



CREEM

Centre for Research into Ecological
and Environmental Modelling

Methods comparison for analysing PAM data to detect a change in cue detection rate of cetaceans related to a potential disturbance

Authors	Cornelia S. Oedekoven ¹ Len Thomas ¹ Catriona M. Harris ¹ Lindesay A.S. Scott-Hayward ¹ Danielle Harris ¹ Magda Chudskinska ¹ Charles G.M. Paxton ¹ Alba Solsona Berger ² Simone Baumann-Pickering ²	¹ University of St Andrews ² University of California, San Diego
Report code	CREEM-2023-02	
Date	31 March 2023	

Please cite this report as: Cornelia S. Oedekoven, Len Thomas, Catriona M. Harris, Lindesay A.S. Scott-Hayward, Danielle Harris, Magda Chudskinska, Charles G.M. Paxton, Alba Solsona Berga and Simone Baumann-Pickering (2023) Methods comparison for analysing PAM data to detect a change in cue detection rate of cetaceans related to a potential disturbance. Report number CREEM-2023-02. Provided to US Office of Naval Research, March 2023 (Unpublished).

Document control

Please consider this document as uncontrolled copy when printed.

Revision	Date	Reason for issue	Prepared by	Checked by
1	8 March 2023	Sent to project team for review	CSO	Team
2	31 March 2023	Final version	CSO	Team
3	19 May 2023	Published online	CH	



University of
St Andrews

The University of St Andrews is a charity registered in Scotland No. SCO13532.

Contents

1	Summary	10
2	Introduction	11
2.1	Why PAM data?	13
3	Cue detection data generated from PAM recordings	14
3.1	Log of detections	14
3.2	Response variables generated from PAM data	14
3.3	Model covariates	16
3.3.1	Potential covariates of interest	16
3.3.2	Additional covariates	17
3.3.3	Setting up covariates	17
4	Statistical Methods	18
4.1	Using discretized time segments to fit generalized linear or generalized additive models	18
4.1.1	Modelling presence-absence data	18
4.1.2	Modelling count data	19
4.1.3	The predictor	19
4.2	Potential Issues	20
4.2.1	Collinearity in covariates	20
4.2.2	Overdispersion	21
4.2.3	Non-independence of consecutive observations	21
4.3	Model selection	22
4.4	R functions for fitting GLMs and GAMs	24

5	Case study: beaked whales in the presence of Navy sonar	26
5.1	Objective	26
5.2	The data	26
5.3	Covariates	28
5.4	Data exploration	29
5.4.1	Violin plots	29
5.4.2	Mean response across covariate ranges	32
5.4.3	Covariate scaling	34
5.4.4	Data-scarce tails of observed covariate values	34
5.4.5	Assessing collinearity	37
5.4.6	Assessing autocorrelation	38
5.4.7	Fitting factor-smooth interactions with missing data	40
5.5	Models fitted with bam	43
5.5.1	Model fitting	43
5.5.2	Code for fitting the <i>maxRLpp:sPres</i> interaction	44
5.5.3	Code to fit full model	45
5.5.4	Forward model selection	46
5.5.5	Backward model selection	47
5.5.6	Best model	48
5.6	Models fitted with MRSea functions	49
5.6.1	Model fitting	49
5.6.2	Code to fit the <i>maxRLpp:sPres</i> interaction	50
5.6.3	Code to fit the full gamMRSea model	51
5.6.4	Forward model selection	52
5.6.5	Backward selection	52
5.6.6	Best model	53
5.7	Models fitted with generalized estimating equations	54
5.7.1	Model fitting	54

5.7.2	Code for fitting the full model	55
5.7.3	Selecting a correlation structure	56
5.7.4	Forward model selection	57
5.7.5	Backward model selection	57
5.7.6	Best model	58
6	Discussion	59
6.1	Acknowledgments	60
	References	61

List of Figures

1	Counts of beaked whale click detections (black) and sonar pings (red) per day in 2014 from recordings on a HARP instrument located in the SOCAL Navy Range. Days the PAM device was operational are shaded gray.	27
2	Violin plots describing the relationship between beaked whale (zc) click counts per 1-min segment and non-sonar variables divided into 20 equi-width bins. Median shown as colored dots indicating the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11), interquartile ranges would show in white but all equal zero.	30
3	Violin plots describing the relationship between beaked whale (zc) click counts per 1-min segment and sonar-related variables. Median shown as colored dots indicating the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11), interquartile ranges would show in white but all equal zero. Vertical red lines indicate a data gap between 0–100 dB for <i>maxRLpp</i> and 0–120 dB for <i>cumSEL</i>	31
4	Mean beaked whale (zc) click counts per 1-min segment in relation to non-sonar variables divided into 20 equi-width bins. Colored dots indicate the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11).	32
5	Mean response of beaked whale (zc) click counts per 1-min segment in relation to sonar-related variables divided into 20 equi-width bins. Median shown as colored dots indicating the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11). Vertical red lines indicate a data gap between 0–100 dB for <i>maxRLpp</i> and 0–120 dB for <i>cumSEL</i>	33
6	Partial fit on the link scale with 95% CIs for covariate sLag in single covariate bam model fitted with thin-plate regression spline 'tp' (left) and alternative option of cubic regression spline 'cr' (right). The dimension of the basis used to represent the smooth term was set to $k = 5$ for both.	35

7	Partial fit on the response scale with 95% CIs for covariate <i>maxRLpp</i> in single covariate gamMRSea model fitted with polynomial spline 'bs' (left) and natural cubic spline 'ns' (right) where data were limited to those observations with presence of sonar (and <i>maxRLpp</i> non-zero).	36
8	Partial fit on the link scale with 95% CIs for covariate <i>sProp</i> (left and middle) and binned <i>sProp2</i> (right) in single covariate bam models. Middle plot shows range of partial fit limited to the same range as right plot.	37
9	Autocorrelation in residuals of initial bam model fitted with all six variables as smooth terms (black vertical lines). Red line indicates the decay in correlation according to an AR1 structure where parameter ρ was set to the estimated autocorrelation in Pearson residuals at lag 1. Right plot shows a zoom in of left plot.	39
10	Robust standard errors of model terms in a single covariate gamMRSea model with <i>timeofd</i> against varying group sizes used for fitting the model. Black: intercept; red, green and blue: three polynomial basis functions.	40
11	Estimated parameter ρ against group sizes used in single covariate geeglm models.	40
12	Partial fits on the response scale for smooth <i>maxRLpp</i> fitted as an interaction with <i>sPres</i> in bam model b1.6.by using the code above. Red dashed lines indicate 95% CIs.	42
13	Partial fits on the response scale for <i>maxRLpp</i> fitted as a polynomial spline as an interaction term with <i>sPres</i> . Red dashed lines indicate 95% CIs.	43
14	Partial fit on the link scale for <i>maxRLpp</i> fitted as a polynomial spline with 2 degrees of freedom in a GLM to data with <i>sPres</i> = 1. Dashed lines indicate pointwise standard errors.	44
15	Partial fits on the response scale with 95% CIs for the bam model with the lowest GCV score. Relative frequency of observations indicated in turquoise.	49

16	Partial fits on the response scale with 95% CIs for the MRSea model with the lowest CV score. Relative frequency of observations is indicated in turquoise.	54
17	Partial fits on the response scale with 95% CIs for the GEE model with the lowest QIC score. Relative frequency of observations is indicated in turquoise.	58

List of Tables

1	Response variable options using cue detections made from PAM data and applicable models	15
2	Comparison of R functions that have been used for fitting GLMs, GEEs and GAMs to PAM data ¹ . Yes indicates package has built-in methods	25
3	Number of 1-min segments in the beaked whale data set; minimum, maximum, median and mean counts of beaked whale clicks (zc) and sonar pings (s) per segment; number of zero and non-zero counts per segment (0 segments and non-0 segments), number of 1-min segments with zero beaked whale click counts when sonar ping counts was non-zero (0 non-0 segments: zc) and opposite (0 non-0 segments: s) and number of segments with zero counts for both (0-0).	28
4	Correlation between potential covariates rounded to nearest decimal point. Off-diagonal 1.0 highlighted in red (diagonal 1.0 in green).	37
5	Models fitted with the ‘bam’ function when conducting forwards model selection. Covariates included are indicated with + symbol. Bold GCV score indicates best model continued with in next round, bold model name indicates best model overall. Rounds separated by horizontal lines. Run time (RT) given in seconds.	47
6	Models fitted with the bam function when conducting backwards model selection. Covariates included are indicated with + symbol. Bold GCV score indicates best model continued with in next round, bold model name indicates best model overall. Rounds separated by horizontal lines. Run time (RT) given in seconds.	48

7	Models fitted with MRSea functions using forwards model selection with cross-validation scores (CV). Covariates included are indicated with + symbol. Run time (RT) given in minutes. Horizontal lines indicate rounds of model fitting. Bold CV scores indicates lowest CV score in a round, bold model name the overall best model.	52
8	Model fitted with MRSea functions using backwards model selection with cross-validation scores (CV). Covariates retained by the automatic model selection are indicated with + symbol. Run time (RT) given in minutes.	53
9	GEE models fitted with the geeglm function with 'working independence' (g6i) and AR1 (g6a) correlation structures and corresponding QIC values. Covariates included are indicated with + symbol. Run time (RT) given in minutes.	57
10	GEE models fitted with the geeglm function using the "working independence" correlation structure. Covariates included are indicated with + symbol. Bold QIC scores indicates best model continued with in next round; bold model name indicates best model overall. Rounds are separated by horizontal lines. Run time (RT) given in seconds.	57
11	GEE models fitted with the geeglm function. Covariates included are indicated with + symbol. Bold QIC score indicates best model continued with in next round; bold model name indicates best model overall. Rounds separated by horizontal lines. Run time (RT) given in seconds.	58

1 Summary

Human-made underwater noise has the potential to cause behavioral disturbance for cetaceans. Noise-making activities include those related to offshore energy (e.g. windfarm (de-)construction or operation), military sonar exercises or seismic explorations (e.g. Evans 1996; Scott–Hayward et al. 2014b; Parsons 2017). Response to disturbance can include animals leaving the general area, redistribution of animals within the area or changes in behavior or vocalization behavior (including changes in cue production rate, where cue refers to any sound that the animals produce with some regularity, such as calls, clicks or whistles) (e.g. Southall et al. 2016; Parsons 2017). While it is often not possible to distinguish between these responses from passive acoustic monitoring (PAM) data, in particular when using sparse arrays of acoustic sensors (e.g. Harris et al. 2018), all of these reactions may lead to a change in the rate at which we detect cues at the PAM stations.

In this study, we focus on statistical analysis methods for detecting a change in cue detection rate caused by a potential disturbance, using PAM data from single acoustic sensors or sparse arrays.

We first review the type of data generated from PAM before considering statistical methods for detecting association between detections of marine mammal vocalizations and detections of anthropogenic noise. We focus on regression modeling approaches (generalized additive models (GAMs) and generalized estimating equations (GEEs) in particular) and highlight some of the challenges for fitting these to time series data. These include collinearity and concavity in covariates, overdispersion, correlation in consecutive observations, model selection and fitting meaningful smooths to covariates with missing data. The latter involves a novel method for fitting factor-smooth interactions.

We present a case study analysis of beaked whale (*Ziphius cavirostris*) click detections made from PAM recordings in areas with concurrent military sonar exercises and apply three different modeling tools for fitting GAMs which each employ different methods for fitting smooth functions and for dealing with correlation as well as use different model selection criteria. Function `mgcv:bam` uses penalized regression splines, is limited to an AR1 correlation structure and GCV scores for model selection (Wood et al. 2016). Func-

tion `MRSea:gamMRSea` fits pure regression splines with optimal knot allocation and robust standard errors and uses 10-fold cross validation for model selection (Scott–Hayward et al. 2014a). The `geepack:geeglm` function (Højsgaard et al. 2005) allows adding pure regression splines manually and choosing among a variety of options for the correlation structure (and allows determining which provides a better fit) and uses QIC for model selection (Cui and Qian 2007).

We fitted each of the three model options to counts of beaked whale detections in almost 500,000 1-min segments and tested a variety of sonar-related and non-anthropogenic covariates. We developed a new approach for fitting factor-smooth interactions for covariates with missing data whereby for a given level of the factor covariate, the corresponding smooth is only fitted to the range of covariate values observed for this level. This method was applied to the sonar-related covariate *maxRLpp* (maximum peak-to-peak received level at the recording device) which was either 0 if no sonar was detected or took values of ≥ 100 dB if sonar was detected – due to the nature of the detection algorithm which only logged a detection of a sonar ping if the received level was ≥ 100 dB.

The covariate composition of the preferred model differed for the `bam` approach highlighting the importance of carefully choosing the modeling approach, correlation structure and model selection criteria. However, which of these modeling approaches is the best to correctly identify the presence (or absence) of an effect of a disturbance on the vocalization rate given it had an effect (given it did not have an effect) could not be determined during this study. This would require a simulation study.

2 Introduction

Human-made underwater noise has the potential to cause behavioral disturbance for cetaceans. Noise-making activities include those related to (e.g. windfarm (de-)construction or operation), military sonar exercises or seismic explorations (e.g. Evans 1996; Scott–Hayward et al. 2014b; Parsons 2017). As these activities become increasingly widespread throughout the oceans, it is important to understand and monitor the effects they have on the cetaceans inhabiting the areas where these activities take place. The type

of potential disturbance we focus on here is mid-frequency active (MFA) sonar emitted during at-sea military exercises; however, our methods are applicable to other types of disturbances as well.

Response to disturbance can include animals leaving the general area, redistribution of animals within the area or changes in behavior or vocalization behavior (including changes in cue production rate, where cue refers to any sound that the animals produce with some regularity, such as calls, clicks or whistles) (e.g. Southall et al. 2016; Parsons 2017). While it is often not possible to distinguish between these responses from passive acoustic monitoring (PAM) data, in particular when using sparse arrays of acoustic sensors (e.g. Harris et al. 2018), all of these reactions may lead to a change in the rate at which we detect cues at the PAM stations. Further factors that affect cue detection rate are distance of the animal from the source and their orientation (for directional signals, e.g. beaked whales Marques et al. (2009)), level of noise which can mask call detection and false positive rate.

In this study, we focus on statistical analysis methods for detecting a change in cue detection rate caused by a potential disturbance, using PAM data from single acoustic sensors or sparse arrays where, for the latter case, no matching of detections was undertaken between sensors recording simultaneously. Instead, we only consider detection data, i.e. detections made either via a manual or automated detection methods (e.g. Oedekoven et al. 2022) which did not undergo any further processing such as identifying buzzes among the detected clicks (e.g. Pirotta et al. 2015) or bearings or ranges estimated. No detection probability is estimated, e.g. via distance sampling methods or sound propagation (e.g. Buckland et al. 2015; Hildebrand et al. 2015) which might have been able to account for some changes in detection rate due to increased noise levels. Further, we do not consider distance of the source of disturbance or the animal producing the cues (or its orientation) to the sound source. Our particular focus is the statistical analysis methods that can be used to detect such changes while correctly accounting for other features of the PAM detection data.

The structure of this report is as follows. We first review the type of data generated from PAM before considering statistical methods for detecting association between detections of marine mammal vocalizations and detections of anthropogenic noise. We focus on regression modeling approaches. We

present a case study analysis of beaked whale click detections made from PAM recordings in areas with concurrent military sonar exercises. This report is designed to be of use to practitioners and as such we discuss modeling tools available in the free software R, as well as presenting code snippets where this helps with the exposition. The code and data used for the case study are available as supplementary materials.

2.1 Why PAM data?

Besides generally being more cost effective than visual surveys, PAM also has the benefit that cetaceans and noise can be monitored in parallel. Depending on the behavioral context, cetaceans produce various forms of sound at somewhat regular rates, e.g. to communicate (Huijser et al. 2020; King et al. 2021) or forage (Pirodda et al. 2015). We refer to these sounds as cues, the rate at which the animals produce these cues as the cue production rate and the rate at which we detect the cues at the sensors as the cue detection rate.

