# EFH project – Master record and List of data files

## 1. Master record

Description and lineage of the overall collection of data products obtained from the EFH project.

### 1.1 Description

These data products describe the distribution of Essential Fish Habitats (EFH) for fish and shellfish species in Scottish waters. This contributes in developing the evidence base on EFH, which will help the Scottish Government in targeting future work towards supporting future planning and (offshore wind) project level assessments in Scottish waters.

This resource is composed of 40 data products, between GIS data layers (as geodatabases and shapefiles) and data spreadsheets (as CSV files), which provide spatial predictions of the EFH distribution (as obtained from EFH modelling and habitat proxy assessment) for 17 species/life stages between fish and shellfish and survey and environmental data to support and validate such predictions.

#### EFH model spatial outputs

EFH models outputs provide evidence on the EFH of 17 species/life stages in UK offshore waters, including: refugia for lesser sandeel (*Ammodytes marinus*) and Norway lobster; nursery habitats for juveniles of plaice, lemon sole, common sole, anglerfish, whiting, blue whiting, hake, sprat, mackerel, and long finned squid (body size criteria and seasonality were used to identify juveniles); spawning grounds for spawning adults of whiting, cod, haddock, Norway pout and mackerel eggs.

The output is a gridded vector dataset (5 x 5 km grid covering UK marine waters) where the variable mapped is a categorical variable combining the EFH model prediction of presence/absence of aggregations of each species/life stage and the associated confidence (High/Low; based on predicted probability of presence). Blank cells are grid cells where no valid model prediction could be obtained (i.e. where one or more environmental variables fall outside the range of variability in the surveys from which the EFH model was calibrated).

The temporal reference for the EFH model outputs (as defined by the temporal reference for the environmental data layers to which the EFH model is applied) includes mean conditions over the period 2010-2020 (for all the species’ outputs), in individual years (namely 2010, 2015 and 2020; for Lesser sandeel, *Nephrops* and juvenile Anglerfish only) and under a climate change scenario. The latter accounts for changes (compared to baseline in year 2015) in the sea bottom temperature (SBT) and/or water depth inshore due to sea level rise (SLR) as per UKCP09 (marine and coastal projections for SBT and/or SLR over a 100 years span, based on a medium emission scenario and medium probability, applied to baseline, as relevant for the species-specific EFH model). Appropriate seasonality, as relevant to the specific species/life stage EFH (e.g. spawning period, period of peak abundance of juveniles) also contributes to defining the temporal refence for the environmental scenarios.

#### EFH habitat proxy spatial outputs

Habitat proxy spatial outputs provide evidence on habitat proxies for EFH functions for 19 species/life stages in UK inshore waters, including: refugia for small sandeel (*Ammodytes tobianus*); spawning for herring; nursery for juvenile plaice, common sole, whiting, cod, saithe, sprat, spotted ray, European lobster, and brown crab; spawning grounds/egg nursery for thornback ray and spurdog; generic habitat for velvet crab, common cockle, dog cockle, razor clam, common whelk, and dog whelk.

The habitat proxy assessment is based on the expert scoring of inshore habitat types (based on EUNIS habitat classification, Levels 3 and 4) according to their ability to provide EFH function for each of the selected species (with associated confidence). The results (EFH habitat proxy scores and associated confidence) are mapped based on the combined distribution of inshore EUNIS habitat types (at Levels 3 and 4) in the UK.

#### Supporting/validating survey spatial data

Point vector layers provide additional lines of evidence on the observed species/life stage distribution from the fish survey catch data that supported the model calibration (summarised as relative frequency of occurrence of aggregations from survey data across the 2010-2020 period) and from additional surveys (see lineage for more details) providing validation of the EFH model predictions.

#### EFH model input datasets

Datasets (as CSV files) used as input for the EFH modelling of aggregations of the species/life stages of interest as indicators of potential nursery EFH in UK waters. The datasets combine key fish/shellfish catch variables (e.g. presence/absence of aggregations) from a variety of fish surveys across the period 2010-2020 and the environmental variables associated with each survey event (as extracted from relevant environmental data layers).

Datasets are provided separately for individual species/life stages, with the file name being a combination of the species (namely Anglerfish, Blue-whiting, Cod, Haddock, Hake, Lemon-sole, Mackerel, Nephrops, Norway-pout, Plaice, Sandeel, Sole, Sprat, Squid, Whiting, Whitin), the EFH function (namely N for nursery, S for spawning, R for refugia) and the indicator life stages (identified by the maximum juvenile body size for nursery, as early egg stage (EG1) or spawning (running) adults for spawning, all life stages for refugia) to which the data refer to.

