2007) (EU project My Ocean)	The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) global Sea Surface Temperature Reanalysis product provides daily gap-free maps of sea surface temperature (referred to as an L4 product) at 0.05deg.x 0.05deg. Horizontal resolution, using in-situ and satellite data from infra-red radiometers. The OSTIA system is run by the UK Met Office. The OSTIA reanalysis uses satellite data provided by the Pathfinder AVHRR project and reprocessed (A)ATSR data together with in-situ observations from the ICOADS data-set, to determine the sea surface temperature. It also uses reprocessed sea-ice concentration data from the EUMETSAT OSI-SAF. The reanalysis data is available from 1985-2007, providing full time series processed consistently with up-to-date knowledge on satellite sensor calibration, characterization and attitude, complete (as far as possible) ancillary data sets, latest versions of models and algorithms. The analysis product has been validated through calculation of mean and RMS statistics of observation-minus-background and observation-minus-analysis. Inter-comparisons with other historical data-sets, e.g. Reynolds OI, HadISST, have been carried out.
Pan European Seas, Ocean Optics Products (monthly average) Reprocessed (1997-2010). (EU project My Ocean)	Ocean Colour "Optics" products are derived from remote sensing (MODIS-Aqua and SeaWIFS sensors). The spectral variations in the light leaving the water surface are related to inherent optical properties (IOPs) including the phytoplankton absorption coefficient (APH). These IOPs can be interpreted in terms of concentrations of optically-significant constituents in the water. Corrections to remove the atmospheric contribution are applied and validation with in situ data has been carried out. The reprocessed data layer covers the period 1997-2010, providing full time series processed consistently with up-to-date knowledge on satellite sensor calibration, characterization and attitude, complete (as far as possible) ancillary data sets, latest versions of models and algorithms. Indication of a possible update is given, but there is no commitment that this will actually happen. Data are provided at a high resolution (2km).

As regards bathymetry data, it is acknowledged that data at a higher resolution than that of EMODnet data layers exist (Defra DEM bathymetry), but these data are spatially restricted to the UK territorial waters. Their use would have limited the analysis of fish datasets to this area, with a consequent reduction in the dataset size used for the model calibration, thus limiting the power of the modelling analysis. In turn, EMODnet bathymetry provides full coverage of English Channel. Although the resolution of these data layers is lower than DEM, it is still higher than the spatial resolution associated with fish data, hence making this dataset suitable for selection (as per criteria described above).

As for habitat data, it is of note that data layers on seabed habitat modelling and mapping in European waters (with associated confidence estimates) are available from the EUSeaMap (2011). Eunis habitat classification (level 3) of seabed is available in this dataset, with habitat types being defined on the basis of abiotic variables including depth zone, substratum type and energy at the seabed. Habitat classification has been obtained from data layers for these variables, by identifying ecologically-relevant thresholds to environmental variables and combining the classified environmental data in GIS by means of multicriteria evaluation (Cameron and Askew, 2011). Rather than using the final Eunis habitat classification provided by EUSeaMap, the choice of using the original environmental datasets at the basis of the habitat classification was made for this project. The use of basic environmental variables rather than combined ones (like Eunis habitat) is considered the best choice to model the relationship between fish distribution and environmental drivers. In fact, this allows discriminating the possible different importance of different variables in affecting the distribution of fish life stages in the marine environment, whereas this would not be possible if variables are combined a priori in a habitat classification. As a result, the data layers used to originate EUNIS habitat classification were considered, namely substratum type and energy at the seabed associated to either waves or tidal currents.

With regard to energy data layers, it is of note that the original data were considered (i.e., continuous energy data layers built on wave and current models calibrated by the National Oceanography Centre, with energy measured in N m^{-2} units) rather than the processed data (after thresholding) that were used for Eunis habitat classification. The use of continuous rather than categorised data (where energy values were distinguished into high, moderate and low categories; Cameron and Askew, 2011) was deemed preferable for the modelling purposes of the project. Using continuous variables, in fact, allows identifying thresholds that are ecologically-relevant to the specific life stage considered in the EFH modelling, rather than being defined *a priory* (based on general literature, not addressing specifically ecological relevance to fish). For the same reason, bathymetry data layers (as obtained from EMODnet) were preferred to the use of depth zones available in EUSeaMap.

It is acknowledged that, using the original energy data layers caused limitations in the ability to assess the confidence on these data, as metadata for the input data layers available in EUSeaMap are mostly related to the processed data layers, with confidence estimates mostly associated to the thresholding results (confidence in boundaries using fuzzy thresholds; Cameron and Askew, 2011).

4. Data Layers Geo-processing

The fish data (BTS, IHLS and WGEGGS datasets) were imported into GIS as point data. Where trawl data was supplied as a series of start and end points the average position for each trawl was calculated and plotted. Fish data records which fell inside the study area (English Channel with longitudinal bounds of -3.690768 and 1.728316 (WGS84)) were extracted from the wider dataset for use in the statistical model calibration. Buffers of 2.5km radius were created for each fish data record. These were designed to cover the area of each 5km trawl line but were also considered a suitable resolution for the eggs and larvae data.

Within each buffer area, the following environmental parameters were extracted from the environmental data layers:

Variable	Theme	description	Type
WDepth	bathymetry	**Original data**: from EMODNET bathymetry map; mean depth (m below surface) produced by a regional Digital Terrain Model (DTM) with a grid size of .25 minute * .25 minute. **Derived data:** mean value calculated within 2.5km from the fish survey station (mean location)	continuous
DomMix	water column, mixing type	**Original data**: from JNCC, Marine water column features (Seasonal) map; vector file with polygons for types of water column mixing. Seasonal source maps matching seasonality of fish data (Summer - BTS; Winter - IHLS and WGEGGS) were used **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location) and exported in excel where dominant type was calculated; dominant type allocated to the buffer area. Following types are included: 1 (a) = well-mixed ROFI (Region of Freshwater Influence); 2 (b) = well-mixed shelf water; 3 (c) = weakly stratified ROFI; 4 (d) = weakly stratified shelf water. This variable was considered a proxy for salinity, with also information on the mixing of water masses of marine and continental origin.	categorical
SST	water column, Sea Surface Temperature (SST)	**Original data**: from maps in EU project My Ocean; APH values at 2km horizontal resolution. Seasonal mean values matching seasonality and year of fish data (July/August - BTS; January - IHLS and WGEGGS) were used. Temperature values were given in Kelvin degrees. **Derived data:** Temperature values transformed into Celsius degrees; mean value of temperature calculated within 2.5km from the fish survey station (mean location)	continuous
APH	water column, Phytoplankton absorption coefficient (APH)	**Original data**: from maps in EU project My Ocean; APH values at 2km horizontal resolution. Monthly mean values matching temporal timeframe of fish data (January - IHLS and WGEGGS) were used within the 2000 – 2012 timeframe **Derived data:** maximum value of APH calculated within 2.5km from the fish survey station (mean location)	continuous
TidE	substratum, tide energy	**Original data**: from JNCC EUSeaMap; Tidal energy (N/m2); data cover the EU Continental Shelf with variable resolution (0.1 to 35 kilometres). **Derived data:** mean value calculated within 2.5km from the fish survey station (mean location)	continuous

Variable	Theme	description	Type
WavE	substratum, wave energy	**Original data**: from JNCC EUSeaMap; Wave energy (N/m2); data cover the EU Continental Shelf with variable resolution (0.1 to 35 kilometres). **Derived data:** mean value calculated within 2.5km from the fish survey station (mean location)	continuous
M-sM	substratum, type, Mud to sandy mud	**Original data**: from JNCC EUSeaMap; vector file with polygons for seabed sediment types, the smallest cartographic unit (polygon) on the map being about 4 km^2 **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location); proportional area (0-1) calculated with respect of the total sum of polygon area within the buffer; proportion of area covered by Mud to sandy mud sediment type.	continuous
S-mS	substratum, type, Sand to muddy sand	**Original data**: from JNCC EUSeaMap; vector file with polygons for seabed sediment types, the smallest cartographic unit (polygon) on the map being about 4 km^2 **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location); proportional area (0-1) calculated with respect of the total sum of polygon area within the buffer; proportion of area covered by Sand to muddy sand sediment type.	continuous
Cs	substratum, type, Coarse sediment	**Original data**: from JNCC EUSeaMap; vector file with polygons for seabed sediment types, the smallest cartographic unit (polygon) on the map being about 4 km^2 **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location); proportional area (0-1) calculated with respect of the total sum of polygon area within the buffer; proportion of area covered by coarse sediment type.	continuous
Mx	substratum, type, Mixed sediment	**Original data**: from JNCC EUSeaMap; vector file with vector file with polygons for seabed sediment types, the smallest cartographic unit (polygon) on the map being about 4 km^2 **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location); proportional area (0-1) calculated with respect of the total sum of polygon area within the buffer; proportion of area covered by mixed sediment type.	continuous
R	substratum, type, Rock or other hard substrata	**Original data**: from JNCC EUSeaMap; vector file with polygons for seabed sediment types, the smallest cartographic unit (polygon) on the map being about 4 km^2 **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location); proportional area (0-1) calculated with respect of the total sum of polygon area within the buffer; proportion of area covered by Rock or other hard substrata type.	continuous
DomSubst	Dominant substratum type	**Original data**: from JNCC EUSeaMap; vector file with polygons for seabed sediment types, the smallest cartographic unit (polygon) on the map being about 4 km^2 **Derived data:** area of different polygons (types) calculated within 2.5km from the fish survey station (mean location); data exported in excel where proportional area (0-1) was calculated with respect of the total sum of polygon area within the buffer and dominant type was identified; dominant type allocated to the buffer area. Dominant types were coded as below: 1=M-sM, 2=S-mS, 3=Cs, 4=Mx, 5=R (substratum type	categorical

Variable	Theme	description	Type
		codes as per variables above)	
Reef	substratum, Presence-absence of reef	**Original data**: from JNCC, Habitats Directive Annex 1 Reefs; vector file with polygons for reef presence-absence category, with also information on potential presence. **Derived data:** Presence-absence of reef calculated within 2.5km from the fish survey station (mean location). Reef presence category takes into account also level of confidence in the reef map: 0 (a) =no reef 1 (b) =reef potentially present (lower confidence) 2 (d)=reef present	categorical

Fish records which had been assigned a full suite of environment variables were exported and taken forward for statistical modelling. Where insufficient data was available for modelling, records were discarded. In the case of Cod/Whiting (within WGEGGS dataset), insufficient data records were retained and the entire data set was discarded.

5. Statistical And Geo-spatial Modelling

5.1 Data exploration

Before applying any statistical analysis or modelling, the datasets (fish survey data and associated environmental predictors) were explored and interrogated to determine the degree of linearity, homogeneity, normality, collinearity and the distribution of the data points and zeros among the various factors (e.g. life stage, sampling time, depth range) (Zuur *et al*., 2009).

Collinearity between explanatory variables in the dataset has been investigated and redundant/collinear environmental variables excluded from the analysis. Collinearity occurs when there are high correlations among predictor variables, leading to unreliable and unstable estimates of the model parameters (Zuur *et al*., 2009)[2]. A Variance Inflation Factor (VIF) was used as an indicator of collinearity and variables with VIF <10 were progressively excluded from the analysis (Zuur *et al*., 2009). As a result, variables that were not further considered in the analysis of the different datasets were the following:

- BTS dataset: Rock or other hard substrata relative coverage (R) and Dominant substratum type (DomSubst) (both variables being negatively correlated with Sand-muddy Sand coverage (S-mS))
- WGEGGS dataset: Coarse sediment coverage (Cs, negatively correlated with Mixed sediment coverage, Mx), Dominant substratum type (DomSubst, correlated with Rock or other hard substrata relative coverage, R) and wave energy (WavE, negatively correlated with depth)
- IHLS dataset: Coarse sediment coverage (Cs, negatively correlated with Mixed sediment coverage, Mx), Dominant substratum type (DomSubst, correlated with Rock or other hard substrata relative coverage, R) and Reef presence category (Reef, correlated with Rock or other hard substrata relative coverage, R).

When a choice was to be made on the exclusion between two collinear variables, dropping categorical variables rather than continuous variables was preferred, and variables obtained from datasets with lower confidence or for which relevant limitations were known were dropped (e.g., reef presence category and wave energy were dropped when a collinearity was highlighted with Rock or other hard substrata relative coverage and depth, respectively) in order to avoid a reduction in the confidence of the final model output.
Also variable M-sM was not considered in the analysis of WGEGGS and IHLS datasets, as this variable did not show any variability within the dataset (all stations in these datasets had an associated M-sM value of 0).

---

[2] As regards specifically classification trees, the exclusion of highly correlated predictors was adopted also by Lawler and Edwards (2002), based on the fact that this analysis only allows one of any set of correlated variables to enter the model at any given split (the variable that best classifies the data is selected) and as the data are split into smaller groups in the modeling process, the relationships among explanatory variables may change, thus affecting the model results.

The resulting environmental predictors included initially in the model analysis of the specified datasets are:

- BTS dataset: depth, sea surface temperature (SST), tidal energy (TidE), wave energy (WavE), type of mixing of the water column (DomMix), Mud-sandy Mud coverage (M-sM), Sand-muddy Sand coverage (S-mS), Coarse sediment coverage (Cs), Mixed sediment coverage (Mx), Reef presence category (Reef)
- WGEGGS dataset: depth, sea surface temperature (SST), phytoplankton absorption coefficient (APH), tidal energy (TidE), type of mixing of the water column (DomMix), Sand-muddy Sand coverage (S-mS), Mixed sediment coverage (Mx), Rock or other hard substrata relative coverage (R), Reef presence category (Reef)
- IHLS dataset: depth, sea surface temperature (SST), phytoplankton absorption coefficient (APH), tidal energy (TidE), wave energy (WavE), type of mixing of the water column (DomMix), Sand-muddy Sand coverage (S-mS) and Mixed sediment coverage (Mx).

To increase the accuracy and predictive power of the model, the number of explanatory variables used should be reduced to a reasonable number (Harrell *et al*., 1996). Given that presumably more explanatory variables increase the variance accounted for, this is a subjective decision that takes into account the balance between explaining more of the variance and not making the models too unwieldy. The pre-selection of the predictors that was operated during the data exploration (to avoid collinearity) contributed to reduce the initial number of predictors included in the analysis. In addition, no polynomial terms or interaction between the predictors have been considered as explanatory variables, again in order to allow an easier and direct interpretation of model results.

The data exploration of the fish catch data highlighted that most species/life stages (in particular juveniles) showed a low frequency of presence in the study datasets (often <25% occurrences) and often low abundance values were recorded. This extreme distribution of the data makes it difficult to apply standard statistical tools (e.g., data transformation). Consequently, a choice was made to reduce fish data to presence and absence data and classification tree models were identified as the most appropriate technique for their analysis.

5.2 Classification tree model calibration

The model calibration consists on the adjustment of the selected mathematical model for the specific data to which the model is applied (Guisan *et al*., 2000). Classification tree models were calibrated on fish presence-absence data.

Classification trees are modelling tools belonging to the CART family (Classification and Regression Trees) that allow the analysis of the relationship between one response variable and several explanatory variables (Guisan *et al*., 2000; Faraway, 2006). CARTs are similar to additive models in that they represent a compromise between the linear model and the completely nonparametric approach. This procedure creates a tree-based classification model, which predicts values of a response variable based on values of explanatory variables. Classification trees are applied when the response variable is a categorical one (e.g., presence/absence).

Explanatory variables in the tree are organised in a hierarchical way (based on their effect on the response variable) into a series of dichotomies (branches of the tree) that allow the user to identify unique combinations of variables associated with specific levels of the categorical response variable and to determine the major influencing variables. The advantage of this method is that the model does not need to be specified at the outset, as the true starting point in this case is the algorithm used in the construction of the tree (Faraway, 2006). Another advantage of CART is that their structure is easier to understand for the technical non-specialist.