Both the human-made noise and cetacean cues can be monitored simultaneously with the same recording devices (Jacobson et al. 2022) or using different devices (e.g. Clausen et al. 2019). Therefore, it lends itself to use PAM to investigate whether, e.g., a change in the cue detection rate can be related to an onset or increase in the human-made noise. To this end, the recordings need to be scanned for the specific cues and human-made signals either manually or automatically (e.g. Oedekoven et al. 2022). Often, these detections are summarized as counts or presence vs absence in discretized time units of a pre-defined length (e.g. in 1-min segments) and related to other metrics or covariates in statistical tests or models.

For example, Williamson et al. (2016) extracted harbor porpoise click detections and compared daily means of varying metrics (detection positive minutes per hour (DPM_{hr}), per day (DPM_d) or intervals (DPI) of 10-90 min length) against predicted density estimates from visual surveys using a Spearman's rank test. Brookes et al. (2013) also used a Spearman's rank test to compare harbour porpoise detection positive hours per day (DPH_d), DPM_d and detection positive days (DPD) with densities predicted from habitat models of animal distribution derived from other data sources. Melcon et al. (2012) investigated hourly presences and absences of blue whale calls in relation to military exercises and ship noise in a logistic regression. Brandt

et al. (2011) analyzed DPM_{hr} of harbor porpoises in a generalized additive model (GAM) in relation to offshore energy. Thompson et al. (2014) also used hourly presence and absences, but for bottlenose dolphins, and analysed these with binomial generalized estimating equations (GEEs) to estimate the probability of dolphin presence in relation to offshore energy activities.

We focus on fitting generalized linear and additive models (GLMs and GAMs, Wood 2017) to data summarized in discretized time segments and discuss various options for fitting these models in the statistical software R [Rcore], including the potential solutions to issues arising from analyzing time series data. Furthermore, we present a new method for fitting smooth functions to variables with missing data via a factor-smooth interaction term. Other methods for analysis of these data are possible, and we return to this in the Discussion.

3 Cue detection data generated from PAM recordings

3.1 Log of detections

A log of detected cetacean cues along with their start times can be generated from PAM recordings using manual detection methods or automatic detectors (e.g., Oedekoven et al. 2022). Manual detection entails observers scanning the spectrograms of the recordings and logging each detected vocalization of the target type along with its time stamp. For automatic detection, a computer algorithm scans the recordings and detects cues (e.g. Thode et al. 2012; Lin et al. 2013). As these algorithms are generally not perfect, a false positive rate needs to be estimated, e.g. by observers manually searching for detections in representative sections of the spectrograms and comparing these with the automatic detections of the corresponding sections. For the purposes of this report, we assume that false positives have either been dealt with appropriately or have a negligible effect on the analysis.

3.2 Response variables generated from PAM data

GLMs and GAMs are types of regression models where the mean value of a response variable (in our case a variable representing acoustic detections) is related to a suite of covariates (also called explanatory variables, such

as Julian date, presence of sonar, etc). This is similar for point process models which model, e.g. temporally varying intensities of events as a function of covariates (Dorazio 2017). For hidden Markov models (HMMs), state-switching probabilities can be modeled as a function of covariates (Zucchini et al. 2017). Depending on the modeling approach, the appropriate response variable needs to be defined.

Based on the log of cue detections described above, various types of response variables can be generated. These include the start times of the individual detected cues, presence vs absence of detections or counts of detections within discretized time segments, or waiting times between cue detections (Table 1). Out of these options, starting times and waiting times (e.g. Dähne et al. 2013; Thompson et al. 2013) have the benefit that they do not require discretizing the time series into somewhat arbitrary segments. However, fitting point process models – which are applicable to start times – is often not straightforward and limited to smaller data sets than those generated in the PAM context Dorazio (2017). Similarly, HMMs (Zucchini et al. 2017) are limited to smaller data sets. When modeling waiting times, the difficulty exists in attributing covariate values on an appropriate time scale.

Hence, we focus on discretized time segments here. Common units for these include 1-min, 10-min or 1-hr segments. They can be used either directly as the unit for the response or as counts of detection positive segments per other, larger time unit, e.g. DPM per 10 min (DPM_{10min} , Nuuttila et al. 2017), per hour (DPM_{hr}) or day (DPM_d , e.g., Benjamins et al. 2017; Sarnocińska et al. 2020; Simonis et al. 2020). Thompson et al. (2013) used DPH_d and Dähne et al. (2013) used detection positive 10-min segments per hour ($DP10M_{hr}$) and per day ($DP10M_d$).

Table 1: Response variable options using cue detections made from PAM data and applicable models

Response	Time unit (with example)	Applicable modeling approaches
Presence-absence of detections	Discretized segments (1-min)	GLM/GEE/GAM, HMM
Counts of detections	Discretized segments (1-min)	GLM/GEE/GAM, HMM
Counts of detection positive sections	Discretized segments (DPM_{hr})	GLM/GEE/GAM, HMM
Start times of detections	Continuous	Point Process Models
Waiting times between detections	Continuous	GLM/GEE/GAM

Additional options of response variables not listed in Table 1 involve further

processing of detections by identifying, e.g., click trains (sequences of odontocete clicks) or encounters of groups. Oswald et al. (2015), for example, grouped detections from a single detector into individual encounters by attributing all successive detections to the same encounter as long as they were not separated by more than 30 min, in which case they were attributed to a new encounter. Further processing – such as localization of the animals when producing the cue or matching detections between detectors – is required for fitting models of absolute call abundance to detections from PAM data (Stevenson et al. 2015; Booth et al. 2017; e.g., plot sampling, distance sampling or spatial capture-recapture models, Oedekoven et al. 2022). In these cases, additional information, such as direction to the detection, signal strength or time of arrival, available from some detectors, has been shown to improve estimates of abundance (e.g., Stevenson et al. 2015). However, none of these options were considered here as we only describe and test methods applied to detection data.

Regardless of choice in response variable, it is important to remember that a change in number of detections during a disturbance can have various causes. For example, fewer detections during periods with high noise levels could be due to animals leaving the area or animals producing cues less frequently when noise levels are up, or due to animals being less detectable due to the increased noise levels.

3.3 Model covariates

3.3.1 Potential covariates of interest

In order to identify a potential impact of the disturbance, we need to define a suitable covariate that allows us to capture any change in detection rate related to the disturbance. A before/during/after disturbance covariate could be set up as a 3-level factor variable referring to the periods *before*, *during* and *after* a disturbance (e.g. pile driving, Brandt et al. 2011). If the response is expected to occur on a shorter timescale, a 2-level factor variable such as *disturbance presence* could be used to refer to periods with presence of the disturbance vs periods with absence. This is the preferred option in cases where the disturbance takes place over large periods and no distinct *before* or *after* periods can be defined. An example for such a covariate is *sonar*

presence in areas where military exercises are conducted throughout the year (see e.g. section 5).

The disturbance covariate could also directly be derived from the received noise levels at the detector. The variable sound exposure level (*SEL*) can be measured per sonar ping or cumulatively over some time frame of interest (Oedekoven and Thomas 2017; Simonis et al. 2020).

Another complication for interpretation is that the distance from the animal to the detectors may be quite large, in particular for those animals producing cues in the lower frequencies as transmission loss is less at lower frequencies. Hence, noise levels received at the detectors may not always be representative of the noise levels received by the animals.

The effect of the disturbance may also continue for some time after the disturbance has stopped as, e.g., animals that may have left the area may take some time to return or animals simply produce fewer cues. To capture a lagged return to normal cue production behavior (or a lagged return into the area), a *lag* covariate could be set up as *time since last sonar*.

3.3.2 Additional covariates

Various covariates can be used which might help explain at least part of the natural variability in the detection data. Examples of these include temporal covariates such as survey year, date (e.g. Julian date: *Julian*) or time of day (e.g. hour), oceanographic variables, e.g. *SST* (sea-surface temperature), or spatial covariates such as *depth* of the ocean or distance to coast. In order to use spatial covariates, PAM data needs to include at least two locations. For oceanographic and some spatial variables it is again important to consider that lower frequency animals may be quite some distance away from the detector. An average over the area from which the animals may be detected from could be used instead of the value measured at the location of the detector.

3.3.3 Setting up covariates

When using discretized time segments for the response variable, the covariates need to be set up so that they represent the segment adequately. When

using oceanographic variables, e.g. *SST*, averages can be taken across the time interval. When using start times, it is important to not only attribute covariate values to the specific start times but also to the periods between the start times. To this end, some form of discretization (e.g. to the nearest second or minute) should be considered.

4 Statistical Methods

4.1 Using discretized time segments to fit generalized linear or generalized additive models

GLMs and GAMs can be fitted to discretized time segments. For these types of models, the response variable is related to covariates via a link function whose purpose is to constrain the model predictions to plausible values, i.e. between 0 and 1 for binomial models (e.g. using presence-absence of cues per minute as the response) and to ≥ 0 for Poisson models (e.g. using number of cues per 1-min segment as the response). In the following we describe the models suitable for the two types of responses – binomial and Poisson models – including their link functions and describe how to build the predictor used in these models.

4.1.1 Modelling presence-absence data

For presence-absence models, the response variable y_i is a 1 or 0 for each of $i = 1, \dots, I$ segments. For each segment with at least one detection $y_i = 1$, otherwise $y_i = 0$, representing the absence of detections. The response variable is then modeled in a binomial model, specifically a Bernoulli, where each segment is a trial, presences are the successful outcomes (absences are the failures) and p_i is the probability of success for presence of call detections in the i th segment. As each segment represents an independent trial, we model the response using:

$$y_i \sim \text{Bernoulli}(p_i) \quad . \quad (1)$$

This probability is commonly modeled via a logit link function:

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad \text{or} \quad \log\left(\frac{p_i}{1 - p_i}\right) = \eta_i, \quad (2)$$

where η_i is the predictor on the scale of the link function for the i th segment (see section 4.1.3).

4.1.2 Modelling count data

When using counts of detections within each segment, the response can be modeled as a Poisson random variable:

$$y_i \sim \text{Poisson}(\mu_i) \quad ,$$

where y_i are the counts of detections in the i th segment and μ_i is the expected value; μ_i is modeled via a link function, usually the log link::

$$\mu_i = \exp(\eta_i) \quad (3)$$

where η_i is the predictor on the scale of the link function (see below).

4.1.3 The predictor

Regardless of response type, the predictor η_i is expressed as a function of covariates:

$$\eta_i = \beta_0 + \sum_q f_q(x_{qi}) \quad , \quad (4)$$

where β_0 is the intercept and $f_q(x_q)$ is a function applied to the q th covariate x_q . The function $f_q(x_q)$ can take various forms depending on whether the covariate is a factor variable (e.g. variable *before/during/after disturbance* with three levels, see above) or a continuous variable (e.g. *depth*). In the latter case we further distinguish between linear relationships between the response and the covariate (linear on the link scale) and smooth terms. For a GLM, $f_q()$ is linear whereas for a GAM, $f_q()$ can take a number of non-linear functional forms using pure or penalized regression splines (e.g. Wood 2017).

In some cases where covariate values vary in a cyclic fashion, e.g. *time of day*, *Julian day* or *month*, cyclic smooths can be the appropriate way to fit.

When using counts as the response, it is important to accommodate for variation in segment lengths. For example, two detections in ten minutes is different to two detections in 5 minutes. An offset term can be included to account for this varying effort by modelling counts per unit effort:

$$\eta_i = \beta_0 + \sum_q f_q(x_{qi}) + \log(t_i) \quad , \quad (5)$$

where t_i is the effort, e.g. duration of the segment. So, when using 10-min segments, the relationship between the linear predictor and the expected response becomes: $\mu_i/t_i = \exp(\eta_i)$ where μ_i is the number of calls we expect to detect in a 10 min segment.

Interaction terms between two covariates can be added to equations (4) and (5) using, e.g.:

$$f_{12}(x_{1i}, x_{2i}) \quad , \quad (6)$$

where x_1 and x_2 can both be factor variables, both continuous variables or one of them a factor and one a linear or smooth term (e.g., Wood 2003). In the case that both are linear/smooth terms, equation (6) is a two-dimensional smooth, suitable, e.g. for modeling spatial effects with x_1 and x_2 as spatial coordinates.

4.2 Potential Issues

4.2.1 Collinearity in covariates

This issue arises if any of the covariates can be expressed as a function of the other covariates in the model and inclusion of these may lead to unstable relationships between the response and covariates and to inflated standard errors. For example, daily sea-surface temperature in the spring may in some cases be expressed as a linear function of date (e.g. Julian date) if temperatures rise steadily throughout the spring. Collinear covariates in a model can be identified using variance inflation factors (e.g. using the `vif` function from the `car` library, Fox and Weisberg 2019).

4.2.2 Overdispersion

Both binomial and Poisson models rely on specific assumptions of a mean-variance relationship. For a binomial model, this relationship is: $V(y) = np(1 - p)$ where n is the sample size, which equals 1 for Bernoulli trials. In the case of extra variability, this assumption is relaxed to $V(y) = \phi np(1 - p)$ where ϕ is the dispersion parameter.

Similarly, in the case of the Poisson model, the general assumption of a mean-variance relationship of $V(y) = \mu$ can be relaxed to $V(y) = \phi\mu$ in the case of overdispersion. Again, ϕ is the dispersion parameter.

This parameter is estimated by dividing the sum of squared residuals by the number of degrees of freedom. In generalized linear and additive models, overdispersed binomial and Poisson data can be modeled using a quasi-likelihood where ϕ is estimated. Alternatives to quasi-Poisson models for overdispersed count data are the negative binomial or Tweedie models (Wood 2017).

4.2.3 Non-independence of consecutive observations

Non-independence of consecutive observations is a frequent issue in time series data as detections are often made in clusters. This is also the case for detections made from PAM data where sequences of segments with absences of any vocalizations (or zero-counts) are followed by sequences of presences (or counts > 0).

This becomes an issue if non-independence remains in the model residuals as we assume independence of the response variable the given covariates in the model. Positive correlation of this type can be identified with the Wald-Wolfowitz Runs test (Wald and Wolfowitz 1943) and autocorrelation function plots (e.g. R Core Team 2022).

Alternatives to GLMs in the presence of correlation are generalized estimating equations (GEEs, Hardin and Hilbe 2013). These models allow fitting various correlation structures including a first-order autoregressive structure (AR1), exchangeable, unstructured, user-defined and ‘working independence’ (Højsgaard et al. 2005). We focus on AR1 and ‘working independence’ as these are often used in the context of spatial or temporal correlation (Thomp-

son et al. 2014; e.g. Scott-Hayward et al. 2015). For AR1, it is assumed that for consecutive observations the correlation decays as a constant rate ρ^l where ρ is the correlation parameter and l is the lag between observations. For both AR1 and ‘working independence’ a grouping structure can be implemented whereby it is assumed that observations within a group are correlated and groups are independent. Using the “working independence” approach, estimates of coefficients are generally the same as using the GLM approach, however, robust standard errors (RSEs) are estimated (Hardin et al. 2003; Zeileis 2004). RSEs can also be estimated for GAMs (Scott–Hayward et al. 2014b, 2015).

4.3 Model selection

Various model selection criteria are available which are often based on the likelihood (or quasi-likelihood) and some form of a penalty term based on the number of parameters of the models. The choice of the appropriate criterion depends on the type of model fitted. They each allow comparisons between models with different plausible sets of covariates as long as the same data are used. Generally the model with the lowest score is the preferred model.

For likelihood based models (e.g. binomial or Poisson GLMs or GAMs with pure splines), the Akaike or Bayesian information criterion (AIC or BIC) can be used (Akaike 1973, 1979):

$$AIC = -2\log Lik + kn_{par} \quad , \quad (7)$$

where $\log Lik$ is the log-likelihood of the observations given the parameters, evaluated at the parameter values that maximise the log-likelihood, $k = 2$ and n_{par} is the number of parameters fitted in the model. For BIC, $k = \log(n)$ with n being the number of observations.

For the quasi-likelihood version of these models, quasi-AIC (QAIC) or quasi-BIC (QBIC) can be used (Schwarz 1978; e.g. Barton 2022):

$$QAIC = -2\log Lik/\phi + kn_{par} \quad , \quad (8)$$

where the log-likelihood $\log Lik$ is obtained by fitting a non-quasi model to the same data; ϕ is the dispersion parameter from the most complex model that

is part of the model comparison and $k = 2$ (QAIC) or $k = \log(n)$ (QBIC).