#### Fish survey data (raw/not processed)

Datasets (as Excel workbooks) containing the fish survey data extracted from relevant survey databases, and subsequently processed and used for the EFH modelling (as input data for the model calibration and/or for subsequent validation of the resulting model maps).

Datasets are provided separately for individual survey programmes (see lineage for details). The datasets in these files include the original data from several source databases, pre-filtered (for selection of the fish/shellfish species of interest, the years between 2010 and 2020, and sample locations within UK waters) and with minor editing (mainly consisting in the addition of fields/columns for indexing and variable detailing).

### 1.2. Lineage

#### EFH model spatial outputs

These spatial outputs are the result of the application of the EFH models generated for selected species/life stages to environmental scenarios, as defined by the selection of environmental data from spatial layers with a specific temporal reference (namely, mean across the period 2010-2020, individual years 2010, 2015 and 2020, and climate change scenario resulting from changes in sea bottom temperature and inshore depth (from sea level rise)). For reference to the source environmental data layers, see “EFH model input datasets” below.

EFH distribution is identified by classification tree model predictions of presence/absence of aggregations of the species/life stages based on the distribution of key environmental predictors. The results (EFH model predictions and associated confidence) are mapped over a spatial grid (5 x 5 km) covering UK waters.

Data are given for 17 species/life stages, named as follows: AnglJ = Anglerfish juvenile; BluewJ = Blue whiting juvenile; CodS = Cod spawning; HaddS = Haddock spawning; HakeJ = Hake juvenile; LemSJ = Lemon sole juvenile; MackJ = Mackerel juvenile; MackEgg = Mackerel eggs; Nephr = Nephrops; NPoutS = Norway pout spawning; PlaiceJ = Plaice juvenile; Sandeel = Lesser (Raitt's) sandeel; SoleJ = Common sole juvenile; SpratJ = Sprat juvenile; SquidJ = Long finned squid juvenile; WhitJ = Whiting juvenile; WhitS = Whiting spawning.

For additional details on methods see project report: Franco A., Smyth K., Thomson S. (2022) Developing Essential Fish Habitat maps for fish and shellfish species in Scotland. Report to the Scottish Government, November 2022.

#### EFH habitat proxy spatial outputs

These spatial outputs were derived from EUNIS habitat layers including:

* The EUSeaMap 2021 Broad-Scale Predictive Habitat Map for Europe A report on the methods used in the 2021 version of EUSeaMap (Vasquez et al., 2021) can be found here: <https://archimer.ifremer.fr/doc/00723/83528/>.
* The EUNIS 2019 habitat maps from surveys (compiled, GB only). More details of the 2019 EUNIS habitats maps from surveys can be found at <https://www.emodnet-seabedhabitats.eu/helpdesk/map-viewer-help/habitat-maps-from-surveys/#EUNIShabitat>.

These spatial layers were clipped to the UK study area and truncated separately to EUNIS Levels 3 and 4. The resulting layers were joined to the score (and associated confidence) assigned (through literature review and expert assessment) by the habitat proxy assessment to each biotope for each species according to the potential habitat suitability for specific EFH functions such as refugia, nursery, and spawning. The score for habitat proxies ranges between 1 (for habitats for which there is evidence of use by the species/life stage, but not with the same intensity (e.g. frequency or abundance) as other habitats) and 3 (habitats that are identified in the literature as important/primary habitats with which the species/life stage more frequently or with higher abundance compared to other habitats). The associated confidence score ranges from 1/Low to 3/High, reflecting the amount and detail of information supporting the species assessment as a whole (across all habitats).

