Appendix 7a: Explanation of data origin and processing

Data on benthic biota present in the Wash and North Norfolk Coast SAC was collated from data requests to the Environment Agency (NR55938, 4 August 2017), which provided Eastern IFCA with a number of datasets (Table 1).

Table 1. Documents provided from the Environment Agency data request NR55938, 4 August 2017	
File Name	**Description of contents**
Ecomaris_Wash Subtidal Report_Nov02	Presentation of results from grab samples surveys in 1991, 1993, 1999, and 2002. Results were presented as thumbnail charts for several species. Some (MDS) analysis of results, but mostly simple statistics. Comparison of effects of 0.5 and 1 mm sieve. Examination of biotopes, and difficulty of using these in The Wash. No raw data in this document.
appendices 2 3 8 and 9	**app2:** Locations of survey stations, and the years in which those stations were sampled. *N.B. Assumption that the latitude and longitude are recorded as degrees minutes decimal minutes.* *Also note: that this worksheet mislabelled the longitude and latitude columns.* *Also note: at least some of the co-ordinates were wrong, as one of the stations was located SW of Cambridge.* *Also note: Conversion of the eastings and northings to longitude and latitude respectively does seem to give reasonable results, so this action was undertaken.* **app3:** Assumption that the station ID, in Row 2, e.g. "W43_91", is in the form "Station_Year", so that "W43_91" is Station Wash 43 sampled in 1991. **App8:** A list of taxa (termed "species" in the worksheet) with in some cases associated genera. It is not clear what the function of this worksheet is. **App 9:** Seemingly results of grab samples with the same station numbering convention as "app3". "App 9" seems to only contain results from 91 and 02. **General:** Seemingly the samples are dated only by year, not by date. This has implications when considering for instance abundance of seasonal species, or juveniles.
APEM 2013 benthic samples	**Wash 2011 Benthic:** Results of grab samples, all taken in 2011. Stations are referred to as "Site Name". Also contains additional line "Sample Name", which seemingly identifies the station for Natural England ("NE" in name) **Wash 2011 PSA:** Percentage fraction of grain sizes for many stations ("Site Name") within The Wash for 2011, plus date (including Day) when sample taken. This does not contain analysis to EUNIS or Folk. Wash survey sites - Locations of survey stations (termed "Site") *N.B. Assumption that the latitude and longitude are written as degrees minutes decimal minutes.* *Also note that this worksheet mislabels the longitude and latitude columns. This was corrected in subsequent data handling.* **2011 Wash Sample Log (edit):** Locations (latitude/longitude, and Easting/Northings) for stations ('Site name') within The Wash. **NNC 2010 Benthic:** Speculated that these are results of grab samples on North Norfolk coast, all taken in 2010. Stations are referred to as 'Site Name'. Also contains additional line "Sample Name, which seemingly identifies the station for Natural England ('NE' in name) **NNC 2010 PSA:** Percentage fraction of grain sizes for many stations ('Site Name'). No analysis to EUNIS or Folk. **NNC survey sites:** Assumption that these are locations (easting/northings) for the North Norfolk sites. **2010 NNC Sample Log (edit):** Locations (latitude/longitude and easting/northings) for stations ('Site name') off the North Norfolk Coast. Also use of the sediment sample (PSA/Biota).