The quasi-likelihood under the independence model criterion (QIC) is also adapted from AIC and based on a quasi-likelihood and with a modified penalty term (Pan 2001; Cui and Qian 2007). It can be used for comparing GEEs models as it accounts for the overdispersion and the correlation structure assumed for the model.

$$QIC = -2Q(\hat{\mu}; I) + 2\text{trace}(\hat{\Omega}_I^{-1}\hat{V}_R) \quad , \quad (9)$$

with the quasi-likelihood $Q(\mu) = \int_y^\mu \frac{y-t}{\phi V(t)}$, expected value $\mu = E(y)$ and variance $Var(y) = \phi V(\mu)$. $\hat{\Omega}_I$ is the negative Hessian of $Q(\hat{\mu}; I)$. I is the independent correlation structure under which the quasi-likelihood is obtained. \hat{V}_R represents an estimate of the variance based on a general working correlation structure R . QIC can also be used to determine the optimal correlation structure for a GEE. Cui and Qian (2007) recommended to first determine the optimal correlation structure by fitting the most complex model with all covariates with different correlation structures. The correlation structure of the model with minimum QIC should be used for all further models fitted to determine the best covariate combination.

For GAMs fitted with the `mgcv` package (Wood et al. 2016), unless a fixed number of knots is specified, an un-biased risk estimator (UBRE) score – or generalized cross validation (GCV) for quasi-likelihood models – is returned which can be used for model selection:

$$\begin{aligned} GCV &= \frac{nDev}{(n - edf)^2} \\ UBRE &= \frac{Dev}{n} + \frac{2\phi edf}{n} - \phi \quad , \end{aligned} \quad (10)$$

where Dev is the deviance under the model and edf the effective degrees of freedom of the model. For more details on penalized regression splines and edf see, e.g., Wood (2011), Wood (2017). Notice that UBRE is effectively just AIC rescaled, but is only used when ϕ is known.

Alternatively, K-fold cross-validation scores provide a relative measure of the predictive power of the models and can be used for model selection. The

data are split into K folds, the model refitted to the data K times excluding one fold (the unseen data) at a time. The unseen data y_u is then compared to their corresponding predicted values \hat{y}_u using a cost function. Package `MRSea`, for example, provides this option using the default cost function of $\Sigma^K (\Sigma(y_u - \hat{y}_u)^2/n_u)$ (Scott–Hayward et al. 2014a).

Marginal testing using p -values from, e.g., likelihood-ratio chi-square, Wald chi-square or F-tests can be used to determine if individual covariates should be retained in the model, as long as the p -values are trustworthy, i.e. any residual correlation present has been accounted for. If residual correlation is present, RSEs must be used to calculate the test-statistics to find the p -values (Hardin et al. 2003; Zeileis 2004).

The number of covariates that can be included determines the total number of candidate models for comparison. For example, if there are two covariates, the total number of models is four including the null model, two single covariate models and one two-covariate model. For three covariates, this number increases to eight, for four to 16, etc. In cases where the number of possible models is too large to consider, step-wise model selection may be used where one covariate is added to or removed from the model at a time and the selection criteria implemented at each step. Alternatively, methods for automatic model selection exist, such as covariate removals by shrinking the coefficients to zero (e.g. Marra and Wood 2011) but are not discussed in detail here. However, collinear covariates should not be included together (section 5.4.5) and models should be plausible (Anderson and Burnham 2004).

4.4 R functions for fitting GLMs and GAMs

GLMs can be fitted using the `glm` function from the `stats` package (R Core Team 2022, Table 2). GEEs can be fitted with the `geeglm` function from the `geepack` package (Højsgaard et al. 2005). Both `glm` and `geeglm` support splines fitted, e.g., via the `bs` function (`splines` package, R Core Team 2022). Smooths can also be fitted with the `gam` and `bam` functions of the `mgcv` package (Wood, Pya and Säfken 2016) or the `gamMRSea` function of the `MRSea` package (Scott-Hayward et al. 2014a). Functions `gam` and `bam` both allow fitting penalized regression splines (Wood 2017), the latter being particularly suited for large data sets due to the potential of much shorter run times. Function `bam` makes some

approximations in order to decrease run times and its memory footprint that are particularly relevant for large datasets such as those commonly generated by PAM data. This function also allows fitting an AR1 correlation structure. By contrast, `gamMRSea` allows fitting RSEs (Zeileis 2004).

Table 2: Comparison of R functions that have been used for fitting GLMs, GEEs and GAMs to PAM data¹. Yes indicates package has built-in methods

Function	glm	geeglm	gam	bam	gamMRSea
R package name	stats	geepack	mgcv	mgcv	MRSea
Factor variables	Yes	Yes	Yes	Yes	Yes
Linear relationships	Yes	Yes	Yes	Yes	Yes
Smooth functions					
• Pure regression splines	Manually ²	Manually ²	Yes	Yes	Yes
• Penalized regression splines	No	No	Yes	Yes	No
Overdispersion	Yes	Yes	Yes	Yes	Yes
Non-independence					
• Robust SE	No	Yes	No	No	Yes
• AR1	No	Yes	No	Yes ³	No

¹Other options are available but not listed here.

²Via e.g. `splines` functions.

³But only global AR1, not panel-based.

Pure and penalized regression splines use basis functions (e.g. polynomial or cubic splines, Wood and Augustin 2002); controlling the wiggleness of the smooth involves choosing the number of basis functions and knots. Knots can be thought of as locations within the covariate range where pieces of the smooth are joined together. The more knots and the more basis functions, the wigglier the smooth. Penalized regression splines generally employ a large number of knots and control the wiggleness of the smooth via a wiggleness penalty. This is the default method for `mgcv` functions `gam` and `bam`. The wiggleness of pure regression splines can be controlled using a spatially adaptive smoothing algorithm (SALSA, Walker et al. 2010), where the number and location of knots is optimized given a set of basis functions and range of knot numbers. This is the default method for `gamMRSea`. Cyclic smooths are recommended when the highest covariate value joins with the lowest, e.g. for covariate *time of day*, where midnight (24:00:00) is followed by 00:00:01.

5 Case study: beaked whales in the presence of Navy sonar

5.1 Objective

In this section, we illustrate the application of the different methods described above to a case study of beaked whales in the presence of MFA sonar in the SOCAL Navy testing range off Southern California. Here, the general research question of interest was: does the rate of detecting beaked whale clicks (cues) change in the presence of Navy sonar? To investigate this question, we related the response variable (number of clicks detected in 1-min segments) to various explanatory covariates, some of these relating to sonar, using the GAM approach. More specifically, we fitted models with the functions `bam`, `gamMRSea` and `geeglm` to illustrate the different options available for each of these model types. We focus on the GAM approach as the data volume (Table 3) and number of covariates considered (section 5.3) rendered both the HMM and point process modelling approaches unfeasible.

5.2 The data

We analyzed detections of Cuvier’s beaked whale (*Ziphius cavirostris*) clicks from PAM recordings made with High-frequency Acoustic Recording Packages (HARPs) deployed in Southern California within the SOCAL Navy testing range (Baumann-Pickering et al. 2014; Solsona Berga et al. 2019). Recordings were available from multiple sites and multiple years (2009-2015). However, we limited our case study to a single site (H) and year (2014) (Figure 1) to ensure run times for all models were manageable, i.e. on the scale of minutes or hours vs days or weeks. We refer to these data as the beaked whale data set in the following. This data set included 244,688 Cuvier’s beaked whale click detections and 108,201 sonar ping detections along with their start times from 348 different days. The discretized version of these data consisted of 495,661 1-min segments for which counts of beaked whale detections, counts of sonar ping detections and measurements of sonar-related covariates were available (Table 3). Covariates are discussed in detail below.

Table 3 highlights various aspects of the data set that could potentially lead to issues during the analysis. 98.9% and 96.1% of all segments contained no beaked whale or sonar ping detections, respectively. This raises the question

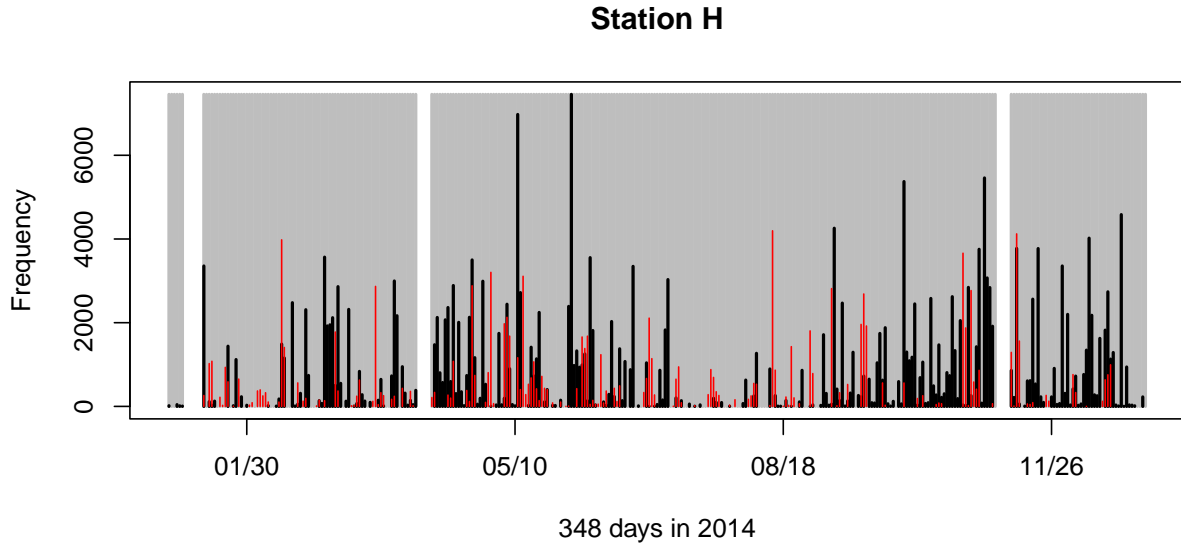


Figure 1: Counts of beaked whale click detections (black) and sonar pings (red) per day in 2014 from recordings on a HARP instrument located in the SOCAL Navy Range. Days the PAM device was operational are shaded gray.

whether the data should be discretized at a coarser scale, e.g. as 10-min or 1-hr segments. However, presences vs absences of clicks and sonar varied on a fine scale: 9.0% of 1-min segments with presence of sonar were followed by at least one segment with absence of sonar. 23.6% of 1-min segments with presence of beaked whale clicks were followed by at least one segment with absence of clicks. 42.3% of breaks between sonar presences were only 1 min (71.4% ≤ 10 min, 82.7% ≤ 1 hr). 19.9% of all sonar events (uninterrupted sequences of 1-min segments with sonar presence) only lasted 1 min (79.2% lasted ≤ 10 min, 96.7% lasted ≤ 1 hr). Using a coarser scale than 1-min segments, we would lose the ability to capture potential fine scale responses by the whales to presences of sonar.

We further note that the data set represents a period during which the sonar was more or less continuously present – at least when considered on a daily basis (Figure 1). Hence, the approach used in, e.g. Jacobson et al. (2022) or Tougaard et al. (2006), where periods before an exposure that could be considered a baseline, were compared with periods of during and after, could not be applied. Nonetheless, we could investigate if the presence of sonar, the duration of sonar exposure, breaks between sonar exposure or attributes of sonar itself had effects on the detection rate of beaked whale clicks.

Table 3: Number of 1-min segments in the beaked whale data set; minimum, maximum, median and mean counts of beaked whale clicks (zc) and sonar pings (s) per segment; number of zero and non-zero counts per segment (0 segments and non-0 segments), number of 1-min segments with zero beaked whale click counts when sonar ping counts was non-zero (0 non-0 segments: zc) and opposite (0 non-0 segments: s) and number of segments with zero counts for both (0-0).

	Clicks	Sonar pings
Number of segments	495,661	
Minimum count per segment	0	0
Maximum count per segment	464	31
Median count per segment	0	0
Mean count per segment	0.49	0.22
Number of zero count segments	490,240	476,285
Number of non-zero count segments	5421	19,376
Number of zero click segments when pings are non-zero	19,147	–
Number of zero ping segments when clicks are non-zero	–	5192
Number of segments with zero clicks and pings	471,093	

5.3 Covariates

The covariates available could be divided into those related to sonar and non-anthropogenic covariates. For the latter, we considered time of day, *Julian* and *month* (Figure 2). For the covariate pertaining to time of day, *timeofd*, the sunrise-sunset data was incorporated into scaled time of day. First, the local time was scaled between $[-1, 1]$, where 0 was the sunset and $-1/1$ were the sunrise. Thus, all the negative values between -1 and 0 were daylight times, and all positive values between 0 and 1 were night times. In this way, the covariate was not affected by annual variation in day length, or clock changes and had the potential to pick up any effects related to, e.g., day vs night or increased clicking during sunrise or sunset.

The detection of sonar pings involved the following process. The sonar tonal contours were detected in the frequency range between 2 and 4.5 kHz with a signal-to-noise ratio above 5 dB and contour durations of at least 200 ms. Detected tonal pings that were close in time and frequency were combined if they were less than 250 ms and 1000 Hz apart. Specific metrics were computed for each ping detection including minimum, maximum, start, and end frequencies, as well as peak-to-peak received level (RLpp, dB re 1 uPa) and SEL (dB re 1 uPa²*s).

Covariates pertaining to sonar included the number of sonar pings detected per 1-min segment (*sCount*) and binary variable sonar presence (*sPres*, 1 = presence vs 0 = absence of sonar pings per 1-min segment). The proportion

of the 1-min segment during which sonar was detected ($sProp$) was used to capture the variability in amount of sonar per 1-min segment. The duration of sonar activity ($sDur$) was measured as the number of consecutive 1-min segments with sonar presence. In order to account for a potential lag in the response to absence of sonar, we defined covariate $sLag$ as the number of 1-min segments of uninterrupted sonar absence. The two covariates describing features of the sonar itself included the maximum peak-to-peak received level ($maxRLpp$) and cumulative sound exposure level ($cumSEL$). Covariate $maxRLpp$ was calculated as the maximum received level among all pings within a given 1-min segment and $cumSEL$ accordingly as the cumulative sound exposure level of all pings within a given 1-min segment.

5.4 Data exploration

5.4.1 Violin plots

Figures 2 and 3 provide a first look at potential patterns between the response and the covariates. From these we were able to draw the following conclusions: 1) the medians and interquartile ranges of the response were zero across the entire range of each covariate; 2) across the range of each covariate, there is variability in the outliers and maximum values of the response; 3) there are long tails with few observations indicated by yellow, amber and red dots for several covariates (see e.g., $sCount$ or $sProp$) which may lead to problems during model fitting, in particular for smooth functions; 4) there were data gaps for two covariates, $maxRLpp$ and $cumSEL$. These existed because sonar pings were only classified as such when the received level exceeded a fixed threshold of 100 dB. Hence, all sonar pings that may have been received at the recorder at <100 dB were not detected and – in case no other sonar ping was detected in the same 1-min segment – covariates $sPres$, $maxRLpp$ and $cumSEL$ falsely set to zero. For the latter two covariates this further resulted in no data existing in the covariate range between 0 and 100 dB. This has consequences for how we were able to fit this covariate (see below for more details).

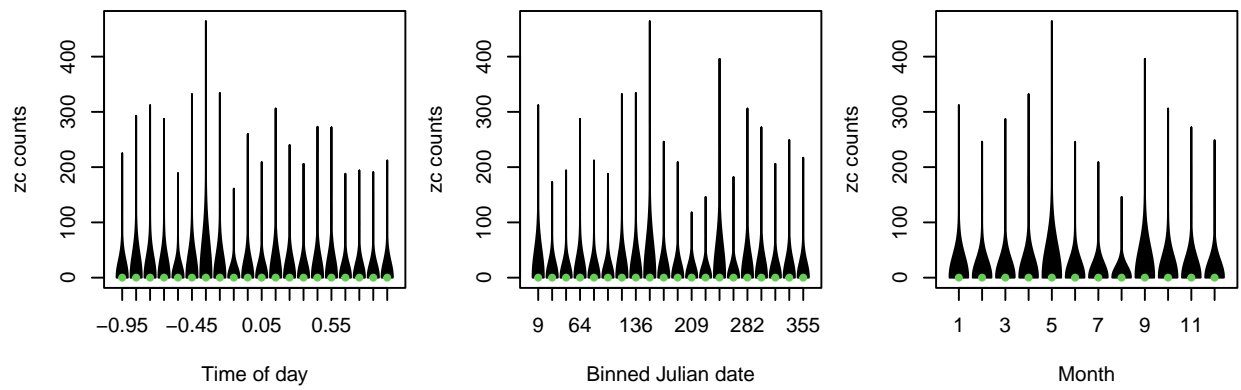


Figure 2: Violin plots describing the relationship between beaked whale (*zc*) click counts per 1-min segment and non-sonar variables divided into 20 equi-width bins. Median shown as colored dots indicating the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11), interquartile ranges would show in white but all equal zero.