#### Supporting/validating survey spatial data

These point vector layers describe observed survey catch data for the species/life stages of interest as derived from different survey programmes, including:

* **Surveys2010-20\_AggrRelFreq\_Allspp\_POINTS:** Spatial vector data layer containing the processed data derived from various input fish surveys on frequence of occurrence of aggregations of 17 species/life stages of fish/shellfish including: AnglJ = Anglerfish juvenile; BluewJ = Blue whiting juvenile; CodS = Cod spawning; HaddS = Haddock spawning; HakeJ = Hake juvenile; LemSJ = Lemon sole juvenile; MackJ = Mackerel juvenile; MackEgg = Mackerel eggs; Nephr = Nephrops; NPoutS = Norway pout spawning; PlaiceJ = Plaice juvenile; Sandeel = Lesser (Raitt's) sandeel; SoleJ = Common sole juvenile; SpratJ = Sprat juvenile; SquidJ = Long finned squid juvenile; WhitJ = Whiting juvenile; WhitS = Whiting spawning.  
  The layer integrates survey data for the period 2010-2020 as derived from different sources, incl. ICES International Bottom Trawl Survey, Beam Trawl Surveys and Mackerel egg surveys; Marine Scotland sandeel dredge surveys; Marine Scotland Northern Shelf Anglerfish Surveys (SIAMISS). See “Fish survey data (raw/not processed)” and “EFH model input datasets” for details on the survey methods.   
  The survey data were gridded on the 5 x 5 km grid used for the EFH model spatial implementation, and the relative frequency of aggregations aggregations of individual fish/shellfish species/life stages was calculated as the proportion of hauls in each grid cell containing aggregations. Values range between 0 and 1, indicating respectively that none or all of the hauls undertaken between 2010 and 2020 within a grid cell contained aggregations of the species/life stage of interest. This shows not only the spatial distribution of aggregations from the surveys, but also informs on the recurrence of such aggregations over the years. The centroid location for each grid cell is also given as lag/long to allow representation of the data as points in the map. (NA is indicated were no survey data are available in a grid cell).
* **Anglerfish\_juv\_SIAMISS2021**:Spatial point dataset containing 2021 data from Marine Scotland Science for the Scottish-Irish anglerfish and megrim industry-science survey (SIAMISS) undertaken to monitor the Northern shelf anglerfish stock in subareas 4 and 6, and division 3.a (North Sea, Rockall and West of Scotland, Skagerrak and Kattegat).This spatial layer was used to validate EFH model maps for juvenile anglerfish.
* **Nephrops\_Burrows2007-16\_AF**: Spatial point data from the Nephrops TV surveys undertaken annually between 2007 and 2016 by Marine Scotland Science in Scottish waters. These video surveys are restricted to suitable Nephrops habitat (identified based on sediment characteristics or fishery VMS data), and they measure Nephrops burrow density rather than individuals, thus providing a direct assessment of the refugia resource. Data on haul characteristics (location and timing), and average density of borrows are included.   
  This spatial layer was used to validate EFH model maps for *Nephrops* refugia.
* **Sandeel Dredge Survey 2021:** Spatial point layer derived from survey data from the sandeel dredge survey undertaken by Marine Scotland in December 2021 at sandeel fishing grounds along the west coast of Scotland. This spatial layer was used to validate EFH model maps in the west coast of Scotland area.
* **WCFD**: Spatial layers containing processed catch data from the West Coast of Scotland Demersal Fish Project (WCDF), a joint industry/science project (managed by Marine Scotland Science and the Scottish Fishermen’s Federation Services Limited) to better understand the distribution and abundance of demersal fish on the west coast of Scotland. This includes point sample/haul catch data from offshore and inshore quarterly trawl surveys undertaken between December 2013 and November 2014 (Ramiro Sánchez et al. 2015).   
  Two layers are distinguished based on the variable mapped, namely a categorical variable identifying the occurrence of juvenile (J) individuals in the catch and a continuous variable identifying the catch CPUE (individuals per haul per hour) of juvenile (J) individuals for fish/shellfish species (Sprat = sprat; Mack = mackerel; Cod = cod; Whit = whiting; Saithe = saithe; BWhit = blue whiting; Hake = hake; Plaice = plaice; Sole = common sole; LSole = lemon sole; Angl = anglerfish; TRay = thornback ray; Sray = spotted ray; Spurd = spurdog; Squid = long finned squid). For each species, the cut-off body size used to identify juveniles in the survey catches are also indicated in the field name (e.g. WhitJ<16 = juvenile whiting with body length <16 cm). The categorical variable was derived from the catch CPUE data as follows: absence, where CPUE (of juveniles) was 0; presence, where CPUE (of juveniles) was >0, but lower than the aggregation threshold (top quartile CPUE); aggregation, where CPUE (of juveniles) was >0 and included in the top quartile of juvenile CPUE. For each species, the relevant season considered to identify juveniles occurrence in the survey catches are also indicated in the field name (as quarters Q1, Q2 etc).  
  Spatial reference: survey haul locations (centroid, lat/long). Temporal reference: data for individual survey events within the year 2013/14.   
  The spatial layer with the categorical variable was used to validate EFH model maps in the west coast of Scotland area, whereas the one based on CPUE was used to validate habitat proxy maps in the west of Scotland area.