	EA Wash 2007 Benthic: Seemingly results of grab samples, all taken in 2007. Stations identified by eastings and northings. Sampling date, including day, recorded. **EA Wash 2007 PSA:** Percentage fraction of grain sizes for many stations ("Site Name") in The Wash coast for 2007, plus date (including Day) when sample taken. Stations identified by Eastings & Northings. No analysis to EUNIS or Folk. **EA Wash 2011 Benthic:** Seemingly results of grab samples in Outer Wash, all taken in 2011. Stations ("Site Name") identified by an unusual labelling format. Sampling date not recorded. **EA Wash 2011 survey sites (sic):** Locations - Eastings & Northings - for numerous "Site Name", "Sample ID" and "Replicate Code" records, but which don't seem to correlate to the "Site Name" in "EA Wash 2011 Benthic". Sample date - including day - included.
WNNC SAC 2015 EA Grab infauna data matrix (*N.B.* This file was labelled "WNNC SAC 2015 EA Grab Survey infauna data matrix" on Huddle but would not download under that name)	**Sample Report Form:** Front piece to the spreadsheet, with details of the analysis. **Sample List:** Sample number and sample date *N.B. All these dates are the same, 12 May 2015, suggesting they collected and processed 46 stations in one day, some in The Wash, some off the North Norfolk Coast.* **Abundance:** Seemingly results of grab samples, collected in 2015 (see note above regarding all samples being taken on one day). Stations possibly identified by "Sample Number" and/or "Site Description". **Notable Taxa:** Seemingly a list of selected taxa. It is not clear what the function of this worksheet is. **Biotopes:** Seemingly stations (described as "Transect") with biotope – JNCC and EUNIS. No indication as to where the biotope allocation came from. Includes locations – eastings and northings.
app6	A PDF with a presentation of one station/page for samples taken in several years: • Thumbnail chart of location • Sediment Description • Habitat classification (JNCC) • List of taxa and abundance This was used for supporting the above datasets rather than to provide data.
Copy of Indication of sensitivity of EA grab stations in the Wash 2011	This spreadsheet from the Environment Agency looked at the number of 'sensitive' species present at stations in the Wash grab surveys. It contains the following worksheets: • Stations and sensitive species • 2011 EA grab survey, The Wash • Abundance at each station • Presenceabsence of sensitive sp • Cefas 2014 epifauna

The datasets (Table 1) were compiled and put into a single format to create a single spreadsheet containing all the data the Environment Agency provided (this data included surveys by the Environment Agency, Natural England and contractors of the two organisations). The final dataset covered a period from 1973 to 2015, although not all years within this period were represented.

The following steps were taken to create the compiled dataset (Table 2).

Table 2. Steps taken to compile the datasets provided (Table 1) to create the compiled dataset
1. Compiled all data from the Environment Agency data release (sourced by the Environment Agency and Natural England) into one spreadsheet. To do this the species lists were compiled to make one large list that encompassed all species recorded at all stations on all dates, in alphabetical order. The metadata list was also made uniform, to include all useful information. Where data or metadata was not available for a species or a station these boxes were left blank.
2. Where particle size data was available but EUNIS or Folk had not been calculated, this was done so using a spreadsheet modified (for the most up-to-date version of Microsoft Excel) from Ocean Ecology for the allocation of sediment data to Folk and EUNIS sediment categories.
3. Where particle size data, EUNIS, or Folk was not provided the EUNIS code from more recent years at the same locations was used to give an indication of the habitat type.
4. Produced common sense edits on the spreadsheet, this included where there was a '?', 'Sp. 1' or 'Sp. A' or any other uncertainty, the values were added to the taxa level above (for example, where it was *Abra sp. 1*, these numbers were added to the *Abra* row, rather than assuming the species).
5. Any names written in all capitals were rewritten out correctly
6. Removed the taxa 'Animalia' as this won't tell us anything specific enough to the assessment
7. The World Register of Marine Species (WoRMS) Taxon Match Service (WoRMS, 2017) was used once the datasets had been compiled to correct for taxa that had been recorded using multiple synonyms, previous names, or with typographical errors in names. Where a single species appeared under multiple names, these were combined to give a single row of data on that species.
8. Added in day grab sizes and mesh sizes as found in reports and original datasets, where it is stated a day grab is used it is here assumed that this is the standard $0.1m^2$ day grab
9. To sort by sieve mesh size used the row 'smallest sieve size' not 'sieve size' as some were sieved by two different meshes. It was assumed that these were sieved sequentially through sieves of two mesh sizes.
10. Calculated sample area for each station, and then normalised data to $0.1m^2$ (this meant multiplying core samples by different amounts depending on their diameter)
11. A final normalised worksheet was compiled for further use and analysis
12. Where presence was noted (as 'P' or 'Frag' where fragments were recorded) rather than abundance, normalisation was not possible. These data entries were retained a 'P'.
13. Data was checked by randomly selecting approximately 40 values from the final normalised worksheet and going back to the original data source to check that all data remained correct. All data checked was correct.
14. This compiled data was used to calculate abundance ratio and number of taxa (Appenidx 8a and 8b)

