

SAMBAH Code File 3

Great Belt Playback Detection Function Analysis

Len Thomas & Louise Burt, CREEM

January 1, 2022



1 Introduction

This document (the `.Rnw` version) contains the `R` code to fit a detection function to the Great Belt (Kerteminde) playback data. The code estimates the playback effective detection area (EDA, $\hat{\xi}^*$), and also estimates variance, using a parametric bootstrap. Output files are (optionally) saved for use in the density analysis in SAMBAH Code File 6. This document is based on SAMBAH internal reports; this version has been created to accompany the paper:

Amundin et al. In press. Estimating the abundance of the critically endangered Baltic Proper harbour porpoise (*Phocoena phocoena*) population using passive acoustic monitoring. *Ecology and Evolution*.

More information about the analysis is given in the methods section of the paper.

The document is a `Sweave` file – i.e., it is a mixture of `LaTeX` and `R` that is designed to be compiled into a report in pdf (or another format such as html). We have tested it using the `Knitr` package in `R` version 4.1.1 (2021-08-10). Readers wishing to see the underlying code should view the version with the `.Rnw` suffix, and look for code chunks starting with `<<`.

2 Exploratory analysis

The playback protocol was to anchor at a set distance from the CPODs, play a recording that consisted of sets of 10 clicks at each of 11 increasing source levels, and then move farther away and repeat. Each set of 10 clicks times 11 source levels (i.e., 110 clicks) is called a playback. There were 85 sets of playbacks in total, on 7 distinct dates. This set up 10813 sets of 10 trials in total, over all source levels, distances and C-PODs (recall that the same click emitted set up a trial for all C-PODs in the podgarden). The largest distance for a successful detection was 271.84, and there were 847 sets of 10 trials at larger distances than this, all of which recorded no detections. A summary of the proportion of successes, by source level and distance is given in Figure 1.

3 Detection function modelling

We used a binary GAM to model the detection function. The final model used was a tensor product of distance and source level, with random effects for CPOD and playback. A view of the model is given in Figure 2

As a partial check of the bivariate smooth, we use the fitted model to predict the EDA (`EDA.pred`) and compare it with the EDA predicted from the GAMs in Figure 1 (which were fit separately to each source level). We use Monte Carlo integration to remove the random effects (i.e., predict on the scale of the link function,

