SAMBAH Code File 2
Great Belt Tracking Experiment 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 data collected on free-swimming porpoises during the Great Belt tracking experiment near Kerteminde. The code estimates the diel effective detection area (EDA, $\hat{\nu}_d^*$), and also estimates variance, using a nonparametric 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 <<.

Note that the code optionally involves running a bootstrap - by default this is turned off and the bootstrap results are loaded from file to save time.

2 Summary of Kerteminde data

The Kerteminde data consist of 36 porpoise encounters when porpoises were tracked acoustically from a boat; these "known location" animals were used to set up trials for a total of 16 C-PODS moored at known locations. For each C-POD and each second of each encounter, we determined whether the porpoise was detected on the C-POD or not (a trial with a successful or unsuccessful outcome, respectively), and also recorded covariates such as distance from porpoise to each POD, animal bearing, etc. In total there were 26207 trials (i.e., seconds times PODs), of which 137 were successful.

In the Figure 1, the trials have been grouped by distance into 100 bins each containing the same number of trials. On the x axis is distance; on the y-axis is proportion of trials that were successful – i.e., a raw empirical estimate of probability of detection.

3 Diel data

Table 1 gives the relative encounter rates by phase of day, taken from the analysis in SAMBAH Code File 1. As outlined in there, we can consider these as an estimate of the relative detectability of porpoises by day phase.

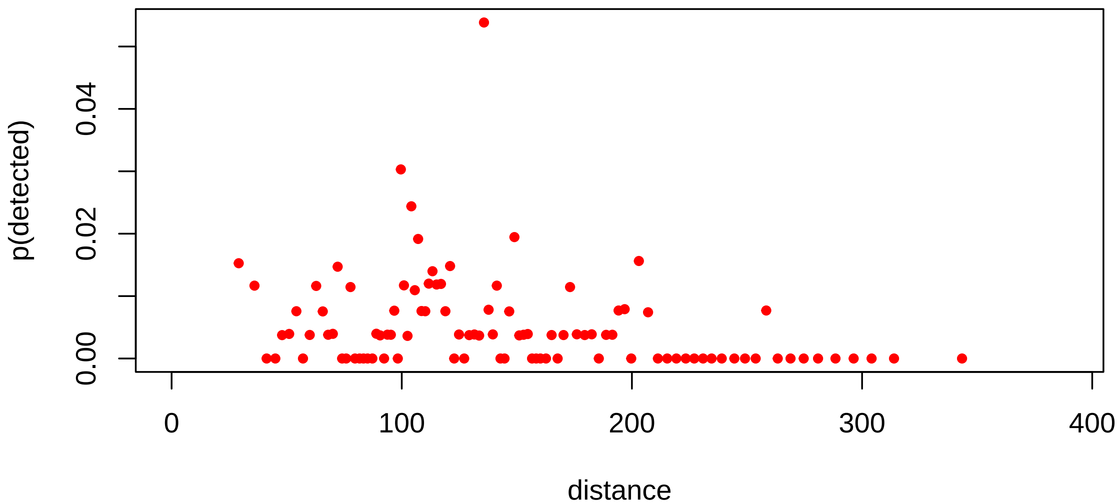

Figure 1: Trials binned into 100 intervals each containing the same number of trials.

(Note that the diel phases dawn and dusk are sometimes referred to as morning and evening – the terminology we used changed late during the drafting of the paper, and some inconsistency remains.)

	phase	rel.er	lcl	ucl	se
1	eve	1.212	0.944	1.555	0.156
2	morn	1.441	1.128	1.841	0.181
3	night	2.084	1.646	2.640	0.254

Table 1: Relative encounter rates by phase of day from model applied to SAMBAH main survey encounter rates (see SAMBAH Code File 1). Day is taken as baseline (i.e., has relative encounter rate of 1.0).

The detections in the Great Belt tracking experiment at Kerteminde do not show the same pattern (apart from Morning) relative to the Day phase. Table 2 shows the number of trials, clicks and proportion of successful trials (a simplistic estimate of p) for the Kerteminde data. However, these raw proportions do not correct for factors such as differences in distances of trials, which can make a large difference to the outcome. Hence we fitted a detection function with diel phase as a factor covariate.

	diel.phase	click	trial	p.success	relative.p
1	day	12	14547	0.00082	1.000
2	eve	16	2695	0.00594	7.197
3	morn	8	5137	0.00156	1.888
4	night	101	3828	0.02638	31.985

Table 2: Number of click positive seconds (click), trial seconds (trial), proportion of click positive seconds by diel phase and proportion relative to the lowest.