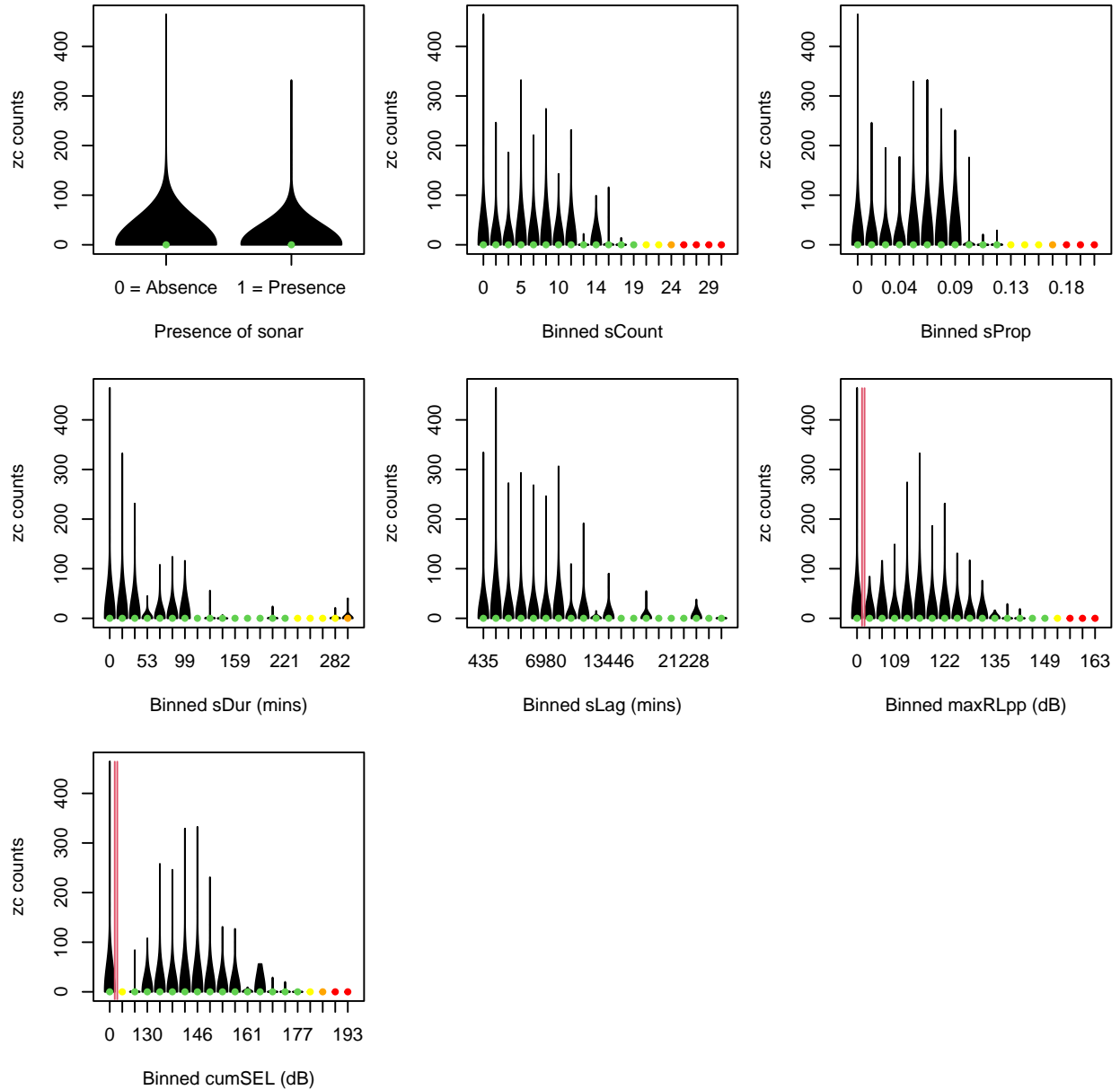


Figure 3: Violin plots describing the relationship between beaked whale (*zc*) click counts per 1-min segment and sonar-related variables. Median shown as colored dots indicating the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11), interquartile ranges would show in white but all equal zero. Vertical red lines indicate a data gap between 0–100 dB for *maxRLpp* and 0–120 dB for *cumSEL*.

5.4.2 Mean response across covariate ranges

Using the same bins as in Figures 2 and 3, plotting the mean response against covariates revealed much better what patterns we may find during model fitting (Figures 4 and 5). For any of the covariates except $sPres$, which is binary, smooth functions will likely be more suitable compared to linear terms.

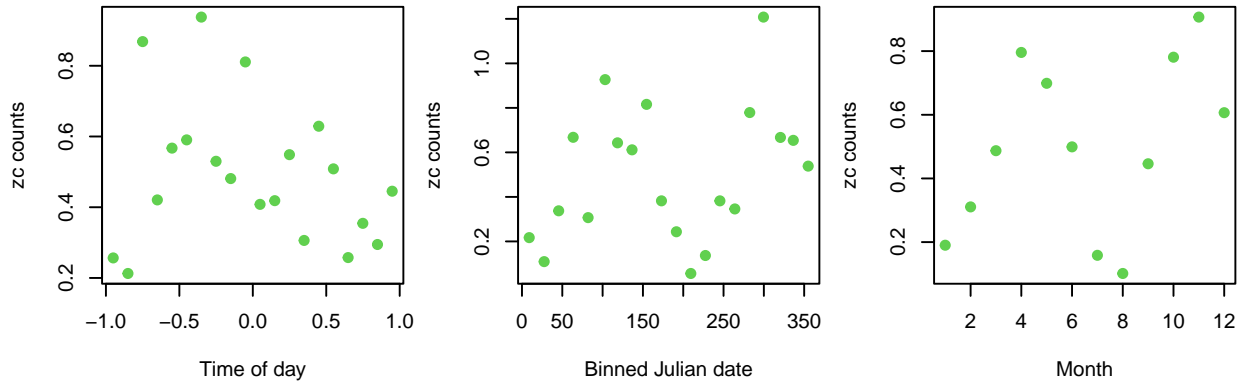


Figure 4: Mean beaked whale (zc) click counts per 1-min segment in relation to non-sonar variables divided into 20 equi-width bins. Colored dots indicate the sample size within the bin (green: >100 , yellow: <101 , amber: <21 , red: <11).

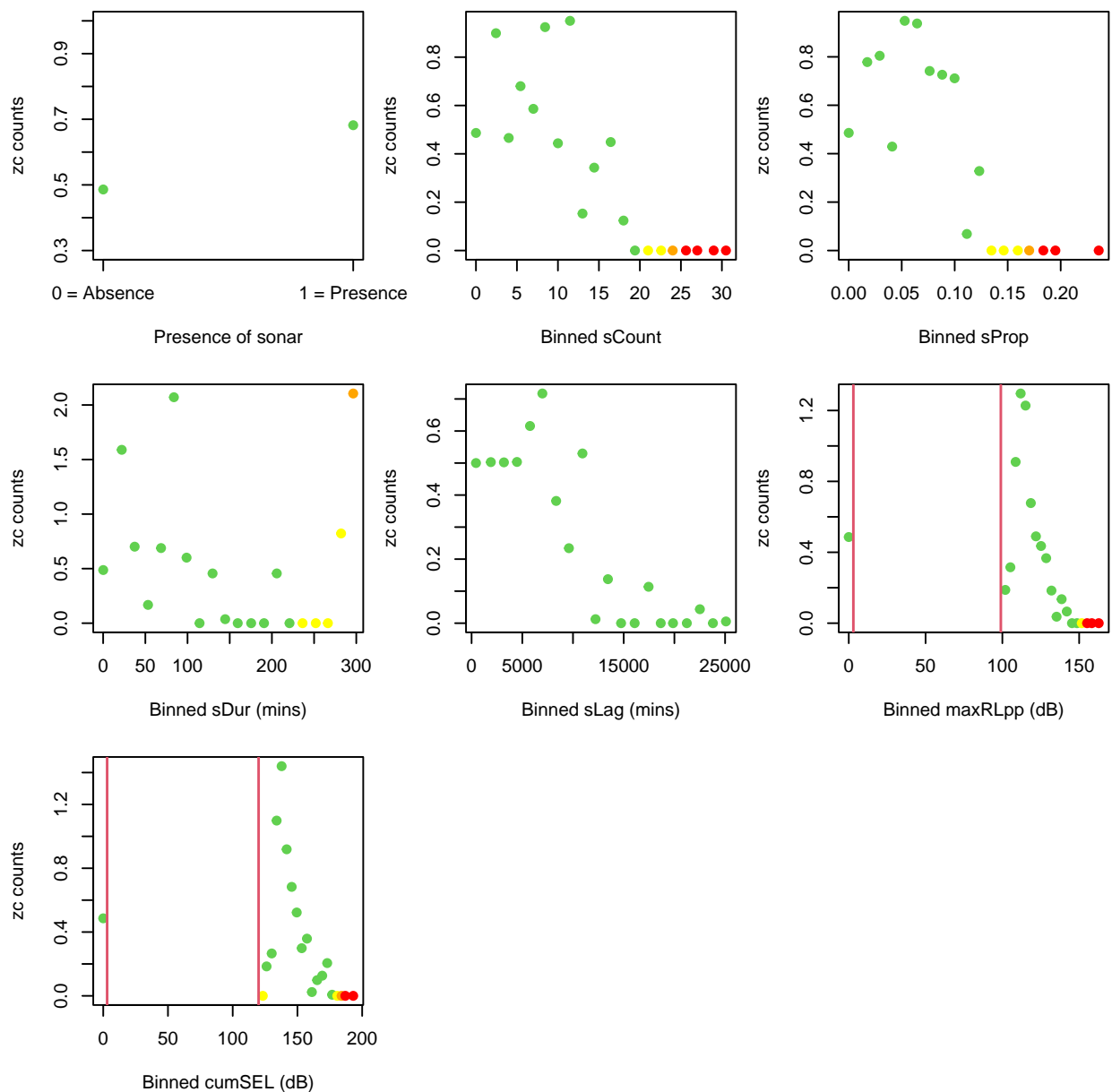


Figure 5: Mean response of beaked whale (*zc*) click counts per 1-min segment in relation to sonar-related variables divided into 20 equi-width bins. Median shown as colored dots indicating the sample size within the bin (green: >100, yellow: <101, amber: <21, red: <11). Vertical red lines indicate a data gap between 0–100 dB for *maxRLpp* and 0–120 dB for *cumSEL*

5.4.3 Covariate scaling

Figures 2 – 5 highlighted that covariates were not on similar scales. For example, $sProp$ ranged from 0–0.236 while $sLag$ ranged from 0–25,371. Using similar scales for covariates has been shown to improve the precision of model predictions (Buntaran et al. 2021). In order to bring all covariates on similar scales, we applied the following transformations:

- $timeofd$ was multiplied by 100, increasing the range to -100 – 100 with -100/100 as sunrise and 0 as sunset
- $sProp$ was multiplied by 100, i.e. turning it from a proportion into a percentage, with a maximum of 23.6%
- $sLag$ was divided by $24*60$, i.e. turning its unit from number of 1-min segments into number of days, with a maximum of 17.9 days

5.4.4 Data-scarce tails of observed covariate values

Figures 3 and 5 also highlighted that some covariates had very few observations in the higher regions of the covariate values (see yellow, amber and red dots). With little data to support the smooth estimation in these regions, the uncertainty estimation maybe unduly large for some types of spline. For example, the mean response for $sDur$ was > 0.5 in the highest two bins due to only three observations with a non-zero response (Figure 5). Smooth functions may try to fit too closely to such patterns in the tails and confidence intervals have the tendency to be very wide in these regions.

A few methods exist to alleviate this issue which can each be considered on its own or in combination with another method. One method is to control the wiggleness of the smooth function. This can be done, e.g., by reducing the maximum number of knots or degrees of freedom allowed for the respective smooth or, in the case of `bam` models, by increasing the penalty for wiggleness (Wood 2017). Furthermore, various types of smooths are available. For example, thin-plate regression splines are the default type of smooth for `bam` models (Wood 2003); however, cubic regression splines (and cyclic-cubic) and p-splines are some of the additional options available. Fitting single covariate models using different types in preliminary analyses may reveal that using

non-default options are better in some cases. For our case study, the only case for which we found a marked difference for `bam` models was covariate $sLag$ (Figure 6). However, in this case, the thin-plate spline was considered the better option due to the much wider confidence interval for the cubic regression spline.

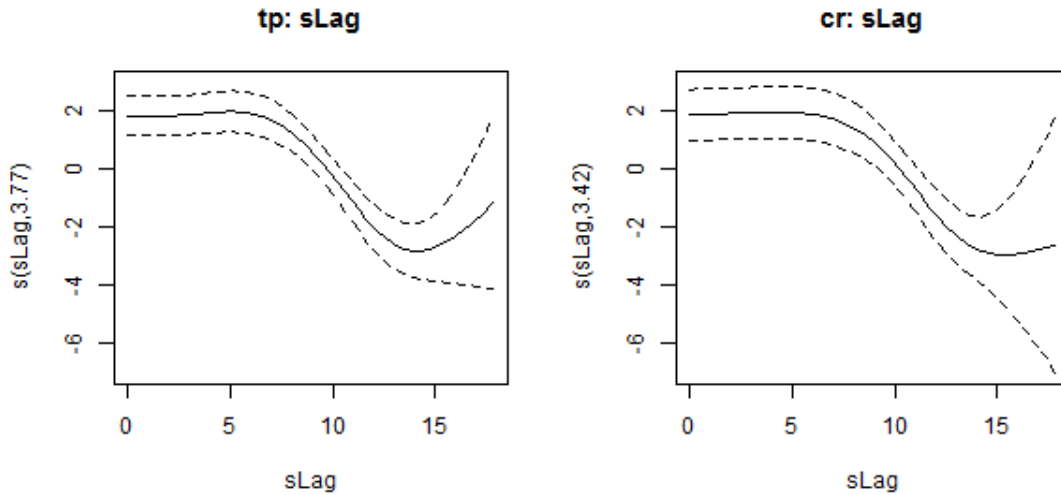


Figure 6: Partial fit on the link scale with 95% CIs for covariate $sLag$ in single covariate `bam` model fitted with thin-plate regression spline 'tp' (left) and alternative option of cubic regression spline 'cr' (right). The dimension of the basis used to represent the smooth term was set to $k = 5$ for both.

For `gamMRSea` models, covariate $maxRLpp$ fitted as a natural cubic spline was preferred over polynomial spline due to the extremely wide confidence intervals (CIs) when using the latter (Figure 7).