The model calibration consists of adjusting the selected mathematical model for the specific data to which the model is applied (Guisan *et al*., 2000). Based on the initial number of explanatory variables (full model), a model selection is carried out by automatic selection procedures in order to find the optimal model, i.e., the best subset of explanatory variables that identify the parameters that best explain the variability in the fish data.

The model selection for the EFH analysis was carried out by means of a combination of cross-validation (estimating the predictive ability of the model) and truncating in order to reduce the tree to a more ‘optimal’ number of terminal nodes (Faraway, 2006). This approach is based on the selection of the complexity (Cp) parameter. Cp parameter controls the size of the tree, and model selection aims at pruning the tree by increasing the Cp parameter (as large tree can cause overfitting). It is of note that pruning the tree (hence using a more simplified tree) usually increases the prediction error, therefore an acceptable compromise must be found between the level of pruning and the model error. The parsimonious “one standard deviation (1-SE)” rule was applied to identify the best level of pruning, i.e. the tree was pruned to a maximum Cp value for which the resulting tree has a rate of error within one standard deviation of the minimum error (Faraway, 2006; Zuur *et al*., 2007). Therefore, this rule allows simplifying the tree while maintaining an acceptable degree of error.

For each leaf of the tree (i.e. predicted category based on a set of environmental rules), the number of observations in the dataset that have been incorrectly classified in that category is given. Based on this, a percentage misclassification error can be associated with the whole model (tree) and to each predicted category (leaf) providing an information on the predictive ability of the classification tree and on the confidence of the single prediction, respectively.

Data exploration and model calibration was carried out in Brodgar 2.7.2, with interface to the statistics package R 2.9.1 (Highland Statistics Ltd., 2000).

5.3 Classification tree interpretation

In order to indicate the tools needed to read the results presented on the classification tree modelling, two examples of these models are provided below, with detailed description of the graphical interpretation of these results.

M1. Overall species habitats model

The predicted habitat categories of this type of model are:

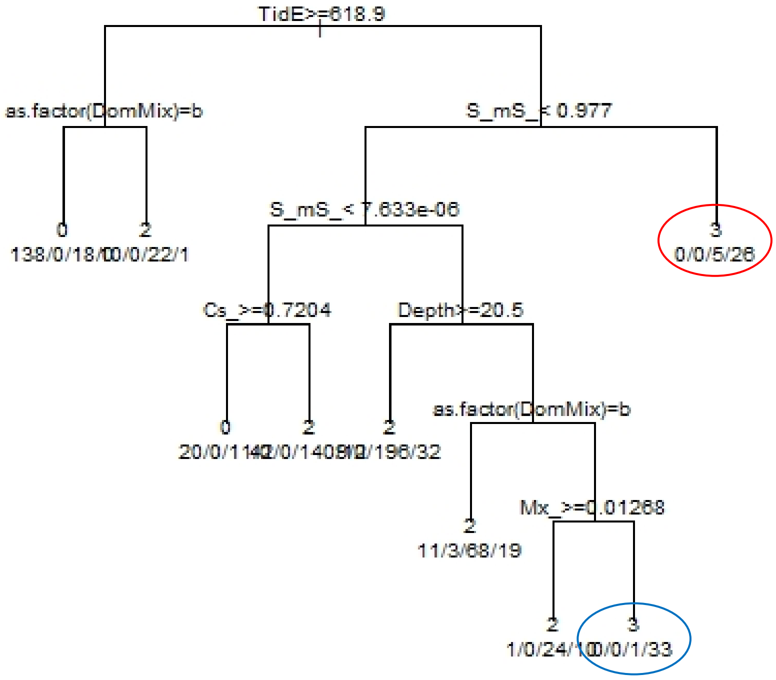
0 = habitats where the species is not present

1 = habitats where juveniles only are present
2 = habitats where adults only are present
3 = habitats where juveniles and adults are present.

Categories 1 and 3 identify the potential nursery habitat of a species (based on juvenile frequency of occurrence). Categories 2 and 3 identify the foraging habitats of adults, and category 0 identifies habitats where the species is less likely to occur (e.g., due to biogeographic distribution of the species or to possible unsuitable habitat conditions).

An example of the obtained result for is given in Figure 2 (Plaice habitats). The category predicted at each leaf (corresponding to a set of environmental conditions) is indicated in the label of the leaf. It is of note that, in this case, the habitat category 1 was not predicted by the model due to the small amount of data regarding this type of habitat in the analysed dataset (BTS, with only 3 fishing events out of 852 where juvenile plaice was observed occurring without any adult present in the catch). The number of observations for the different habitat categories that the model allocated to each prediction (leaf) is also indicated in the leaf label.

Figure 2: Classification tree for Plaice habitats.



The environmental variables selected by the model as relevant in predicting the distribution of plaice habitats are indicated at each node of the tree, with the importance of the variable decreasing while going down the tree. The results for plaice habitats, for example, suggest that tidal energy at the bottom (TidE) is the most important environmental predictor of plaice habitats among those included in the model, and it is followed by the category for mixing of the water masses (DomMix) and by the proportional coverage of sand to muddy sand on the seabed (S-mS).

The environmental conditions indicated at each node (e.g., Tidal energy ≥ 618.9 N/m^2 at the first (upper) node) indicate the conditions predicted along the branch on left hand side of the node, whereas opposite conditions (e.g., Tidal energy < 618.9 N/m^2) are associated with the right hand branch.

Based on the above rules, for example, the leaf on the top right of the tree (circled in red, Figure 2) indicates that the occurrence of potential nursery habitat for the species (habitat 3) is predicted where low-moderate tidal energy conditions (TidE<618.9) combine with higher proportional coverage of sand to muddy sand substrata on the seabed (S_mS≥0.98). This group is defined by 31 observations, 26 of which belonging to category 3 and 5 to category 2, whereas no observations for categories 0 and 1 are included in this group. Considering that the model predicts this group as a habitat category 3, this leads to misclassification error of this prediction of 5/31 = 0.16. Similarly, the occurrence of potential nursery habitat is predicted also in other environmental conditions (as defined by the branches of the tree leading to the group in blue circle in Figure 2). In this case, the prediction for this habitat is associated with a lower misclassification error (1/34 = 0.03), indicating a higher confidence in this prediction compared to the previous one.

The combination of misclassification error and total number of observations used to predict each leaf of the tree has been used to rank the confidence in the predicted categories. This allowed the spatial representation of these confidence levels associated with the habitat predictions when mapping the EFH.

The total misclassification error for the model can be calculated by summing the number of observations misclassified in each prediction (leaf) and dividing it by the total number of observations used in the model. In this case, this model has a total misclassification error of 0.22 (185 misclassified observations out of 852), denoting a relatively high confidence in the model predictive ability (see Section 5.6 for details on confidence criteria).

M2. Nursery/Spawning habitats model

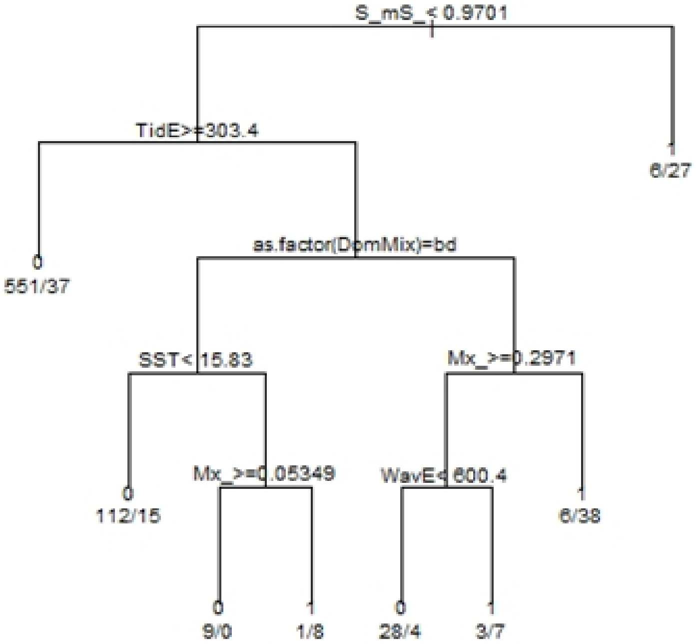
The predicted habitat categories of this type of model (on juvenile presence-absence in BTS dataset) are

0 = absence of nursery habitat
1 = presence of nursery habitat.

Similar habitat categories, but identifying presence-absence of spawning grounds, were predicted when modelling data on eggs (plaice) and larvae (herring).
The example of the obtained result is given in Figure 3 (Plaice primary nursery habitat).

The tree interpretation follows the same rules as described above. For example, the leaf on the top right of the tree indicates that the occurrence of potential nursery habitat for the species where there is a higher proportional coverage of sand to muddy sand substrata on the seabed (S_mS≥0.97) (Figure 3). This prediction is given with an associated misclassification error of 0.18 (6 misclassified observations out of 33 included in the predicted group), while the total misclassification error for the whole model is 0.08, indicating a relatively high confidence in the predictive ability of this model. In this model the ratio between the number of occurrences of juveniles and the total number of observations included in each group (leaf) can be used to calculate the probability of presence of juveniles associated with the environmental conditions defined for that prediction. In the example above (leaf on the top right of the tree), the condition of high proportional coverage of sand to muddy sand substrata on the seabed (S_mS≥0.97) allows predicting the occurrence of plaice 0-group (hence its primary nursery habitat) with a high probability of presence (0.82).

Figure 3: Classification tree for the probability of occurrence of Plaice nursery habitats (0-group).



5.4 Classification tree model results

In this section the graphic results obtained for the species models are presented (Figure 4-13).

These results, together with the tables shown in the results section of the Final Report (summarising the combination of environmental conditions leading to the prediction of EFH for a species) constitute the algorithm defined by each model and used for the model implementation in GIS.

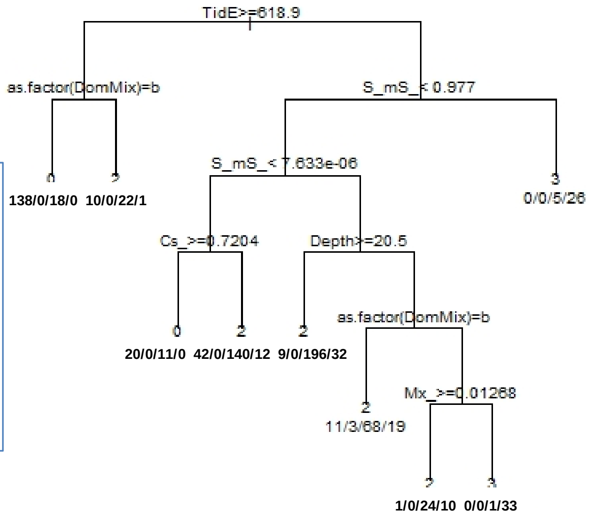
Figure 4: Classification tree models calibrated for plaice.

Plaice

a) M1_species habitats

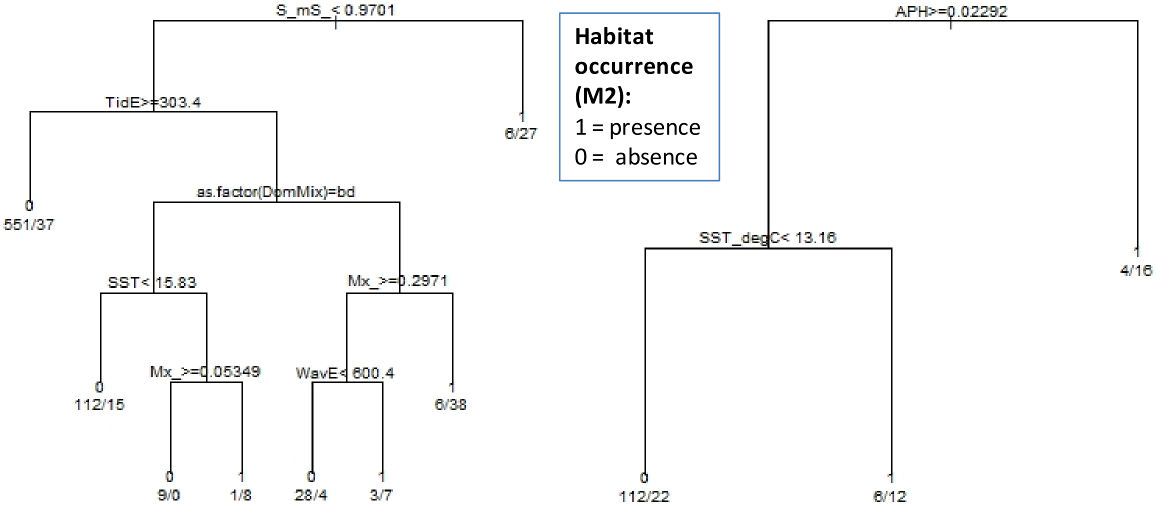
Habitat categories (M1):
0 = species absent
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom)



b) M2_ nursery habitat

c) M2_spawning habitat



Habitat occurrence (M2):
1 = presence
0 = absence

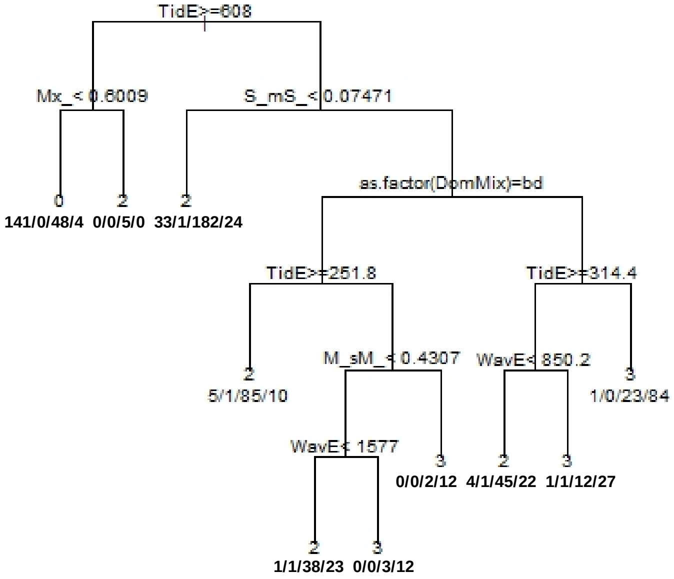
Figure 5: Classification tree models calibrated for sole.

Sole

a) M1_species habitats

Habitat categories (M1):
0 = species absent
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom)



b) M2_ nursery habitat

Habitat occurrence (M2):
1 = presence
0 = absence

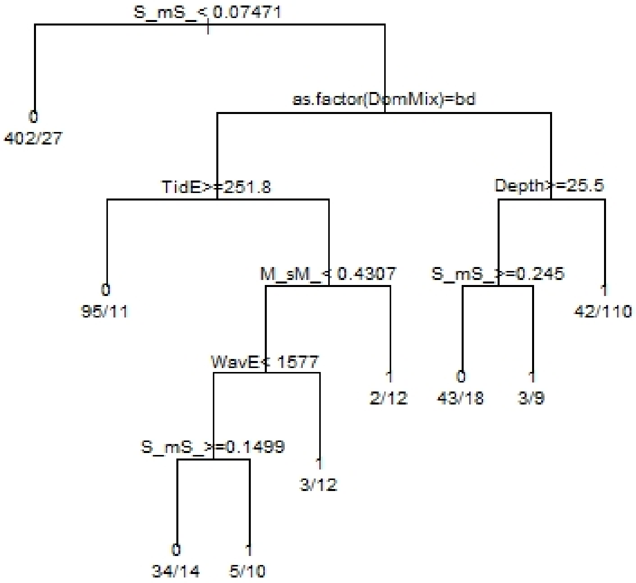


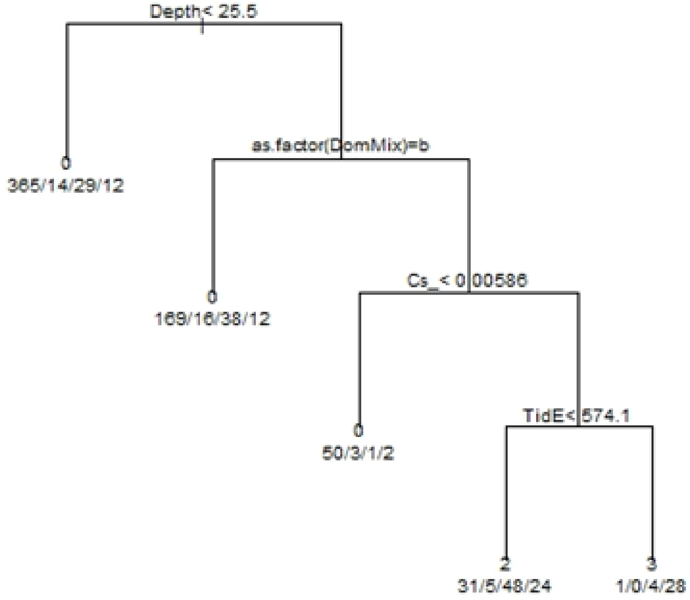
Figure 6: Classification tree models calibrated for lemon sole.