4 Detection function fitting

As documented in Amundin et al. (in press), we fitted the detection function using a binomial GAM with hand-selected knot points and a cubic regression spline basis. We used knot points at 100, 300 and 500m. (The

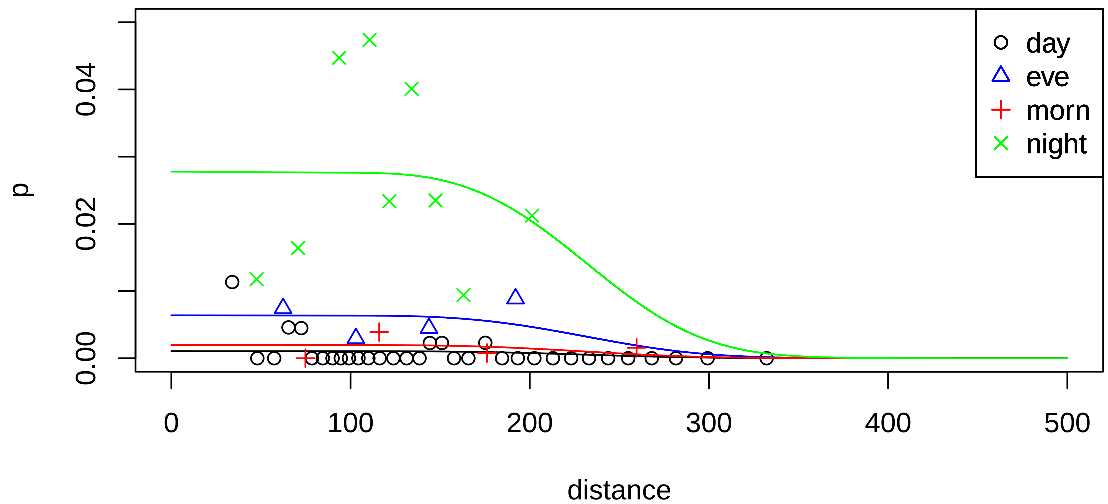

Figure 2: Trials binned into 50 distance intervals, in proportion to the amount of time in the day each diel phase, and spread evenly through the distances in each diel phase (dots). Binomial GAM (cubic regression spline with 3 hand-placed knots) fit, with additive term for diel phase is also shown.

fitted function was reasonably robust to small changes in number and location of knots, although it did vary with larger changes in knot positioning; the Effective Detection Area, however, did not vary much.) The reason for a small number of knots was to achieve a very smooth function. Given the above observed differences in proportion of positive trials per diel phase, we used diel phase as a factor covariate. The resulting fit is shown in Figure 2.

The estimated effective detection areas and radii are shown in Table 3. The overall average effective detection area for Kerteminde (averaging over diel phases) is estimated to be 1101.7 with corresponding effective detection radius 18.727. Table 3 also shows the estimated EDA and EDR for the SAMBAH region. These are the values that will be used in estimating density in the SAMBAH region.

	diel	nu.kert	rho.kert	nu.SAMBAH	rho.SAMBAH	p300.SAMBAH	p500.SAMBAH
1	day	187.6068	7.7277	887.8456	16.8110	0.00986	0.00355
2	eve	1137.6045	19.0292	1075.6865	18.5041	0.01195	0.00430
3	morn	350.8693	10.5681	1279.5269	20.1813	0.01422	0.00512
4	night	4972.5648	39.7846	1850.4984	24.2700	0.02056	0.00740

Table 3: Estimated values of EDA (nu) and EDR (rho) by diel phase for Kerteminde (.kert columns), and for the SAMBAH region (.SAMBAH colmns). Also given are the estimated probability of detecting a porpoise during a 1 second period given that it is within a circle of 300m or 500m around a CPOD in the SAMBAH region.

5 Variance estimation

We have implemented variance estimation via a non-parametric bootstrap. The sampling unit was the encounter, and we conditioned on the number of encounters per diel phase – these numbers are shown in Table 4. The total number of bootstraps was 1000. We added a large number (25,000 per diel phase) of structural zeros at 500m to ensure the bootstrap replicate detection functions declind to zero by 500m – without this, many actually