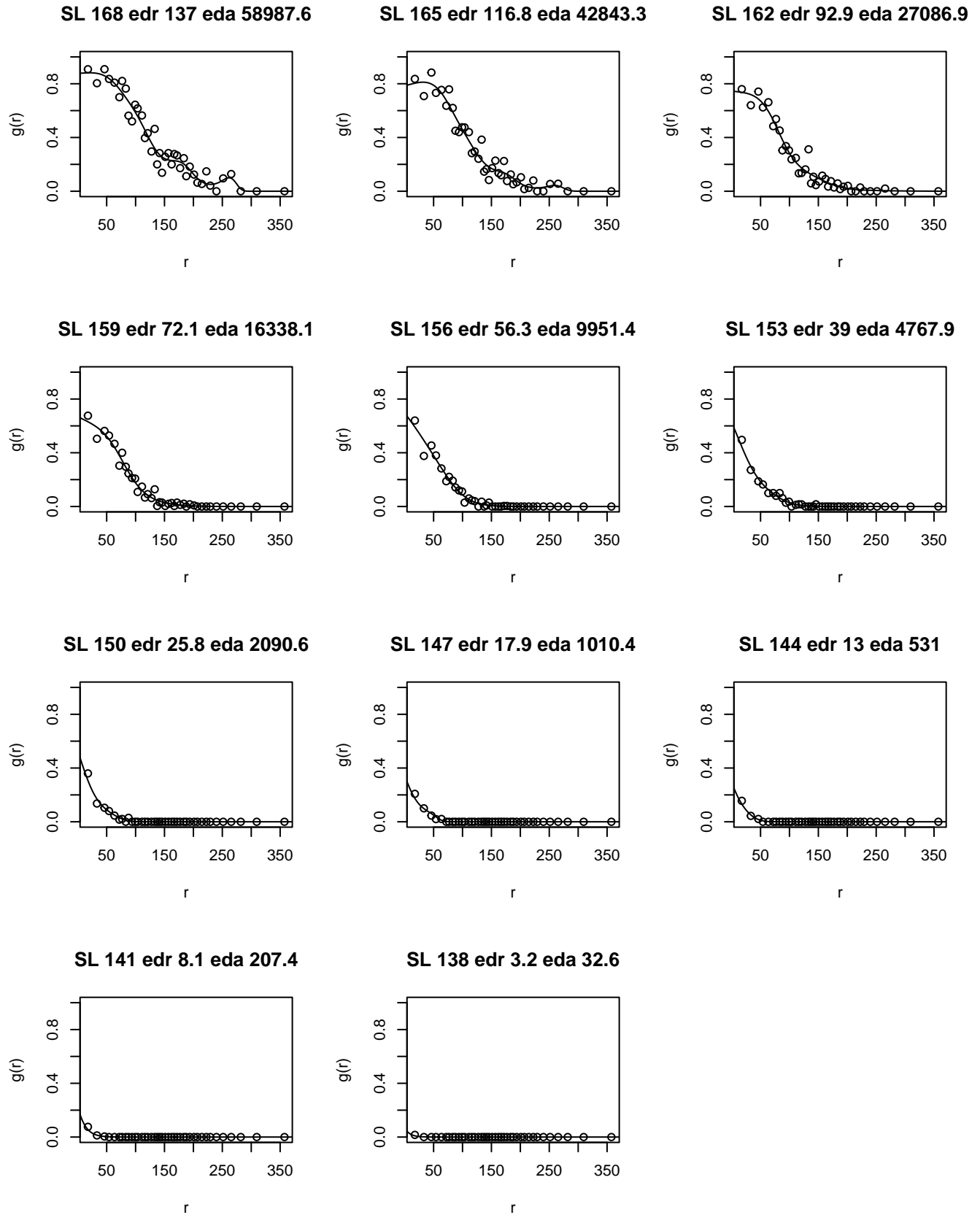


Figure 1: Proportion of successes in 40 bins (circles) plotted for each source level, with a binomial GAM fit separately to each source level (line).