Lemon sole

a) M1_species habitats

Habitat categories (M1):
0 = species absent
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom)



b) M2_ nursery habitat

Habitat occurrence (M2):
1 = presence
0 = absence

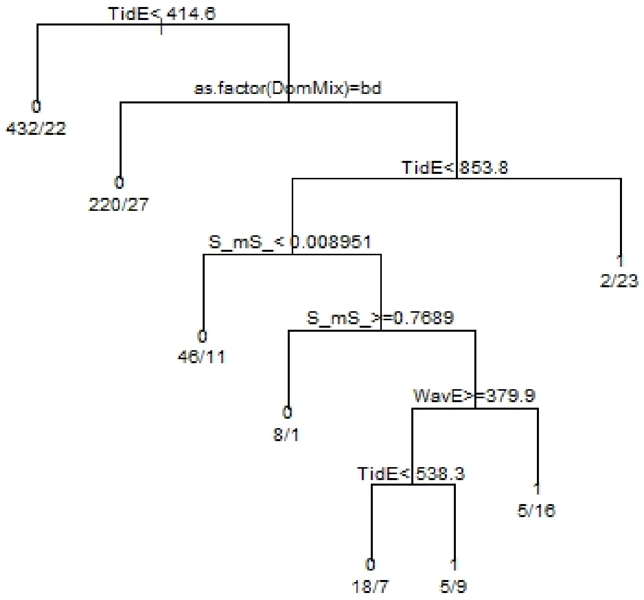


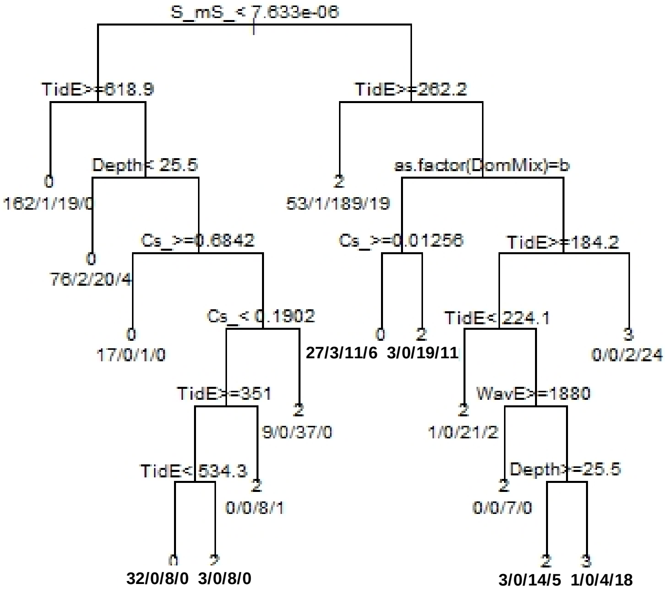
Figure 7: Classification tree models calibrated for dab.

Dab

a) M1_species habitats

Habitat categories (M1):
0 = species absent
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom)



b) M2_ nursery habitat

Habitat occurrence (M2):
1 = presence
0 = absence

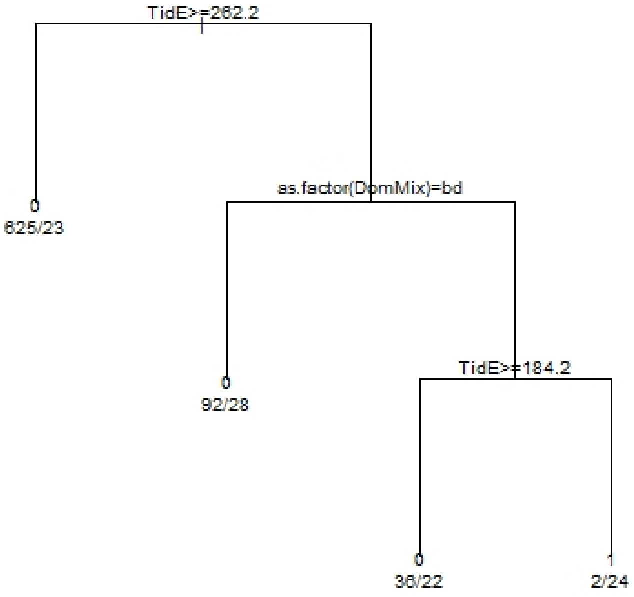


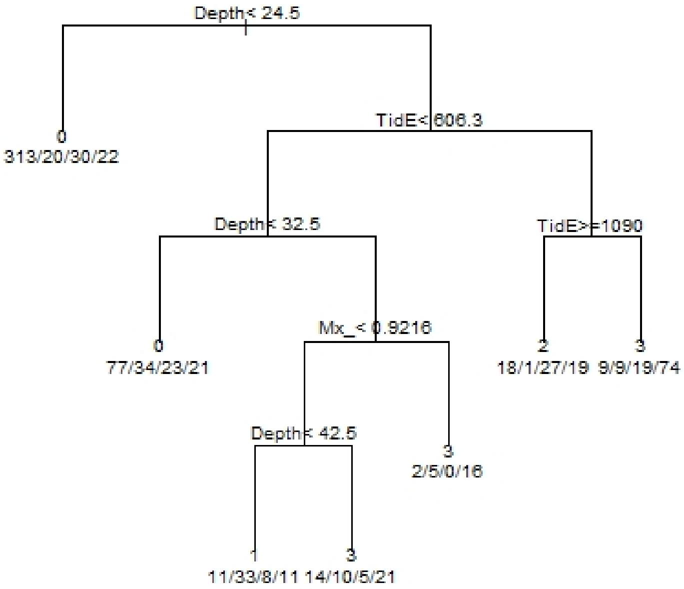
Figure 8: Classification tree models calibrated for red gurnard.

Red gurnard

a) M1_species habitats

Habitat categories (M1):
0 = species absent
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom)



b) M2_ nursery habitat

Habitat occurrence (M2):
1 = presence
0 = absence

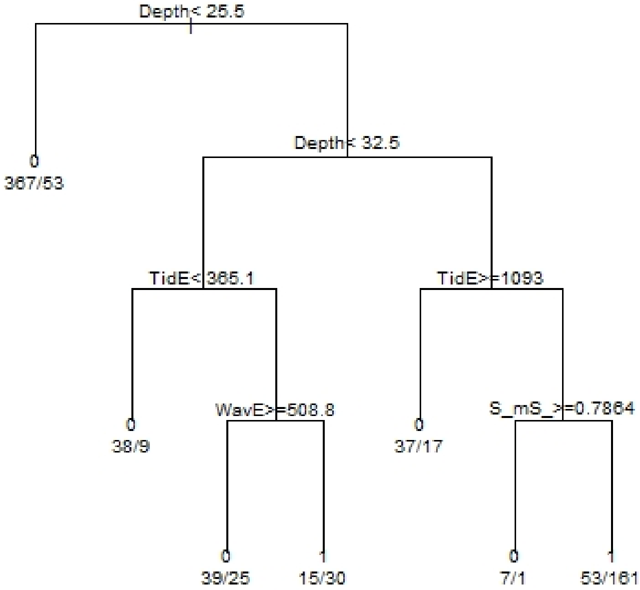
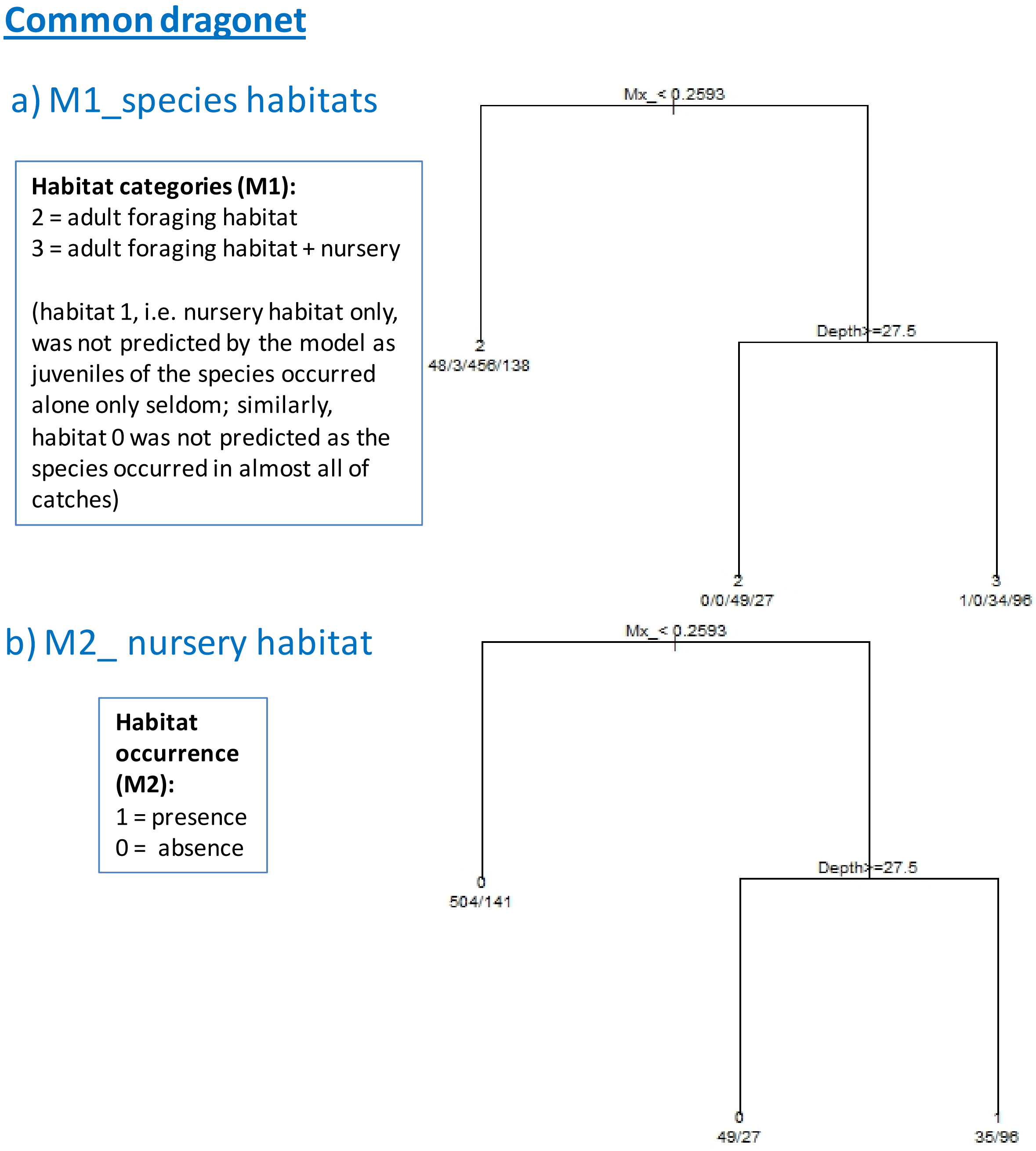


Figure 9: Classification tree models calibrated for common dragonet.



Common dragonet

a) M1_species habitats

Habitat categories (M1):
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom; similarly, habitat 0 was not predicted as the species occurred in almost all of catches)

b) M2_ nursery habitat

Habitat occurrence (M2):
1 = presence
0 = absence

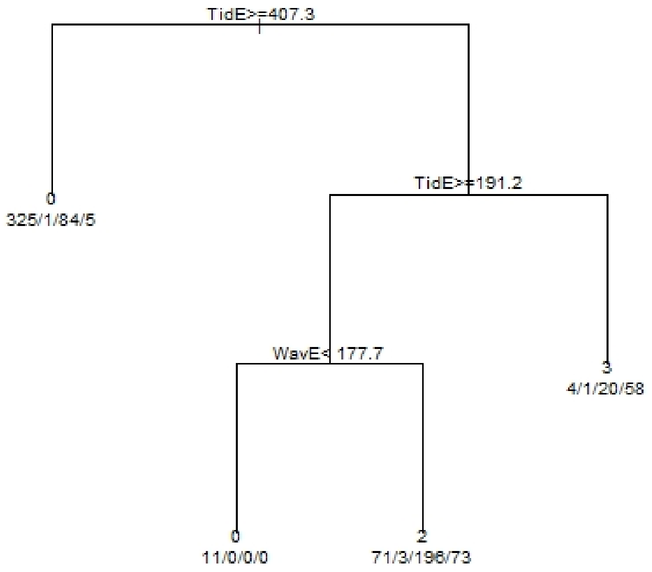
Figure 10: Classification tree models calibrated for solenette.

Solenette

a) M1_species habitats

Habitat categories (M1):
0 = species absent
2 = adult foraging habitat
3 = adult foraging habitat + nursery

(habitat 1, i.e. nursery habitat only, was not predicted by the model as juveniles of the species occurred alone only seldom)



b) M2_ nursery habitat

Habitat occurrence (M2):
1 = presence
0 = absence

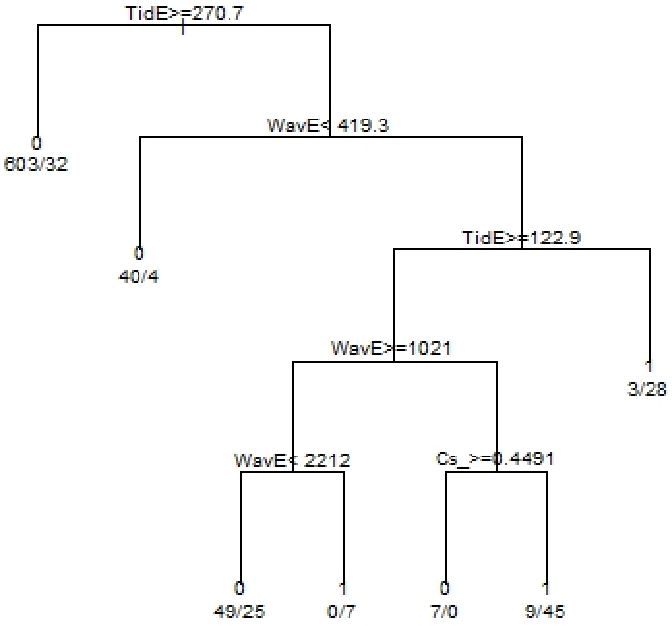


Figure 11: Classification tree model calibrated for thickback sole.