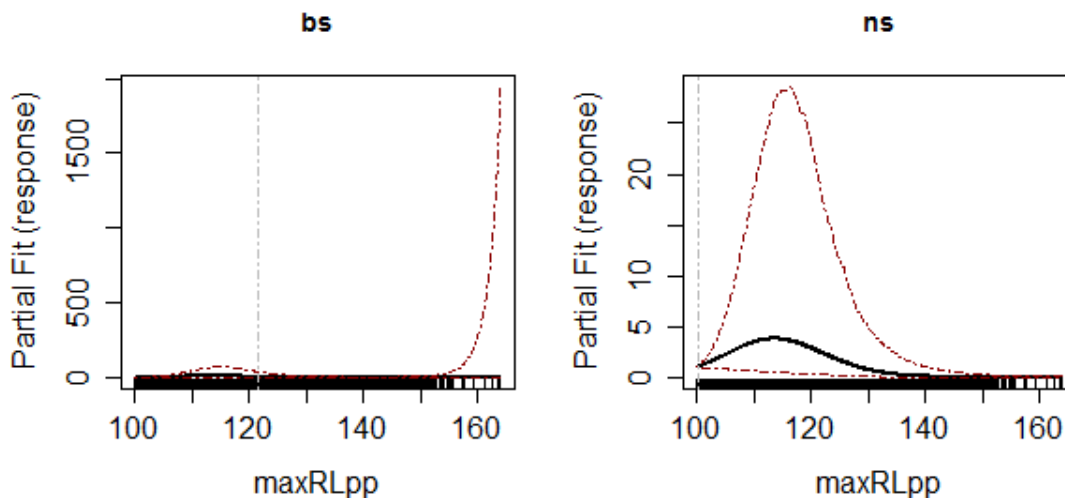


Figure 7: Partial fit on the response scale with 95% CIs for covariate $maxRLpp$ in single covariate gamMRSea model fitted with polynomial spline 'bs' (left) and natural cubic spline 'ns' (right) where data were limited to those observations with presence of sonar (and $maxRLpp$ non-zero).

Binning outliers is another option to alleviate the data-scarce tails problem: this requires changing some of the observed values of the covariates and should only be considered as a last resort. More specifically, if the tail is at the high end of the observed covariate range (right tail), those covariate values higher than the highest value at which inference is desired could be changed to this value. The equivalent method can be applied for data scarce left tails. For example, all covariate values that are considered outliers could be changed to the highest observed value not considered an outlier (or lowest in case of a data-scarce left tail). Outlier definitions can be based on how far values are from the interquartile range (IQR) of observed values. For example, all values beyond $1.5 \times$ the width of the IQR from the IQR are defined as outliers by the `boxplot` function (R Core Team 2022). For our case study, binning $sProp$ in this manner would entail changing 485 outlier values (out of the 495,661 1-min segments) to the highest value not considered an outlier for binned $sProp2$ (11.47, Figure 8). In this case, the interpretation for predicted responses at $sProp2 = 11.47$ then changes from the expected mean response at $sProp = 11.47$ when not binning to the expected mean response at $sProp2 \geq 11.47$. An alternative to binning is to limit model predictions (and inference) to the range of covariate values at which sufficient data are available. This was the preferred option for our case study as it did not require altering observed values.

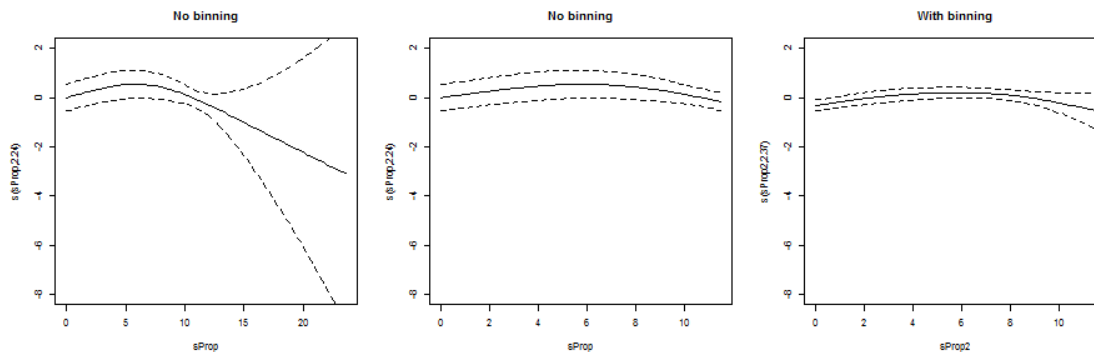


Figure 8: Partial fit on the link scale with 95% CIs for covariate $sProp$ (left and middle) and binned $sProp2$ (right) in single covariate bam models. Middle plot shows range of partial fit limited to the same range as right plot.

5.4.5 Assessing collinearity

Strong correlation existed between several pairs of covariates, in particular *Julian* and *month*, *sCount* and *sProp*, *sPres* and *maxRLpp*, *sPres* and *cumSEL*, *maxRLpp* and *cumSEL* (Table 4).

Table 4: Correlation between potential covariates rounded to nearest decimal point. Off-diagonal 1.0 highlighted in red (diagonal 1.0 in green).

	timeofd	Julian	month	sCount	sPres	sProp	sLag	sDur	maxRLpp	cumSEL
timeofd	1.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.1
Julian	0.0	1.0	1.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Month	0.0	1.0	1.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
sCount	-0.1	0.0	0.0	1.0	0.8	1.0	-0.1	0.6	0.8	0.8
sPres	-0.1	0.0	0.0	0.8	1.0	0.8	-0.2	0.6	1.0	1.0
sProp	-0.1	0.0	0.0	1.0	0.8	1.0	-0.1	0.6	0.8	0.8
sLag	0.0	0.1	0.1	-0.1	-0.2	-0.1	1.0	-0.1	-0.2	-0.2
sDur	-0.1	0.0	0.0	0.6	0.6	0.6	-0.1	1.0	0.6	0.6
maxRLpp	-0.1	0.0	0.0	0.8	1.0	0.8	-0.2	0.6	1.0	1.0
cumSEL	-0.1	0.0	0.0	0.8	1.0	0.8	-0.2	0.6	1.0	1.0

To properly assess collinearity between covariates in the model we used the `vif` function from the `car` library which calculates variance-inflation and generalized variance-inflation factors (VIFs and GVIFs) (Fox and Weisberg 2019). VIFs can be used for comparing linear terms, GVIFs for factor or smooth terms (where the nonlinear equivalent to collinearity is called concavity). We tested both fitting each covariate as a polynomial spline and as a linear term. Starting with a full model where each of the 10 covariates from Table 4 was fitted in a GLM, we eliminated the covariate with the highest GVIF (or VIF when comparing linear terms) and refitted the model until all

GVIFs and VIFs were below 5. After this process six covariates remained for model fitting: *timeofd*, *Julian*, *sProp*, *sLag*, *sDur* and *maxRLpp*.

5.4.6 Assessing autocorrelation

Two approaches for dealing with autocorrelation were considered for our case study: estimating RSEs and fitting an AR1 correlation structure. Models fitted with `gamMRSea` allow the former while `bam` allows the latter; `geeglm` allows both as well as a comparison of which provides a better fit to the data.

For any of these models, consecutive observations that are considered to be correlated need to be grouped, while groups are considered independent. We note that this may seem artificial for our case study as there is no natural division into groups that may be considered independent from each other. However, calculating an AR1 correlation structure across all 495,661 observations generally caused the software to give out-of-memory errors. In practice correlation decreases to a negligible value at some point, and so the grouping can be seen as a pragmatic approximation. Guidance on what constitutes a negligible value can be taken from the 95% CIs from the output of the `acf` function its plot (see next paragraph and Figure 9). Also, within a range of sensible group sizes the results are unaffected. However, if group sizes are too small (or too large) can lead to falsely retaining covariates in the model (or omitting covariates from the model). We discuss now how adequate group sizes can be established.

We begin by fitting a `bam` model with all available covariates fitted as smooths without specifying any correlation structure (call it `bam6`). We used the `acf` function to determine the correlation in `bam6` residuals across lags of 0 – 50 observations (Figure 9). At lag 50, the correlation first dipped within the 95% CIs of zero autocorrelation. Hence, we used groups of 50 consecutive observations. The correlation at lag 1 from Figure 9 can be used to obtain the parameter ρ for fitting an AR1 structure in `bam` models (function `bam` requires input of this parameter while `geeglm` estimates ρ as part of the model fitting process).

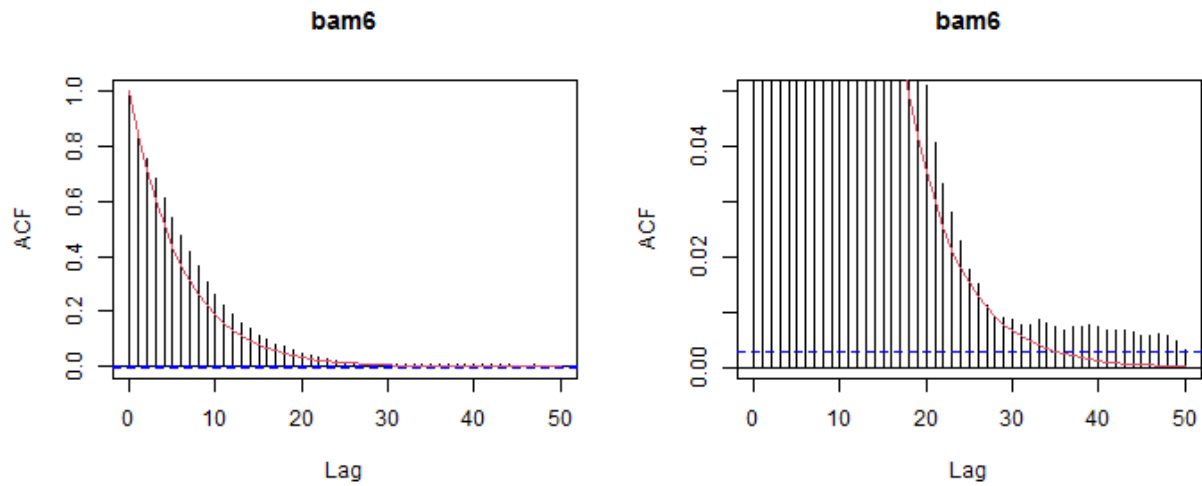


Figure 9: Autocorrelation in residuals of initial bam model fitted with all six variables as smooth terms (black vertical lines). Red line indicates the decay in correlation according to an AR1 structure where parameter ρ was set to the estimated autocorrelation in Pearson residuals at lag 1. Right plot shows a zoom in of left plot.

To determine whether a group size of 50 was adequate for estimating RSEs, we refitted a single covariate (*timeofd*) model with varying group sizes. As RSEs for all the coefficients in the model leveled off at 50, we deemed this a suitable group size for RSEs (Figure 10).

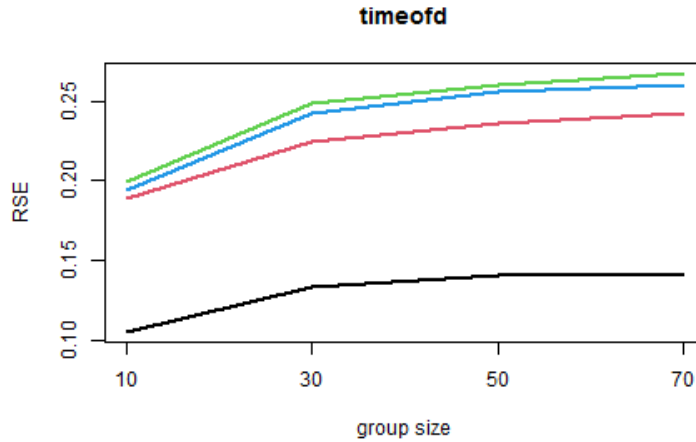


Figure 10: Robust standard errors of model terms in a single covariate `gamMRSea` model with *timeofd* against varying group sizes used for fitting the model. Black: intercept; red, green and blue: three polynomial basis functions.

To determine if 50 was a suitable group size for fitting an AR1 structure, we refitted multiple single covariate `geeglm` models with varying group sizes and extracted the estimated parameter ρ . As ρ was reasonably stable at group size 50 for all six single covariate models, we deemed it appropriate for use (Figure 11).

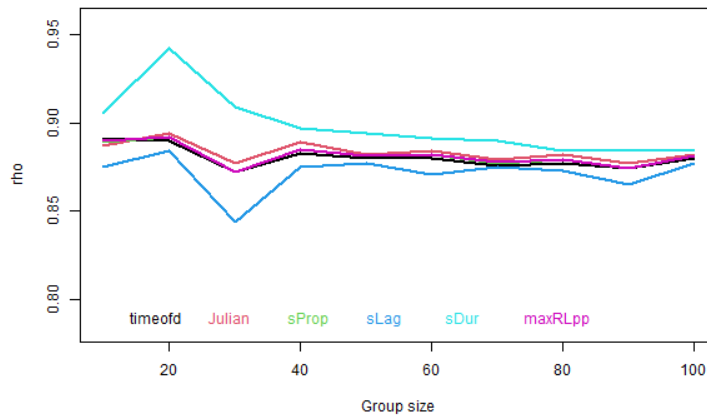


Figure 11: Estimated parameter ρ against group sizes used in single covariate `geeglm` models.

5.4.7 Fitting factor-smooth interactions with missing data

The particular challenge for this case study existed in finding a suitable way to fit a smooth function to covariate *maxRLpp* where a range of values were

missing or unobserved: $maxRLpp$ was equal to 0 dB when $sPres = 0$ and ranged between 100 and 164 when $sPres = 1$. The issue lies in the fact that there was no data to model the relationship between the response and $maxRLpp$ for $maxRLpp$ values between 0 and 100 dB. A similar issue would exist when attempting a linear term, however, we focus on smooth functions.

We started by exploring various options available for fitting a factor-smooth interaction. Smooth functions defined via the `s` function in `bam` models can be set up as interactions with factor variables using, e.g., example 1: `s(maxRLpp, by = as.factor(sPres))` or example 2: `s(maxRLpp, as.factor(sPres), bs = "fs")` (Pedersen et al. 2019; Wood 2017). Both options allowed the smooths fitted to each level of the factor covariate to differ: the former allowed group specific smoothers with different wiggleness, the latter allowed different shapes of the smoothness (but same wiggleness). However, none of these options recognized that the range of $maxRLpp$ differed for the two levels of $sPres$ and, hence, fitted a smooth function across the entire range from 0 to 160 dB for both levels of $sPres$. The following code fits smooth $maxRLpp$ as an interaction with $sPres$ using example 1 (Figure 12). Argument `shift` includes the intercept (β_0 from equation (4)) in the linear predictor η_i and argument `trans` transforms the linear predictor onto the response scale (e.g. from η_i to μ_i in equation (3)).

```
rho <- 0.8400991
b1.6.by <- bam(response ~ s(maxRLpp, by = as.factor(sPres), k = 5),
               scale = 0, rho = rho, AR.start = NULL, discrete=TRUE,
               data=zc14, family = quasipoisson(), na.action = "na.fail")
plot(b1.6.by, trans = family(b1.6.by)$linkinv, shift = coef(b1.6.by)[1])
```

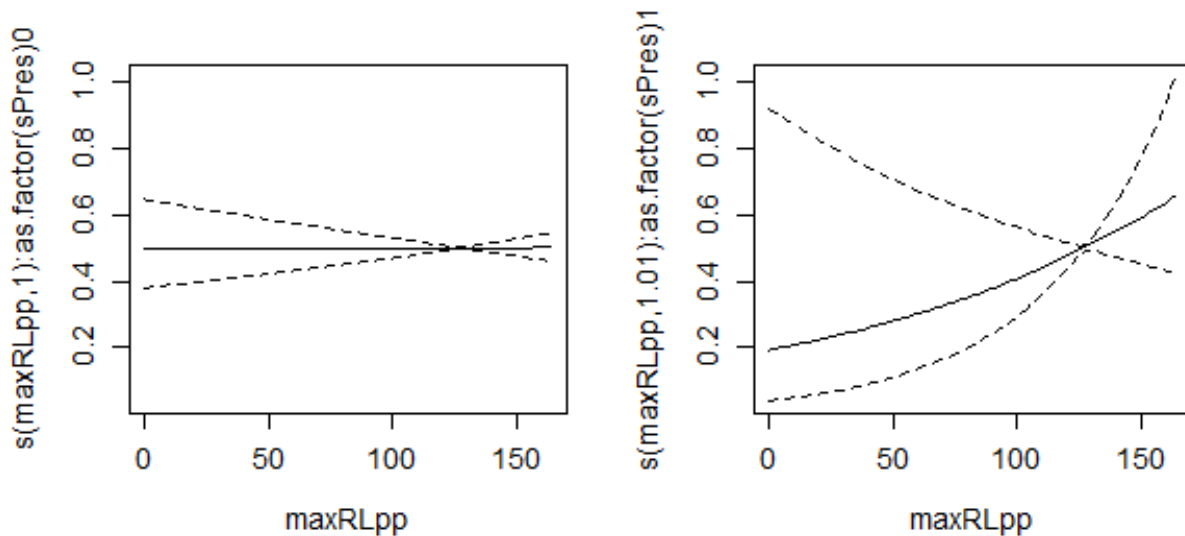


Figure 12: Partial fits on the response scale for smooth $maxRLpp$ fitted as an interaction with $sPres$ in bam model b1.6.by using the code above. Red dashed lines indicate 95% CIs.