For additional details on methods used to process and map the data, see report: Franco A., Smyth K., Thomson S. (2022) Developing Essential Fish Habitat maps for fish and shellfish species in Scotland. Report to the Scottish Government, November 2022.

#### EFH model input datasets

The **fish data** in the EFH model input datasets originate from different survey databases including:

* For juvenile Anglerfish: The Scottish-Irish anglerfish and megrim industry-science survey (SIAMISS), a trawl survey undertaken to monitor the Northern shelf anglerfish stock in subareas 4 and 6, and division 3.a (North Sea, Rockall and West of Scotland, Skagerrak and Kattegat). Data pre-selected for the surveys within the period 2010-2020 (data sourced from Marine Scotland).
* For juveniles of whiting, blue whiting, mackerel, hake, sprat, and long finned squid; for spawning adults of cod, whiting, haddock, and Norway pout; and for individuals (any life stage) of *Nephrops*: International Bottom Trawl Surveys (IBTS), undertaken by different countries in waters around the UK (including different combinations of the EVHOE, IE-IGFS, NS-IBTS, SCOROC, SCOWCGFS survey programmes, depending on the species), under ICES coordination. Data pre-selected for the surveys within the period 2010-2020 (data sourced from ICES DATRAS database <http://datras.ices.dk/Home/Default.aspx>).
* For juveniles of flatfish species (lemon sole, plaice, common sole): International Beam Trawl Surveys (BTS) undertaken by different countries in waters around the UK, under ICES coordination (data sourced from ICES DATRAS database <http://datras.ices.dk/Home/Default.aspx>).
* For mackerel eggs: Mackerel and Horse Mackerel Egg Survey (MEGS) undertaken by different countries in the Northeast Atlantic under ICES coordination. Data pre-selected for the surveys within the period 2010-2020 (data sourced from ICES eggs and larvae surveys online database, <https://eggsandlarvae.ices.dk/Download.aspx>)
* For sandeel: the annual sandeel dredge survey undertaken by Marine Scotland in December at sandeel fishing grounds off the Firth of Forth and Turbot Bank (in Sandeel Area 4, East coast of Scotland, where the largest of the sandeel stocks in Scottish waters is and there is an active fishery for it). Data pre-selected for the surveys within the period 2010-2020 (data sourced from Marine Scotland).

The fish survey catch data (CPUE by species by length class) were processed to obtain presence/absence of aggregations (field PAaggr) based on CPUE data for the relevant species/life stage in the season relevant to that life stage (e.g. quarter Q1 for spawning season of cod, haddock, Norway pout and haddock; quarter Q3 for post-spawning season for juvenile flatfish). Catches for juvenile life stages only were selected based on body size (e.g. <28 cm, for 0- and 1-group anglerfish). Catches for spawning ('running') individuals only were selected based on the proportion of adults in spawning stage as derived from spawning maturity- age-length keys (SMALK) available for the selected catch data (from Datras database). Aggregations of the species/life stage of interest were identified where the associated CPUE was included in the top quartile of the overall CPUE distribution for the species/life stage.

Variables derived from the fish survey data for each survey event/haul include: CPUEtot (total individuals of the species per hour per haul), CPUEjuv (juvenile individuals of the species per hour per haul), PA (presence/absence, as 1/0, of juveniles of the species in the haul), and PAaggr (presence/absence of aggregations (i.e. at top quartile CPUE) of juveniles of the species in the haul).

**Environmental variables** associated with the individual survey event/haul were extracted from different environmental data layers according to survey location (Lat, Long as central point location for the survey haul) and timing (date or month, year). These included (in all datasets):