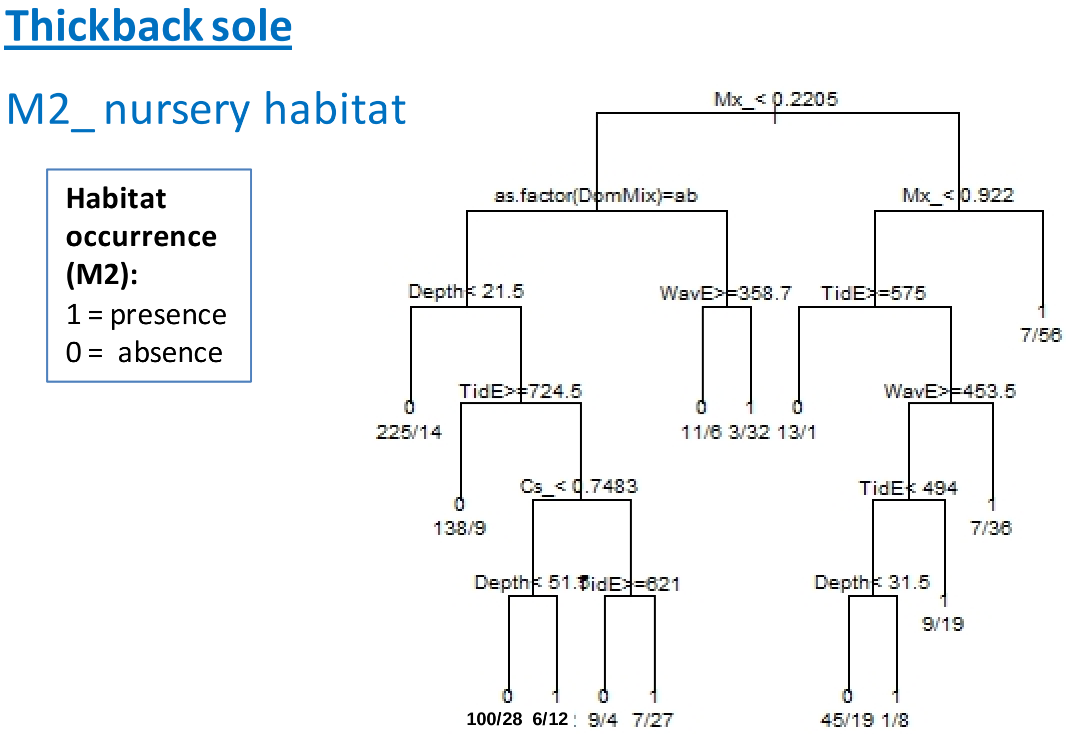


Figure 12: Classification tree model calibrated for thornback ray.

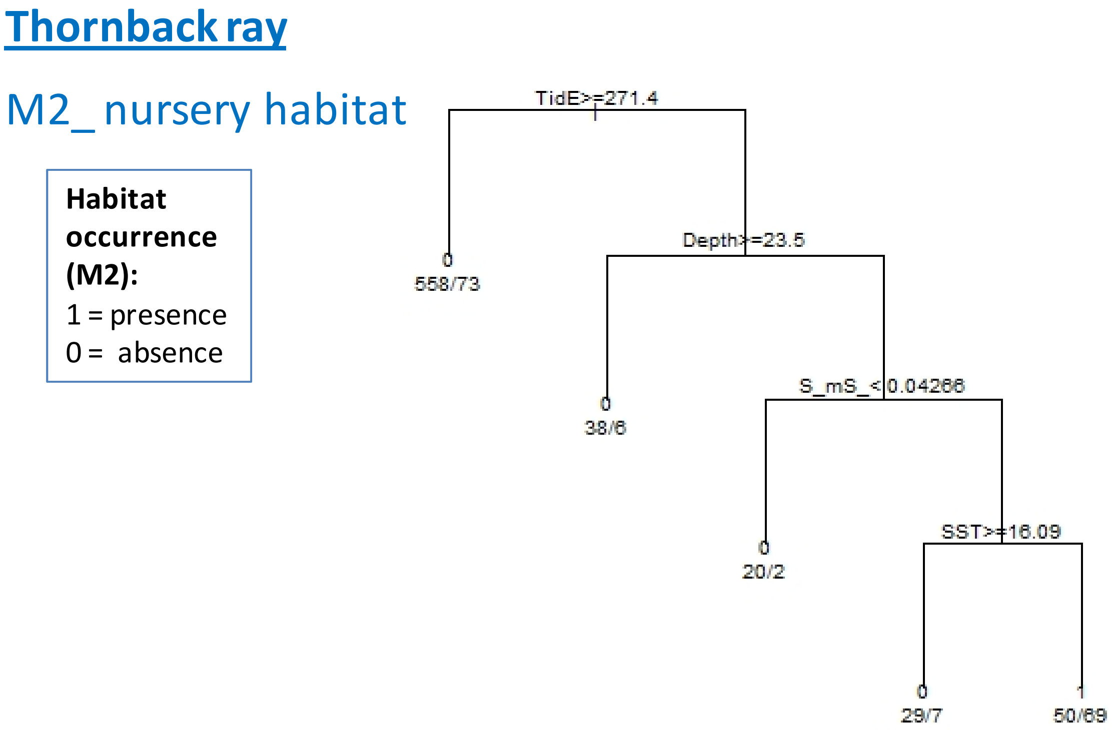
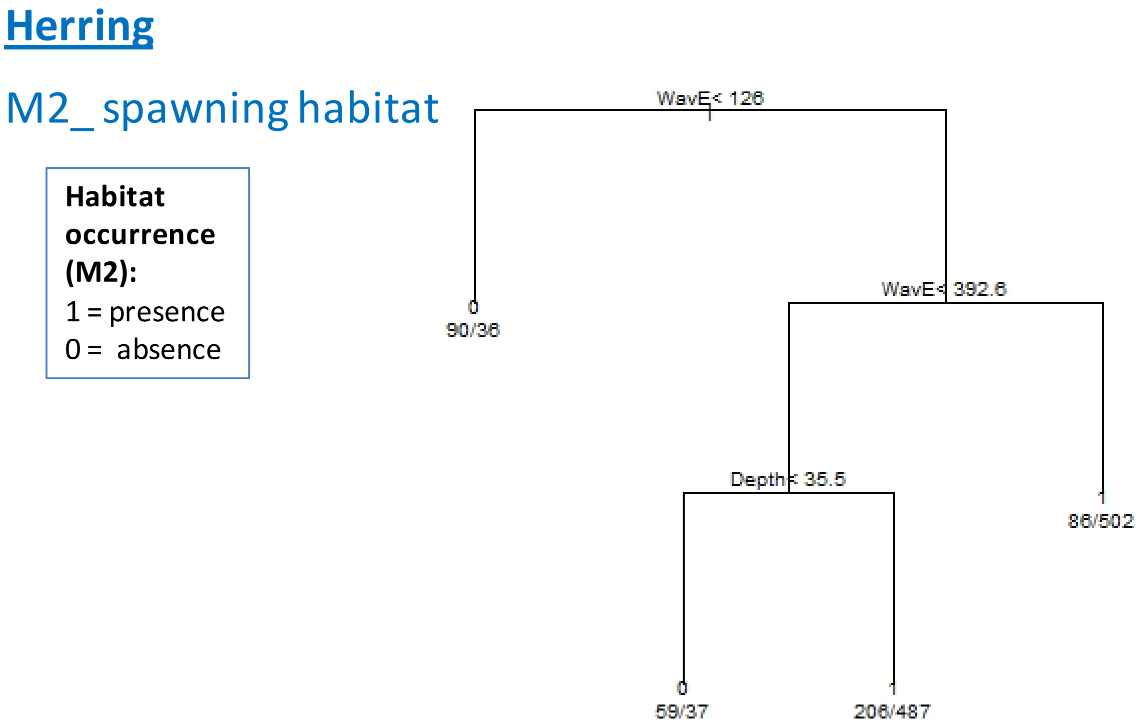


Figure 13: Classification tree model calibrated for herring.



5.5 Model prediction (GIS implementation)

In order to implement the results of the model a 5x5km vector grid of the study area was created, this resolution again reflecting the size of the buffers created for the fish data. As before, the mean, maxima or dominant (as appropriate) environmental variables relevant to the model outputs were extracted for each grid cell within the study area. Where data was temporally variable (temperature and phytoplankton absorption coefficient), data available for the years within the 2000 – 2012 timeframe were derived for the grid cells (considering the appropriate months) and the mean values calculated over the years to obtain a single value for each variable within the grid cell.

Grid cells which matched the combination of environmental criteria generated by the model were then highlighted as being potential EFH for the species analysed. A colour code identifying different habitat types (spatial predictions of model M1) or different probability of presence of a certain life stage (spatial predictions of model M1) was assigned to the grid. Although the environmental criteria generated by the model are defined in a way that allows prediction of EFH also where the environmental conditions fall outside the range of variability of the environmental data used for the model calibration (see Table 6), no confidence was associated to these predictions. Therefore, grid cells which showed environmental values falling outside the range of variability of the data used for model calibration were identified and shown as blank cells in the maps to indicate areas outside the model range. The resulting layers were then clipped to the study area.

For most original data layers a full coverage of data was available at a resolution higher than the 5x5km grid. However, in some cases data was missing and these areas of no coverage were highlighted as ‘No Data’ or ‘Not Assessed’ in the final output.

The temporal reference of the prediction is associated with the season of the calibration dataset which originates the model and to the average seasonal conditions observed within the study period (2000-2012). For these datasets with a temporal element, a standard deviation of the data across years could also be calculated as an indicator of confidence (see Section 5.6). Standard deviation was also calculated for the depth variable to indicate areas of variable depth (e.g. slopes or shelf edges).

The spatial implementation of the EFH model was carried out in ArcGIS v10.0, with also MapInfo Professional v9.0 and Microsoft Office Excel 2007 used for part of the data processing (see Data Processing Templates for details). As the size of some of the datasets exceeded the processing power of ArcGIS, these datasets were processed as subsets, and the results then re-merged into a single file. Apart from being time consuming this did not contribute an error to the result.

5.6 Output confidence

A confidence value was associated with each map output based on the combination of the confidence in the elements and processes that determined the map. These were identified as:

- **Cm**: the confidence in the modelling process for the species/life stage, as indicated by the specific model predictive ability
- **Cf**: the confidence in the fish survey data used to calibrate the model, with reference to the species/life stage considered in the model (see Table 4 in the Final Report)
- **Ce**$_j$: the confidence in each of the j-th environmental data layers used for the calibration and implementation of the EFH model for a species/life stage
- **re**$_j$: the ranked importance of each of the j-th environmental data layers in the model prediction of EFH (as defined by their hierarchy in the tree model) for a species/life stage.

The confidence in the model predictive ability (Cm) was measured as 1-relative misclassification error (i.e., the proportion of cases correctly classified by the model on the whole) and rated according to the scale 1-to-3 following the criteria indicated in Table 2.

The confidence on the input environmental data layers (Ce$_j$) was determined using the criteria and ratings defined by the MMO assessment standards. In particular, the MMO Quality Assurance Data Template was used to assign a total confidence value based on the confidence allocated to methodology, timeliness, spatial confidence, completeness, and production quality standards (the results of this assessment are summarised in Table 3, with details on the reasons behind the assessment being given in the excel files for the QA templates). As the total confidence obtained within the template was given as a percentage, it was back transformed into ratings 0 to 3 based on the criteria shown in (Table 4).

The confidence of the fish survey data (Cf) was calculated as the mean of the confidence ratings accounting for different characteristics of the data (Table 4 in the Final Report). These included:

- the ability of the survey method to capture abundance/occurrence of a species/life stage depending on
  - gear selectivity (Elliott and Hemingway, 2002; Ellis *et al*., 2010a)
  - seasonality of the sampling in relation to the life cycle/spawning periods of the species (see Table 1b in the Final Report)
  - spatial distribution/coverage of the sampling design in relation to the knowledge on distribution of nursery/spawning habitats of a species (Ellis *et al*., 2010)
  - taxonomic standards (Defra, 2010; Ellis *et al*., 2012)
- the ability of a certain life stage to indicate the distribution of an EFH (depending also on how the life stage has been identified; e.g. early vs. later egg stages; 0-group juveniles vs. immature juveniles, likelihood of also including fish >1yrs old).

A confidence rating between 1 and 3 (low to high) was assigned to each characteristic. The resulting mean value was assigned a confidence level using the criterion provided in the MMO Quality Assurance Data Template (the %score assigned in the MMO template was transformed into the corresponding rating value by considering that the rating values vary between 0 (0%) and 3 (100%); the resulting criteria is given in Table 5).

The ranked importance of the environmental data included in the model (re_j) was also used to weight the confidence in the single input environmental data layers and calculate the weighted arithmetic mean of confidence associated with the set of environmental data used by the model (so that the contribution of each environmental data layer to the final confidence is proportional to its importance in affecting the species life stage distribution).

The confidence rating associated with the model predictive ability (Cm) was weighted by the average confidence rating associated with the model input data (Ci) using the following formula:

$$C = Cm \cdot \frac{Ci}{3}$$

where

$$Ci = \frac{1}{2}\left[Cf + \frac{\sum_j (Ce_j \cdot re_j)}{\sum_j re_j}\right]$$

The resulting total confidence rating was associated with a confidence level following the criterion defined in Table 5.

Table 5 reports in detail the resulting total confidence associated with each model output together with the confidence ratings of the elements that contributed to the total confidence estimate (overall confidence ratings of input data layers, ranked importance of the different environmental predictors, as determined by the model and confidence associated with the model predictive ability) based on the calculation described above.

Table 2: Confidence in the classification tree model predictive ability (Cm).

Rating	Confidence	Definition
1	Low	Low confidence in the model predictive ability. Total model predictive ability ≤ 0.5, equivalent to Total misclassification error ≥ 0.5 (i.e., less than 50% of the data in the calibration dataset is correctly classified by the model)
2	Moderate	Moderate confidence in the model predictive ability. Total model predictive ability ≤ 0.75 and > 0.5, equivalent to Total misclassification error ≥ 0.25 and < 0.5 (i.e., between 50 and 75% of the data in the calibration dataset is correctly classified by the model)
3	High	High confidence in the model predictive ability. Total model predictive ability > 0.75, equivalent to Total misclassification error < 0.25 (i.e., more than 75% of the data in the calibration dataset is correctly classified by the model)

Table 3: Summary of confidence scores assigned to the input environmental data layers as from MMO Quality Assurance Data Templates.

Data layer	Methodology Confidence	Timeliness Confidence	Spatial Confidence	Completeness Confidence	Confidence in Quality Standards	Overall Quality Assessment
Bathymetry (EMODnet)	3 High Confidence	2 Moderate Confidence	3 High Confidence	2 Moderate Confidence	2 Moderate Confidence	80% Moderate High Confidence
Seabed substratum type (EMODnet for EUSeaMap)	1 Low Confidence	0 Unable to Assess	2 Moderate Confidence	1 Low Confidence	0 Unable to Assess	27% Low Confidence or Unable to Assess
JNCC EuSeaMap North and Celtic Seas Energy data layers (EUSeaMap)	1 Low Confidence	0 Unable to Assess	0 Unable to Assess	0 Unable to Assess	0 Unable to Assess	7% Low Confidence or Unable to Assess
Marine Water Column Features (JNCC)	3 High Confidence	3 High Confidence	2 Moderate Confidence	3 High Confidence	0 Unable to Assess	73% Moderate Confidence
Global Ocean OSTIA Sea Surface Temperature and Sea Ice analysis REPROCESSED (1985-2007) (EU project My Ocean)	3 High Confidence	2 Moderate Confidence	2 Moderate Confidence	2 Moderate Confidence	2 Moderate Confidence	73% Moderate Confidence
Pan European Seas, Ocean Optics Products (monthly average) Reprocessed (1997-2010) (EU project My Ocean)	3 High Confidence	2 Moderate Confidence	2 Moderate Confidence	2 Moderate Confidence	2 Moderate Confidence	73% Moderate Confidence

Table 4: Rating criterion for the assessment of confidence in the input environmental data layers based on the % confidence obtained within the MMO Quality Assurance Data Template. The criterion takes into account the thresholds used in the MMO template to relate %score to confidence level.