	n.ids.byphase
day	21
eve	5
morn	4
night	6

Table 4: Number of encounters by diel phase

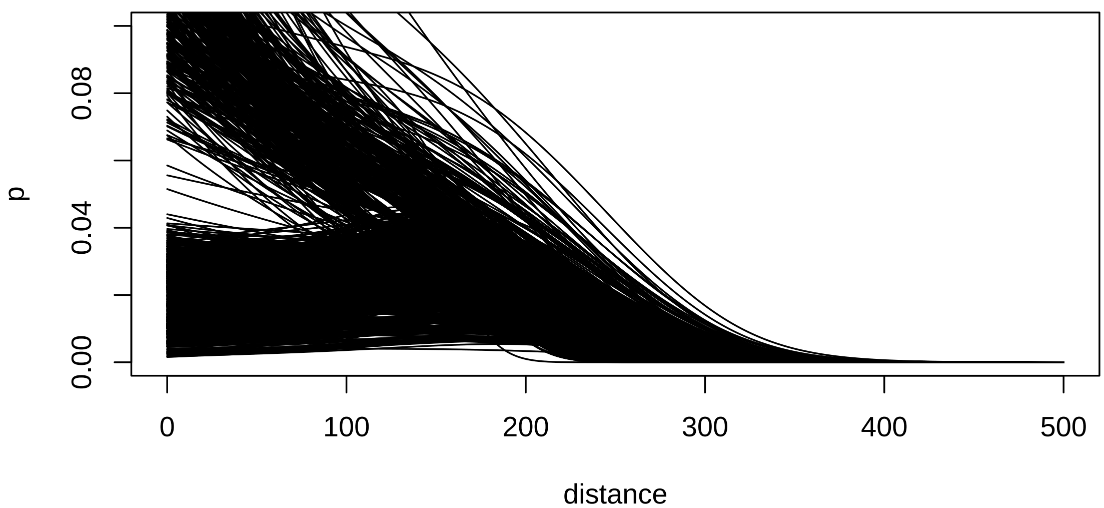

Figure 3: Some example bootstrap replicate predictions for the night diel phase.

increased outside the range of the data (i.e., 380m and above) in some bootstrap replicates.

The random seed used for generating bootstraps was 384762.

The first 1000 bootstrap replicate predictions for the night diel phase are shown in Figure 3. A histogram of the bootstrap estimates of EDA is shown in Figure 5. The bootstrap mean is 1217 (compared to the original mean of 1101.7), while the CV (bootstrap se/original mean) is 0.44808 and a 95% percentile CI is (557.94, 2403).

	diel	nu.kert	se.nu.kert	e.d	nu.SAMBAH	se.nu.SAMBAH
1	day	187.61	75.78	1.00	887.85	397.82
2	eve	1137.60	252.11	1.21	1075.69	481.99
3	morn	350.87	224.52	1.44	1279.53	573.33
4	night	4972.56	2924.13	2.08	1850.50	829.17
5	wtd.mean	1101.72	493.65		1101.72	493.65

Table 5: Estimated values of EDA for Kerteminde (nu.kert), and for the SAMBAH region (nu.SAMBAH). Also given the relative encounter rate (e.d, relative to day), and uncertainty estimates (SEs) on the EDAs