One could argue that we could simply restrict our predictions to $maxRLpp = 0 | sPres = 0$ and to $maxRLpp = 100-164 | sPres = 1$. However, the shape of the relationship for $maxRLpp > 100$ dB in Figure 12 was influenced strongly by the mean response at 0 dB and did not reflect the pattern seen in the data exploration plots (Figure 5). Thus we conclude that neither of these options are suitable for modelling a covariate of this nature. We introduce a new way of fitting an interaction that limits the smooth functions to the observed range of the covariate for a given factor level. This approach is applicable all models for which the design matrix can be manually altered (e.g., bam, gamMRSea, geeglm). In a first step, we create basis functions that respect the nature of maxRLpp using options from the splines package (R Core Team 2022). This is done in the following steps:

- 1) First, we filter out the data for when $maxRLpp$ is not zero ($sPres = 1$).
- 2) We fit a spline to these data in a single covariate GLM.
- 3) We extract the b basis functions from the model matrix. For spline specification either manually select knot number and location bam models, Wood (2017)] or use SALSA from MRSea to do this (Scott-Hayward et al. 2014b).

- 4) Merge the basis functions back into the full data inserting a 0 for all rows where $maxRLpp = 0$.
- 5) Add an intercept-like column to allow estimation of when $maxRLpp = 0$. This column is 1 for $maxRLpp = 0$ and zero otherwise.
- 6) Fit a new model to the full data with the b basis columns and extra intercept column as linear terms. For each of these, a coefficients is estimated and, when combined together, create a discrete-smooth relationship for $maxRLpp$ (Figure 13). Note, the wide CIs for $maxRLpp > 150$ correspond to the small sample size in these regions (Figure 3). Other terms may be added in the standard way for the different methods.

Examples specific to the models are in the sections that follow.

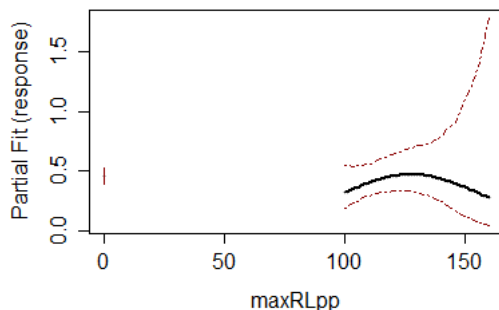


Figure 13: Partial fits on the response scale for $maxRLpp$ fitted as a polynomial spline as an interaction term with $sPres$. Red dashed lines indicate 95% CIs.

5.5 Models fitted with bam

5.5.1 Model fitting

Model fitting was based on the six candidate covariates: $timeofd$, $Julian$, $sProp$, $sLag$, $sDur$ and $maxRLpp$. The latter was included using the interaction with $sPres$ described in the previous section, referred to as $maxRLpp:sPres$ in the following. Specifically, the smooth part of $maxRLpp$ was fitted with a quadratic polynomial. $Julian$, $sProp$, $sLag$ and $sDur$ were fitted with thin-plate regression splines (Wood 2003), $timeofd$ as a cyclic cubic regression spline. As only one year of data were included in this case study, we opted out of fitting $Julian$ as a cyclic spline. We used the model

bam6 from above (Figure 9) to determine ρ (0.84) which was used for fitting all following bam models. We fitted all bam models with an AR1 correlation structure. As our models were not Gaussian, this required setting `discrete = TRUE`. This entails that covariate values for a given smooth are discretized – which is done by the function algorithm – before model fitting (Wood 2017). The AR1 structure is applied to the working residuals, thereby approximating a GEE. Model selection was based on minimum GCV scores (equation 10).

5.5.2 Code for fitting the *maxRLpp:sPres* interaction

As described in section 5.4.7, this involves filtering the data for *sPres* = 1 and fitting a smooth for *maxRLpp* to these data (Figure 14).

```
zc14$datetime <- as.POSIXct(zc14$Start, format="%d-%b-%Y %H:%M:%S")
# Create smooth in non-zero portion
testdat <- filter(zc14, maxRLpp != 0)
test_b1.6 <- glm(response ~ bs(maxRLpp, degree = 2),
                 data=testdat, family = quasipoisson())
termplot(test_b1.6)
```

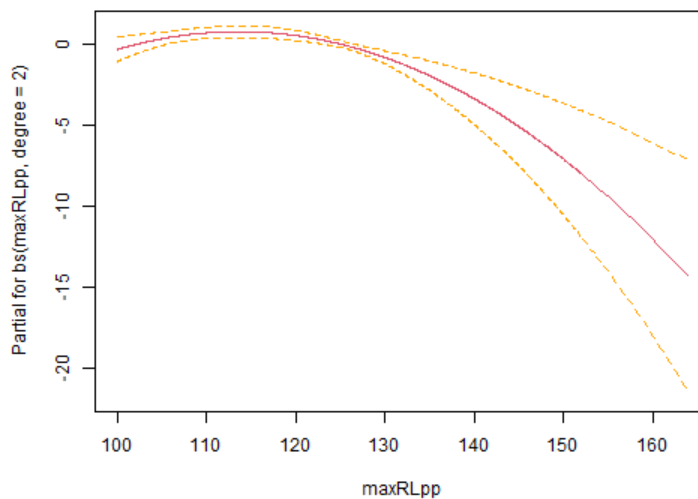


Figure 14: Partial fit on the link scale for *maxRLpp* fitted as a polynomial spline with 2 degrees of freedom in a GLM to data with *sPres* = 1. Dashed lines indicate pointwise standard errors.

Then the two baseline functions are extracted from the model matrix and merged with the full data set (including segments with $sPres = 0$). This involved generating an intercept column for estimating the expected response when $maxRLpp = 0$, and fitting a `bam` model with additional smooth terms.

```
# fit spline to larger data.
modmat.maxRLpp <-
  model.matrix(test_b1.6)[,(2:length(test_b1.6$coefficients))]
mergedata <-
  data.frame(datetime = testdat$datetime, modmat.maxRLpp)
names(mergedata) <-
  c("datetime", "maxRLpp2.1", "maxRLpp2.2")
# add intercept column for estimating mean at maxRLpp=0
# this is zero where smooth applies and 1 otherwise
mergedata$maxRLpp2.0 <- 0
# create new variables in dataset
zc14 <- left_join(zc14, mergedata) %>%
  replace_na(list(maxRLpp2.0 = 1, maxRLpp2.1 = 0, maxRLpp2.2 = 0))
# Fitting non-zero smooth to full data
# (with smooth intercept) as a GLM
glmb1.6 <- glm(response ~ maxRLpp2.0 + maxRLpp2.1 + maxRLpp2.2,
               data = zc14, family=quasipoisson)
# fit as a bam with additional term smooth Julian
rho <- 0.8400991
b2.6 <- bam(response ~ maxRLpp2.0 + maxRLpp2.1 + maxRLpp2.2
            + s(Julian, k = 5), scale = 0, rho = rho,
            AR.start = NULL, discrete = TRUE,
            data=zc14, family = quasipoisson())
```

5.5.3 Code to fit full model

The remaining four covariates are added as smooth terms. Note that *timeofd* is fitted as a cyclic smooth. Argument `k` selects the dimension of the basis function, thereby allowing for some control over the wigginess of the smooths. This could be controlled further via the `gamma` argument of the `bam` function which manipulates the amount of penalization with values larger than the

default of 1 enforce smoother fits.

```
# fitting full bam model
b6 <- bam(response ~ maxRLpp2.0 + maxRLpp2.1 + maxRLpp2.2
  + s(timeofd, bs = "cc", k = 5) + s(Julian, k = 5)
  + s(sProp, k = 5) + s(sLag, k = 5)
  + s(sDur, k = 5), scale = 0, rho = rho,
  AR.start = NULL, discrete=TRUE, data=zc14,
  family = quasipoisson(), na.action = "na.fail")
```

5.5.4 Forward model selection

Starting with a null model, in each round of model fitting, one covariate was added at a time, i.e. 6 models in round 1. The covariate with the lowest GCV score was carried forward into the next round where, again, each of the remaining covariates were added in individual models. If any of these models scored a lower GCV, the covariate combination was carried forward to the next round. This process was continued until adding covariates no longer lowered the scores or until the model contained all available covariates (Table 5). The model with covariates *timeofd*, *sLag*, *sDur* and *maxRLpp:sPres* was deemed best from the forwards selection process (model b4.3 in Table 5).

Table 5: Models fitted with the ‘bam’ function when conducting forwards model selection. Covariates included are indicated with + symbol. Bold GCV score indicates best model continued with in next round, bold model name indicates best model overall. Rounds separated by horizontal lines. Run time (RT) given in seconds.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (sec)	GCV
b1.1	+						13	1296553
b1.2		+					14	1307249
b1.3			+				6	1297164
b1.4				+			9	1296120
b1.5					+		10	1300254
b1.6						+	5	1863808
b2.1	+			+			8	1294585
b2.2		+		+			11	1308850
b2.3			+	+			11	1296139
b2.4				+	+		25	1296717
b2.5				+		+	11	1295876
b3.1	+	+		+			22	1310282
b3.2	+		+	+			12	1294565
b3.3	+			+	+		29	1294222
b3.4	+			+		+	13	1294352
b4.1	+	+		+	+		25	1309789
b4.2	+		+	+	+		28	1294386
b4.3	+			+	+	+	30	1293245
b5.1	+	+		+	+	+	79	1309327
b5.2	+		+	+	+	+	81	1309327

5.5.5 Backward model selection

We also conducted backward model selection (Table 6). This was a similar approach, however, starting with the full model, each round of model fitting determined which covariate should be omitted from the next round based on which model scored the lowest GCV (Table 6).

Table 6: Models fitted with the bam function when conducting backwards model selection. Covariates included are indicated with + symbol. Bold GCV score indicates best model continued with in next round, bold model name indicates best model overall. Rounds separated by horizontal lines. Run time (RT) given in seconds.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (sec)	GCV
b6	+	+	+	+	+	+	79	1309327
b5.1		+	+	+	+	+	59	1310693
b5.2	+		+	+	+	+	43	1294010
b5.3	+	+		+	+	+	46	1310097
b5.4	+	+	+		+	+	40	1307526
b5.5	+	+	+	+		+	24	1310143
b5.6	+	+	+	+	+		39	1308448
b4.1			+	+	+	+	23	1294953
b4.2	+			+	+	+	30	1293245
b4.3	+		+		+	+	28	1296035
b4.4	+		+	+		+	15	1294342
b4.5	+		+	+	+		31	1294386
b3.1				+	+	+	47	1297464
b3.2	+				+	+	26	1297858
b3.3	+			+		+	14	1294352
b3.4	+			+	+		35	1294222

5.5.6 Best model

Both forwards and backwards model selection arrived at the same combination of covariates: *timeofd*, *sLag*, *sDur* and *maxRLpp:sPres* (Tables 5 and 6). Confidence intervals were wide for all sonar related covariates, in particular in those areas of the covariates with diminishing sample sizes (Figure 15).

The partial fits indicated highest beaked whale click counts for *timeofd* values near -40, i.e. early afternoon (0 = sunset and -1/1 = sunrise). Lowest click counts were expected for *timeofd* values near 90, i.e. just before sunrise. For covariate *sLag*, click counts decreased with increasing *sLag* after a peak near *sLag* = 4, although 95% CIs for this relationship were very wide throughout the range of the covariate. Similarly, 95% CIs were very wide for *sDur* and *maxRLpp* making any inference about the relationship between the response and the covariate doubtful. Nonetheless, the model seems to have been driven by *sDur* as judged by the largest scale in the partial fits (compare y-axes in Figure 15).

Model run times were favorably short with the longest run time of 79 seconds for the full model b6 (Tables 5 and 6).

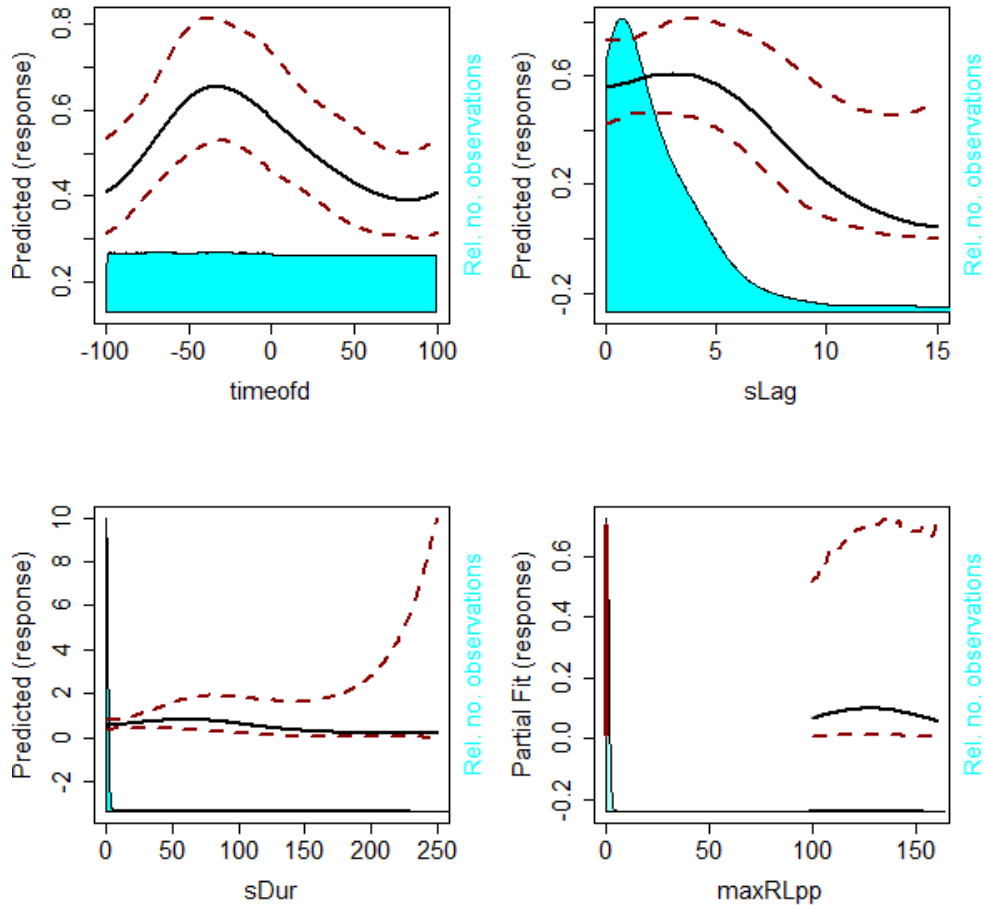


Figure 15: Partial fits on the response scale with 95% CIs for the bam model with the lowest GCV score. Relative frequency of observations indicated in turquoise.

5.6 Models fitted with MRSea functions

5.6.1 Model fitting

Model fitting was based on the six candidate covariates: *timeofd*, *Julian*, *sProp*, *sLag*, *sDur* and *maxRLpp*. The latter was included using the interaction with *sPres* referred to as *maxRLpp:sPres* in the following. *Julian*, *sProp*, *sLag* and *sDur* were fitted with quadratic B-splines and *timeofd* as a cyclic cubic regression spline. The smooth part of *maxRLpp:sPres* was selected to be a natural cubic spline with two degrees of freedom using SALSA for knots selection. QBIC was used to determine the flexibility of all smooth terms via the SALSA algorithm (Walker et al. 2010). 10-fold cross-validation was used to determine the best models using first a forwards selection procedure and

then a back wards selection process. Note that when using MRSea, backwards selection can be performed by setting the option `removal = TRUE`. Residual correlation is accounted for by providing the argument `panelid`. This results in the estimation of robust standard errors. The panel structure was defined as groups of 50 consecutive observations as described in section 5.4.6.