% Confidence (from MMO template)	Confidence	Rating
0	Unable to assess	0
<46.66%	Low	1
<59.99%	Low-Mod	1.5
<79.99%	Moderate	2
<93.32%	Mod-High	2.5
≥93.32%	High	3

Table 5: Rating criterion for the assessment of confidence in the fish survey data and in the classification tree model predictive ability (range 1-3). The criterion has been derived by comparison with the criterion defined as per MMO Quality Assurance Data Template.

% Confidence (from MMO template)	Mean Rating (Cf)	Confidence
0	0	Unable to assess
<46.66%	< 1.4	Low
<59.99%	≥1.4 and < 1.8	Moderate - Low
<79.99%	≥ 1.8 and < 2.4	Moderate
<93.32%	≥ 2.4 and < 2.8	Moderate - High
≥93.32%	≥ 2.8	High

Table 6: Ranges of variability of environmental variables in the calibration datasets. Categories for DomMix indicate: 1 = well-mixed ROFI (Region of Freshwater Influence); 2 = well-mixed shelf water; 3 = weakly stratified ROFI; 4 = weakly stratified shelf water. Categories for Reef indicate: 0=no reef; 1=reef potentially present (lower confidence); 2=reef present. "-" indicates variables excluded from the model.

		range in the calibration dataset		
variable	unit	BTS	WGEGGS	IHLS
Depth	m	8-81	14-71	17-75
DomMix	(category)	1-4	1-3	1-3
M-sM	(proportion)	0-0.77	-	-
S-mS	(proportion)	0-1	0-0.42	0-1
Cs	(proportion)	0-1	-	-
Mx	(proportion)	0-1	0-1	0-1
R	(proportion)	0-1	0-1	0-1
Reef	(category)	0-2	0-2	-
TidE	N/m^2	31.31-1498.47	248.51-1952.72	208.12-1768.51
WavE	N/m^2	60.35-5361.51	-	58.01-1824.10
SST	degrees C	13.65-17.47	11.9-13.67	10.65-14.57
APH	m^{-1}	-	0.015-0.110	0.015-0.203

Table 7: Total confidence associated with each model output and confidence ratings of input data layers and of model predictive ability that contributed to the total confidence estimate.

		Total confidence		Fish survey data	Input environmental predictors										Model predictive ability	
					Dom Mix	SST	APH	Depth	M-sM	S-mS	Cs	Mx	TidE	WavE		
					Ranked importance in the model											
Species	**Model output**	**Rating**	**Class**	**Cf**	**re1**	**re2**	**re3**	**re4**	**re5**	**re6**	**re7**	**re8**	**re9**	**re10**	**Pred. ability**	**Cm score**
Plaice	M1_species habitat	2.1	Moderate	2.8	5	0	0	3	0	5	3	1	6	0	0.78	3
	M2_nursery habitat	1.4	Low	2.8	3	2	0	0	0	5	0	2	4	1	0.59	2
	M2_spawning habitat	1.5	Moderate-Low	2.6	0	1	2	0	0	0	0	0	0	0	0.56	2
Sole	M1_species habitat	1.3	Low	2.6	3	0	0	0	1	4	0	4	5	1	0.74	2
	M2_nursery habitat	1.4	Low	2.6	4	0	0	3	2	5	0	0	3	1	0.69	2
Lemon sole	M1_species habitat	2.3	Moderate	2.6	3	0	0	4	0	0	2	0	1	0	0.77	3
	M2_nursery habitat	0.7	Low	2.6	5	0	0	0	0	3	0	0	6	1	0.41	1
Dab	M1_species habitat	2.1	Moderate	2.8	4	0	0	4	0	6	3	0	5	1	0.77	3
	M2_nursery habitat	0.7	Low	2.8	1	0	0	0	0	0	0	0	2	0	0.25	1
Red gurnard	M1_species habitat	1.5	Moderate-Low	2.8	0	0	0	4	0	0	0	1	3	0	0.66	2
	M2_nursery habitat	1.5	Moderate-Low	2.8	0	0	0	4	0	1	0	0	2	1	0.65	2
Dragonet	M1_species habitat	1.4	Low	2.6	0	0	0	1	0	0	0	2	0	0	0.71	2
	M2_nursery habitat	0.7	Low	2.6	0	0	0	1	0	0	0	2	0	0	0.36	1
Solenette	M1_species habitat	1.2	Low	2.6	0	0	0	0	0	0	0	0	3	1	0.69	2
	M2_nursery habitat	1.2	Low	2.6	0	0	0	0	0	0	1	0	5	4	0.57	2
Thickback sole	M2_nursery habitat	1.3	Low	2.4	4	0	0	3	0	0	1	5	3	3	0.70	2
Thornback ray	M2_nursery habitat	0.6	Low	2.3	0	1	0	3	0	2	0	0	4	0	0.44	1
Herring	M2_spawning habitat	2.1	Moderate	2.6	0	0	0	1	0	0	0	0	0	2	0.93	3

Confidence rating									
Ce1	**Ce2**	**Ce3**	**Ce4**	**Ce5**	**Ce6**	**Ce7**	**Ce8**	**Ce9**	**Ce10**
2	2	2	2.5	1	1	1	1	1	1

A similar process was applied to calculate a relative confidence to the EFH spatial predictions, in order to identify areas of lower confidence in the output maps.

In this case, the confidence rating associated with the model predictive ability was obtained for each of the categories predicted by the tree model by using the misclassification error associated with each leaf of the tree (with an approach similar to the one used to obtain Cm). The confidence was rated according to the scale 1-to-3 following the criteria indicated in Table 6. Also, a confidence value of 0 was assigned to cells where the model prediction extrapolated beyond the range of environmental variability of the calibration dataset, taking into account only the set of environmental predictors that resulted relevant for the specific model (Table 4). The resulting confidence maps based solely on the model predictive ability are shown for the different model outputs in Figure 14 to31.

Confidence maps for the input environmental data layers were used together with the information on the ranked importance of the environmental predictors as identified by the model (r_{e1} to r_{e10} in Table 7). In particular, the following confidence indicators were used for the environmental input data layers:

- Depth (EMODnet Gridded Bathymetry): coefficient of variation (CV[3]) associated to the calculation of the mean water depth within each grid cell, as calculated during the data processing
- Sea Surface Temperature (My Ocean Global Sea Surface Temperature and Sea Ice (reprocessed), 1985-2007): coefficient of variation (CV) associated with the calculation of the mean seasonal SST for the study period 2000-2012 based on the mean seasonal SST for each year available within the study period, as calculated during the data processing
- Phytoplankton absorption coefficient APH (My Ocean Pan European Seas Optics Products, 1997-2010): coefficient of variation (CV) associated with the calculation of the mean monthly APH value for the study period 2000-2012 based on the spatial maximum APH value for January each year (as available within the study period), as calculated during the data processing.

In order to use the above errors to indicate confidence, the CV values obtained for the grid cells were standardised to a variability range 0-1 (by subtracting the minimum value in the map and dividing by the range), and the complement number (1-value) calculated. In this way, a relative score was obtained showing increasing value with higher confidence (the resulting confidence maps associated to the environmental data layers are shown in Figure 32 to 35). No confidence maps/estimates were available for the other input environmental data layers relevant to the EFH model outputs (Figure 36 to 42). In these cases, a value of 0 (correspondent to the confidence class 'unable to assess', based on the MMO criteria) was assigned to all the cells as a measure of confidence associated with these data layers. The confidence assigned to the input fish survey data was the one indicated in Table 7 (Cf), with the same value assigned to all the cells.

The same formula for the overall confidence assessment was then applied to each grid cell and the resulting spatial confidence was represented as a relative value

---

[3] Coefficient of variation (CV) is defined as the ratio between the standard deviation and the mean value calculated within each grid cell.

(higher to lower confidence) to be read in relation to the total confidence value associated with the output map which is assumed to represent the overall confidence for the map.

Table 8: Confidence in the classification categories predicted by the EFH model.

Rating	Confidence	Definition
1	Low	Low confidence in the model prediction for the habitat categories (leaves of the tree). Model predictive ability < 0.6, equivalent to Misclassification error >0.4 (i.e., less than 60% of the data in the predicted category is correctly classified by the model)
2	Moderate	Moderate confidence model prediction for the habitat categories (leaves of the tree). Model predictive ability < 0.8 and ≥ 0.6, equivalent to Misclassification error > 0.2 and ≤ 0.4 (i.e., between 60 and 80% of the data in the predicted category is correctly classified by the model)
3	High	High confidence in the model prediction for the habitat categories (leaves of the tree). Model predictive ability ≥ 0.8, equivalent to Misclassification error ≤ 0.2 (i.e., more than 80% of the data in the calibration dataset is correctly classified by the model)

Figure 14: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Plaice.

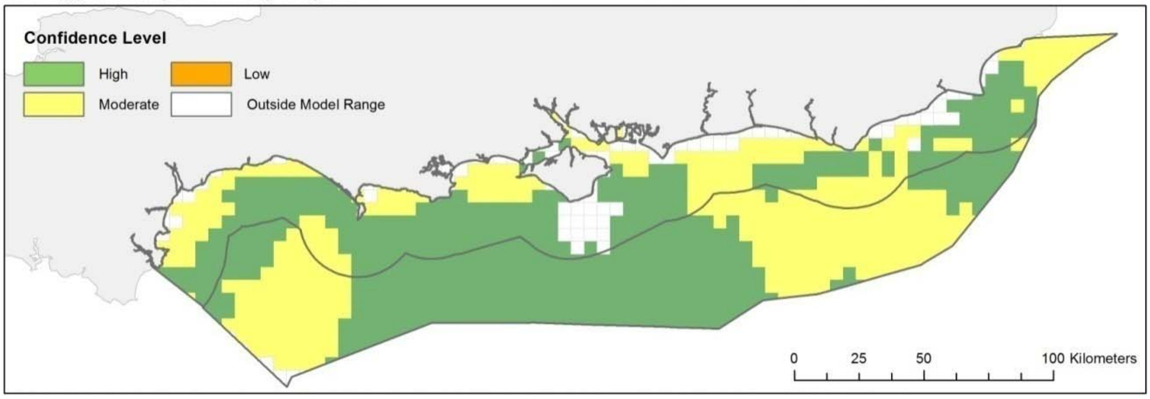
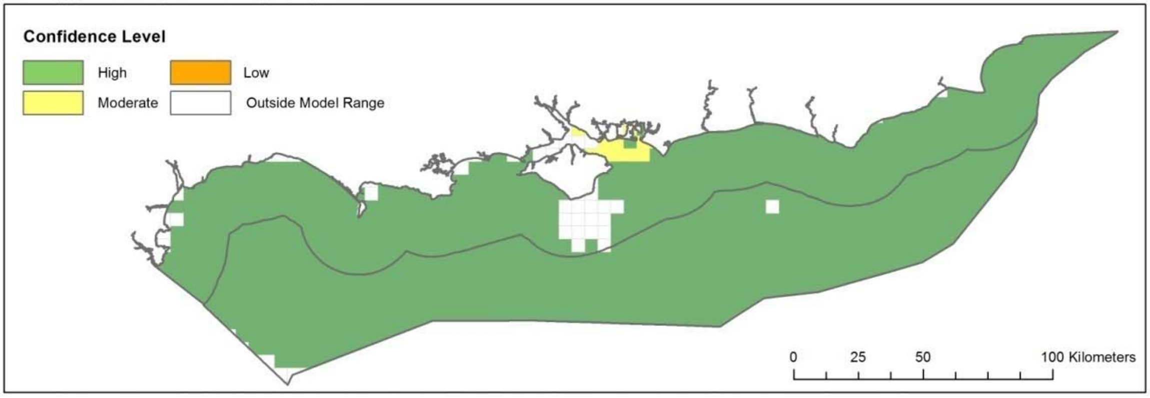


Figure 15: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Plaice.



Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 16: Spatial confidence in the predictive ability of the EFH model M2 (spawning habitat) for Plaice.

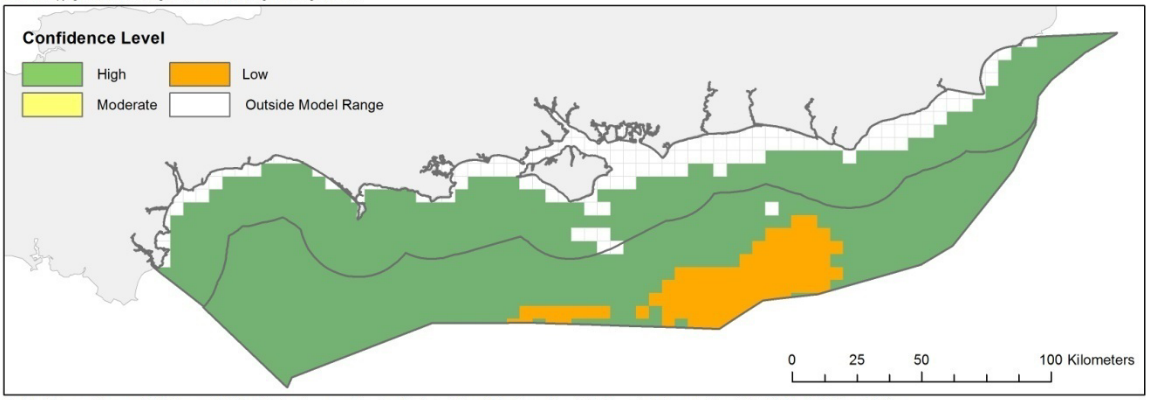


Figure 17: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Sole.

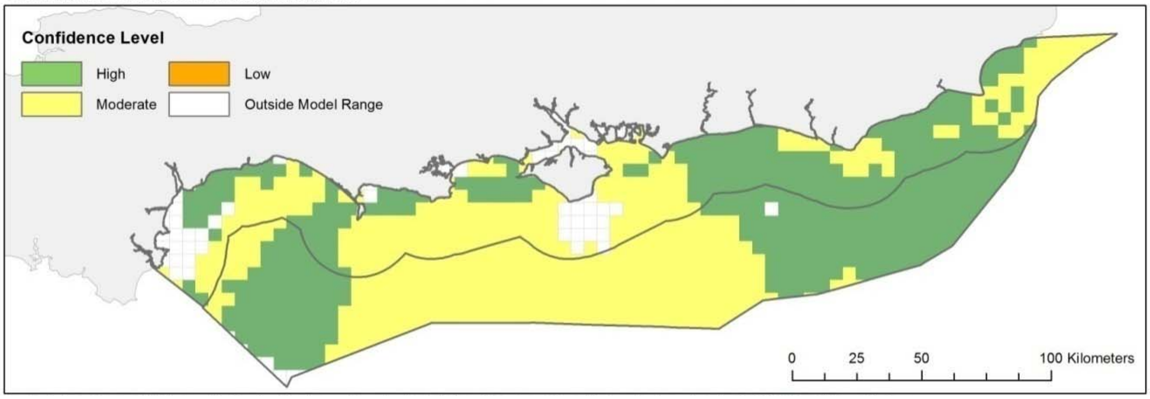
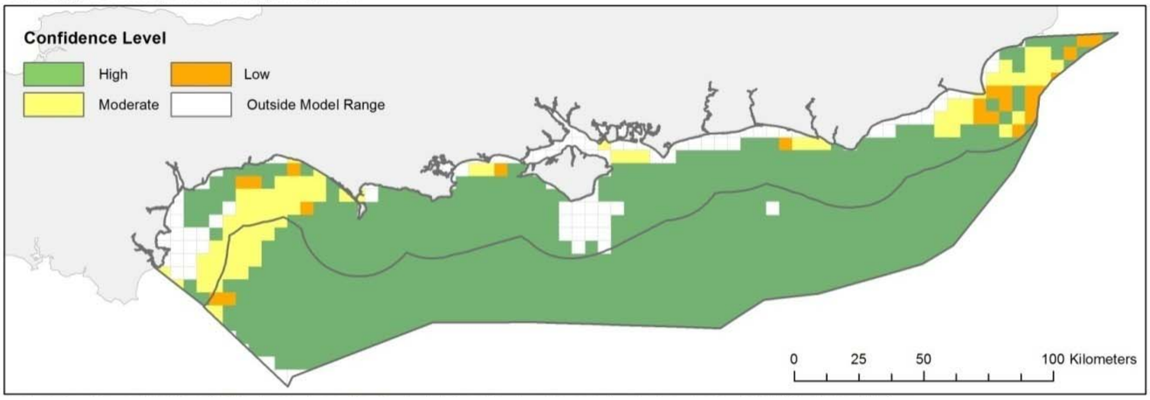


Figure 18: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Sole.



Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 19: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Lemon sole.

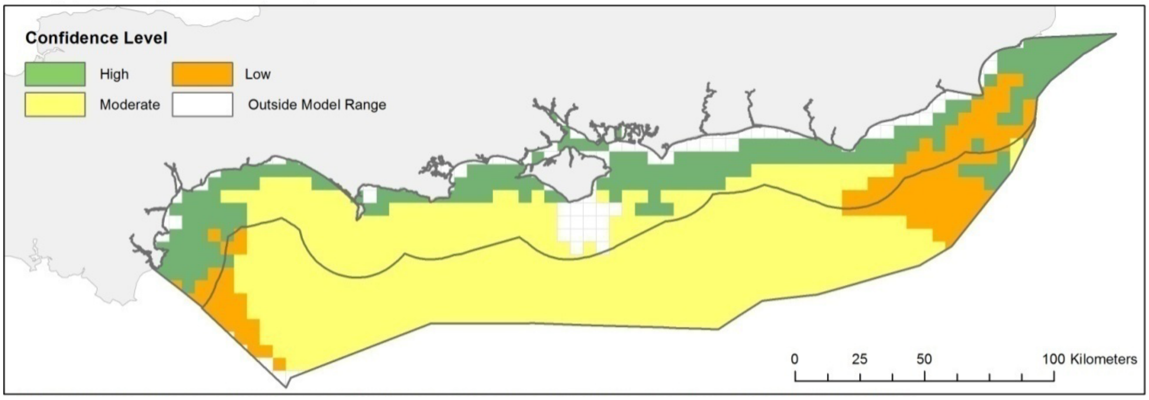


Figure 20: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Lemon sole.

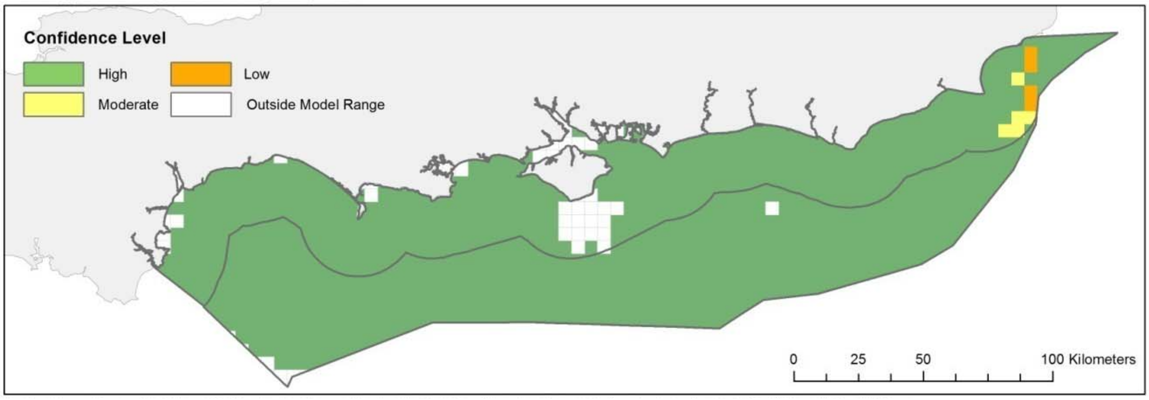
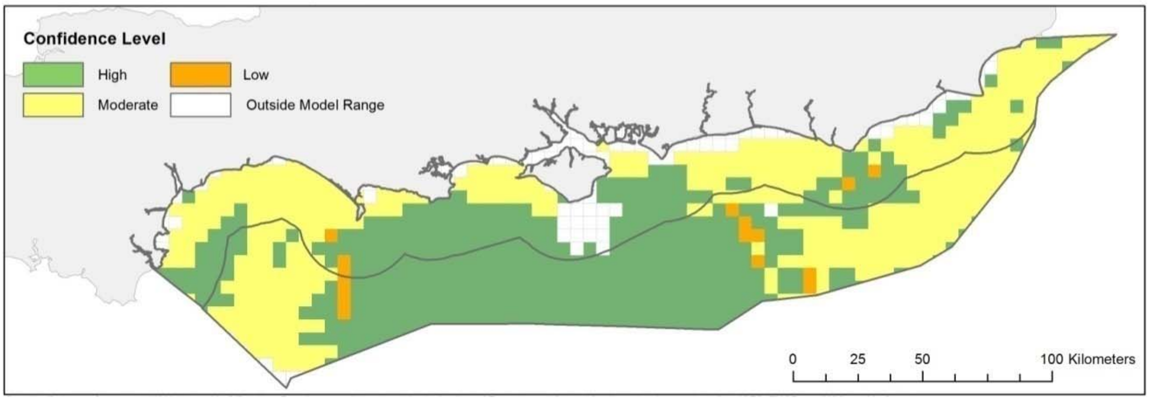


Figure 21: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Dab.



Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 22: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Dab.

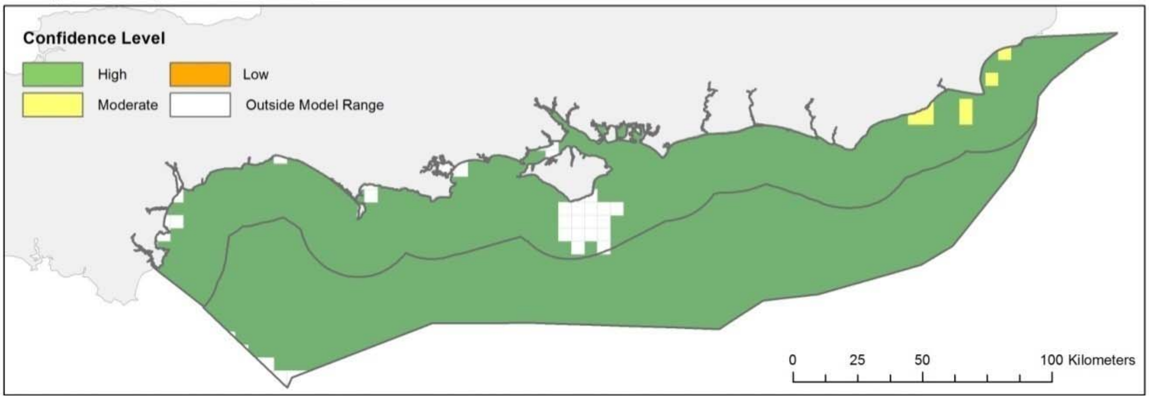


Figure 23: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Red gurnard.

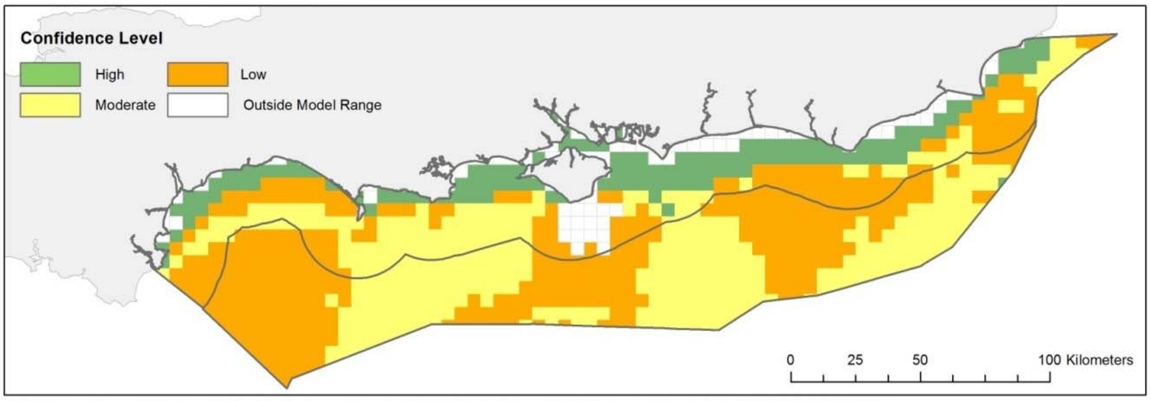
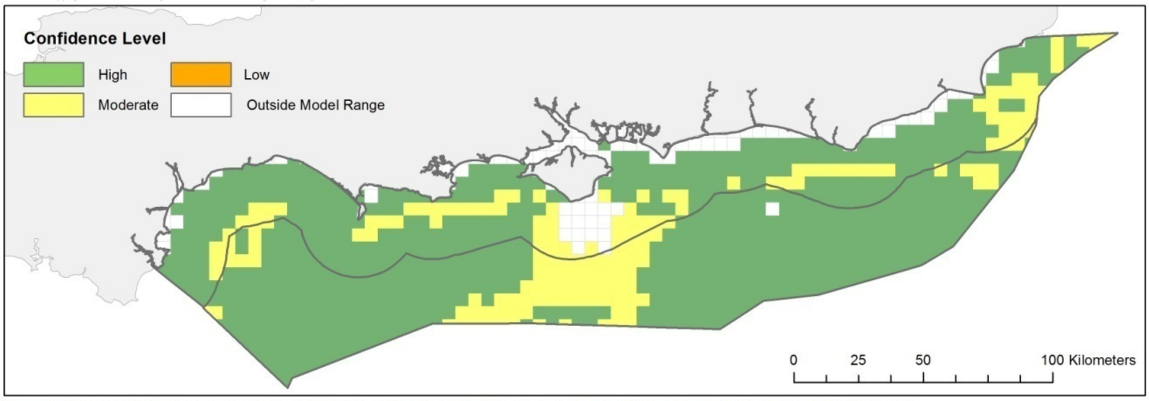


Figure 24: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Red gurnard.



Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 25: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Common dragonet.

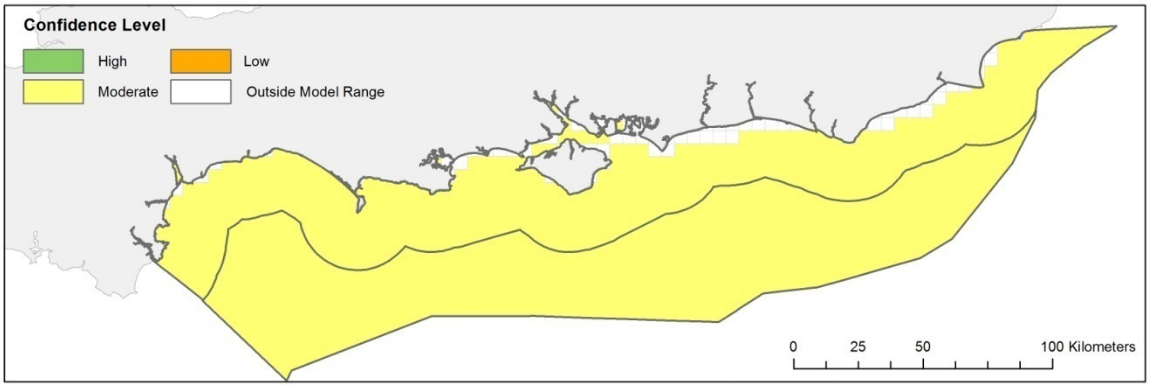


Figure 26: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Common dragonet.

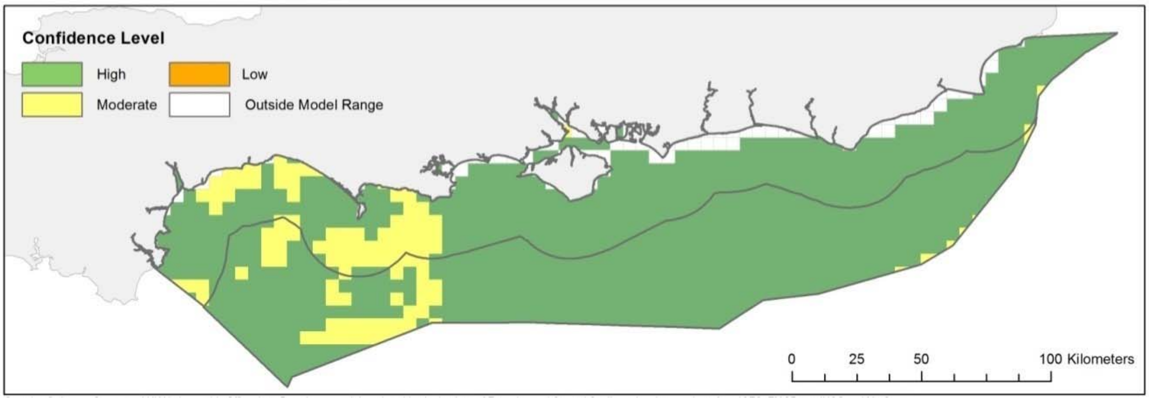
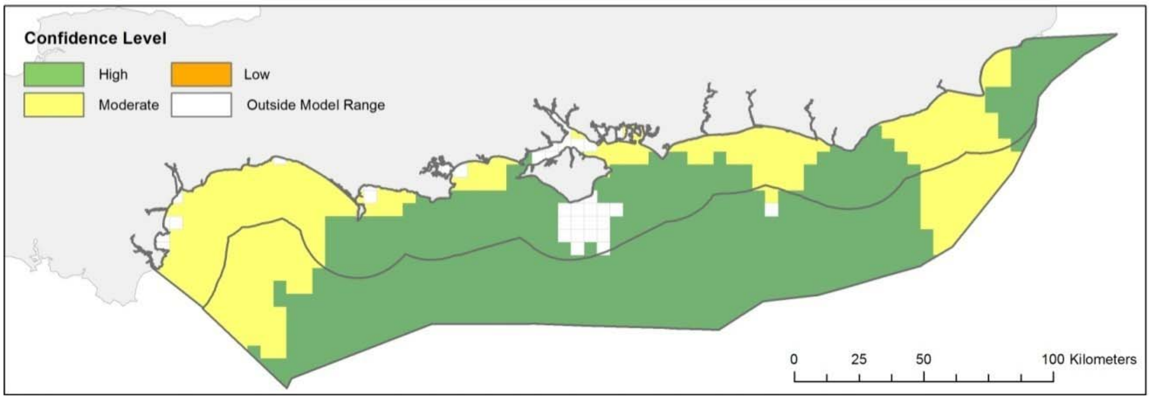


Figure 27: Spatial confidence in the predictive ability of the EFH model M1 (species habitats) for Solenette.



Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 28: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Solenette.

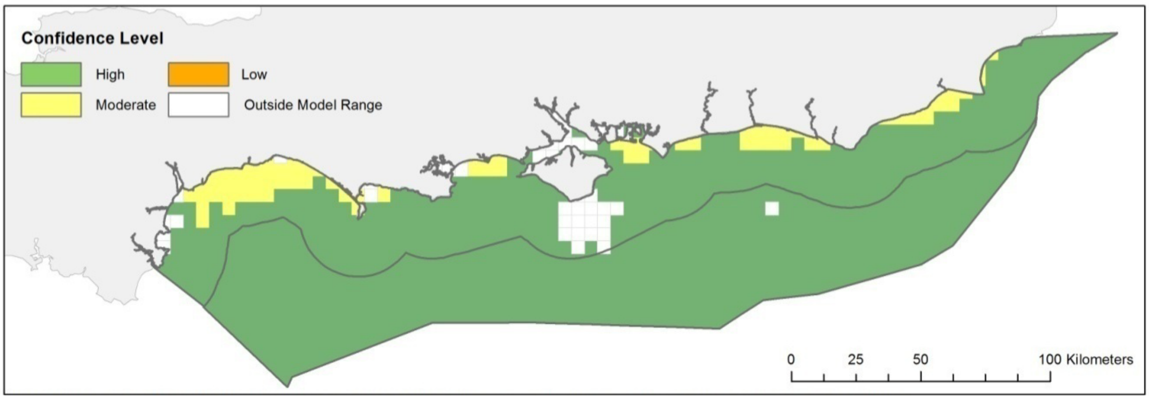


Figure 29: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Thickback sole.