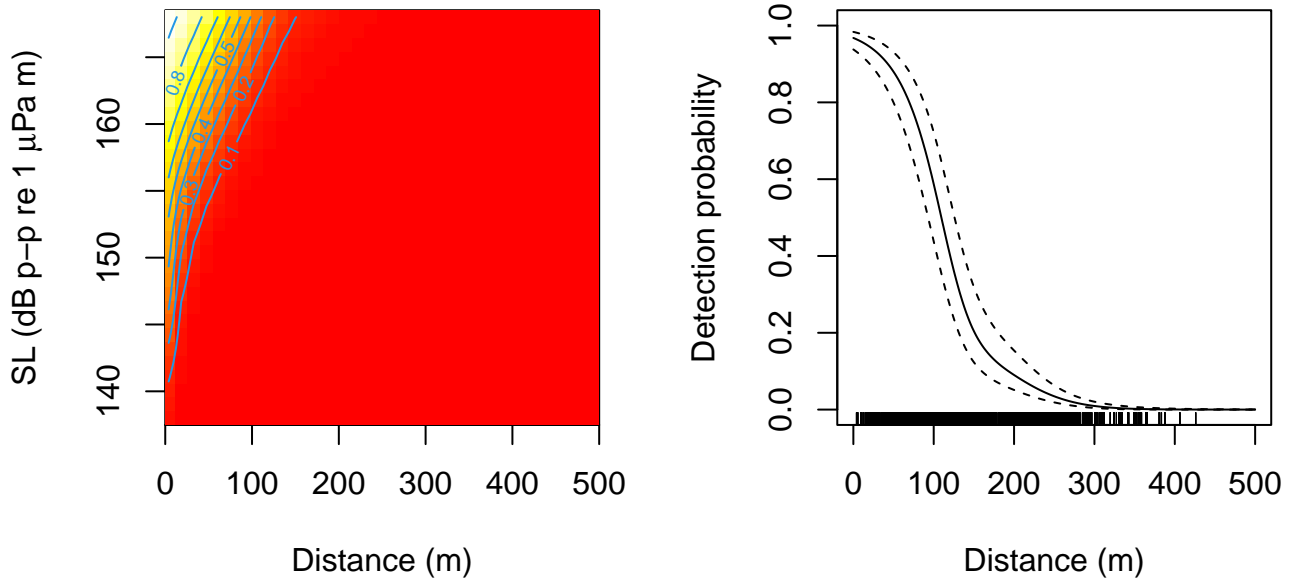


Figure 2: View of detection function model fit to Kerteminde playback data. The right hand plot shows predictions at a SL of 168 dB.

add random samples from the random effect (i.e., a normal distribution with the correct sd) to the predictions, back transform, and take the mean). Results (with 1000 samples) are shown in Figure 3.

The goal is to estimate EDA for both these and the main SAMBAH data using a single source level. It was decided to use the maximum available, 168.

Using 168 and the integration method, we obtain an estimated EDA of 62235 and EDR of 140.75. This is the EDA that will be used in producing the final density estimate, together with the bootstrap distribution derived below for variance estimation.

4 Variance estimation

We used a parametric bootstrap to produce a variance estimate, with each replicate of the parametric bootstrap integrating out the random effect. (Note - we constrained the prediction for distance to 400m as the smooth sometimes predicted wild values at 450m onwards – unsurprising given the largest data point was 425m.)

We should check that a truncation at $w = 400$ is far enough that we don't miss any significant mass during the integration, so we tracked the value of $g(w)$ - the maximum was 0.014102.

The resulting bootstrap EDAs look reasonable – see Figure 4. The bootstrap mean is 62550.26 (compared to the original mean of 62234.73), while the SE is 8606.3 and the CV (bootstrap se/original mean) is 13.829% and a 95% percentile CI is (47560.95, 81141.49).

