**Introduction**

This procedure step-by-step guide to the Generalized Additive Mixed-effects Models (GAMMs) from the main text of this paper. It contains instructions for uploading and filtering data, model formulation, and interpretation of results. The whale shark passive acoustic and visual census data from the main text of this paper are used as a worked-through example, complete with the necessary R scripts. Readers wishing to follow along with the example analyses can find all of the necessary data and programming script here.

**Data**

**Acoustic monitoring**

The first step to any analysis is to acquire and format the necessary data. Building the desired GAMMs from passive acoustic detections requires several pieces of information including the raw detection record, the raw receiver events, the deployment history of the acoustic array, and metadata for each tagged shark.

* The raw acoustic record is the CSV file exported from the Vemco User Environment (VUE) (see example below). For the Al Lith whale shark analysis this file is listed as “AllDetections\_Lith.csv” and consists of 124322 detection records. This file contains ALL detections of tags which were eventually deployed on whale sharks in Al Lith, including detections from range tests and other extraneous records that need to be removed before the data can be analyzed.
* The raw events record is another VUE export containing the complete daily logs for every receiver. For the Al Lith whale shark analysis this file is listed as “AllEvents\_Lith.csv” and consists of 406521 log entries. This file contains the complete daily logs for ALL receivers which were ever deployed in Al Lith, including logs from other studies at other sites. As with the raw detections, these data need to be filtered before being analyzed.
* The deployment history of the array is a record of when specific receivers were deployed to and retrieved from each station. For the Al Lith whale shark analysis this file is listed as “ArrayEvents\_Lith.csv” and consists of 974 deployment/retrieval records.
* The shark metadata lists each detected tag and the characteristics of the shark on which it was deployed. The characteristics include animal sex, estimated size, tagging date, and tag number. For the Al Lith whale shark analysis this file is listed as “WSTags\_Lith.csv” and includes the individual shark information for 100 tags that were detected in the array.
* The station metadata containing information about its its location and depth “Stations\_Lith.csv”

**Visual census**

Building the GAMMs based on visual census required the raw encounter data as well as a record of survey effort.

* The raw encounter data is a CSV file downloaded from the Wildbook for Whale Sharks online database (www.whaleshark.org) which lists all reported whale shark encounters for a specific area. For the Al Lith whale shark analysis this file is listed as “AllEncounters\_Lith.csv” and consists of XX encounter records from around Shib Habil. The file includes ALL encounter records, including those of untagged sharks, and need to be filtered before analysis.
* The survey effort record is a list of days and times spent searching for the target species. For the Al Lith whale shark analysis this file is listed as “Survey\_Effort\_Lith.csv” and includes information on XX shark surveys at Shib Habil.

Readers looking to use this approach on their own data should consult the formatting guide provided with the example data (“GAMM\_Formatting.xls”).

**Uploading and Filtering Data**

After the necessary data has been acquired and formatted. The next step is upload the files onto R and to refine the raw data for analysis. This includes removing extraneous detections from the acoustic record, removing extraneous encounters from the visual census data, correcting receiver records for temporal drift/ time-zone inconsistencies, and deriving effort from receiver deployment/survey records.

**Model Formulation**

**Calculating the response variable**

The filtered data can now be properly analyzed. Our approach quantifies animal residency as a binomial occupancy metric for each individual. These calculations are performed independently for each shark’s visual census and passive acoustic datasets and are based on the following pair of equations:

Where is 𝑁 the number of surveys performed since the individual was first captured and the corresponding presence-absence vector 𝐹 is defined for each set of lags 𝜏, where 𝜏 is the difference between the previous recapture event 𝑗 and the posterior recapture event 𝑘, as 1 if the individual belongs to the set of individuals recaptured on both occasions ( and ) and defined as 0 otherwise.

**Calculating explanatory variables**

**Smoothed variables**

Smoothed variables included temporal lag () and time-of-year. Lag was already quantified to calculate the individual occupancy metrics as the set of all possible time-differences between potential capture events for a given animal. Time-of-year was quantified as week of the year, so sightings/detections from January 1st through the 7th would have a time-of-year value equal to 1, January 8th-15th would have a value of 2 and so on. Both of these terms were included as smooths in the models. Time of the year was included as a cyclic cubic regression spline (as the want the value at the end of the year to match the value at the beginning of the year). Lag was included as a low rank isotropic smoother, which is the default in our modelling framework.

**Linear variables**

Size was based off of the visual estimates reported for each shark. All estimates were rounded to the nearest half meter. Sex was quantified as 1 if claspers were present (male) and 0 if they were absent (female). Survey Effort was quantified differently for the visual and acoustic GAMMs. Visual census effort was quantified as the number of survey hours per week and acoustic effort was quantified as the weekly average number of receivers in the inshore and offshore arrays.

**Random variables**

Using every possible pair of potential capture events allows us to compute a lag effect which is not influenced by time-of-year, but also introduces pseudo-replication into the model. Fortunately, the resulting pseudo-correlation can easily be accounted for by using the date of first capture (initial photographic/acoustic tagging) as a random variable. Similarly, the effects of individual variation can be accounted for by including shark identification numbers as another random variable.

**Error structure and Linking function**

As our response variable is a binary variable, of 1 if the individual was resighted and 0 if it wasn’t, we used a binomial error structure, and a logit link function.

**Fitting the Models**

Now that we have the general form of the models, the permutations can be fit to the visual and acoustic data. Every possible permutation that includes both smoothed terms (temporal lag and time-of-year) is assessed in this way.

**Model Selection and Interpretation of Results**

**Selection by AIC**

Once all model permutations are fit, the most parsimonious visual and acoustic GAMMs can be selected based on AIC. The selected visual GAMM included lag, time-of-year, size, and survey effort as explanatory variables. The selected acoustic GAMM included lag, time-of-year, size, and survey effort as explanatory variables.

**Calculating recapture odds/probabilities**

The recapture odds are simply the predicted values from the GAMM. Odds are excellent to understand the effect of a particular predictor by itself because they show how the likelihood can increase or decrease relative to a baseline. So, in Figure 2A we can see the effect that time of the year alone has on the likelihood of resighting a shark. In Figure 2B we can see how the number of years along affects the likelihood, this is without considering the seasonality. That is why we do not see any seasonal pattern in Fig. 2A because it shows how the odds change irrespective of the other variables. Although odds are good to understand the effect of different variables, it is harder to understand how they play together. For that, we use probabilities, which are shown in Fig. 3. We calculate the recapture probability by adding together the odds of different variables and the intercept and applying the inverse-logit function.

**Code:**

The code to perform wrangle data and calculate model variables can be found in `code/01\_prepare\_data.R` for the acoustic detections and in `code/11\_prepare\_data\_visual.R` for the visual encounters.

Code to calculate the lags can be found in `code/02\_calculate\_lags.R`

Code to run and fit the models can be found in `code/ 03\_acoustic\_models\_random\_effects.R`, `code/ 04\_acoustic\_models\_fixed\_effects.R`, `code/ 12\_visual\_models\_random\_effects.R`, and `code/ 13\_visual\_models\_fixed\_effects.R`.

Code for model selection and interpretation can be found in code/outputs.R