Processing flow for use of the compiled dataset from Table 1 and 2 to look at sensitive species

Following on from calculations of number of taxa and abundance ratio, a copy of the worksheet was used to look more specifically at species identified as sensitive. This worksheet included all of the original metadata as well as depth range (below chart datum), EUNIS code, whether or not the location was within the SAC, the total number of individuals (not counting those recorded as present (P)), the total number of taxa, abundance ratio (number of individuals/number of taxa present) (using only taxa where the number of individuals were counted).

Subsequently, the following data was added into columns for each species:

- Total number of individuals recorded for each species (obtained by adding up all the counts, not including where only presence (P) is noted)

- Total number of Ps recorded for each species
- The total number of records of a species - this was the sum of the number of times a figure was recorded for a species and the number of times it was recorded as present (P)

The results of a compilation of data available on the BIOTIC and MarLIN system, which pulled together information on characteristics and sensitivities of species, were used as an overview of sensitivities of species.

Key information from the first worksheet, including essential data and metadata was copied and transposed to produce a table laid out as follows:

Ref no[1]	Sample year	Lat	Long	Depth range (m BCD)	EUNIS code	Inside SAC?	Tot co individuals (not counting P)	Tot no of taxa recorded as P (Colonial)	Tot no of taxa where individuals counted	Tot no taxa counted + P	Abundance ratio (No individuals /No taxa where individuals counted)

N.B. Some superfluous data was removed at this point but can be found using reference numbers[1]. Data removed included sample month, sample date, site ID, site/station name, survey code, sample ID, replicate code, sample method, sample area, core diameter, total area sampled, normalisation value to 0.1m², % gravel, % sand, % mud, sieve size (mm), smallest sieve size (mm), sample depth in grab, source, data source, and folk.

- Where data on both *Station lat / long* and *Sample lat / long* were available, the *Sample lat / long* was used. *Lat* and *long* in the table therefore refer to *Sample lat / long.*
- Data in the column EUNIS code were the EUNIS code recorded during surveys wherever available, however where this was not available the EUNIS code was generated by examination of habitat occurring at the same location in the latest Natural England data release. Values originating in the survey data were more representative of conditions prevailing at the time of the survey.
- Rows for which no latitude or longitude data was available were removed, this affected 16 rows.
- Rows that provided data for outside of the SAC were also removed (79 stations). Of these 79, all but 6 had the EUNIS code "non-designated", indicating that no habitat type data for the location exists on the latest Natural England data release. Therefore, removal of these rows did not occasion a loss of significant amounts of useful data.

This data was used in a pivot table on Microsoft Excel to generate data used to produce figures to represent the data (Appendix 8; Appendix 8a-i).

A column containing data for all brittlestars was added, populated by summing the numbers in columns containing data for: *Ophiocten affinis, Ophiothrix fragilis, Ophiura,*

---

[1]Note reference numbers were used to ensure that after superfluous metadata was removed from this table, it could still be found on previous worksheets.

Ophiura albida, Ophiura ophiura, Ophiuridae, and *Ophiuroidea*. This data was used for all subsequent analysis in the same way as other taxa.

A new worksheet and pivot table was set up to deal with colonial species (those entered into the dataset as P for present), which the original pivot table could not cope with as it calculated averages and standard deviations from numeric data. This table was used to calculate the number of samples in which taxa were recorded as present each year, expressed as a percentage of the total number of samples taken in a given year.

References

WoRMS, 2017. WoRMS Taxon Match Service. Available online at: http://www.marinespecies.org/aphia.php?p=match. Accessed on 6 December 2017.

Ocean Ecology, n.d. Allocation of sediment data to Folk and EUNIS sediment categories for sublittoral sediments.