* Distance from coast (high water level) (Dist, m), calculated for each survey location from EMODnet layer;
* Depth (m), from EMODnet Bathymetry 2020 source layer (also including a measure of confidence, Bathy\_Conf, provided with the bathymetry layer);
* Slope (degrees), calculated from EMODnet Bathymetry 2020 layer as mean slope within 5 km buffer from survey location;
* Substratum type (Substr), calculated as predominant substratum category within 5 km of survey location from EMODnet Seabed Habitats 2019 data layer (also including a measure of confidence associated with the substratum type allocation);
* Presence of Sandbank Habitat (SandbH01, class 0/1), Presence of Structured Habitats (StructH01, class 0/1), Presence of Vegetated Habitats (VegH01, class 0/1), calculated as presence of relevant habitats within 5 km of survey location based on a combination of source data layers (OSPAR 2020, EMODnet Seabed Habitats 2019, GEMS 2019)
* Kinetic energy of currents at seabed (CUR, N m2/s), obtained for each survey location from EMODnet 2017 data layer;
* Kinetic energy of waves at seabed (WAV, N m2/s), obtained for each survey location from EMODnet 2019 data layer;
* Sea surface temperature (SST, degrees Celsius), Sea Bottom Temperature (SBT, degrees Celsius), Sea Surface Salinity (SSS, PSU), Net Primary Production (NPPV, Carbon per unit volume of seawater, mg C m-3 day-1);
* Mixed layer thickness (MLT, m), calculated for each survey location as mean value for relevant month and year of survey, from E.U. Copernicus Marine Service data layers.

#### Fish survey data (raw/not processed)

Separate datasets have been derived for individual survey programmes, including:

* **NS-IBTS**: Bottom trawl surveys (offshore) undertaken by different countries in the North Sea under ICES coordination, using standardised design and methods (see Manual SISP 10, IBTSWG 2020, <http://doi.org/10.17895/ices.pub.7562>). Broad scale, standardised sampling with GOV-trawl, with catch data available for pelagic, demersal and benthic fish, but most efficient to capture demersal species (juveniles and adults). Original raw data available in ICES Datras website since 1965, include data on haul characteristics, CPUE (by species by length class), ALK (age-length keys, for some species), SMALK (sex-maturity age-length keys, for some species). The dataset here is an extract of the data downloaded from the Datras online database, combining data on haul characteristics and CPUE (by species by length class) for the selected species of interest for the project, with selection of survey data available within the period 2010-2020.
* **SCOWCGFS**: Bottom trawl surveys (offshore) undertaken by Scotland as part of the Northeast Atlantic International Bottom Trawl surveys under ICES coordination, using standardised design and methods (see Manual SISP 15, IBTSWG 2017, <http://doi.org/10.17895/ices.pub.3519>). Broad scale, standardised sampling with GOV-trawl, with catch data available for pelagic, demersal and benthic fish, but most efficient to capture demersal species (juveniles and adults). Original raw data available in ICES Datras website since 2011, including data on haul characteristics, CPUE (by species by length class), ALK (age-length keys, for some species), SMALK (sex-maturity age-length keys, for some species). The dataset here is an extract of the data downloaded from the Datras online database, combining data on haul characteristics and CPUE (by species by length class) for the selected species of interest for the project, with selection of survey data available within the period 2010-2020.
* **SCOROC**: Bottom trawl surveys (offshore) undertaken by Scotland as part of the Northeast Atlantic International Bottom Trawl surveys under ICES coordination, using standardised design and methods (see Manual SISP 15, IBTSWG 2017, <http://doi.org/10.17895/ices.pub.3519>). Broad scale, standardised sampling with GOV-trawl, with catch data available for pelagic, demersal and benthic fish, but most efficient to capture demersal species (juveniles and adults). Original raw data available in ICES Datras website since 2011, including data on haul characteristics, CPUE (by species by length class). The dataset here is an extract of the data downloaded from the Datras online database, combining data on haul characteristics and CPUE (by species by length class) for the selected species of interest for the project, with selection of survey data available within the period 2010-2020
* **IE-IGFS**: Bottom trawl surveys (offshore) undertaken by Ireland as part of the Northeast Atlantic International Bottom Trawl surveys under ICES coordination, using standardised design and methods (see Manual SISP 15, IBTSWG 2017, <http://doi.org/10.17895/ices.pub.3519>). Broad scale, standardised sampling with GOV-trawl, with catch data available for pelagic, demersal and benthic fish, but most efficient to capture demersal species (juveniles and adults). Original raw data available in ICES Datras website since 2003, including data on haul characteristics, CPUE (by species by length class). The dataset here is an extract of the data downloaded from the Datras online database, combining data on haul characteristics and CPUE (by species by length class) for the selected species of interest for the project, with selection of survey data available within the period 2010-2020
* **EVHOE**: Bottom trawl surveys (offshore) undertaken by France in the Celtic Sea and Bay of Biscay as part of the Northeast Atlantic International Bottom Trawl surveys under ICES coordination, using standardised design and methods (see Manual SISP 15, IBTSWG 2017, <http://doi.org/10.17895/ices.pub.3519>). Broad scale, standardised sampling with GOV-trawl, with catch data available for pelagic, demersal and benthic fish, but most efficient to capture demersal species (juveniles and adults). Original raw data available in ICES Datras website since 1997, including data on haul characteristics, CPUE (by species by length class), ALK (age-length keys, for some species), SMALK (sex-maturity age-length keys, for some species). The dataset here is an extract of the data downloaded from the Datras online database, combining data on haul characteristics and CPUE (by species by length class) for the selected species of interest for the project, with selection of survey data available within the period 2010-2020.
* **BTS**: Beam trawl surveys (offshore) undertaken by different countries in the North Sea, English Channel, Celtic Sea and Bay of Biscay under ICES coordination, using standardised design and methods (see Manual SISP 14, WGBEAM 2019, <http://doi.org/10.17895/ices.pub.5353>). Broad scale, standardised sampling with Beam trawls (of different size, depending on country). Catch data available for demersal and benthic fish, but most efficient to capture bentho-demersal species (flatfishes, juveniles and adults). Original raw data available in ICES Datras website since 1985, including data on haul characteristics, CPUE (by species by length class), ALK (age-length keys, for some species), SMALK (sex-maturity age-length keys, for some species). The dataset here is an extract of the data downloaded from the Datras online database, combining data on haul characteristics and CPUE (by species by length class) for the selected species of interest for the project, with selection of survey data available within the period 2010-2020.
* **MEGS**: Mackerel and Horse Mackerel Egg Survey (MEGS) undertaken by different countries in the Northeast Atlantic under ICES coordination, using standardised design and methods (see Manual SISP 6, WGMEGS 2019, <http://doi.org/10.17895/ices.pub.5140>). Original raw data available in ICES Datras website since 2010, including data on haul characteristics, egg count (by species by egg stage).
* **SIAMISS**: Marine Scotland Science for the Scottish-Irish anglerfish and megrim industry-science survey (SIAMISS) undertaken to monitor the Northern shelf anglerfish stock in subareas 4 and 6, and division 3.a (North Sea, Rockall and West of Scotland, Skagerrak and Kattegat). Bottom trawl survey data available since 2013, including data on haul characteristics, Anglerfish CPUE by length class. For survey details see: ICES (2018). Report of the Benchmark Workshop on Anglerfish Stocks in the ICES Area (WKANGLER), 12–16 February 2018, Copenhagen, Denmark. ICES CM 2018/ACOM:31, 177 pp.
* **Sandeel dredge survey**: Annual sandeel dredge survey undertaken by Marine Scotland in December at sandeel fishing grounds off the Firth of Forth and Turbot Bank (in Sandeel Area 4, East coast of Scotland, where the largest of the sandeel stocks in Scottish waters is and there is an active fishery for it) between 2008 and 2020. The dataset also contains data from surveys on the west coast of Scotland in 2021. Survey targeting sandeel grounds in Scottish waters, using standardised sampling with a modified scallop dredge. Data including haul characteristics (timing and location) and sandeel CPUE (individuals per hour). For survey details see: ICES (2010) Report of the Benchmark Workshop on Sandeel (WKSAN), 6–10 September 2010, Copenhagen, Denmark. ICES CM 2010/ACOM:57, 201 pp.
* ***Nephrops* burrow TV survey**: annual *Nephrops* TV surveys undertaken by Marine Scotland Science in Scottish waters. These video surveys are restricted to suitable *Nephrops* habitat (identified based on sediment characteristics or fishery VMS data), and they measure *Nephrops* burrow density rather than individuals, thus providing a direct assessment of the refugia resource. Ongoing surveys, since 2007, including data on haul characteristics (location and timing), and average density of borrows.
* **WCFD**: West Coast of Scotland Demersal Fish Project (WCDF), a joint industry/science project (managed by Marine Scotland Science and the Scottish Fishermen’s Federation Services Limited) to better understand the distribution and abundance of demersal fish on the west coast of Scotland. Offshore and inshore quarterly trawl surveys were undertaken between December 2013 and November 2014 (Ramiro Sánchez et al. 2015). The dataset includes data on haul characteristics and CPUE (by species by length class), separately from inshore and offshore surveys. For some species, catch data are also available as abundance and biomass by age class (estimates are based on age-length keys that were used for the surveys).