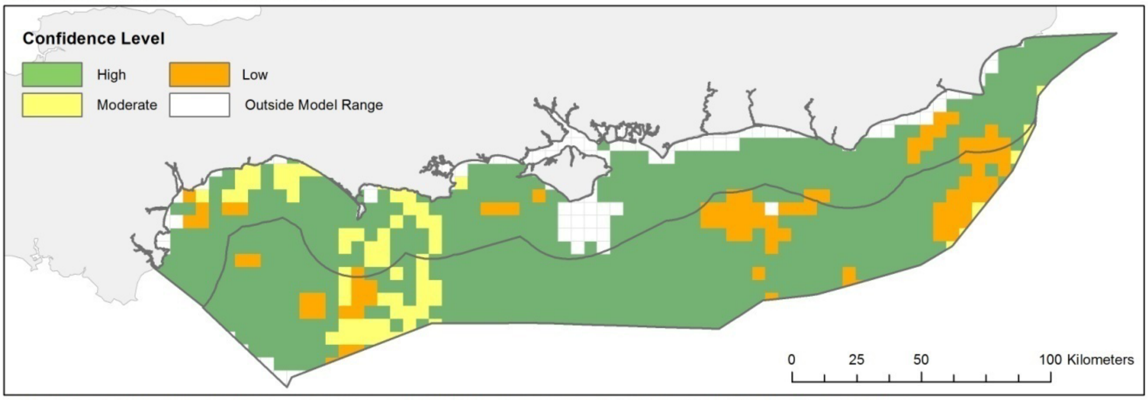
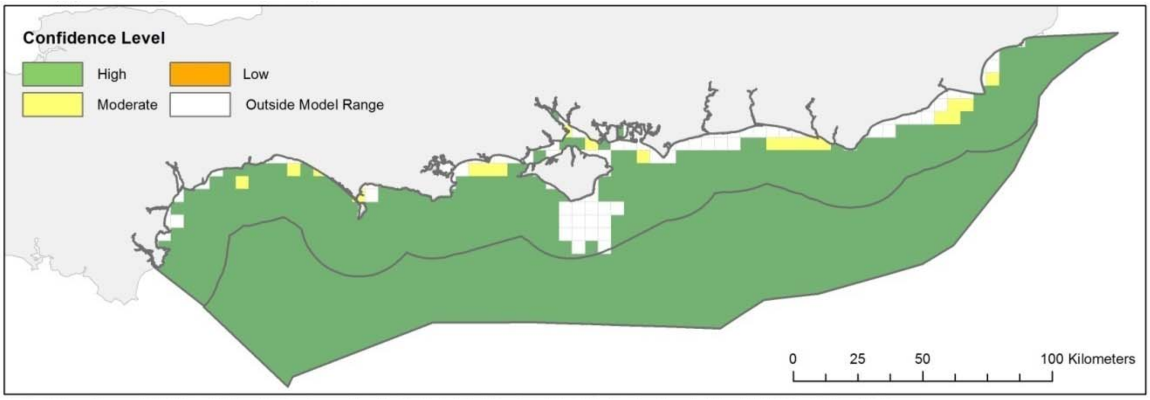
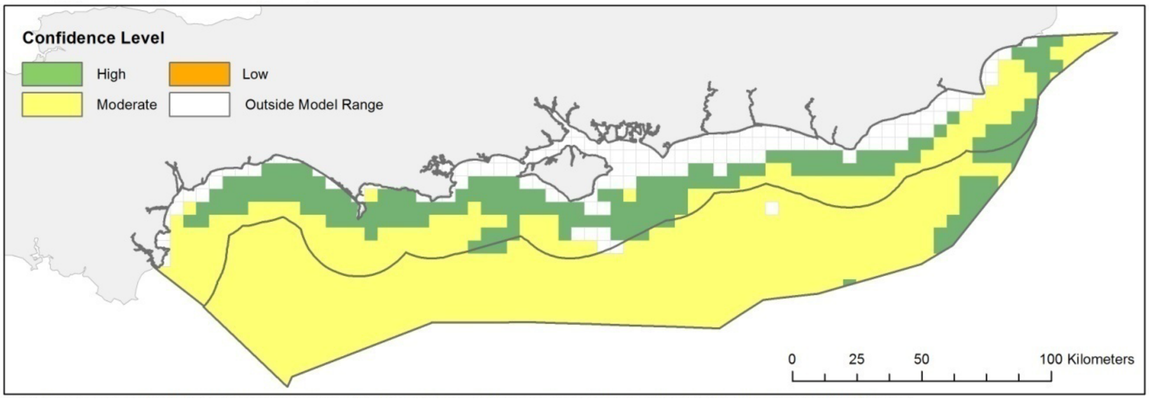


Figure 30: Spatial confidence in the predictive ability of the EFH model M2 (nursery habitat) for Thornback ray.



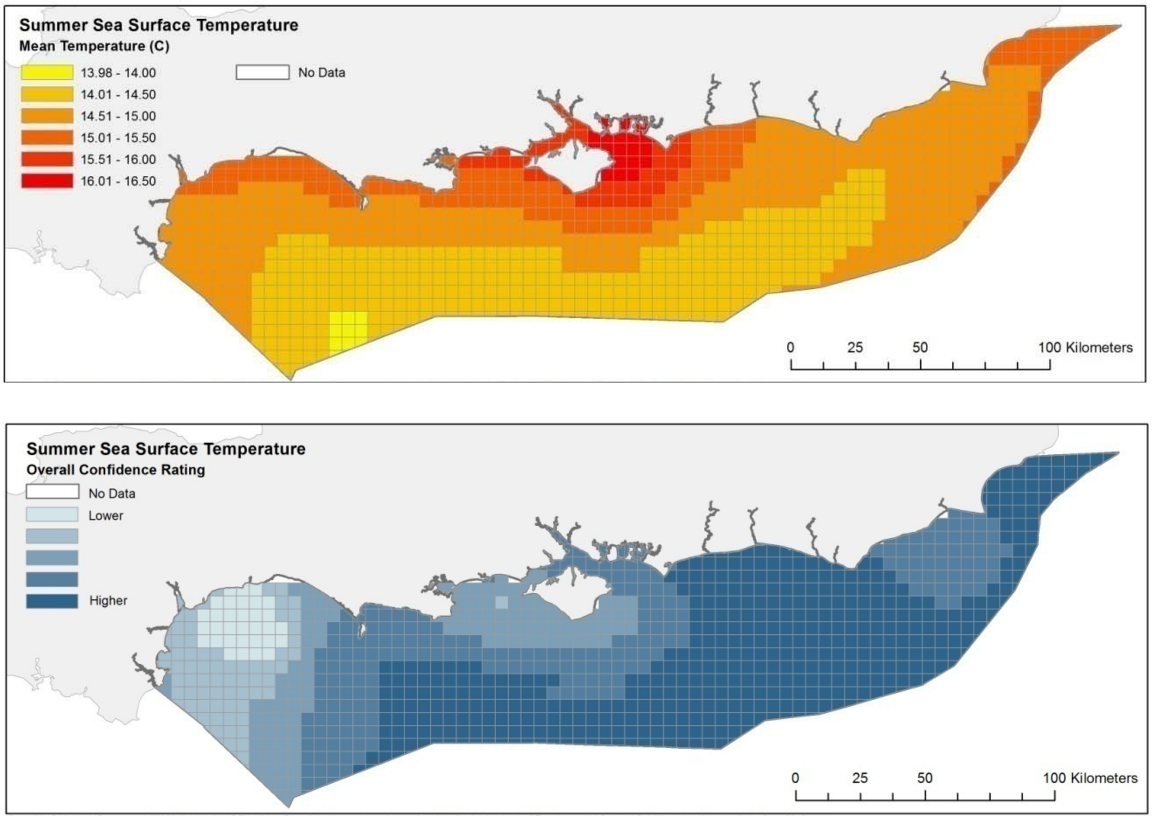
Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 31: Spatial confidence in the predictive ability of the EFH model M2 (spawning habitat) for Herring.



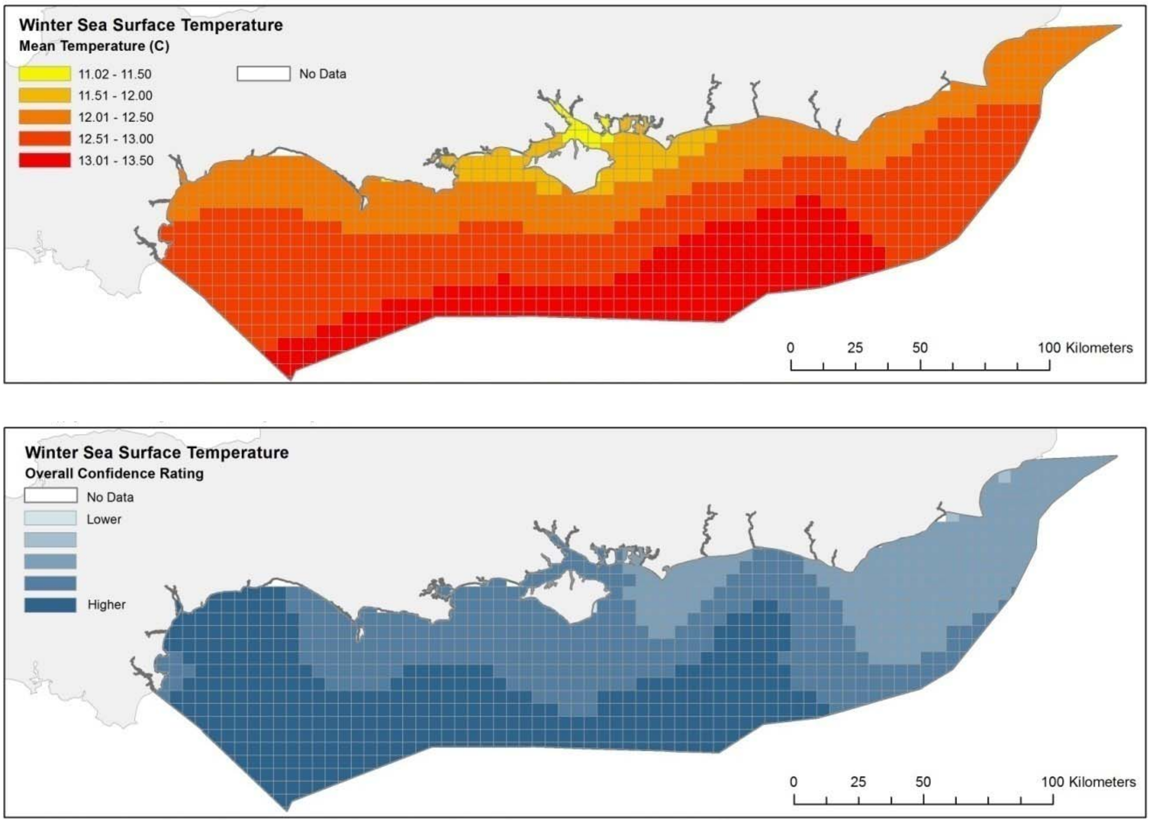
Contains Ordnance Survey and UK Hydrographic Office data. Based on a model produced by the Institute of Estuarine and Coastal Studies using data products from ICES, EMODnet, JNCC and My Ocean. © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 32: Input data layer for sea surface temperature, SST (Summer, °C) used for the EFH model implementation and associated relative confidence.



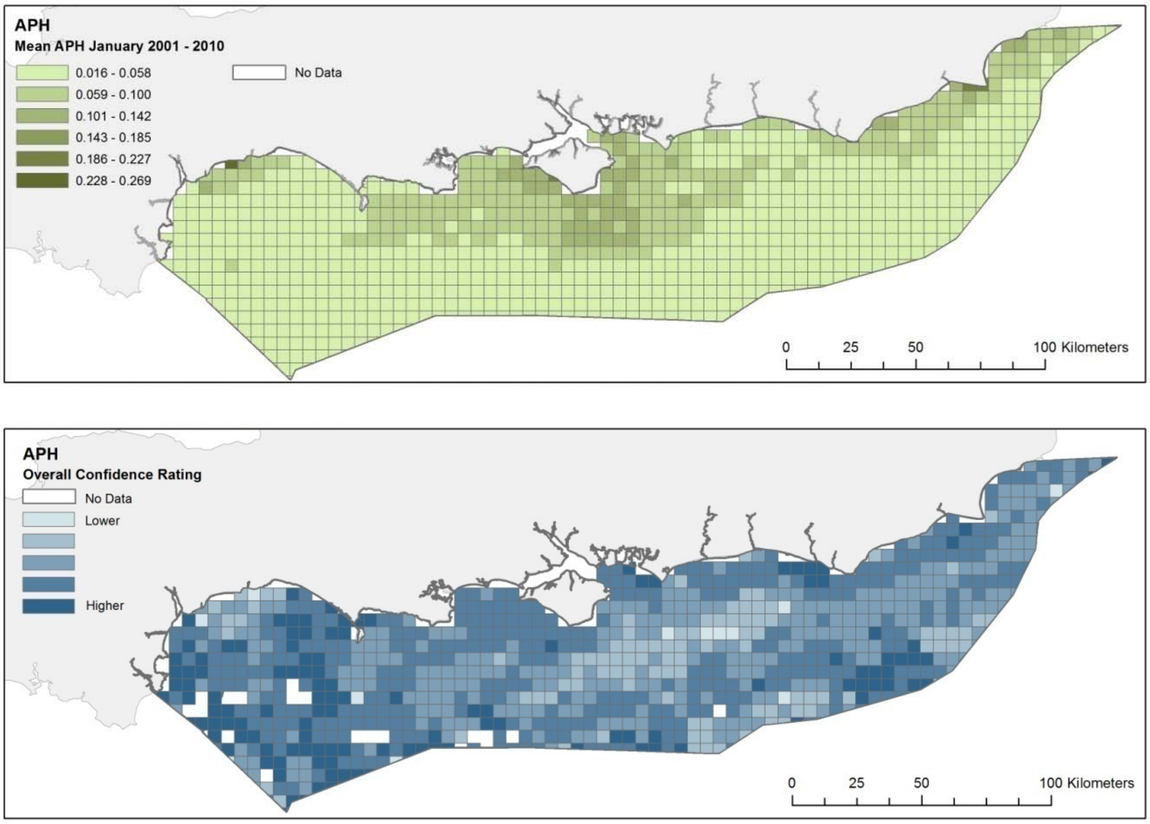
Contains Ordnance Survey and UK Hydrographic Office data. Based on data processing carried out by the Institute of Estuarine and Coastal Studies using data products from MyOcean © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 33: Input data layer for sea surface temperature, SST (Winter, °C) used for the EFH model implementation and associated relative confidence.



Contains Ordnance Survey and UK Hydrographic Office data. Based on data processing carried out by the Institute of Estuarine and Coastal Studies using data products from MyOcean © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 34: Input data layer for phytoplankton absorption coefficient, APH (January, m-1) used for the EFH model implementation and associated relative confidence.



Contains Ordnance Survey and UK Hydrographic Office data. Based on data processing carried out by the Institute of Estuarine and Coastal Studies using data products from MyOcean © Crown copyright and database right 2013. © Marine Management Organisation.

Figure 35: Input data layer for Depth (m) used for the EFH model implementation and associated relative confidence.