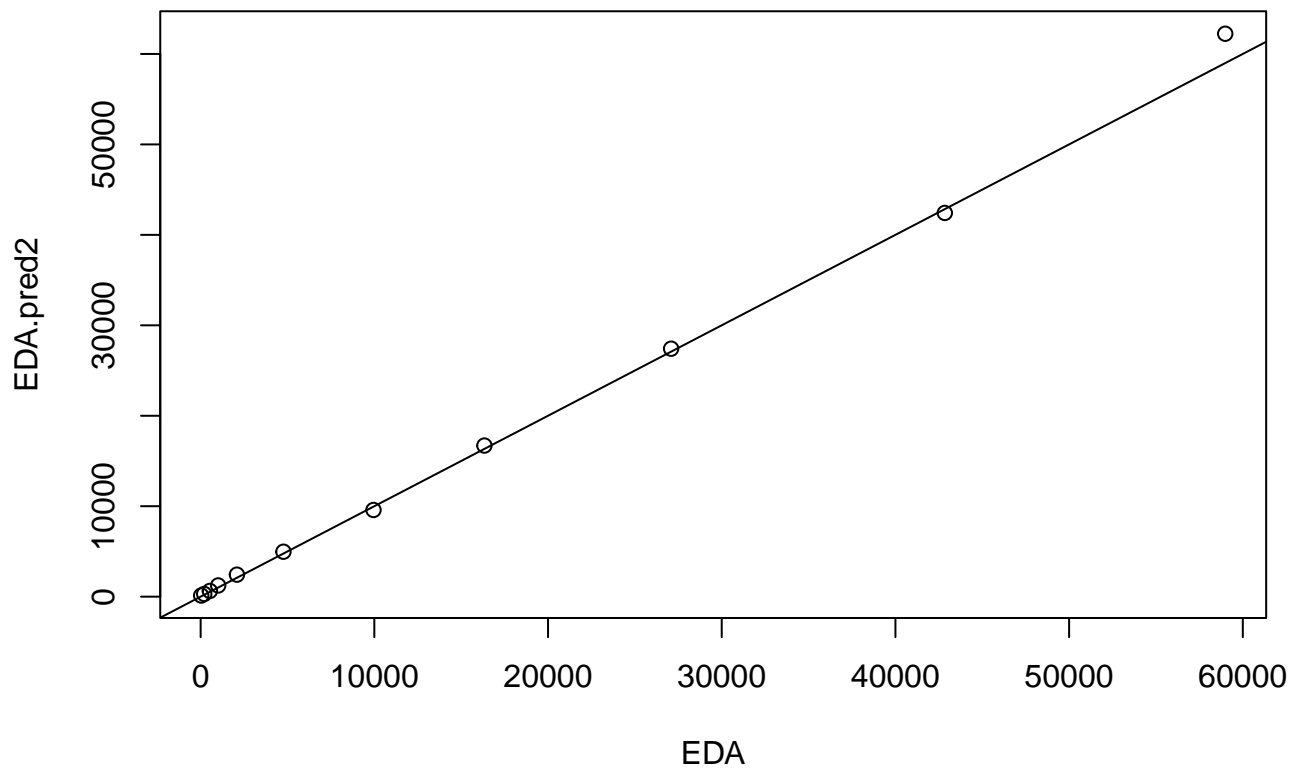


Figure 3: EDA predicted from best fitting model (y-axis) against that from the GAMS fit to each source level. Predictions integrate out the random effect on the scale of the link function. The 1:1 line is also shown.

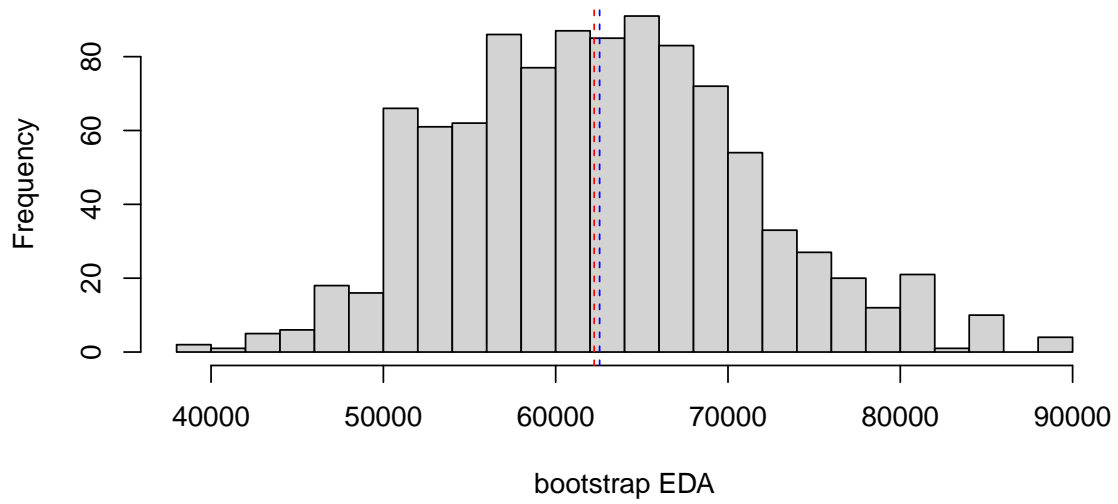


Figure 4: Histogram of bootstrap estimates of effective detection area. The bootstrap mean, and the mean from the original data are shown as blue and red vertical lines, respectively.