5.6.2 Code to fit the *maxRLpp:sPres* interaction

The steps involved for fitting the interaction term with MRSea are similar to those described for `bam` in section 5.5.2.

```
zc14$datetime <- as.POSIXct(zc14$Start, format="%d-%b-%Y %H:%M:%S")

##### fitting maxRLpp requires a 2-step process
# create smooth in non-zero portion
# filter for when maxRLpp not = 0
testdat <- filter(zc14, maxRLpp != 0)

varlist = "maxRLpp"
initialModel <- glm(response ~ 1, data = testdat, family=quasipoisson)

salsaldlist <- list(fitnessMeasure = "cv.gamMRSea",
                   minKnots_1d = rep(1, length(varlist)),
                   maxKnots_1d = rep(2, length(varlist)),
                   startKnots_1d = rep(1, length(varlist)),
                   degree = rep(2, length(varlist)),
                   gaps = rep(0, length(varlist)),
                   splines = c("ns"), # natural cubic splines
                   cv.opts=list(cv.gamMRSea.seed=1, K=10))

test1.6 <- runSALSA1D(initialModel, salsaldlist, varlist = varlist,
                     predictionData = NULL, datain = testdat,
                     panelid = testdat$blockid, removal = TRUE)

# extract the basis functions from model fitted to
# non-zero maxRLpp data
```

```

# fit spline to larger data.
testmod <- test1.6$bestModel
# extract the basis functions for maxRLpp from testmod (2 df, so 2 basis
modmat <- model.matrix(testmod)[,(2:length(testmod$coefficients))]
# merge these basis functions with datettime
# (used later to merge with original data)
mergedata <- data.frame(datettime=testdat$datettime, modmat.maxRLpp)
names(mergedata) <- c("datettime", "maxRLpp2.1", "maxRLpp2.2")
# add intercept column for estimating mean at maxRLpp=0
# this is zero where smooth applies (where maxRLpp is non-zero)
# and 1 otherwise (where maxRLpp = 0)
mergedata$maxRLpp2.0 <- 0
# create new variables in dataset (replace_na is required
# to fill in the new variable for maxRLpp = 0)
zc14 <- left_join(zc14, mergedata) %>%
  replace_na(list(maxRLpp2.0 = 1, maxRLpp2.1 = 0, maxRLpp2.2 = 0))

# Fitting non-zero smooth to full data (with smooth intercept)
m1.6 <- glm(response ~ maxRLpp2.0 + maxRLpp2.1 + maxRLpp2.2,
            data = zc14, family=quasipoisson)

```

5.6.3 Code to fit the full gamMRSea model

```

# adding smooth covariates timeofd, Julian, sProp, sLag and sDur
# to m1.6 in the order of CV ranking determined
# in forward model selection
# setting up salsalDlist
seqx <- c(2, 4, 3, 5, 1) # order establ. via CV of single cov. models
varlist = c("timeofd", "Julian", "sProp", "sLag", "sDur")[c(seqx)]
salsaldlist <- list(fitnessMeasure = "cv.gamMRSea",
                  minKnots_1d = rep(1, length(varlist)),
                  maxKnots_1d = c(2,2,2,2,2),
                  startKnots_1d = rep(1, length(varlist)),
                  degree = rep(2, length(varlist)),
                  gaps = rep(0, length(varlist)),

```

```

      splines = c("cc", "bs", "bs", "bs", "bs")[seqx],
      cv.opts=list(cv.gamMRSea.seed=1, K=10))
# using runSALSA1D to fit gamMRSea model
m6 <- runSALSA1D(m1.6, salsaldlist, varlist = varlist,
      predictionData = NULL, datain = zc14,
      panelid = zc14$blockid, removal = TRUE)

```

5.6.4 Forward model selection

Using forward model selection, we determined that a single covariate model with *Julian* fitted as a smooth was the preferred model based on lowest CV scores from rounds 1 (m1 models with one covariate) and 2 (m2 models with two covariates) (Table 7).

Table 7: Models fitted with MRSea functions using forwards model selection with cross-validation scores (CV). Covariates included are indicated with + symbol. Run time (RT) given in minutes. Horizontal lines indicate rounds of model fitting. Bold CV scores indicates lowest CV score in a round, bold model name the overall best model.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (mins)	CV
m1.1	+						4.0	53.880
m1.2		+					9.3	53.827
m1.3			+				4.5	53.877
m1.4				+			5.6	53.870
m1.5					+		5.0	53.877
m1.6						+	0.1	53.870
m2.1	+	+					13.2	53.828
m2.2		+	+				10.8	53.834
m2.3		+		+			7.8	53.854
m2.4		+			+		9.7	53.834
m2.5		+				+	4.6	53.864

5.6.5 Backward selection

Backward selection was performed using the full model with option `removal = TRUE` to take advantage of the inbuilt model selection tool of the `runSALSA1D` function. For a given smooth term, after having found the optimal knot numbers and locations, the fit is compared to a linear term for the covariate and to omitting the covariate (Scott–Hayward et al. 2013). For simplicity, we limited the maximum number of knots to two for each

covariate and used the preferred option of CV scores for determining the flexibility of the smooths.

As none of the covariates were removed from the model, the preferred model contained all six covariates (Table 8).

Table 8: Model fitted with MRSea functions using backwards model selection with cross-validation scores (CV). Covariates retained by the automatic model selection are indicated with + symbol. Run time (RT) given in minutes.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (mins)	CV
m6	+	+	+	+	+	+	166	53.808

5.6.6 Best model

The CV score of the full model (Table 8) was lower (53.808) than that of the preferred model from forwards model selection (53.827 for model `m1.2` with *Julian*, Table 7). Hence, the overall preferred model fitted was the full model. For all covariates, except *maxRLpp*, confidence intervals were wide in relation to the best fitting line across the whole range of covariates (Figure 16). However, in comparison to the partial fits from the best `bam` model (Figure 15), CIs were narrower for *sDur* and *maxRLpp* in the MRSea model.

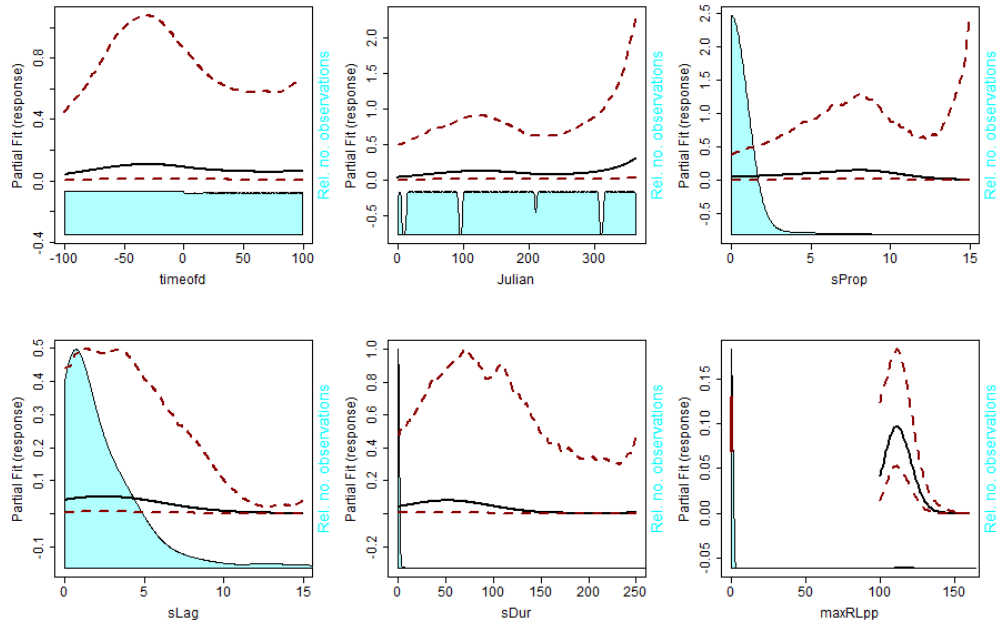


Figure 16: Partial fits on the response scale with 95% CIs for the MRSea model with the lowest CV score. Relative frequency of observations is indicated in turquoise.

5.7 Models fitted with generalized estimating equations

5.7.1 Model fitting

As for `bam` and `MRSea` models, fitting for GEE models was based on the six candidate covariates: `timeofd`, `Julian`, `sProp`, `sLag`, `sDur` and `maxRLpp`, each fitted as a smooth. The latter was included using the interaction with `sPres` referred to as `maxRLpp:sPres` as before.

Models fitted with the `geeglm` function allow incorporating smooths, e.g., as polynomial splines via the `bs` function (or as natural cubic regression splines via the `ns` function). Slightly more complicated is incorporating cyclic smooths via the `mgcv::s` function.

We make use of the methods for incorporating `maxRLpp` described for `gamMRSea` models in section 5.6.2, i.e. using a natural cubic spline with two degrees of freedom using `SALSA` for knots selection for the smooth part of `maxRLpp:sPres`. We also make use of the `gamMRSea` model with `timeofd` for extracting the call to fit a cyclic smooth with the `s` function and obtaining optimal knot locations.

Optimal knot locations could also be obtained for the polynomial splines for

Julian, *sProp*, *sLag* and *sDur*; however, when tested on the full model with an AR1 structure, it caused the model to crash. Hence, *Julian*, *sProp*, *sLag* and *sDur* were fitted with quadratic B-splines without specifying knot locations.

Group sizes for the correlation structure entered the model via argument `id`. Should there be missing observations within a group, indexing observations via the `wave` argument can be useful. This was not an issue for our case study.

5.7.2 Code for fitting the full model

5.7.2.1 Using robust standard errors (“working independence”) Here we fit the full model with all six covariates with the “working independence” structure. This entails obtaining the basis functions for the *maxRLpp:sPres* interaction first, i.e. *maxRLpp2.0*, *maxRLpp2.1*, *maxRLpp2.2* (section 5.6.2). This further involves fitting the `gamMRSea` model `m1.1` first and extracting the formulation to fit the cyclic smooth for *timeofd* (section 5.6.3).

```
# obtaining the call and knot locations for fitting cyclic timeofd
m1.1$bestModel$call
# fitting the model with 'working independence'
g6i <-
  geeglm(response ~ maxRLpp2.0 + maxRLpp2.1 + maxRLpp2.2
    + bs(Julian)
    + bs(sLag)
    + bs(sProp)
    + bs(sDur)
    + smooth.construct(s(timeofd, bs = "cc",
      k = (length(m1.1$splineParams[[2]]$knots)) + 2),
      knots = list(timeofd = as.numeric(c(m1.1$splineParams[[2]]$bd[1],
        m1.1$splineParams[[2]]$knots, m1.1$splineParams[[2]]$bd[2]))),
      data = data.frame(timeofd)$X[, -1],
      data = zc14, family = poisson, id = zc14$blockid,
      corstr = "independence")
```

5.7.2.2 Using an AR1 correlation structure We go through the same steps as in section 5.7.2.1. However, we specify an AR1 correlation structure.

```

# obtaining the call and knot locations for fitting cyclic timeofd
m1.1$bestModel$call
# fitting the model with AR1
g6a <-
  geeglm(response ~ maxRLpp2.0 + maxRLpp2.1 + maxRLpp2.2
    + bs(Julian)
    + bs(sLag)
    + bs(sProp)
    + bs(sDur)
    + smooth.construct(s(timeofd, bs = "cc",
      k = (length(m1.1$splineParams[[2]]$knots)) + 2),
      knots = list(timeofd = as.numeric(c(m1.1$splineParams[[2]]$bd[1],
      m1.1$splineParams[[2]]$knots, m1.1$splineParams[[2]]$bd[2]))),
      data = data.frame(timeofd)$X[,-1],
      data = zc14, family = poisson, id = zc14$blockid,
      corstr = "ar1")

```

5.7.3 Selecting a correlation structure

To determine which correlation structure was preferred, we used the `geepack::QIC` function to obtain QIC scores for the `g6i` and `g6a` models (Højsgaard et al. 2005).

```

# model comparison based on QIC
QIC(g6i, g6a)

```

The model with the ‘working independence’ (`g6i`) structure returned a lower QIC compared to the model with the AR1 structure (`g6a`). Therefore, we concluded that using the ‘working independence’ and estimating RSEs was the better approach (Cui and Qian 2007).

Table 9: GEE models fitted with the `geeglm` function with 'working independence' (g6i) and AR1 (g6a) correlation structures and corresponding QIC values. Covariates included are indicated with + symbol. Run time (RT) given in minutes.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (mins)	QIC
g6i	+	+	+	+	+	+	0.5*	778294
g6a	+	+	+	+	+	+	3.8*	778887

*Runtimes of MRSea models m1.1 and m1.6 need to be added (Table 7)

5.7.4 Forward model selection

Using forward model selection based on minimum QIC, we arrived at the full model via the six model fitting rounds shown in Table 10.

Table 10: GEE models fitted with the `geeglm` function using the "working independence" correlation structure. Covariates included are indicated with + symbol. Bold QIC scores indicates best model continued with in next round; bold model name indicates best model overall. Rounds are separated by horizontal lines. Run time (RT) given in seconds.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (secs)	QIC
g1.1	+						17	822210
g1.2		+					14	819745
g1.3			+				14	832922
g1.4				+			14	820239
g1.5					+		15	831243
g1.6						+	15	829186
g2.1	+	+					18	806851
g2.2		+	+				17	817810
g2.3		+		+			18	800079
g2.4		+			+		17	816249
g2.5		+				+	17	814129
g3.1	+	+		+			22	786867
g3.2		+	+	+			20	798419
g3.3		+		+	+		20	796917
g3.4		+		+		+	21	794822
g4.1	+	+	+	+			22	785615
g4.2	+	+		+	+		22	783992
g4.3	+	+		+		+	22	781842
g5.1	+	+	+	+		+	29	780682
g5.2	+	+		+	+	+	29	779488
g6	+	+	+	+	+	+	30	778294

5.7.5 Backward model selection

Using backward model selection, we concluded that the full model was the best fitting model (Table 11).

Table 11: GEE models fitted with the `geeglm` function. Covariates included are indicated with + symbol. Bold QIC score indicates best model continued with in next round; bold model name indicates best model overall. Rounds separated by horizontal lines. Run time (RT) given in seconds.

Model	timeofd	Julian	sProp	sLag	sDur	maxRLpp:sPres	RT (secs)	QIC
m6	+	+	+	+	+	+	30	778294
m5.1		+	+	+	+	+	26	791221
m5.2	+		+	+	+	+	28	798778
m5.3	+	+		+	+	+	29	779488
m5.4	+	+	+		+	+	26	797880
m5.5	+	+	+	+		+	33	780682
m5.6	+	+	+	+	+		27	783344

5.7.6 Best model

Both forward and backward model selection arrived at the full model as the preferred model fitted with `geeglm`. Again, confidence intervals were wide, in particular in regions of the covariates with few observations. In comparison to the `gamMRSea` models, CIs were wider for the GEE approach for three covariates, *sProp*, *sDur* and *maxRLpp* (compare partial fits in Figures 16 and 17).

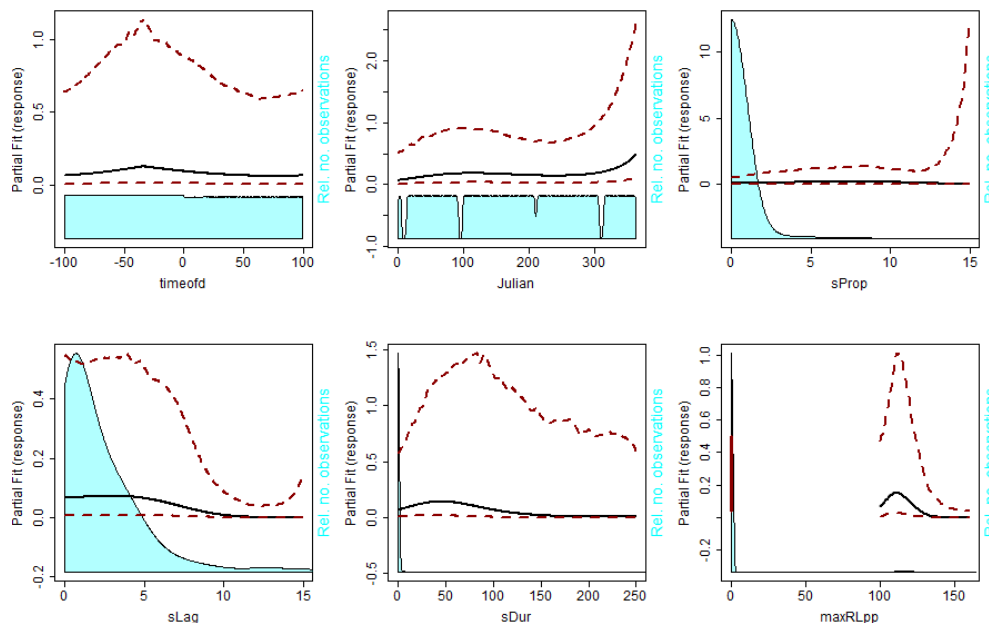


Figure 17: Partial fits on the response scale with 95% CIs for the GEE model with the lowest QIC score. Relative frequency of observations is indicated in turquoise.

