

Project Title: SIORC Shark Tagging Database Internship

Project Scope: This project was built to overhaul the UK Shark Tagging Programme (UKSTP) database. The project comprises of three stages: to standardise the data (stage 1), to deliver the data on a GIS platform (stage 2) and to present the information to the MASTS community at a half day workshop (stage 3).

Stage Report: Stage 1- Standardisation Report

Project Supervisors: Dr James Thorburn and Dr Lea-Anne Henry

Intern: Amie Williams

1) Standardisation

The standardisation of the database was broken down into a series of processes due to the number of records (9,000-10,000). The three main standardisation priorities were:

i.) Measurement parameters

Many of the measurements present were recorded in mixed metric and imperial systems, reporting in, but not limited to, lbs, kg, inch, cm, m and ft. A key process was creating new excel columns in order to insert those identified as inconsistent metrics (i.e. those not reporting in inches or pounds) into the appropriate metric columns. Furthermore, all length measurements were converted into centimetres in a new column.

As a number of data had only a reporting of just length, a predicted weight column was created in order to gain a predicted measurement using the following formula:

$$W = aL^b \tag{Equation 1}$$

where L is the fish length (cm) and a and b are fixed parameters. Such parameters were either consistent with those used to produce the Scottish Sea Angling Conservation Network’s (SSACN) length: weight charts or obtained from literature.

In total, measurement parameters consisted of 6 columns (Figure 1).

Length (inch)	Length (cm)	Weight (Raw)	Weight (lbs)	Weight (kg)	Predicted Weight (lbs)
------------------	----------------	-----------------	-----------------	----------------	---------------------------

Figure 1| Depiction of the database headings for length and weight measurements. The column Weight (Raw), keeps the reported weight in its original format before it was duplicated either into the respective lbs or kg column. This was done to preserve the original data in the case that an error has occurred when interpreting measurement units.

ii.) Qualitative locations

Locations of reports were selected as a priority for standardisation due to the lack of consistency across recordings, therefore making it difficult to isolate similar locations when a filter was applied. In order to maximise the inclusion of the original data, the locations were initially divided

in two separate headings: "Location Area" and "Specified Area". The data was then formatted to ensure the location area was general (e.g. Southwest Scotland) and the specified area was more specific to the descriptions provided (e.g. Portpatrick (East, 4 miles)). The format for the specified areas remained constant throughout in order for the data to be more readily available (i.e. location, N/S/E/W, miles) when filters were applied.

Further to this, the data was also allocated a ICES Major Fishing Area and the respective subdivision (FAO, 2017) using the latitude and longitude data, adding two more columns to the database.

iii.) Latitude and longitude

The final major standardisation exercise was to ensure latitude and longitude reporting was correct. This included:

- Checking latitude and longitude correspond with descriptive locations
- Checking for land based locations (that were not possibly shore caught)
- Ensuring north, south, east and west reports were correctly assigned
- Standardisation of latitude and longitude columns (i.e. removal of N,S,E,W letters and degree symbols, ensure decimals were in the correct places and ensure formulas which converted degree, minute, second - degree minutes were correct)
- Convert latitude and longitude data into decimal degrees in preparation for stage 2

Other standardisation measures were taken but were not as time consuming as the priorities. Applying column filter for database, ensuring consistent reporting of sexes, species, gravid, date, etc.

2) Next Steps

The next steps of this project is to progress to stage 2, where QGIS will be utilised as the GIS platform to display the information on the database before the information is presented to MASTS (stage 3). A further consideration for this database is to investigate those data points which were removed due to reporting inconsistencies'.

References

FAO (2017). FAO Major Fishing Areas. Available at: <http://www.fao.org/fishery/area/search/en>. Accessed 8th February 2017.