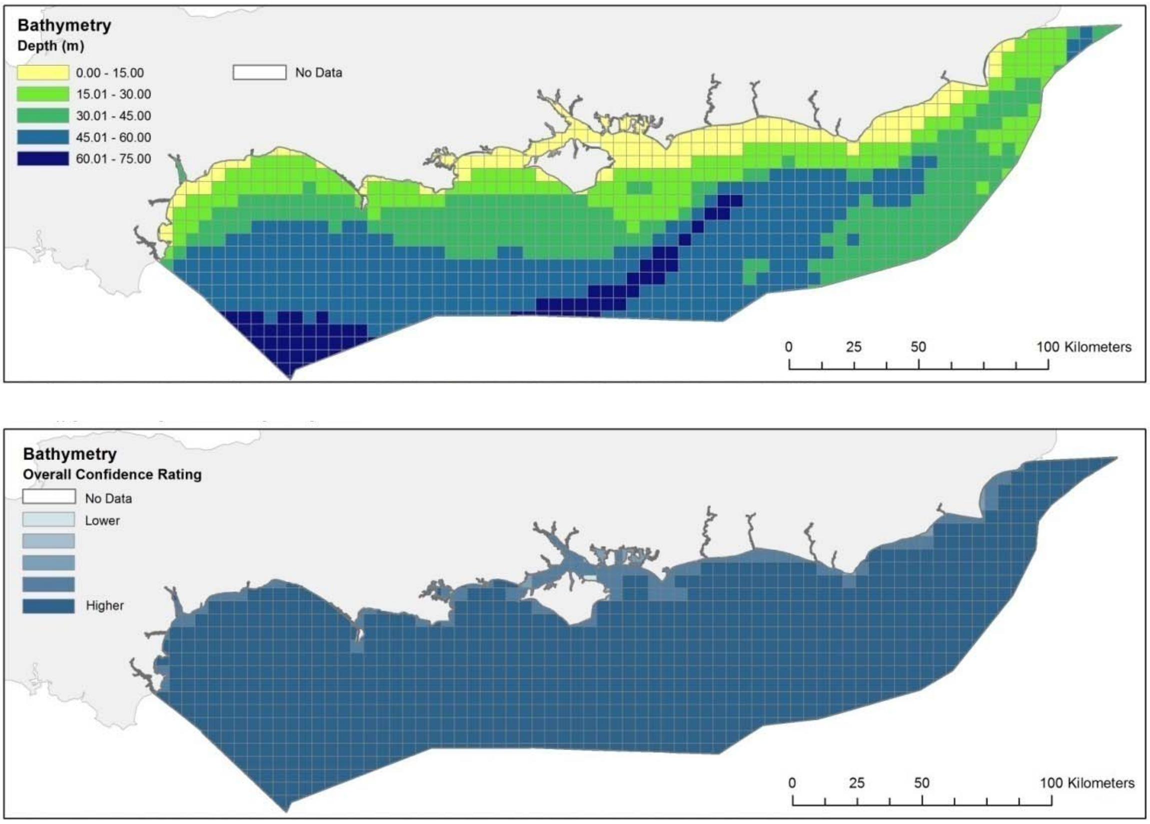


Figure 36: Input data layer for type of mixing of the water column, DomMix (Summer) used for the EFH model implementation.

Figure 37: Input data layer for tidal energy, TidE (N m^{-2}) used for the EFH model implementation.

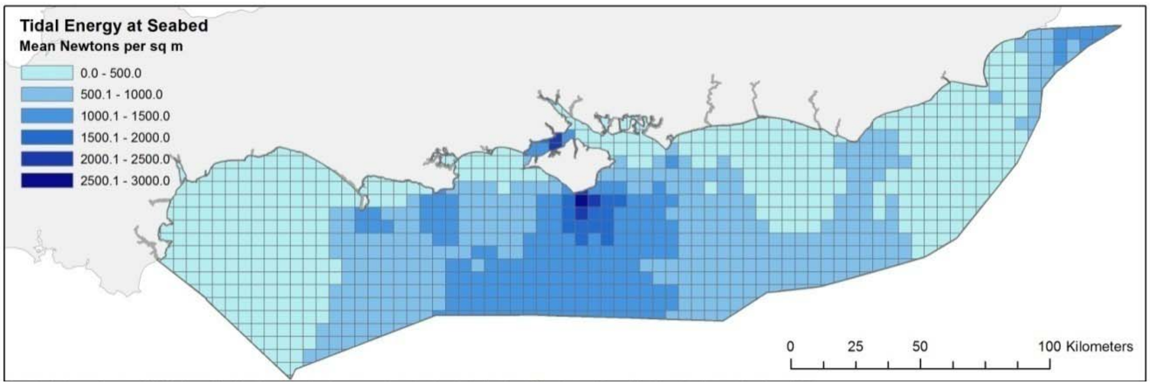
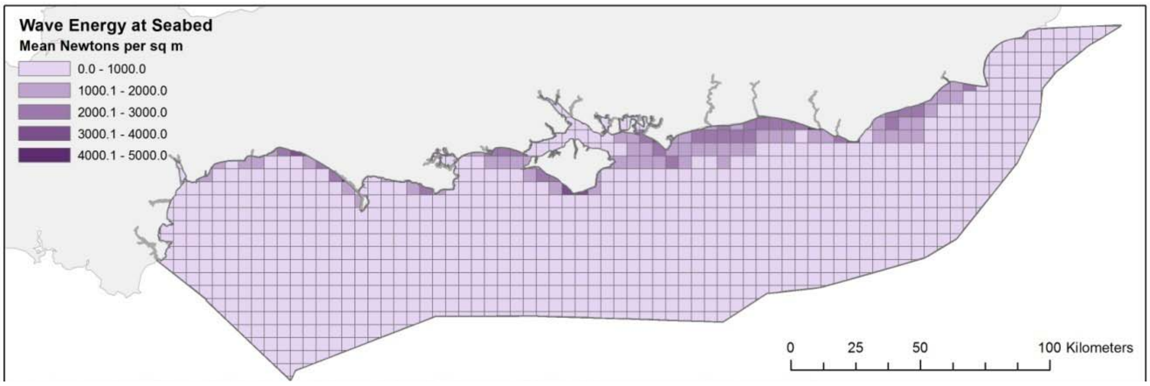
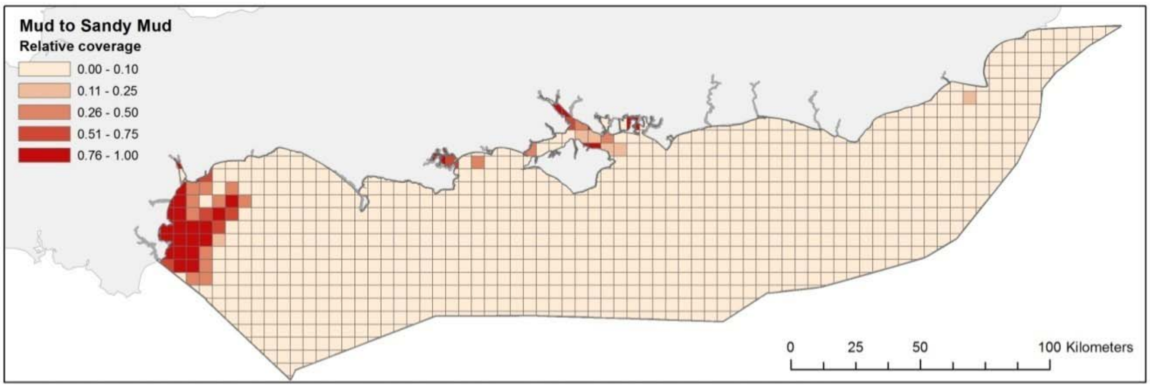


Figure 38: Input data layer for wave energy, WavE (N m^{-2}) used for the EFH model implementation.



Contains Ordnance Survey and UK Hydrographic Office data. Based on data processing carried out by the Institute of Estuarine and Coastal Studies using data products from JNCC © Crown copyright and database right 2013. Marine Management Organisation.

Figure 39: Input data layer for mud to sandy mud relative seabed coverage, M-sM used for the EFH model implementation.



Contains Ordnance Survey and UK Hydrographic Office data. Based on data processing carried out by the Institute of Estuarine and Coastal Studies using data products from EMODnet © Crown copyright and database right 2013. Marine Management Organisation.

Figure 40: Input data layer for sand to muddy sand relative seabed coverage, S-mS used for the EFH model implementation.

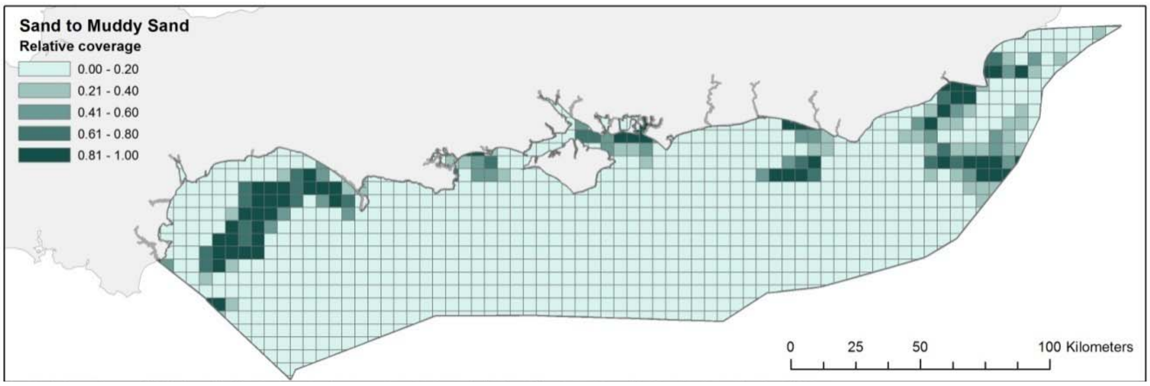


Figure 41: Input data layer for coarse sediment relative seabed coverage, Cs used for the EFH model implementation.

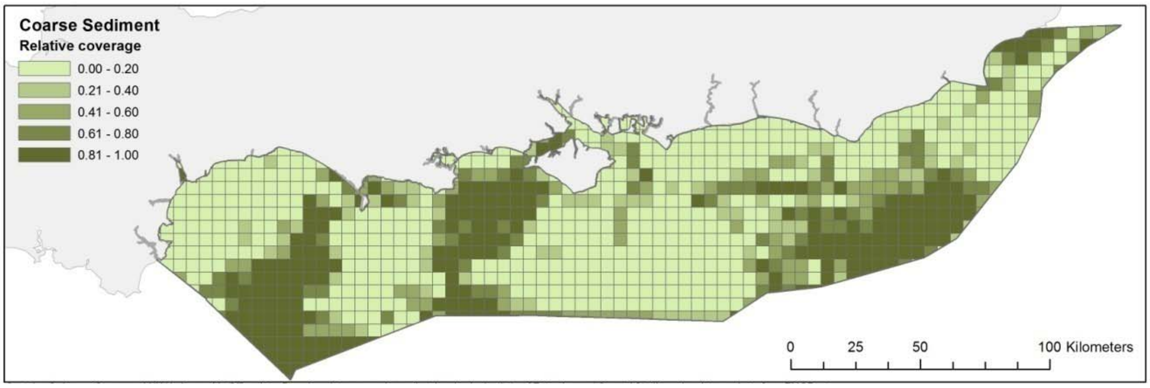
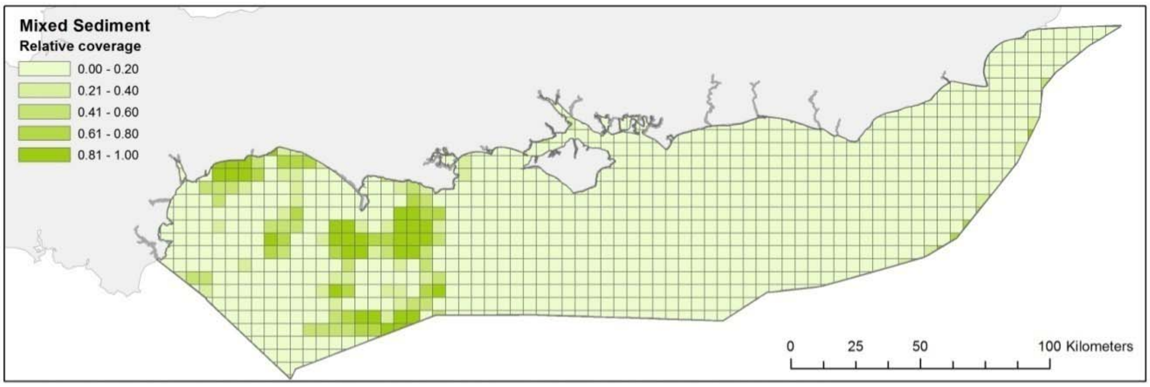


Figure 42: Input data layer for mixed sediment relative seabed coverage, Mx used for the EFH model implementation.



Contains Ordnance Survey and UK Hydrographic Office data. Based on data processing carried out by the Institute of Estuarine and Coastal Studies using data products from EMODnet © Crown copyright and database right 2013. Marine Management Organisation.

6. References

Cameron, A. and Askew, N. (eds.) (2011). EUSeaMap - Preparatory Action for development and assessment of a European broad-scale seabed habitat map final report. Available at http://jncc.gov.uk/euseamap

Elliott, M., Hemingway, K.L. (2002). Fishes in estuaries. Blackwell Science.

Ellis, J.R., Readdy, L., South, A. (2010a). MB0102 Accessing and developing the required biophysical datasets and data layers for Marine Protected Areas network planning and wider marine spatial planning purposes. Report No 15: Task 2B Distribution of highly mobile species. Centre for Environment, Fisheries and Aquaculture Science (Cefas), report to Defra, August 2010.

Ellis, J.R., Milligan, S., Readdy, L., South, A., Taylor, N., Brown, M. (2010b). MB5301 Mapping spawning and nursery areas of species to be considered in Marine Protected Areas (Marine Conservation Zones). Report No 1: Final Report on development of derived data layers for 40 mobile species considered to be of conservation importance. Centre for Environment, Fisheries and Aquaculture Science (Cefas) report to Defra, August 2010.

Ellis, J.R., Milligan, S., Readdy, L., Taylor, N., Brown, M. (2012). Spawning and nursery grounds of selected fish species in UK waters. Sci. Ser. Tech. Rep., Cefas Lowestoft, 147: 56 pp.

Faraway, J.J. (2006). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman and Hall/CRC, 312 pp.

Froese, R. and D. Pauly. Editors. (2013). FishBase. World Wide Web electronic publication. www.fishbase.org, version (04/2013).

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

Harrell, F.E., Lee, K.L., Mark, D.B., (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387.

Highland Statistics Ltd. (2000). Brodgar. Software package for Multivariate Analysis and Multivariate Time Series Analysis. Version 2+. www.brodgar.com

Lauria V., Vaz S., Martin C, Mackinson S., Carpentier A. (2011). What influences European plaice (*Pleuronectes platessa*) distribution in the eastern English Channel? Using habitat modelling and GIS to predict habitat utilization. ICES Journal of Marine Science, 68: 1500-1510.

Lawler J.J., Edwards, Jr. T. (2002). Landscape patterns as habitat predictors: building and testing models for cavity-nesting birds in the Uinta Mountains of Utah, USA. Landscape Ecology, 17: 233–245.

Stephens D., Kral F., Maxwell TAD, Robinson P., Engelhard G.H. (2010). Action 8: Cartography and habitat modelling. Report of the CHARM project - CHannel integrated Approach for marine Resource Management, INTERREG IV A France (Channel) – England cross-border European cooperation programme.

Zuur, A.F., Ieno, E.N., Smith, G.M. (2007). Analyzing Ecological Data. Springer, 672 pp.

Zuur, A.F., Ieno, E.N., Elphick, C.S. (2009). A protocol for data exploration to avoid common statistical problems. Methods in Ecology & Evolution, 1: 3-14.