6 Discussion

Our case study has shown that correlation in consecutive observations can be a major issue when analyzing PAM data with GAMs and dealing with it appropriately is still an ongoing area of research. Similarly, choosing among the available model selection methods for a given modeling approach can be difficult. We aimed to give some guidance on this topic in section 4.3.

Analyses of our case study revealed that the preferred `bam` model had a different covariate composition compared to the `gamMRSea` and `geeglm` models. The best `bam` model only contained three covariates and the interaction term – *timeofd*, *sLag*, *sDur* and *maxRLpp:sPres* – while for the other two methods (models fitted with `gamMRSea` and `geeglm`) the best models contained all six available covariates – including also *Julian* and *sProp*. This discrepancy may have been due to the different correlation structure fitted for `bam` models compared to the other two methods. An alternative reason for the different covariate composition in the best `bam` model may have been using GCV for model selection as opposed to 10-fold CV for `gamMRSea` or QIC for the `geeglm` model. Whether this difference was due to the different model selection criteria used for the three different approaches or to different correlation structures could not be determined as part of this study.

We also found a discrepancy in the best model chosen by forward vs backward selection using the `gamMRSea` method: the best model via backwards selection was deemed better than the best model using forwards selection. This highlights the importance of exploring the model space thoroughly. Ideally all possible covariate combinations should be tested – as long as they are plausible; however, this is generally too time consuming (six covariates yield 64 combinations). Furthermore, the order in which the covariates enter the model can be important. For the `gamMRSea` model using backwards selection, terms entered the model in the order of their CV scores when fitted in single covariate models. This was due to the fitting process with function `runsalsalD` which sequentially finds the best knot locations for each smooth term given the other covariates already in the model (factor and linear terms as well as smooth terms for which this process was completed).

With few exceptions, confidence intervals were generally very wide in partial fits for all modelling approaches. For `bam` models, CIs were particularly

wide in comparison to the bandwidth of the smooth itself for the two covariates with most observations at zero ($sDur$ and $maxRLpp:sPres$, Figure 15). For `gamMRSea` and `geeglm` models, the effects of all covariates were small compared to the width of the CIs – with the exception of the $maxRLpp:sPres$ interaction (Figures 16, 17). These wide CIs were likely due to the high variability in beaked whale click counts per 1-min segments, an issue that might be alleviated by using presence/absence of clicks in 1-min segments as the response in binomial models (e.g. Solsona Berga et al. 2019).

Differences in the shape of smooths existed between the three modelling approaches which may have been due, in part, to differences in the maximum number of knots (e.g. compare the smooths for $timeofd$ for `bam`, `gamMRSea` and `geeglm`, Figures 15, 16, 17). The covariate of interest, $maxRLpp$ was best modelled by `gamMRSea` where the partial fit for the interaction term $maxRLpp:sPres$ closely resembled the pattern shown in Figure 5, likely an effect of the optimal knot locations determined with SALSA (Walker et al. 2010). A better result for `bam` models may have been achieved by allowing more degrees of freedom when setting up the baseline functions for the $maxRLpp:sPres$ interaction.

In summary, each modelling approach has its benefits. Models fitted with `bam` had the shortest run times; however, they were limited to fitting an AR1 correlation structure. Models fitted with `gamMRSea` took longer to fit; however, they allowed fitting RSE – which do not rely on specifying the correlation structure correctly – and using SALSA. The `geeglm` models allowed determining the best correlation structure. However, it is not very user friendly when fitting complex smooth functions, such as cyclic smooths. Which of these modeling approaches is the best to correctly identify the presence (or absence) of an effect of a disturbance on the vocalization rate given it had an effect (given it did not have an effect) could not be determined during this study. This would require a simulation study.

6.1 Acknowledgments

Funding to conduct this study was provided by the US Office of Naval Research (ONR Award Number: N00014-18-1-2807). The US Navy Pacific Fleet, Chief of Naval Operations Environmental Readiness Division (CNO

N45), Office of Naval Research, and Naval Postgraduate School all contributed funding for data collection and analysis that went into the outcomes presented in this report. We thank colleagues from the University of St Andrews and other institutions for their input – in particular Prof. David Borchers, Dr. David Miller, Prof. Finn Lindgren (University of Edinburgh), Dr. Julie Oswald, Prof. Vincent Janik.

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. Budapest: Akademiai Kiado. Budapest: Akademiai Kiado.
- Akaike H (1979) A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* 66 (2):237–242
- Anderson D, Burnham K (2004) Model selection and multi-model inference. Second NY: Springer-Verlag 63:10
- Bachl FE, Lindgren F, Borchers DL, Illian JB (2019) Inlabru: An r package for bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution* 10:760–766
- Barton K (2022) MuMIn: Multi-model inference
- Baumann-Pickering S, Roch M, Brownell R, et al (2014) Spatio-temporal patterns of beaked whale echolocation signals in the north pacific. *PloS one* 9:e86072. <https://doi.org/10.1371/journal.pone.0086072>
- Benjamins S, van Geel N, Hastie G, et al (2017) Harbour porpoise distribution can vary at small spatiotemporal scales in energetic habitats. *Deep Sea Research Part II: Topical Studies in Oceanography* 141:191–202
- Booth CG, Oedekoven CS, Gillespie D, et al (2017) Assessing the viability of DensityEstimation for cetaceans from passive acoustic fixed sensors throughout the life cycle of an offshore e&PField development: Report number: SMRUC-OGP-2017-001. Submitted to IOGP sound and marine life joint industry programme (unpublished). SMRU Consulting, St Andrews
- Brandt MJ, Diederichs A, Betke K, Nehls G (2011) Responses of harbour porpoises to pile driving at the Horns Rev II offshore wind farm in the danish north sea. *Marine Ecology Progress Series* 421:205–216
- Brookes KL, Bailey H, Thompson PM (2013) Predictions from harbor porpoise habitat association models are confirmed by long-term passive acous-

- tic monitoring. *The Journal of the Acoustical Society of America* 134:2523–2533
- Buckland ST, Rexstad E, Marques TA, Oedekoven CS (2015) Distance sampling: Methods and applications. *Methods in statistical ecology*. Springer
- Buntaran H, Forkman J, Piepho H-P (2021) Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: Accuracy and precision. *Theoretical and Applied Genetics* 134:1513–1530
- Clausen KT, Tougaard J, Carstensen J, et al (2019) Noise affects porpoise click detections—the magnitude of the effect depends on logger type and detection filter settings. *Bioacoustics* 28:443–458
- Cui J, Qian G (2007) Selection of working correlation structure and best model in GEE analyses of longitudinal data. *Communications in statistics-Simulation and computation*® 36:987–996
- Dähne M, Gilles A, Lucke K, et al (2013) Effects of pile-driving on harbour porpoises (*phocoena phocoena*) at the first offshore wind farm in germany. *Environmental Research Letters* 8:025002
- Dorazio KU Robert M. AND Karanth (2017) A hierarchical model for estimating the spatial distribution and abundance of animals detected by continuous-time recorders. *PLOS ONE* 12:1–18. <https://doi.org/10.1371/journal.pone.0176966>
- Evans PGH (1996) Human disturbance of cetaceans. In: Taylor VJ, Dunstone N (eds) *The exploitation of mammal populations*. Springer Netherlands, Dordrecht, pp 376–394
- Fox J, Weisberg S (2019) *An R companion to applied regression, Third*. Sage, Thousand Oaks CA
- Hardin JW, Fomby TB, Hill RC (2003) The sandwich estimate of variance. In: Fomby TB, Carter Hill R (eds) *Maximum likelihood estimation of misspecified models: Twenty years later: Advances in econometrics*. Emerald Group Publishing Limited, Bingley, pp 45–73
- Hardin JW, Hilbe JM (2013) *Generalized estimating equations (2nd ed.)*. Chapman; Chapman; Hall/CRC
- Harris DV, Miksis-Olds JL, Vernon JA, Thomas L (2018) Fin whale density and distribution estimation using acoustic bearings derived from sparse arrays. *The Journal of the Acoustical Society of America* 143:2980–2993
- Hildebrand JA, Baumann-Pickering S, Frasier KE, et al (2015) Passive acous-

- tic monitoring of beaked whale densities in the gulf of mexico. *Scientific reports* 5:16343
- Højsgaard S, Halekoh U, Yan J (2005) The R package geepack for generalized estimating equations. *Journal of Statistical Software* 15:1–11
- Huijser LAE, Estrade V, Webster I, et al (2020) Vocal repertoires and insights into social structure of sperm whales (*physeter macrocephalus*) in mauritius, southwestern indian ocean. *Marine Mammal Science* 36:638–657. <https://doi.org/https://doi.org/10.1111/mms.12673>
- Jacobson EK, Henderson EE, Miller DL, et al (2022) Quantifying the response of blainville’s beaked whales to US naval sonar exercises in hawaii. *Marine Mammal Science*
- King SL, Guarino E, Donegan K, et al (2021) Evidence that bottlenose dolphins can communicate with vocal signals to solve a cooperative task. *Royal Society Open Science* 8:202073
- Lin T-H, Yu H-Y, Chen C-F, Chou L-S (2013) Automatic detection and classification of cetacean tonal sounds from a long-term marine observatory. In: 2013 IEEE international underwater technology symposium (UT). pp 1–6
- Marques TA, Thomas L, Ward J, et al (2009) Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville’s beaked whales. *The Journal of the Acoustical Society of America* 125:1982–1994
- Marra G, Wood SN (2011) Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis* 55:2372–2387. <https://doi.org/https://doi.org/10.1016/j.csda.2011.02.004>
- Melcon ML, Cummins AJ, Kerosky SM, et al (2012) Blue whales respond to anthropogenic noise. *PLoS One* 7:e32681
- Nuuttila H, Courtene-Jones W, Baulch S, et al (2017) Don’t forget the porpoise: Acoustic monitoring reveals fine scale temporal variation between bottlenose dolphin and harbour porpoise in cardigan bay SAC. *Marine Biology* 164:
- Oedekoven CS, Marques TA, Harris D, et al (2022) A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring. *Environmental and Ecological Statistics* 29:101–125
- Oedekoven CS, Thomas L (2017) A modelling approach for acoustic detec-

- tions of blue whale d calls and beaked whale clicks collected with high-frequency acoustic recording packages (HARP). CREEM report 15622-1. Provided to scripps institution of oceanography, university of california, san diego, la jolla, CA USA. September 2017 (unpublished). University of St Andrews, Scotland
- Oswald JN, Oedekoven CS, Yack TM, et al (2015) Development of statistical methods for examining relationships between odontocete vocal behavior and navy sonar signals. Final report. Prepared for u.s. Fleet forces command. Submitted to naval facilities engineering command atlantic, norfolk, virginia, under contract no. N62470-10-3011, task order 39, issued to HDR inc., Virginia beach, virginia. 30 march 2015. Biowaves, Inc. Encinitas, CA; CREEM/University of St Andrews, Scotland
- Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics* 57:120–125. <https://doi.org/https://doi.org/10.1111/j.0006-341X.2001.00120.x>
- Parsons E (2017) Impacts of navy sonar on whales and dolphins: Now beyond a smoking gun? *Frontiers in Marine Science* 295
- Pedersen EJ, Miller DL, Simpson GL, Ross N (2019) Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ* 7:e6876
- Pirotta E, Merchant ND, Thompson PM, et al (2015) Quantifying the effect of boat disturbance on bottlenose dolphin foraging activity. *Biological Conservation* 181:82–89. <https://doi.org/https://doi.org/10.1016/j.biocon.2014.11.003>
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Sarnocińska J, Teilmann J, Balle JD, et al (2020) Harbor porpoise (*phocoena phocoena*) reaction to a 3D seismic airgun survey in the north sea. *Frontiers in Marine Science* 6:
- Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6:461–464
- Scott–Hayward LA, Mackenzie ML, Ashe E, Williams R (2015) Modelling killer whale feeding behaviour using a spatially adaptive complex region spatial smoother (CReSS) and generalised estimating equations (GEEs). *Journal of agricultural, biological, and environmental statistics* 20:305–322
- Scott–Hayward LA, Mackenzie ML, Walker CG, et al (2014a) Complex region spatial smoother CReSS. *Journal of Computational and Graphical*

Statistics 23:340–360

- Scott–Hayward LA, Oedekoven CS, Mackenzie ML, et al (2013) User guide for the MRSea package: Statistical modelling of bird and cetacean distributions in offshore renewables development areas. University of St. Andrews
- Scott–Hayward LA, Oedekoven CS, Mackenzie ML, Walker CG (2014b) MRSea package: Statistical modelling of bird and cetacean distributions in offshore renewables development areas. University of St. Andrews: Contract with Marine Scotland: SB9 (CR/2012/05)
- Simonis AE, Brownell RL, Thayre BJ, et al (2020) Co-occurrence of beaked whale strandings and naval sonar in the mariana islands, western pacific. *Proceedings of the Royal Society B: Biological Sciences* 287:20200070
- Solsona Berga A, Trickey J, Rice A, et al (2019) Potential impact of mid-frequency active sonar on whales from passive acoustic monitoring data. *The Journal of the Acoustical Society of America* 146:2939–2939
- Southall BL, Nowacek DP, Miller PJ, Tyack PL (2016) Experimental field studies to measure behavioral responses of cetaceans to sonar. *Endangered Species Research* 31:293–315
- Stevenson BC, Borchers DL, Altwegg R, et al (2015) A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution* 6:38–48
- Thode AM, Kim KH, Blackwell SB, et al (2012) Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. *The Journal of the Acoustical Society of America* 131:3726–3747
- Thompson PM, Brookes KL, Cordes LS (2014) Integrating passive acoustic and visual data to model spatial patterns of occurrence in coastal dolphins. *ICES Journal of Marine Science* 72:651–660
- Thompson PM, Brookes KL, Graham IM, et al (2013) Short-term disturbance by a commercial two-dimensional seismic survey does not lead to long-term displacement of harbour porpoises. *Proceedings of the Royal Society B: Biological Sciences* 280:20132001
- Tougaard J, Tougaard S, Jensen RC, et al (2006) Harbour seals on horns reef before, during and after construction of horns rev offshore wind farm. Vattenfall A/S
- Wald A, Wolfowitz J (1943) An exact test for randomness in the non-parametric case based on serial correlation. *Annals of Mathematical Statistics* 14:378–388

- Walker CG, Mackenzie ML, Donovan CR, O’Sullivan MJ (2010) SALSA - a spatially adaptive local smoothing algorithm. *Journal of Statistical Computation and Simulation* 81:179–191
- Williamson LD, Brookes KL, Scott BE, et al (2016) Echolocation detections and digital video surveys provide reliable estimates of the relative density of harbour porpoises. *Methods in Ecology and Evolution* 7:762–769
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73:3–36
- Wood SN (2003) Thin-plate regression splines. *Journal of the Royal Statistical Society (B)* 65:95–114
- Wood SN (2017) *Generalized additive models: An introduction with R*, 2nd edn. Chapman; Hall/CRC
- Wood SN, Augustin NH (2002) GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological modelling* 157:157–177
- Wood SN, Pya N, Säfken B (2016) Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association* 111:1548–1575
- Zeileis A (2004) Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software* 11:1–17. <https://doi.org/10.18637/jss.v011.i10>
- Zucchini W, MacDonald IL, Langrock R (2017) *Hidden markov models for time series: An introduction using r*. CRC press