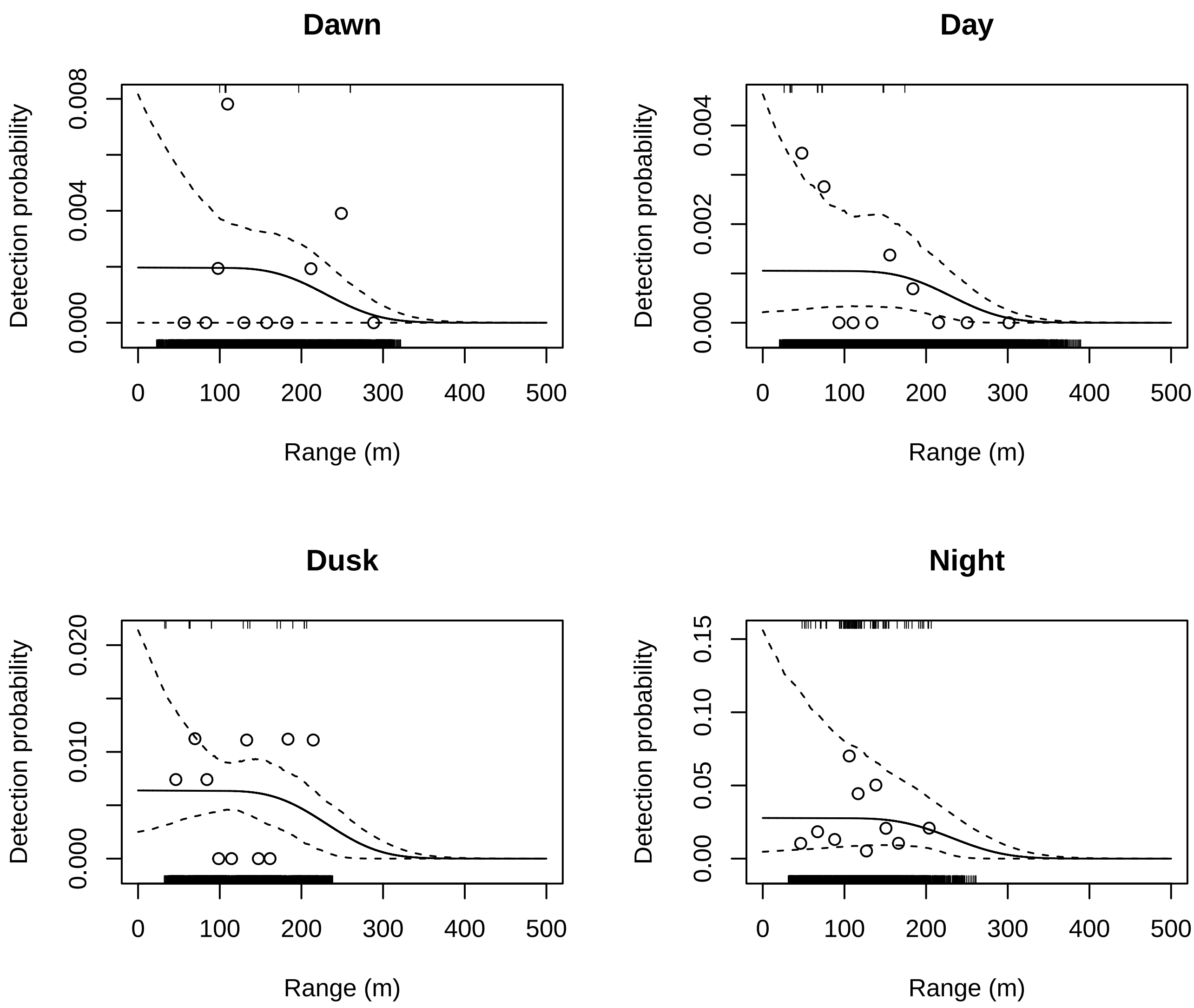

Figure 4: Predicted det prob for each phase, with 95% CI, binned proportion of successes and rug plots showing successes and failures.

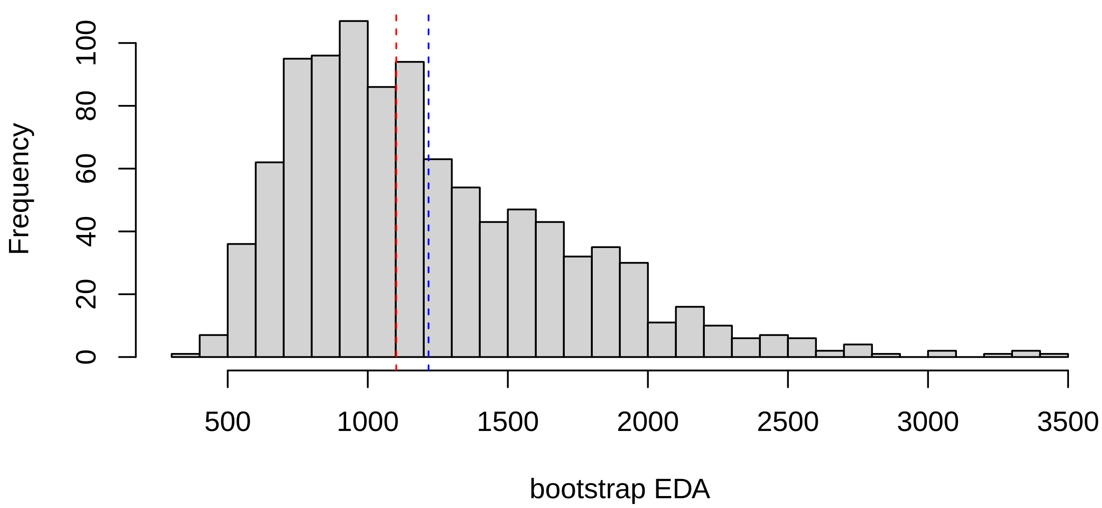

Figure 5: Histogram of bootstrap estimates of effective detection area. The bootstrap mean, and the mean from the original data are shown as blue and red vertical lines, respectively.