## 2. Project files – list

### 2.1 Project spatial outputs

#### 2.1.1 EFH model spatial outputs

These are the spatial data layers (as GIS geodatabases or shapefiles) of the EFH predictions resulting from EFH models.

* **Pred\_Confid\_Mean2010\_20\_AllsppV2\_GRID\_CLIPPED**
* **Pred+Confid\_2010+15+20\_Allspp**
* **Pred\_Confid\_2015\_CC\_Allspp\_clippedScot\_30\_06\_22\_CLIPPED\_GRID**

#### 2.1.2 EFH Habitat proxy spatial outputs

These are the spatial data layers (as GIS geodatabases) of the EFH predictions resulting from habitat proxy assessment (based on EUNIS habitats, Level 3 and 4).

* **EUSMAA\_CLIPPED\_MS\_Albers\_L4\_JOINED**
* **EUSMAA\_CLIPPED\_L3\_JOINED**
* **Survey\_Data\_GB\_Combined\_L3\_JOINED**
* **Survey\_Data\_GB\_Combined\_L4\_JOINED**

#### 2.1.3 Supporting/validating survey spatial data

These are the spatial data layers (as GIS shapefiles) summarising the fish survey data behind the EFH model predictions (surveys 2010-20) and from additional surveys which contribute to the validation of the EFH maps.

* **Surveys2010-20\_AggrRelFreq\_Allspp\_POINTS**
* **Anglerfish\_juv\_SIAMISS2021**
* **Nephrops\_Burrows2007-16\_AF**
* **Sandeel\_DredgeData2021\_AF**
* **WCDF2013-14\_JuvCombinedClasses\_AF**
* **WCDF2013-14\_JuvCPUE\_ALL seasons**

### 2.2 EFH model input data

These are the datasets including the key variables on fish/shellfish distribution and associated environmental conditions, as derived respectively from source fish surveys and environmental data layers), and used as input data for the EFH modelling of fish/shellfish species/life stages.

* **Anglerfish\_N\_Juv28cm**
* **Blue-whiting\_N\_Juv19cm**
* **Cod\_S\_running**
* **Haddock\_S\_running**
* **Hake\_N\_Juv19cm**
* **Lemon-sole\_N\_Juv21cm**
* **Mackerel\_N\_Juv22-24cm**
* **Mackerel\_S\_EG1**
* **Nephrops\_R\_all**
* **Norway-pout\_S\_running**
* **Plaice\_N\_Juv12cm**
* **Sandeel\_R\_all**
* **Sole\_N\_Juv25cm**
* **Sprat\_N\_Juv9-9.5cm**
* **Squid\_N\_Juv15cm**

### 2.3 Fish survey data (raw/not processed)

These are the datasets extracted from fish survey databases to be processed and used for the EFH modelling (as input data for the model calibration and/or for subsequent validation of the resulting model maps). The datasets in these files (raw data) include the original data from the source databases, pre-filtered (for selection of the fish/shellfish species of interest, the years between 2010 and 2020, and sample locations within UK waters) and with minor editing (mainly consisting in the addition of fields/columns for indexing and variable detailing).

* **1.IBTS\_NS-IBTS data\_Datras-ICES**
* **2.IBTS\_SCOWCGFS data\_Datras-ICES**
* **3.IBTS\_SCOROC data\_Datras-ICES**
* **4.IBTS\_IE-IGFS data\_Datras-ICES**
* **5.IBTS\_EVHOE data\_Datras-ICES**
* **6.BTS data\_Datras-ICES**
* **7.MEGS 2010-19-ICES**
* **8.Anglerfish SIAMISS data 2013-21\_MS**
* **9.Sandeel dredge survey data 2008-21\_MS**
* **10.Nephrops burrow TV survey 2007-16\_MS**
* **11.WCDF2013-14\_MS**

Some of these datasets (#1 to #8) include two spreadsheets: one containing data characterising the survey hauls, one with survey catch data (CPUE etc). When saved in CSV format, these files have been named as above, with the addition of the suffix “a” or “b” to identify the two spreadsheets, respectively. The dataset provided for #11 (WCFD) was the original dataset as provided by Marine Scotland (as an Excel file containing multiple spreadsheets). Therefore, this has not been transformed into CSV format, nor a README file has been added. For these, please refer to the original survey database (with Marine Directorate).